The Design and Analysis of Experiments Under Inferference Using Partial Network data

Steve Shane Alex Arun Tyler

May 29, 2023

1 Introduction

We present methods and assumptions for accounting for interference using partial network data. Interference is the phenomenon in which treatments applied to a particular observational unit cause changes in other units. The term "interference" was introduced by Cox [1958] in their foundational work and was formalized by Rubin [1990] in the "Stable Unit Treatment Value Assumption" (SUTVA), which combines the assumptions of causal consistency and absence of interference.

Numerous application areas where SUTVA is violated include the study of infectious diseases Hudgens and Halloran [2008], Tchetgen and VanderWeele [2012], public policy (eg. public heath insurance) Malani et al. [2021], Imai et al. [2021], online platforms Saveski et al. [2017], Pouget-Abadie et al. [2018, 2019] and online marketplaces Ha-Thuc et al. [2020], Johari et al. [2022].

The presence of interference leads to novel causal effects beyond the average treatment effect, which is not adequately defined when interference is present. Some include the global average treatment effect (or average total treatment effect) which contrasts the average outcome under treatment assignments of giving everyone the treatment vs. giving no one the treatment. Others included average spillover effects. under the Early considerations of the lack of SUTVA in vaccine

trials were discussed in Halloran and Struchiner [1995] and Hudgens and Halloran [2008].

Early methods for dealing with interference involved partitioning the data into interfering blocks, and using two-stage randomized designs for these effects Hudgens and Halloran [2008]. However, when interference propagates along a single network as whole such partitions cannot be made.

Aronow and Samii [2017] make the contribution of defining the exposure, an analogue to the treatment in the setting of interference, by defining the potential outcomes which can exist as a result of different treatments of neighbours (or higher order interactions) in a network. They also introduce unbiased estimators for the random effects (which may suffer from high variance). Defining an exposure mapping is a modelling decision left to the analyst, though Yuan et al. [2021] makes the first steps in a data driven version.

More recent approaches to estimating causal effects under interference have broadly been partitioned into two streams. Those which develop novel estimators for particular causal effects, and those which introduce novel experimental designs, primarily to reduce variance.

Some of the recent estimators include Chin [2019] who considers regression approaches to estimating global average treatment effects and [Yu et al., 2022, Cortez et al., 2022a,b] which consider various model-based assumptions to estimate the GATE, as well as direct and indirect effects.

Other design based approaches include graph cluster randomization Ugander et al. [2013] which developed concorrently with Aronow and Samii [2017] and the later extensions Ugander and Yin [2020]. Candogan et al. [2021] discusses computational methods for near optimal design in partitioning non-interfering blocks. And Brennan et al. [2022] considers minimizing bias in the case difference in mean estimators in bipartite networks. Doudchenko et al. [2020] and Bajari et al. [2023] also introduce designs for this special kind of interference.

Detection of interference is another area that has drawn focus Saveski et al. [2017], Pouget-Abadie et al. [2019], often involving randomization tests as is seen in Basse et al. [2019], Puelz et al. [2019], Athey et al. [2018].

Most of the methods thus-far have required full graph information, with the exception of Yu et al. [2022], Cortez et al. [2022a] which substitutes this graph information for model assumptions and a staggered roll-out design. Sävje et al. [2021] also studies a similar problem for Bernoulli designs. Full network data is, however, often difficult or expensive to collect. A large literature has focused on survey methods that make network data collection more feasible and accessible. One survey method involves the collection of Aggregated Relational Data (ARD). Rather than collecting all edges or relationships in a network of interest, a researcher conducting an ARD survey asks respondents "How many people do you know with trait X?" for various traits. ARD was originally proposed to estimate the size of hard-to-reach populations, such as the number of HIV-positive men in the U.S. Killworth et al. [1998], Scutelniciuc [2012], Jing et al. [2014]. ARD has been shown to be 70-80% cheaper to collect than full network data [Breza et al., 2020] and has been shown to be useful in many network inference problems [Breza et al., 2020, 2023]. Though much of the discussion in this work focuses on using aggregated relational data to understand the structure of the unobserved network, our methods easily apply to other empirically relevant data cases. See Section 2.4 for more details.

Our paper discusses three main topics. The first is assumptions regarding nonparameteric causal identification. We introduce a novel assumption, that when controlling for graph statistics in addition to the exposure, allows for the identification of causal effects. This is in contrast to methods like Veitch et al. [2019], Cristali and Veitch [2022], McFowland III and Shalizi [2023] which involve confoundedness through latent traits. We take a regression approach to identify causal effects of interest, rather than focus solely on the GATE as in Chin [2019]. The second then involves the use of partial network data and models in order to identify causal effects in the case of the stochastic-block-model. We discuss identifiability and inference using saturated randomized designs and elaborate on optimal design strategies under different modeling and error assumptions.

2 Methods

2.1 Data Generating Model

We suppose that there are n individuals denoted by a set $N = \{1, 2, \ldots, n\}$. Let $Y_i \in \mathbb{R}$ denote an outcome of interest, let $A_i \in \{0, 1\}$ denote an individual's binary treatment and $V_i \in \mathcal{V}$ denote their treatment **exposure** as in Aronow and Samii [2017]. The exposure generalizes the treatment from the binary case, to a higher dimensional input, based on other's treatment in the population. We take the standard assumption that interference is propagated over some network $\mathcal{G} = (N, E)$ where N is the set of individuals in the population and $E \in N \times N$ are the set of edges. The graph $\mathcal{G}$ can also be represented by the adjacency matrix G_n. In general, G_n can be directed, but we leave this discussion to the directed case.

We suppose that there exists some function f_V such that $V_i = f_V(\boldsymbol{a}, \theta_i)$ where θ_i is the local graph information for individual i. This is commonly the i^{th} row of neighbors, but it need not be restricted to immediate neighbors. Finally, we assume that the graph G_n is generated by an exchangeable random graph with parameters Z_i for each individual. In practice, this could be the stochastic blockmodel where Z_i is discrete, but nonparametric identification need not require this. Additionally, we suppose that there may be additional covariates H_i (which may be of interest for heterogeneous effects) which may be related to the outcome, but do not necessarily contribute to the graph. Let $Y_i(\boldsymbol{a})$ be the potential outcome for some outcome of interest under a treatment vector $\boldsymbol{a} \in \{0, 1\}^n$. We will consider parameters of interest which are functions of the potential outcomes in the graph.

2.2 Nonparametric Identification

We are primarily considering parameters of the form $\Psi(\boldsymbol{a}, h, G_n) = \frac{1}{n} \sum_{i=1}^{n} P(Y_i(\boldsymbol{a}) | H_i = h, G_n)$. These are the conditional average treatment effects of assigning a treatment vector $\boldsymbol{a}$, possibly conditional on h. We consider two sets of assumption and discuss the data needed to estimate each of these models. Let S_i denote some individual level graph parameter that may affect the

outcome. For example, this may be individual degree, as we may suspect that larger degree individuals may have different outcomes than smaller degree individuals. In order to estimate the corresponding parameters of interest, we need to first introduce some assumptions.

Definition 2.1 (Exposure Weak Ignorability)**.** *We say that an exposure assignment is* ***weakly ignorable*** *if the following holds:*

$$Y_i(v) \perp\!\!\!\perp \{V_i = v\} | S_i$$

This simply states that once we condition on the parameters which generate the graph, then the potential outcome for the observed exposure v is independent of the actual exposure given.

Definition 2.2 (Exposure Consistency)**.** *Exposure consistency holds if*

$$V_i = v \implies Y_i = Y_i(v)$$

where $Y_i(v)$ is the potential outcome of individual i for the exposure v.

The last assumption we need is the independence of the graph from the outcome, conditional on the generative model parameters and the exposure.

Definition 2.3 (Conditional Independence of the Graph and Outcome)**.** *We assume that the outcome is conditionally independent of the outcome conditional on the exposure and the graph generative parameters*

$$Y_i \perp\!\!\!\perp G_{ij} | V_i, S_i \text{ for all } i, j$$

Therefore we can derive the causal effects of interest via the following.

$$\begin{aligned}
P(Y_i(\boldsymbol{a})|S_i = s, G_n) &= P(Y_i(v)|S_i = s, G_n) \text{ By the exposure mapping assumption} && (1)\\
&= P(Y_i(v)|V_i = v, S_i = s, G_n) \text{ By Exposure Weak Ignorability} && (2)\\
&= P(Y_i|V_i = v, S_i = s, G_n) \text{ By Consistency} && (3)\\
&= P(Y_i|V_i = v, S_i = s) \text{ Graph Conditional Independence} && (4)
\end{aligned}$$

Therefore, causal effects of interest can be estimated as long as we model properly $P(Y_i|V_i = v, S_i = s)$. In particular, we may be concerned with the mean regression $\mathbb{E}[Y_i|V_i = v, S_i = s] = m(v, s)$

Lastly, if we want to include auxillary covariates H_i, we suppose that each of these properties above are not affected when also conditioning on H_i. In the Supplement in section 6.1 we discuss a structural causal model interpretation.

2.3 Types of Available Data

We consider saturation randomization design experiments. The saturation randomized design partitions the dataset into C clusters of size n_c respectively, then assigns a random τ_k fraction of each of the clusters to the given treatment, for a total number of treated of $n_{a=1} = \sum_{k=1}^{K} \tau_k n_k$. In practice, this design is particularly useful when there are a limited number of treatments one can assign. If τ_k are random as proposed in Hudgens and Halloran [2008], this design generalizes the completely randomized design and cluster randomized design that are common in practice and is known as *randomized saturation randomized design*. Alternative versions where the saturation level τ_k is fixed at each cluster are known as *deterministic saturation randomized design*, and have been explored in Cai et al. [2022]. Under strict budgeting constraints on the number of treatments available, the deterministic version is a more natural object of study.

2.4 Partial Network Data

Though the methodology for causal inference under interference drawn a lot of interest in the recent years. A fundamental challenge is that of actually measuring network data.

We consider datatypes for which a generative model Π_0, can be estimated. For example, latent space models and stochastic block models

In this section, we discuss the different types of network data that is typically available to researchers in applied settings and how they might use this data to run experiments on networked

populations.

The focus of this work is on networks with clustering induced by types. From a modeling perspective,

Suppose that the graph has K communities (this can come from a stochastic block model or some other model in which node memberships introduce homophily). We are interested in estimating $P_{kk'} = P(g_{ij} = 1 | i \in G_k, j \in G_{k'})$, where G_k denotes the set of nodes in community k. Suppose that we have a consistent estimate of this probability. We are supposing that we know communities and they are fixed for now. We might try to estimate them in the first step and then proceed with estimating $\widehat{P}$, given estimated communities $\widehat{G}_1, \ldots, \widehat{G}_K$.

We now detail several examples.

Example 2.1 (Aggregated Relational Data)**.** *Suppose we collect ARD about the K communities. Let Y_{ik} denote the ARD response from node i about community k. Let G'_k denote the nodes in our sample in group k, and let n_k denote the number of nodes in the graph in group k. Then*

$$\widehat{P}_{kk'} = \frac{1}{|G'_k| \times n_{k'}} \sum_{i \in G'_k} Y_{ik} \ .$$

satisfies the assumptions of Theorem **??**.

Example 2.2 (Subgraph sampling)**.** *We sample $m \leq n$ of nodes in the graph randomly, with at least one node from each of the K communities. Let G' be the sub-graph induced by these m nodes. Let G'_k denote the set of sampled nodes in community k, assumed to be positive for each k. Let*

$$\widehat{P}_{kk'} = \frac{1}{|G'_k||G_{k'}|} \sum_{i \in G'_k} \sum_{j \in G'_{k'}} G'_{ij} \ .$$

Then $\widehat{P}_{kk'}$ satisfies the conditions of Theorem **??**.

Example 2.3 (Edges missing)**.** *Suppose that edges are missing according to some distribution. Let G' be the observed graph, and suppose that $P(G'_{ij} = 1 | X_{ij})$ is the probability of observing the edge G'_{ij}, given dyad-level covariates X_{ij} and the edge G_{ij}. Suppose that we have a consistent*

estimator of this probability. Then,

$$\widehat{P}_{kk'} = \frac{1}{|G'_k||G_{k'}|} \sum_{i \in G'_k} \sum_{j \in G'_{k'}} \frac{G'_{ij}}{\widehat{P}(G'_{ij} = 1|X_{ij})} .$$

We now show that by collecting respondent-driven sampling, the researcher can also estimate $\{P_{kk'}\}$. This example is based on Roch and Rohe [2018].

Example 2.4 (Respondent-driven Sampling)**.** *Suppose we sample RDS from a walk on the graph G. Here a walk is defined as in . Then, let $Q_{kk'}$ denote the probability that a node in community k refers a node in community k', and let $\widehat{Q}_{kk'}$ denote the estimate of Q based on the sample. Then, $\widehat{P}_{kk'} = \widehat{Q}_{kk'}/m$ is an unbiased estimate of $P_{kk'}$ that satisfies Theorem* **??***, where $m = \mathbf{1}^T \widehat{Q} \mathbf{1}$.*

2.5 Choice of Exposure Mapping

As we saw in Section 2.2, identification of causal effects of interest relies on properly specifying the regression models $E[Y_i|Z_i = z, V_i = v] = m(z, v)$, or $E[Y_i|Z_i = z, V_i = v, X_i = x] = m(z, v, x)$. This will rely on a choice of the exposure mapping $V_i = f_V(\boldsymbol{a}, \theta_i(G_n))$. The choice of the exposure mapping will be a modelling decision left to the practitioner. Choosing the correct one may be based on domain knowledge.

We next discuss some options for exposure maps. Let $G_i^{(k)}$ denote the graph of connections to individuals of group k. Some simple choices for the regression function may include:

$$\text{Fraction of Treated Neighbors } V_i = (a_i, \frac{\boldsymbol{a}^T G_i}{\boldsymbol{a}^T \mathbf{1}}) =: (a_i, q_i) \tag{5}$$

$$\text{Fraction of Treated Neighbors by Group } V_i = (a_i, \frac{\boldsymbol{a}^T G_i^{(1)}}{\boldsymbol{a}^T \mathbf{1}}, \frac{\boldsymbol{a}^T G_i^{(2)}}{\boldsymbol{a}^T \mathbf{1}}, \ldots, \frac{\boldsymbol{a}^T G_i^{(K)}}{\boldsymbol{a}^T \mathbf{1}}) =: (a_i, q_i^{(1)}, \ldots, q_i^{(K)}) \tag{6}$$

[SWR: highlight the fact that this by group part can be like in group and out group etc.] For each of these, since we do not observe q_i directly, we must approximate it using the

model. Let $\widetilde{V}_i = \mathbb{E}[V_i|\boldsymbol{a};\Pi_0]$ where Π_0 are the model parameters, and D_i is the data available about individual i (i.e. ARD). In the typical fashion of other ARD papers, we will replace the V_i term with $\widetilde{V}_i$ as in Breza et al. [2020, 2023]. Similarly, we may want to replace S_i by $\mathbb{E}[S_i|\boldsymbol{a};\Pi_0]$ though this will be identical to the methods used in previous examples.

The assumptions in the model shift the challenge of estimating causal parameters, to a problem of feature engineering and modelling.

Example 2.5 (Linear Outcome Stochastic Block Model)**.** *We consider first a simple regression model. Suppose that* $S_i = (d_i)$ *and that* $V_i = (a_i, \frac{\boldsymbol{a}^T G_i^{in}}{\mathbf{1}^T G_i^{in}}, \frac{\boldsymbol{a}^T G_i^{out}}{\mathbf{1}^T G_i^{out}})$*. Where in and out refer to the nodes either within the same cluster or across clusters of the stochastic blockmodel.*

We suppose that $Z_i \in \{1, 2, \ldots, K\}$ *refer to the latent parameters of the stochastic block model with across block probability matrix* P*. Suppose also that* $\widehat{P}$ *and* $\widehat{Z}_i$ *refer to the estimated versions of these parameters, and suppose that:*

$$\mathbb{E}[Y_i|V_i = v, S_i = s] = \alpha + \beta_d d_i + \beta_a a_i + \beta_{in} \frac{\boldsymbol{a}^T G_i^{in}}{\mathbf{1}^T G_i^{in}} + \beta_{out} \frac{\boldsymbol{a}^T G_i^{out}}{\mathbf{1}^T G_i^{out}}$$

Then we can estimate the parameters using OLS from the model.

Under this parameterization, the GATE τ_{GATE} *is* $\beta_a + \beta_{in} + \beta_{out}$*.*

2.6 Linear Feature Models

For this paper, we consider models which are linear in some feature f of the node covariates H_i, the graph statistics S_i and the exposure V_i.

$$Y_i = \beta^T f(H_i, S_i, V_i) + \epsilon_i$$

where f

In many cases, after a model is assumed, any parameter of interest can be written as a function

of the parameter vector $\theta = g(\beta)$

[SWR: - Do the example of the response model in Ugander]

2.6.1 Model Based Inference

Given a model Π_0 of the graph data generating mechanism. Then we can let $\widetilde{f}_i = \mathbb{E}[f(H_i, S_i, V_i)|\boldsymbol{a}; \Pi_0]$ denote the average vector of the features under the data generating mechanism Π_0. This lets us write $f_i = \widetilde{f}_i + \eta_i$ where $\mathbb{E}[\eta_i] = 0$.

Then the noise η_i can be passed to the error term, allowing for a consistent ordinary least squares estimator

$$\begin{aligned} \widehat{\beta}_{ols} &= [\widetilde{f}^T \widetilde{f}]^{-1} \widetilde{f}^T \boldsymbol{Y} \\ &= [\widetilde{f}^T \widetilde{f}]^{-1} \widetilde{f}^T \boldsymbol{\eta}\beta + [\widetilde{f}^T \widetilde{f}]^{-1} \widetilde{f}^T \boldsymbol{\epsilon}. \end{aligned} \tag{7}$$

We assume that the noise due to the graph is independent $\boldsymbol{\eta} \perp\!\!\!\perp \boldsymbol{\epsilon}$. And therefore we can express the variance of $\widehat{\beta}_{ols}$ as the sum of two terms.

Lemma 2.4 (Inference for OLS Model)**.** *If $\boldsymbol{\eta} \perp\!\!\!\perp \boldsymbol{\epsilon}$, then the OLS estimate in equation (7) is unbiased with variance:*

$$\text{Var}(\widehat{\beta}) = V_{\Pi_0}(\beta) + V_{\boldsymbol{\epsilon}} \tag{8}$$

Where: $W = [\widetilde{f}^T \widetilde{f}]$, $V_{\eta,\beta} = \text{Var}\,[\eta\beta]$

$$\begin{aligned} V_{\Pi_0}(\beta) &= W^{-1} \widetilde{f}^T V_{\eta,\beta} \widetilde{f} W^{-1} \\ V_{\boldsymbol{\epsilon}} &= W^{-1} \widetilde{f}^T \Sigma \widetilde{f} W^{-1} \end{aligned}$$

Two important features of this model must be considered before we proceed with inference. The

first is the assumption on the noise $\boldsymbol{\epsilon}$. In an ideal scenario, we would have the errors independent, which would correspond to the case that we have modelled all residual correlation of the outcomes with our model. However, it is natural to think there may be additional correlation. The second is whether the noise contribution of ϵ dominates that of the random graph covariates.

We first continue with the negligibility of the additional covariate noise.

Definition 2.5 (Negligible covariate noise)**.** *We say a feature f has **negligible covariate noise** if Let $\theta_n \in \mathbb{R}^{n\times p}$ for some parameter size p. The noise η will be asymptotically negligible if for any sequence of matrices θ_n such that:*

$$\theta_n^T \Sigma_n \theta_n \to C \in \mathbb{R}^{p\times p}$$

Then:

$$\theta_n^T V_{\eta,\beta} \theta_n \to 0 \in \mathbb{R}^{p\times p}$$

This definition simply captures the notion that for any estimate which has an asymptotic variance C, then the contribution from the noise of η will not be included.

Example 2.6 (Degree density in a graphon model)**.** *Consider a random graph model based on a graphon model*[1] *i.e. there is some latent trait Z_i for each node, and for some function $\nu(Z_i, Z_j)$ then $p_{ij} := P(G_{ij} = 1|Z_i, Z_j) = \nu(Z_i, Z_j)$. Then $\mathrm{Var}(\frac{d_i}{n}) = \frac{1}{n^2}\sum_{j=1}^{n} p_{ij}(1-p_{ij}) = O\left(n^{-1}\right)$. Secondly $\mathrm{Cov}(\frac{d_i}{n}, \frac{d_j}{n}) = p_{ij}(1-p_{ij})/n^2$.*

So we see that features like the degree density in a graphon model are ones which have a Covariance matrix with entries decreasing on order at least $O(n^{-1})$.

Lemma 2.6 (Sufficient condition for negligible covariate noise)**.** *Let $||C||$ denote the spectral norm of C. If*

- $\lambda_{min}(\Sigma) > 0$

[1] It is not necessary here to consider if Z_i are fixed themselves or random

- *Given the population* β, $||V_{\eta,\beta}|| = o(1)$.

Then the covariate noise is negligible.

Proof. Since by definition, $\theta_n^T V_{\eta,\beta} \theta_n$ is positive semidefinite, then a sufficient condition for convergence to the 0 matrix is to show that the largest eigenvalue (the spectral norm) converges to 0.

Firstly the minimum eigenvalue of Σ enforces constraints on θ_n

$$\begin{aligned} ||\theta_n^T \Sigma \theta_n|| &\to ||C|| \\ \implies \lim_{n\to\infty} ||\theta_n^T \theta_n|| \lambda_{\min}(\Sigma) &\leq ||C|| \\ \implies ||\theta_n^T \theta_n|| &= O(1) \end{aligned}$$

This lets us bound the eigenvalue of $||V_{\eta,\beta}|| = o(1)$

$$||\theta_n^T V_{\eta,\beta} \theta_n|| \leq ||\theta_n^T \theta_n|| ||V_{\eta,\beta}|| = o(1)$$

and the proof is complete. □

This result can immediately be applied to the examples like the edge density, since the spectral norm is bounded by the infinity norm.

[SWR: TODO: Comment on the connection to Arun, Matt, Tyler and Vydourie's paper regarding central limit theorems. This will be essential for actually doing inference here if we want to use a CLT]

2.6.2 Approximation to the population graphon

[SWR: TODO:] Connection with Gao et al. [2015] so that if $f^* = \mathbb{E}[f(H_i, S_i, V_i)|\mathcal{G}]$ where $\mathcal{G}$ is the population graphon and $\widetilde{f} = \mathbb{E}[f(H_i, S_i, V_i)|\Pi_0^n]$ is the approximation using a stochastic blockmodel at a particular sample size n, then we control the approximation error of $|\widetilde{f} - f^*|$.

2.7 Optimal Design in Linear Feature Models

We consider optimal design for linear combinations of the parameter vector. $\theta = \phi^T \beta$. Since we assume that the covariate noise η is negligible, finding the optimal design, will amount to minimizing the variance $V_{\boldsymbol{\epsilon}}$

We will consider settings where wee have a set of treatment cluster labels $c_i \in \{1, \ldots, C\}$. These clusters need not be the same as the clusters in a stochastic-blockmodel. Depending on the draw, some treatments may fall into different blocks in the cluster, affecting the vector of draws $\widetilde{f}$.

We consider the following procedure for an optimal design matrix. Let $\boldsymbol{\tau} \in [0,1]^C$ denote a set of treatment saturation levels for the saturation randomized design.

Algorithm 1: Optimal design for saturation randomization design.

Input: $\mathcal{T}$ (grid of τ values), $f(H_i, S_i, V_i)$: feature vector, Π_0: Model, ϕ linear combination of coefficients, J: number of re-samples from the sample, M: number of simulated graph samples.

for $\boldsymbol{\tau} \in \mathcal{T}$ **do**

- Sample $\boldsymbol{a}$ via saturated randomized design. **for** $m \in \{1, 2, \ldots, M\}$ **do**
  - Sample $G^{(m)} \sim \Pi_0$
  - Compute $H_i^{(m)}, V_i^{(m)}, S_i^{(m)}$ from $G^{(m)}$
  - Compute $f^{(m)} = f(H_i^{(m)}, V_i^{(m)}, S_i^{(m)})$
- Compute $\widetilde{f} = \frac{1}{M} \sum_{m=1}^{M} f^{(m)}$
- Compute the variance of the draw $V_j(\boldsymbol{\tau}) = \phi^T [\widetilde{f}^T \widetilde{f}]^{-1} [\widetilde{f}^T \widetilde{f}]^{-1} \phi$

return $\boldsymbol{\tau}^* = \arg\min_{\boldsymbol{\tau} \in \mathcal{T}} \frac{1}{J} \sum_{j=1}^{J} V_j(\boldsymbol{\tau})$

Though in principle, we can do such a grid search. This problem is usually non-convex and therefore computationally challenging.

Example 2.7 (Linear Treatment Fraction)**.** *Suppose we have a conditional model:*

$$Y_i = \alpha + \delta \sum_{j=1}^{n} G_{ij} a_j + \epsilon_i.$$

Then δ is a peer effect parameter, and there is no direct effect in this model. We will derive the

optimal weights of τ for estimating the parameter β. Suppose that the data are generated according to a stochastic-blockmodel with parameters P, Z with n_k denoting the number of individuals in group k. Additionally assume that $\text{Var}[\boldsymbol{\epsilon}] = \sigma I$ *Let* $\widetilde{f}_i = \mathbb{E}[f_i(S_i, V_i)|\tau; \Pi_0] = (1, \sum_{k'=1}^{K} P_{k_i,k'} n_{k'} \tau_{k'})$.

Therefore letting M_i be a matrix such that:

$$M_i = (n_1 P_{k_i,1}, \ldots, n_K P_{k_i,K})$$

Then

$$[\widetilde{f}^T \widetilde{f}] = \begin{pmatrix} \mathbf{1}^T\mathbf{1} & \mathbf{1}^T M\boldsymbol{\tau} \\ \mathbf{1}^T M\boldsymbol{\tau} & \boldsymbol{\tau}^T M^T M\boldsymbol{\tau} \end{pmatrix}$$

$$\implies \text{Var}[\widehat{\delta}] = \frac{n}{n\boldsymbol{\tau}^T M^T M\boldsymbol{\tau} - \boldsymbol{\tau}^T M^T \mathbf{1}\mathbf{1}^T M\boldsymbol{\tau}}$$

Therefore minimizing this quantity is equivalent to solving:

$$\min_{\boldsymbol{\tau}} \boldsymbol{\tau}^T \left(M^T \mathbf{1}\mathbf{1}^T M - nM^T M\right) \boldsymbol{\tau}$$

Since $M^T\mathbf{1}\mathbf{1}^T M - nM^T M$ is the difference between two PSD matrices, this is a non-convex problem and may therefore be quite challenging to solve.

2.7.1 Algorithms for Optimal Design

Here we will introduce gradient descent and greedy algorithms for optimal design.

3 Simulations

3.1 Estimating the Model Using ARD

3.1.1 Fractional Treatment model

In our first example, we consider a parameter denoting the differences in the rate of change of the neighbouring treatments.

$$\theta = \int_0^1 \left(\frac{d}{dp} \mathbb{E}_{a_j \sim \text{Bern}(p)} \left[\frac{1}{n} \sum_{i=1}^{n} Y_i(a_i = 1, \boldsymbol{a}_j) \right] - \frac{d}{dp} \mathbb{E}_{a_j \sim \text{Bern}(p)} \left[\frac{1}{n} \sum_{i=1}^{n} Y_i(a_i = 0, \boldsymbol{a}_j) \right] \right) dp$$

In this model, we assume that there are two parameters γ_0 and γ_1 reflecting the average peer effect from the fraction of treated neighbours when treated and not treated respectively.

$$Y_i(\boldsymbol{a}) = \alpha + \beta a_i + \eta \frac{d_i}{\overline{d}} + \gamma_0(1 - a_i) \frac{\sum_{j \in [n]} G_{ij} a_j}{d_i} + \gamma_1(a_i) \frac{\sum_{j \in [n]} G_{ij} a_j}{d_i} + \sigma \epsilon_i. \tag{9}$$

In this example, we are interested in the difference of the peer effect parameters $\gamma_1 - \gamma_0$.

[SWR: nonparametrically define what this parameter should be] We simulate a set of stochastic block models with parameters

$$\begin{aligned} n &\in \{100, 316, 1000, 3162, 10000\} \\ K &= \lceil \sqrt{n} \rceil \\ P_{ii} &= 0.2 \\ P_{ij} &= 0.02 \quad i \neq j \end{aligned}$$

We assume that the latent clusters are the observed traits, but in Breza et al. [2023] it is shown that this is not entirely necessary for consistency of the blockmodel.

3.1.2 Add Arun's Example

3.1.3 Modelling the GATE via a regression approach

In Ugander and Yin [2020] the authors emphasize that much of the work for estimating treatment effects under is done using unrealistic models. The authors consider a generative model for which they claim is more realistic, due to it's covariate variation (H) across the network as well as the scale variation by degree.

$$
\begin{aligned}
Y_i(\mathbf{0}) &= \frac{d_i}{\overline{d}}\left(\alpha + bh_i + \sigma\epsilon_i\right) \\
Y_i(\boldsymbol{a}) &= Y_i(\mathbf{0})(1 + \delta a_i + \gamma\frac{\sum_{j\in[n]} G_{ij}a_j}{d_i})
\end{aligned}
$$

Where ϵ_i is some independent noise ($N(0,1)$), and h_i is some covariate which varies in the network. The (expected)-GATE in this model is $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i(\mathbf{0})](\delta+\gamma)$, whereas the GATE for a single draw of the network is $\frac{1}{n}\sum_{i=1}^{n}Y_i(\mathbf{0})(\delta+\gamma)$. We will consider the deviation of the estimate to each of these quantities (either the average parameter, or the)

We of course can write the outcome as a linear model

$$
\begin{aligned}
Y_i(\boldsymbol{a}) = \alpha\frac{d_i}{\overline{d}} + b\frac{d_i}{\overline{d}}h_i + \alpha\delta a_i\frac{d_i}{\overline{d}} + b\delta a_i\frac{d_i}{\overline{d}}h_i + \alpha\gamma\frac{\sum_{j\in[n]} G_{ij}a_j}{d_i}\frac{d_i}{\overline{d}} + b\gamma\frac{\sum_{j\in[n]} G_{ij}a_j}{d_i}\frac{d_i}{\overline{d}}h_i \\
+ \sigma\left(\frac{d_i}{\overline{d}}\left[1 + \delta a_i + \gamma\frac{\sum_{j\in[n]} G_{ij}a_j}{d_i}\right]\right)\epsilon_i
\end{aligned}
$$

Hence this can be expressed as a heteroskedastic error linear model.

In our case, we take a more explicit regression modelling approach to this problem. Though graph cluster randomization approaches will guarantee a method of estimating the GATE. If we are willing to use a modelling assumption, we can more effectively use the whole data, giving better

estimates.

We will use 4 methods. We use first a simple difference in means estimator between treated and untreated individuals. Though this is often still effective in cluster randomizations, such a method will inherit some bias. A further study of cluster randomization in the context of bipartite graphs is seen in Brennan et al. [2022] using a difference in means estimator.

We assume that the trait H_i is observed in the network and can be adjusted for. We generate H_i according to

$$H_i = \begin{cases} Z_i & \text{if } i \text{ is even} \\ Z_i + 1/2 & \text{if } i \text{ is odd} \end{cases}$$

Hence H_i is a covariate which is correlated within clusters in the graph.

We see that the full data regression model performs the best, though this requires knowledge of the full network data. The ARD version however, performs nearly as well.

3.2 Optimal Design Under Model Assumptions

[SWR: todo: This is a nonconvex problem. We need a good way of computing this as opposed to grid search because the number of clusters will grow exponentially. Will do a gradient descent or a greedy algorithm.]

4 Application

4.1 Semi-synthetic data

Pick a network. Cluster it. Generate ARD from the network clusters. Fit a blockmodel. Then estimate the regression model.

We use the dataset from Banerjee et al. [2013a] ,Banerjee et al. [2013b] and estimate a stochastic-blockmodel.

We find the number of clusters is

5 Conclusions

We introduce a framework for identifying all causal effects under interference by modelling the outcome as a function of exposure mapping and graph statistics. We highlight conditions for where a model can be learned by these data under the stochastic-block-model.

We allow for a model-based approach which will let a user identify the causal effects of interest.

We consider methods for optimal design.

We also consider the

We lastly apply this method to a semi-synthetic data

References

P. M. Aronow and C. Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics*, 11(4):1912–1947, 2017. ISSN 1932-6157. doi: 10.1214/16-AOAS1005. URL https://projecteuclid.org/euclid.aoas/1514430272.

S. Athey, D. Eckles, and G. W. Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

P. Bajari, B. Burdick, G. W. Imbens, L. Masoero, J. McQueen, T. S. Richardson, and I. M. Rosen. Experimental design in marketplaces. *Statistical Science*, 1(1):1–19, 2023.

A. Banerjee, A. Chandrasekhar, E. Duflo, and M. Jackson. Diffusion of microfinance. *Science*, 341(6144):1–7, 2013a.

A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. The Diffusion of Microfinance, 2013b. URL https://doi.org/10.7910/DVN/U3BIHX.

G. W. Basse, A. Feller, and P. Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019.

J. Brennan, V. Mirrokni, and J. Pouget-Abadie. Cluster randomized designs for one-sided bipartite experiments. *arXiv preprint arXiv:2210.16415*, 2022.

E. Breza, A. G. Chandrasekhar, T. H. McCormick, and M. Pan. Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 2020.

E. Breza, A. G. Chandrasekhar, S. Lubold, T. H. McCormick, and M. Pan. Consistently estimating network statistics using aggregated relational data. *Proceedings of the National Academy of Sciences*, 120(21):e2207185120, 2023.

C. Cai, J. Pouget-Abadie, and E. M. Airoldi. Optimizing randomized and deterministic saturation designs under interference. *arXiv preprint arXiv:2203.09682*, 2022.

O. Candogan, C. Chen, and R. Niazadeh. Near-optimal experimental design for networks: Independent block randomization. *Chicago Booth Research Paper*, (21-17), 2021.

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

M. Cortez, M. Eichhorn, and C. Yu. Staggered rollout designs enable causal inference under interference without network knowledge. In *Advances in Neural Information Processing Systems*, 2022a.

M. Cortez, M. Eichhorn, and C. L. Yu. Exploiting neighborhood interference with low order interactions under unit randomized design. *arXiv preprint arXiv:2208.05553*, 2022b.

D. R. Cox. Planning of experiments. 1958.

I. Cristali and V. Veitch. Using embeddings for causal estimation of peer influence in social networks. *arXiv preprint arXiv:2205.08033*, 2022.

N. Doudchenko, M. Zhang, E. Drynkin, E. Airoldi, V. Mirrokni, and J. Pouget-Abadie. Causal inference with bipartite designs. *arXiv preprint arXiv:2010.02108*, 2020.

C. Gao, Y. Lu, and H. H. Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, pages 2624–2652, 2015.

V. Ha-Thuc, A. Dutta, R. Mao, M. Wood, and Y. Liu. A counterfactual framework for seller-side a/b testing on marketplaces. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2288–2296, 2020.

M. E. Halloran and C. J. Struchiner. Causal inference in infectious diseases. *Epidemiology*, pages 142–151, 1995.

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

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

L. Jing, C. Qu, H. Yu, T. Wang, and Y. Cui. Estimating the sizes of populations at high risk for HIV: a comparison study. *PloS ONE*, 9(4):e95601, 2014.

R. Johari, H. Li, I. Liskovich, and G. Y. Weintraub. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 68(10):7069–7089, 2022.

P. D. Killworth, C. McCarty, H. R. Bernard, G. A. Shelley, and E. C. Johnsen. Estimation of seroprevalence, rape, and homelessness in the united states using a social network approach. *Evaluation Review*, 22(2):289–308, 1998.

A. Malani, P. Holtzman, K. Imai, C. Kinnan, M. Miller, S. Swaminathan, A. Voena, B. Woda, and G. Conti. Effect of health insurance in india: a randomized controlled trial. Technical report, National Bureau of Economic Research, 2021.

E. McFowland III and C. R. Shalizi. Estimating causal peer influence in homophilous social networks by inferring latent locations. *Journal of the American Statistical Association*, 118(541): 707–718, 2023.

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

J. Pouget-Abadie, V. Mirrokni, D. C. Parkes, and E. M. Airoldi. Optimizing cluster-based randomized experiments under monotonicity. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2090–2099, 2018.

J. Pouget-Abadie, G. Saint-Jacques, M. Saveski, W. Duan, S. Ghosh, Y. Xu, and E. M. Airoldi. Testing for arbitrary interference on experimentation platforms. *Biometrika*, 106(4):929–940, 2019.

D. Puelz, G. Basse, A. Feller, and P. Toulis. A graph-theoretic approach to randomization tests of causal effects under general interference. *arXiv preprint arXiv:1910.10862*, 2019.

S. Roch and K. Rohe. Generalized least squares can overcome the critical threshold in respondent-driven sampling. *Proceedings of the National Academy of Sciences*, 115(41):10299–10304, 2018.

D. B. Rubin. Formal mode of statistical inference for causal effects. *Journal of statistical planning and inference*, 25(3):279–292, 1990.

M. Saveski, J. Pouget-Abadie, G. Saint-Jacques, W. Duan, S. Ghosh, Y. Xu, and E. M. Airoldi. Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1027–1035, 2017.

F. Sävje, P. Aronow, and M. Hudgens. Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2):673, 2021.

O. Scutelniciuc. Network scale-up method experiences: Republic of kazakhstan. *Consultation on estimating population sizes through household surveys: Successes and challenges (New York, NY)*, 2012.

E. J. T. Tchetgen and T. J. VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

J. Ugander and H. Yin. Randomized Graph Cluster Randomization. 9 2020. doi: 10.48550/arxiv.2009.02297. URL https://arxiv.org/abs/2009.02297v1.

J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: Network exposure to multiple universes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Part F128815:329–337, 8 2013. doi: 10.1145/2487575.2487695.

V. Veitch, Y. Wang, and D. M. Blei. Using Embeddings to Correct for Unobserved Confounding in Networks. *Advances in Neural Information Processing Systems*, 32, 2 2019. URL https://arxiv.org/abs/1902.04114v2.

C. L. Yu, E. M. Airoldi, C. Borgs, and J. T. Chayes. Estimating the total treatment effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*, 119(44):e2208975119, 2022.

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

6 Supplement

6.1 Structural Causal Model Intepretation

As an alternative view of the problem, and relating it to other work in the interference literature. We can consider the nonparametric structural equation model such as those in Ogburn et al.

[2017], Cristali and Veitch [2022]. Assuming for now that covariates are not included.

$$
\begin{aligned}
Z_i &= f_Z[\epsilon_{Z_i}] \\
G_{ij} &= f_G[Z_i, Z_j, \epsilon_{ij}] \\
A_i &= f_A[Z_i, \epsilon_{A_i}] \\
V_i &= f_V(\boldsymbol{A}, \theta_i(G_n)) \\
S_i &= f_S(\phi_i(G_n)) \\
Y_i &= f_Y[V_i, S_i, \epsilon_{Y_i}]
\end{aligned}
$$

where f_W are the nonparamteric structural equations and ϵ_W are exogenous noise variables. The graph in this case is just generated by a simple graphon model based on node level parameters Z_i.

This assumption relies on assigning treatment, conditional on the network generating parameters Z_i. We can equivalently represent this This structural equation model, can be interpreted using the causal DAG (directed acyclic graph) in Figure 2.

Assumption set 1 can be described using the following DAG. This model assumes that the differences in the observed outcomes are going to be determined specifically by their local graph statistics S_i and the exposure treatment given V_i. Though it may be natural to think about clustering the graph based on the Z_i parameters, for example, in a stochastic block model, this model shows how this is not in fact necessary to do so. Though we later discuss the optimal design approach for certain parameters of interest.

A concern one may have is whether there is additional variation in the outcome which is unaccounted for by the local graph statistics alone. In this case, there may be an additional arrow from $Z_i \to Y_i$ which incorporates this assumption. This draws similarities more closely to what is assumed in Veitch et al. [2019] and Ogburn et al. [2017], in that the generative parameters of the network confound the outcome itself.

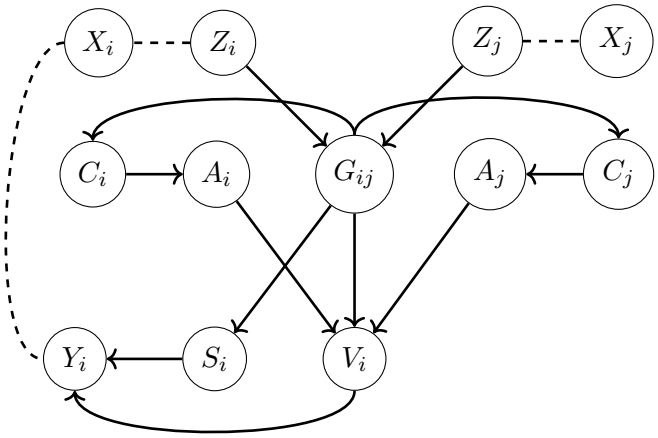

Figure 1: Interference DAG for assumption set 1. The dashed lines refer to unknown edges in the network, so long as a DAG is still encoded in the model.

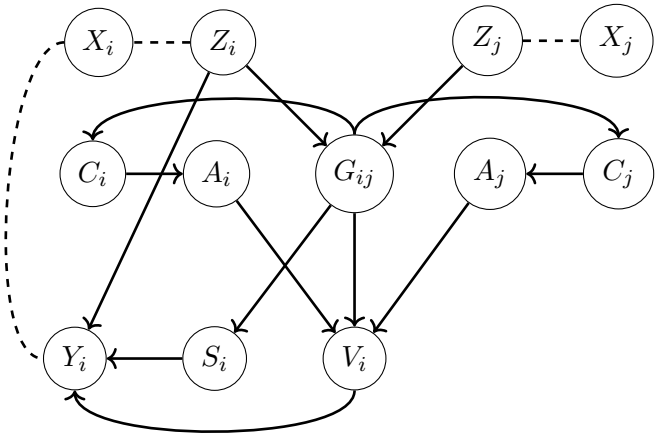

Figure 2: Interference DAG for assumption set 2. In this case, we simply add another line from the latent model parameter Z_i to the outcome Y_i.