Community Informed Experimental Design

Heather Mathews and Alexander Volfovsky
Duke University

Abstract

Network information has become a common feature of many modern experiments. From vaccine efficacy studies to marketing for product adoption, stakeholders aim to estimate global treatment effects — what happens if everyone in a network is treated versus if no one is treated. Because individual outcomes are potentially influenced by the treatments or behaviors of others in the network, experimental designs must condition on the underlying network. Social networks frequently exhibit homophilous community structure, meaning that individuals within observed or latent communities are more similar to each. This observation motivates the development of community aware experimental design. This design recognizes that information between individuals likely flows along within community edges rather than across community edges. We demonstrate that this design reduces the bias of a simple difference in means estimator, even when the community structure of the graph needs to be estimated. Further, we show that as the community detection problem gets more difficult or if the community structure does not affect the causal question, the proposed design maintains its performance.

Keywords: networks, causal inference, community detection, A/B testing

1 INTRODUCTION

Across industries, experiments provide important evaluation of new hypotheses and potential future directions (Kohavi et al., 2013; Xu, Chen, Fernandez, Sinno, & Bhasin, 2015). Whether it is testing the efficacy of a new drug or evaluating a change in the output of an algorithm, the quantity of interest must guide the experimental design. The classical experiment considers individuals in a population and a set of potential outcomes for each individual — these potential outcomes are indexed by treatments that might affect them. As such, an experiment between two alternative treatments (call them a treatment arm and a control arm) will randomly assign individuals to each arm, with the goal of quantifying a contrast between the potential outcomes of an individual had they been assigned to treatment versus had they been assigned to control. This is operationalized by first specifying several simplifying assumptions (discussed in detail below) about the causal process, specifying an estimand or quantity of interest and an estimator. A general estimand that we will consider in this paper is the *global average treatment effect (GATE)*:

$$\tau = \frac{1}{N} \sum_{i=1}^{N} \Big[Y_i(\mathbf{Z} = \mathbf{1}) - Y_i(\mathbf{Z} = \mathbf{0}) \Big], \tag{1}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_N)$ is the vector of treatment assignments for all individuals in the sample and $Y_i(\mathbf{Z})$ is the potential outcome for individual i under the assignment vector $\mathbf{Z}$. This quantity represents the difference in outcome had everyone in the sample been treated versus not treated at the same time point. It is fairly clear that without additional assumptions this quantity is not estimable from data since assigning everyone to one of the arms of the experiment would not allow us to estimate anything. A further complication is that the potential outcome is indexed by the full vector of treatments — this suggests that an individual's outcome might be affected by a treatment that is assigned to someone else, a notion referred to as *interference*.

The GATE is of particular interest in settings where individuals are *networked* since interference is likely. As a result, for GATE estimation, reasonable assumptions need to be made about how interference manifests in a network. The typical assumption is that of neighborhood interference: an individual is assumed to be impacted by his or her immediate neighbors and no one else (formally defined in the next section). While this assumption accounts for the basic structure of the network, it fails to account for other potentially influential features. For example, homophily, the notion that individuals that are similar to one another tend to connect and act in similar ways, has long been demonstrated to be a driving force in many social networks (Igarashi, Takai, & Yoshida, 2005; Shrum, Cheek Jr., & Hunter, 1988).

As such, it is natural to believe that it is not just connections between individuals that drive interference patterns but rather connections between similarly behaving individuals that drive interference patterns. Homophilous behavior frequently leads to the formation of latent community structure in networks (Fortunato, 2010; Girvan & Newman, 2002; McPherson, Smith-Lovin, & Cook, 2001). As such, a more realistic mechanism for interference is one that accounts for the fact that same-community connections are more likely to interfere with an individual's outcome than cross-community connections. We note that homophily will only affect interference patterns and not induce other cross-unit dependence in the outcomes which is known to lead to identifiability issues (Ogburn, Sofrygin, Diaz, & Van Der Laan, 2017; Shalizi & Thomas, 2011).

Below we discuss several common simplifying assumptions and the designs they motivate. We then formalize our own assumption that is motivated by the literature on community influence in social networks and develop an experimental design that can lead to high quality, yet easily interpretable estimates of the GATE.

To fix notation, throughout, we will represent a network in terms of its adjacency matrix $\boldsymbol{A}$: a binary $N \times N$ matrix where $A_{ij} = 1$ if individual i is connected to individual j. A network can be directed or undirected $(A_{ij} = A_{ji})$

1.1 Classical Assumptions and Designs

The most common assumption in causal inference, termed the Stable Unit Treatment Value Assumption (SUTVA) or individualistic treatment assignment, (Manski, 1995; Rubin, 1990), states that the treatment assigned to an individual can only affect their potential outcome. Under SUTVA the GATE reduces to the standard average treatment effect:

$$\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N}\Big[Y_i(Z_i = 1) - Y_i(Z_i = 0)\Big]. \tag{2}$$

This estimand motivates a very simple experimental design and estimation procedure: assign individuals to treatment independently, and estimate τ_{ATE} using the naive difference in means estimator,

$$\hat{\tau} = \frac{1}{N_T}\sum_{i=1}^{N} Y_i Z_i - \frac{1}{N_C}\sum_{i=1}^{N} Y_i(1 - Z_i) \tag{3}$$

where, for notational convenience, N_T and N_C are the fixed number of treated and control individuals respectively.

Since we know that SUTVA is likely not a viable assumption in networked populations, the literature has proposed several alternative assumptions that limit the influence of treatments across the network according to some notion of distance between individuals (Aronow & Samii, 2017; Jagadeesan, Pillai, & Volfovsky, 2020; Sävje, Aronow, & Hudgens, 2021; Sussman & Airoldi, 2017; Toulis & Kao, 2013).

In particular, these **Network or Neighborhood SUTVA** type assumptions can be written as follows: for treatment allocation vectors $\boldsymbol{Z}, \boldsymbol{Z}'$ with $g_i(\boldsymbol{Z}) = g_i(\boldsymbol{Z}')$ we have $Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}')$ where the function $g_i(z)$ extracts the components of the vector z that relate to the individuals in the network who can interfere with unit i. For example, the Neighborhood Interference assumption of Sussman and Airoldi (2017) and Awan et al. (2020) takes $g_i(\boldsymbol{Z}) = (Z_i, \{Z_j : A_{ij} = 1\})$ or the simplified neighborhood exposure assumption takes $g_i(\boldsymbol{Z}) = (Z_i, \sum_{\{j:A_{ij}=1\}} Z_j > 0)$. Note that this class of assumptions has been developed in order to study estimands that mimic τ_{ATE} rather than the more general GATE, and there is now a separate literature on different estimation techniques that are agnostic to the experimental design (Aronow & Samii, 2017; Sävje, 2021).

Some designs focus on only estimating the direct effect of treatment which results in the goal of minimizing the impact of interference in the design. That is, consider how the GATE can be decomposed into a direct effect of treatment on an individual and an effect of neighborhood influence. However, in direct effect estimation, only the former is desired. When this is the goal, the design proposed below is sub-optimal (unless direct and interference effect can easily be decoupled which is generally not true).

To obtain high quality estimates of the GATE, network aware experimental design is crucial. A natural idea is to identify subsets of the network that are far enough apart but that represent the interference pattern of the full network. Assigning such subsets to treatment or control provides a particular view of the GATE. This approach was formalized by Eckles, Karrer, and Ugander (2016); Ugander, Karrer, Backstrom, and Kleinberg (2013) as graph cluster randomization (GCR) .

Formally, the justification for these methods rests in the notion that within a large network there are small sub-networks that are separated from each other by sufficient network distance so as to not influence the treatment effect of individuals across them. Following this, graph cluster randomization approaches partition the network into clusters using some type of clustering method (e.g. epsilon-net and one-hop max in Ugander and Yin (2020)). Letting $C(i)$ indicate the cluster that node i is assigned to, each cluster is assigned to treatment with probability q. These methods lead to a provable reduction in bias and variance when compared to classic randomization schemes. However, these experimental designs fail to leverage additional network information.

1.2 Our approach: Community Interference Assumptions and Design

The previously used network interference assumptions effectively treat every edge or connection in a network equivalently. That is, the presence of an edge implies that interference must occur. However, *not all connections are created equal* (Aukett, Ritchie, & Mill, 1988; Bail et al., 2018; Chamberlain, Kasair, & Rotheram-Fuller, 2007; Staber, 1993), which should be reflected in the design. Specifically, we consider the setting where network connections are formed with higher probability between similar individuals and so while connections between dissimilar individuals are possible, they are going to be deemed less strong and less likely to lead to interference.

We will formalize this by considering networks formed according to community structure: each individual i belongs to one of K communities and edges within communities are more likely than those across communities.

It is natural to represent community membership as a K dimensional binary vector $\boldsymbol{U}_i$ where $U_{ik} = 1$ if individual i belongs to community k and $\sum_{l=1}^{K} U_{il} = 1$. This framing suggests the definition of the **community network interference** function as $Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}')$ if $g_i(\boldsymbol{Z}) = g_i(\boldsymbol{Z}')$ where

$$g_i(\boldsymbol{Z}) = (Z_i, \{Z_j : A_{ij} = 1 \text{ and } \boldsymbol{U}_i = \boldsymbol{U}_j\}).$$

As such, interference occurs only within communities among individuals who share connections. To further motivate this assumption, in Section 5, we consider data from an anti-bullying experiment performed in middle schools where network data is collected (Paluck, Shepherd, & Aronow, 2016, 2020). This experiment is set up with an exposure mapping that matches our within community interference assumption since individuals are blocked by gender and grade. As mentioned, the notion that individuals within a community act similarly and are influenced by those in their own community is supported throughout the community detection literature (Aukett et al., 1988; Chamberlain et al., 2007; Staber, 1993). However, typically these community labels are latent and the mechanisms driving them are also unobserved.

Designing a community aware mechanism for treatment assignment is thus crucial to capture and balance community level differences while also accurately capturing the impact of interference. In the following sections, we show how leveraging community structure, either known or estimated, improves experimental design, leading to better estimation of the GATE. As the community detection problem becomes more challenging, our methods reduce to that of standard graph cluster randomization.

Our proposed approach is as follows: identify the communities that exist in the graph and perform randomized graph cluster randomization within these communities. There are thus two crucial steps in this design that can be notationally and conceptually confusing: (1) identifying communities that inform which units share behavior and (2) identifying design-relevant clusters which assist in randomization. Both of these steps technically use forms of clustering algorithms. However, an important distinction between these is the existence of ground truth. When we speak of community labels we are postulating that there exist true community labels that we must try to estimate — these communities inform along which connections interference will happen. As such, the quality of estimation of these communities is important and so we discuss nodes with incorrectly estimated labels if they are estimated to be in the wrong community. On the other hand, the clusters that are used to operationalize graph cluster randomization are purely technical artifacts and so there is no notion of ground truth.

1.3 Paper Outline

The rest of this manuscript is structured as follows: In the next section, we provide a detailed outline of our proposed procedure. In Section 3 we characterize the bias of the naive difference in means estimator under three designs: independent assignment, GCR, and our proposed community informed design (CID). We demonstrate both analytically and empirically why we expect our procedure to lead to a reduction in bias under community driven interference. Section 4 describes the behavior of our method when the first stage of the community detection problem becomes more difficult. Section 5 showcases the empirical performance of our approach on several simulated datasets (including under misspecification of the interference assumption). We also implement our design using real network data.

2 METHODS

We develop a procedure that yields a high quality design for conducting experiments in networks with community structure. To illustrate this, Figure 1 shows examples of clusters that will be assigned to treatment or control under community informed design (left) versus standard GCR (right) on a network where nodes belong to one of two communities, denoted by the shape of the node. Recall that for networks with community interference, only within community ties are meaningful for GATE estimation. As a result, clusters of all same community nodes are desirable. Red nodes indicate that all nodes in a cluster belong to community 1 while blue indicates that all nodes in a cluster belong to community 2. Purple nodes indicate that a cluster contains nodes from both communities.

Notice that while standard GCR can produce clusters that fall into a single community, it also yields clusters that cross community boundaries. When these multi community clusters exist, individuals are less likely to share treatment with their within community neighbors. The consequences of this are explored in later sections. In contrast, community informed design guarantees that all clusters only consist of same community nodes thus capturing the true underlying interference structure. This figure also illustrates the potential of our approach even if the community structure is not informative of interference as the clusters identify potential GCR clusters.

To implement our method, community informed design can be divided into three main parts 1) estimate community labels 2) find clusters of closely connected individuals within community level sub-graphs and 3) perform randomization at the cluster level for treatment allocation. This process is summarized in Algorithm 1, visually demonstrated in Figure 2, and each step is discussed in detail below.

Estimating Community Labels

First consider the algorithm choice for community label recovery, $f(\boldsymbol{A}, K)$. An abundance of algorithms exist in the literature for this task (Abbe, 2017; Bhattacharyya & Bickel, 2014; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Bruna & Li, 2017; Mathews, Mayya, Volfovsky, & Reeves, 2019; Rohe, Chatterjee, Yu, et al., 2011), any of which could be used here. However given the recent theoretical and empirical results in information theory and statistics, we implement spectral methods (Krzakala et al., 2013; Mayya & Reeves, 2019; Reeves, Mayya, & Volfovsky, 2019). At a high level, spectral based methods produce an embedding of the observed network in a lower dimensional space using an eigen-decomposition of the adjacency matrix. There are several (nearly equivalent) approaches for doing this. We implement one from

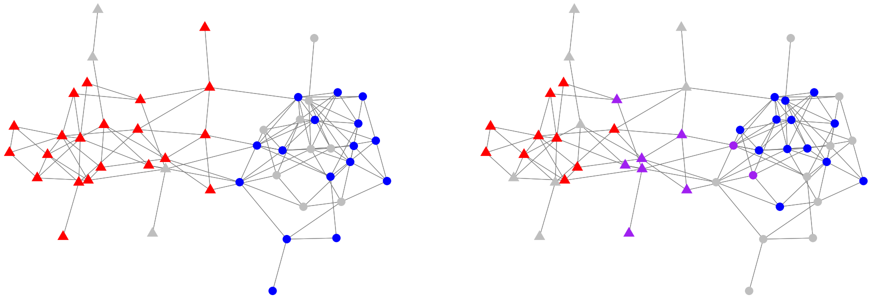

Fig. 1 This figure shows two versions of the same network where each has a clustering design implemented. The left plot shows example cluster assignments under community informed design while the right shows standard graph cluster randomization. This network is generated from a 2 community stochastic blockmodel. The true community labels are indicated by the node shape (triangles for community 1, circles for community 2). Red nodes indicate clusters where all nodes in a cluster are in community 1 whereas blue nodes indicate all nodes in a cluster are in community 2. Purple nodes indicate clusters that have nodes from both community 1 and 2. Not all clusters are colored; hence why some nodes are gray

Algorithm 1 Community Informed Design

```
INPUT:
Adjacency matrix $\boldsymbol{A}$,
Number of communities $K$,
Community detection algorithm $f(\cdot,\cdot)$,
Clustering algorithm $h(\cdot,\cdot)$,
Treatment assignment $s(\cdot)$
OUTPUT: $Z$
Estimate community labels, $\hat{\boldsymbol{U}}$, using a community detection algorithm of choice, denoted $f(\boldsymbol{A}, K)$. Specification of $K$ a priori may not be necessary for all algorithms
Create a community induced sub-graph for each community based off of $\hat{\boldsymbol{U}}$ that contains only within community edges
Run a clustering algorithm of choice, denoted $h(A, \hat{\boldsymbol{U}})$, on each sub-graph independently such that each sub-graph is partitioned into clusters, $C_{\hat{\boldsymbol{U}}}$. This ensures that only nodes belonging to the same community can be in the same cluster.
Assign treatment according to some design, $s(C_{\hat{\boldsymbol{U}}})$. This could denote a pairing design, randomization designs with covariate information, etc.
```

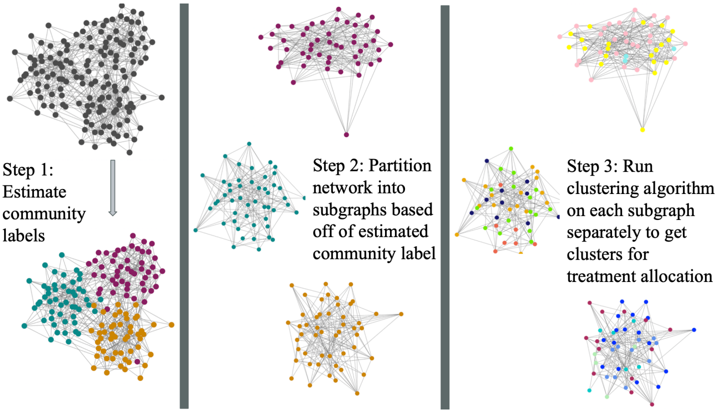
Step 1: Estimate community labels
Step 2: Partition network into subgraphs based off of estimated community label
Step 3: Run clustering algorithm on each subgraph separately to get clusters for treatment allocation

Fig. 2 This figure is a visualization representation for steps 1-3 in Algorithm 1

Rohe et al. (2011) that considers the normalized graph Laplacian, a summary of the network that better captures clustering behavior, especially for sparse networks. This is defined as

$$\boldsymbol{L} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$$

where $\boldsymbol{D}$ is a diagonal matrix of degrees in adjacency matrix $\boldsymbol{A}$, with $D_{ii} = \sum_j A_{ij}$.

Using a K dimensional eigen-decomposition, we write that $\boldsymbol{L} \approx \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$ where K represents the number of communities one expects to observe in the network (this value can be adaptively chosen by considering the size of the non-zero eigenvalues of L, otherwise known as identifying the eigenvalues that fall far from the 'bulk' (Krzakala et al., 2013)). After the embedding, the k-means algorithm with multiple restarts is applied to the top K eigenvectors, $\boldsymbol{V}$. The resulting labels are used to create $\hat{\boldsymbol{U}}$.

While this work focuses on when K is known, we note potential consequences when K is estimated incorrectly. If K is underestimated, then some communities might be combined. If K is overestimated, overall balance for GATE estimation might still be maintained however this could lead to only having access to partial community level treatment effects.

As a note, if covariate information is available, this can be incorporated into the community detection step introduced in later sections (Binkiewicz, Vogelstein, & Rohe, 2017). Even when there is covariate information available, it often does not directly map to communities (Mathews & Volfovsky, 2021).

Determining Clusters

After the labels are estimated, a community induced sub-graph is created for each of the K communities. As a result, each sub-graph contains only within community ties. For determining clusters for treatment, $h(\boldsymbol{A}, \hat{\boldsymbol{U}})$ is chosen to be the 3-net (generically epsilon net) algorithm which is implemented on each sub-graph separately. Unlike the previous step where the goal is to find ground truth communities that define interference, the goal of this step is to find sets of closely connected individuals where each cluster is far from the other clusters. While community detection algorithms used in the previous step could be implemented, again the goals of the previous and current step differ, and, as such, different algorithms are better suited for each task.

We use 3-net due to its positive performance in the literature as demonstrated in Ugander and Yin (2020). Further, it provides useful theoretical properties for estimating network exposure probabilities that is leveraged in later sections. At a high level, 3-net clustering generates a random ordering of all nodes in the graph. Based on this ordering, a maximal distance 3 hop independent set is created, and the nodes in this set are called seed nodes. The number of seed nodes determines the number of clusters that will exist in the experiment. Each node is assigned to its closest seed node based off of minimal graph distance, and these assignments then form clusters. Since the seed nodes are guaranteed to be network exposed to either control or treatment, this algorithm makes estimation of exposure probabilities plausible through Monte Carlo simulation. Ugander et al. (2013) demonstrate the theoretical and empirical properties of 3-net GCR which we adapt to our community informed design. An explicit outline of the implementation of the community epsilon net is in the supplement.

Randomization of Clusters

Finally, for this work, $s(C_{\hat{\boldsymbol{U}}})$ is defined to be independent cluster level randomization where each cluster is treated with probability $q = 0.5$. However, this randomization scheme can be altered depending on the desired estimand or if the design has constraints on randomization schema.

Note that community informed design generalizes other discussed methods. If community labels are known, Algorithm 1 can be adjusted such that $\hat{\boldsymbol{U}} = \boldsymbol{U}$ and thus step 1 can be skipped. Also, when $K = 1$, Algorithm 1 reduces to standard graph cluster randomization, and further, if $K = 1$ and the number of clusters is set to N then Algorithm 1 becomes standard independent randomization.

3 BIAS REDUCTION

We study the bias of using the naive difference in means estimator of Eq. (3) for estimating the GATE under community driven interference.

In this setting, specifically let $g_i = \sum_{j=1}^{N} Z_j A_{ij} \mathbb{1}[\boldsymbol{U_i} = \boldsymbol{U_j}]$ and write the potential outcome as $Y_i(Z_i, g_i)$. The potential outcomes can be decomposed as $Y_i(Z_i, g_i) = C_i(Z_i) + B_i(g_i)$ (similar to the decomposition in Jagadeesan et al. (2020); Karwa and Airoldi (2018); Sussman and Airoldi (2017)) where $C_i(Z_i)$ captures the direct effect of treatment on node i and $B_i(g_i)$ captures the effect of the treated within community neighbors of node i. The true GATE, τ, can be represented as:

$$\frac{1}{N}\sum_{i=1}^{N}\Big[(C_i(1)+B_i(d_i^c))-(C_i(0)+B_i(0))\Big] \tag{4}$$

where d_i^c is the number of within community neighbors of node i.

The derivation for the expectation of the estimator, $\hat{\tau}$, then follows as:

$$\begin{aligned}
E_{\boldsymbol{Z}}[\hat{\tau}] &= E_Z\left[\frac{1}{N_T}\sum_{i=1}^{N} Y_i Z_i - \frac{1}{N_C}\sum_{i=1}^{N} Y_i(1-Z_i)\right] \\
&= E_{\boldsymbol{Z}}\left[\frac{1}{N_T}\sum_{i=1}^{N} Y(1,g_i)Z_i - \frac{1}{N_C}\sum_{i=1}^{N} Y(0,g_i)(1-Z_i)\right] \\
&= E_{\boldsymbol{Z}}\left[\frac{1}{N_T}\sum_{i=1}^{N}(C_i(1)+B_i(g_i))Z_i - \frac{1}{N_C}\sum_{i=1}^{N}(C_i(0)+B_i(g_i))(1-Z_i)\right].
\end{aligned}$$

Combining the above, assume that a community detection algorithm learns equally sized communities and that a clustering algorithm creates equally sized clusters. Under these assumptions, $N_T = N_C = N/2$ can be fixed and the bias is given by:

$$\begin{aligned}
b &= E_{\boldsymbol{Z}}[\hat{\tau}] - E[\tau] \\
&= \sum_{i=1}^{N}\left(C_i(1)\pi_i(1) + \pi_i(1,d_i^c)B_i(d_i^c) + \pi_i(1,0)B_i(0) + \sum_{g_i \neq \{0,d_i^c\}} \pi_i(1,g_i)B(g_i)\right) \\
&- \sum_{i=1}^{N}\left(C_i(0)\pi_i(0) + \pi_i(0,d_i^c)B_i(d_i^c) + \pi_i(0,0)B_i(0) + \sum_{g_i \neq \{0,d_i^c\}} \pi_i(0,g_i)B(g_i)\right) \\
&- \left(\frac{1}{N}\sum_{i=1}^{N}(C_i(1)+B_i(d_i^c))-(C_i(0)+B_i(0))\right)
\end{aligned}$$

where we define the following weights as:

$$\begin{aligned}
\pi_i(1) &= \frac{1}{N_T}E[\mathbb{1}[Z_i=1]] = \frac{1}{N_T}\mathbb{P}(Z_i=1) \\
\pi_i(0) &= \frac{1}{N_C}E[\mathbb{1}[Z_i=0]] = \frac{1}{N_C}\mathbb{P}(Z_i=0)
\end{aligned}$$

and

$$\pi_i(1,g_i) = \frac{1}{N_T}E[\mathbb{1}[Z_i=1, G_i=g_i]] = \frac{1}{N_T}\mathbb{P}(Z_i=1, G_i=g_i)$$

$$\pi_i(0,g_i) = \frac{1}{N_C}E[\mathbb{1}[Z_i=0,G_i=g_i]] = \frac{1}{N_C}\mathbb{P}(Z_i=0,G_i=g_i).$$

The values of the above weights are determined by the chosen design and accompanying interference assumptions.

The bias can be rewritten as:

$$b = \sum_{i=1}^{N} C_i(1)\left(\pi_i(1) - \frac{1}{N}\right) - C_i(0)\left(\pi_i(0) - \frac{1}{N}\right) \tag{5a}$$

$$+ \sum_{i=1}^{N} B_i(d_i^c)\left(\pi_i(1,d_i^c) - \pi_i(0,d_i^c) - \frac{1}{N}\right) \tag{5b}$$

$$+ \sum_{i=1}^{N} \sum_{g_i \neq \{0,d_i^c\}} B_i(g_i)\left(\pi_i(1,g_i) - \pi_i(0,g_i)\right). \tag{5c}$$

This calculation follows from the results of Karwa and Airoldi (2018) for general interference. Also note that these results are easily generalized to the more flexible setting when N_T and N_C are random by noting the inequality $E_Z[1/\sum_{i=1} Z_i] \geq 1/E_Z[\sum_{i=1} Z_i]$.

Above, $g_i \neq \{0,d_i^c\}$ is equivalent to $g_i \in \{1,...,(d_i^c-1)\}$ and thus indicates the set of all possible observed g_i under our interference assumption except for those where either no within community neighbors are treated or those where all within community neighbors are treated. Anytime the assignment is different from $(1,d_i^c)$ or $(0,0)$, estimation of the counterfactuals that contribute to the GATE suffer.

As a result, a design is needed where the probability of observing these is small. Below we show that CID leads to smaller bias than GCR. There are three components to the bias. The first part, Eq. 5a, can be controlled by choosing a design that treats nodes with probability $q=0.5$. The second component, Eq. 5b, can also be controlled by the design: under community informed design, $\pi_i^{CID}(0,d_i^c) = 0$ since each individual belongs to a cluster within their own community and must share treatment status with at least one of their within community neighbors. We can further control $\pi_i^{CID}(1,d_i^c)$ which is given in Eq. (6). In contrast, under GCR, we will show that $\pi_i^{GCR}(0,d_i^c) \geq 0$ and $\pi_i^{GCR}(1,d_i^c) \leq \pi_i^{CID}(1,d_i^c)$ which leads to:

$$\left|\sum_{i=1}^{N} B_i(d_i^c)\left(\pi_i^{CID}(1,d_i^c) - \pi_i^{CID}(0,d_i^c) - \frac{1}{N}\right)\right| \leq$$

$$\left|\sum_{i=1}^{N} B_i(d_i^c)\left(\pi_i^{GCR}(1,d_i^c) - \pi_i^{GCR}(0,d_i^c)) - \frac{1}{N}\right)\right|,$$

suggesting that CID reduces the impact of Eq. 5b.

To demonstrate the above, we consider GCR and CID with the 3-net clustering algorithm. Let d^c_{max} denote the maximal within community degree across all nodes in A. Following from Ugander and Yin (2020), the network exposure probabilities can be bounded for both designs. Tighter bounds are derived in the supplement, however for simplicity, consider the worst case scenario. For community informed design, $\mathbb{P}(Z_i = 1, G_i = d^c_i) \geq \frac{.5}{d^c_{max} \times (1 + d^c_{max})}$. To obtain $\pi^{CID}_i(1, d^c_i)$, we simply divide $\mathbb{P}(Z_i = 1, G_i = d^c_i)$ by $E[\sum_i^N \mathbb{1}[Z_i = 1]] = N/2$:

$$\frac{1}{Nd^c_{max} \times (1 + d^c_{max})} \leq \pi^{CID}_i(1, d^c_i) \leq \frac{1}{2N} \leq \frac{1}{N}, \tag{6}$$

assuming that a node has at least 1 within community neighbor which is reasonable under assortative community structure.

Again, under CID when true communities are known, $\pi^{CID}_i(0, d^c_i) = 0$ and $\pi^{CID}_i(1, d^c_i) < 1/N$. However, we are not guaranteed that $\pi^{GCR}_i(0, d^c_i) = 0$. Further, under GCR, $\mathbb{P}(Z_i = 1, G_i = d^c_i) \geq \frac{.5}{d_{max} \times (1 + d_{max})}$. Since $d_{max} \geq d^c_{max}$, these exposure probabilities under GCR are lower than that under our design. As a result, our design yields a probability closer to $1/N$; therefore CID yields lower bias in the second term. Also note, that under CID, given the higher values of $\mathbb{P}(Z_i = 1, G_i = d^c_i)$ and $\mathbb{P}(Z_i = 0, G_i = 0)$, the probability of ending up with less desirable exposure is sufficiently reduced.

The last component of the bias calculation, Eq. 5c, contributes a fairly small amount to the total bias. Note that ideally $\pi_i(1, g_i)$ should be large when $g_i \approx d^c_i$ and small otherwise. Similarly $\pi_i(0, g_i)$ should be large when $g_i \approx 0$ and small otherwise. We show this empirically in Figure 3. If $B_i(g_i)$ is relatively flat in g_i, this will lead to terms canceling in Eq. 5c. Alternatively, if $B_i(g_i)$ is increasing in g_i, only the terms associated with large g_i will contribute to the bias. While the exact ordering between GCR and CID depends on the $B_i(g_i)$, due to the fact that $\mathbb{P}(1, d^c_i)$ is larger for CID, we expect the contribution to the bias to be smaller under CID.

4 COST OF ESTIMATING COMMUNITY LABELS

Thus far we have abstractly discussed the existence and identification of communities. Since these communities form the foundation of CID, it is crucial to understand how community label estimation impacts the design and GATE estimation. In this section, we derive the expected bias incurred for GATE estimation that is due to incorrectly estimating the community label of individuals in the first stage of the algorithm. Define a node, i, to have an incorrect community label when $\hat{\boldsymbol{U}}_i \neq \boldsymbol{U}_i$ up to permutation of the labels (estimated community label does not match the true community label). In order to explicitly derive the cost of mislabelled nodes we take a model assisted approach (Basse & Airoldi, 2018; Särndal, Swensson, & Wretman, 2003). Consider the

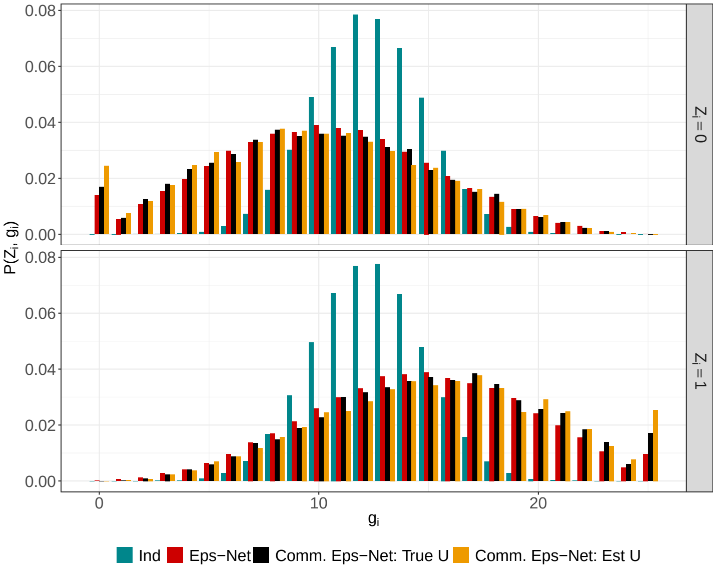

Fig. 3 Simulated $\mathbb{P}(Z_i, g_i)$ over 4000 simulations. This figure demonstrates how community informed design puts higher probability on more desirable potential outcomes

following outcome model:

$$Y_i = \alpha + \beta Z_i + [(\boldsymbol{U\Gamma U^T})_{i,} \circ \boldsymbol{A}_{i,}]\boldsymbol{Z}^T + \epsilon_i \tag{7}$$

$$= \alpha + \beta Z_i + \sum_{j=1}^{N} \boldsymbol{U_{i,}\Gamma U_{j,}^T} \times A_{ij} Z_j + \epsilon_i, \tag{8}$$

where α is a baseline effect, β is overall direct effect, $\boldsymbol{Z}$ is the treatment indicator vector, and $\boldsymbol{\Gamma}$ is a $K \times K$ matrix describing the effect of treated neighbors based on community membership and ϵ represents individual variability. Further, $\boldsymbol{A}$ is the adjacency matrix of the network, generated from a stochastic blockmodel (SBM) (Holland, Laskey, & Leinhardt, 1983). Under the SBM, the probability of a connection between nodes i and j depends solely upon the community memberships of those nodes. For a SBM with K communities,

$$A_{ij} | \boldsymbol{U}_i, \boldsymbol{U}_j \sim Bern(\boldsymbol{U_i^T Q U_j})$$

where $\boldsymbol{U} \sim P_{\boldsymbol{U}}$ is again a $N \times K$ membership matrix drawn such that each row contains one 1 indicating which community a node belongs to. Let $\boldsymbol{Q}$ be a $K \times K$ probability matrix describing the relations between communities where $\boldsymbol{Q}$ is parameterized by 2 probabilities, a and b, such that $A_{ij}|\boldsymbol{U}_i, \boldsymbol{U}_j \sim Bern(a)$ if $\boldsymbol{U}_i = \boldsymbol{U}_j$ or $A_{ij}|\boldsymbol{U}_i, \boldsymbol{U}_j \sim Bern(b)$ if $\boldsymbol{U}_i \neq \boldsymbol{U}_j$. We evaluate how well the GATE is estimated in a setting where $\boldsymbol{U}_i = \boldsymbol{U}_j$ but $\hat{\boldsymbol{U}}_i \neq \hat{\boldsymbol{U}}_j$.

Studying the effect of mislabelled nodes can be reduced to asking how close an observed Y_i is to the actual counterfactual of interest. It is important to note that only the neighborhood interference aspect of bias is impacted by mislabelled nodes. Thus the direct effect of the bias is ignored in this section. For example, when individual i is treated, we would want his or her true within community connections to also be treated, thus consider:

$$E\Big[Y_i(\boldsymbol{Z}=1) - Y_i|Z_i = 1\Big] =$$

$$E\Big[\boldsymbol{\gamma}\boldsymbol{U}_i\sum_j A_{ij}\times\mathbb{1}[\boldsymbol{U}_i=\boldsymbol{U}_j] - \gamma\boldsymbol{U}_i\sum_j A_{ij}\times\mathbb{1}[\boldsymbol{U}_i=\boldsymbol{U}_j]\times Z_j|\hat{\boldsymbol{U}}\Big]$$

where the expectation is with respect to the model uncertainty and the design but conditioned on $\hat{\boldsymbol{U}}$. Note that $\boldsymbol{\gamma}$ denotes the diagonal elements of $\boldsymbol{\Gamma}$. Now consider the situation when the label of node i is incorrectly estimated. Formally, let T_{ij} be the event that $\boldsymbol{U}_i = \boldsymbol{U}_j$ and $\hat{\boldsymbol{U}}_i \neq \hat{\boldsymbol{U}}_j$. As such, the difference between the counterfactual and observed value for a misclassified treated node is equal to

$$E\left[\sum_{j=1}^{N} A_{ij}Z_j\boldsymbol{\gamma}\boldsymbol{U}_i\mathbb{1}_{T_{ij}}\right].$$

Note that conditional on T_{ij}, A_{ij} is independent of $\boldsymbol{U}_i$ and Z_i is independent of Z_j. Hence, we can write:

$$E\left[\sum_{j=1}^{N} A_{ij}Z_j\boldsymbol{\gamma}\boldsymbol{U}_i\mathbb{1}_{T_{ij}}\right] = \mathbb{P}(T_{ij})\times N\times a\times q\times\sum_{k=1}^{K}\gamma_k\boldsymbol{P}_k \qquad (9)$$

where $\boldsymbol{P}_k$ is the probability that a node belongs to community k. Recall $P[Z_j|T_{ij}] = q$ and $P[A_{ij}|T_{ij}] = a$. The probability of T_{ij} depends heavily on the network structure and algorithm f (for label recovery). However, this probability can easily be represented in terms of the sizes of each community and the number of mislabelled nodes within each community. A full derivation of Eq. (9) is provided in the supplement. Let D_k be the number of mislabelled nodes in community k and N_k be the number of nodes in true community k. For $K = 2$:

$$\mathbb{P}(T_{ij}) = \tfrac{E[\sum_{k=1}^{2}(N_k - D_k)\times 2D_k]}{N^2}.$$

Details on $\mathbb{P}(T_{ij})$ for general K can be found in the supplement along with empirical validation of Eq. (9). Note that if GCR is used instead of CID, this is equivalent to setting $\mathbb{P}(T_{ij}) = 1/2$ under the equal two community example (since this is the worst case performance for community detection).

5 SIMULATIONS

We now demonstrate the empirical performance of our proposed design coupled with the difference in means estimator in several scenarios. Since community informed design generalizes GCR and independent assignment, we label the methods by the algorithm, $h(\cdot, \cdot)$, used in step 3 of Algorithm 1. Throughout, we focus on comparing the following:

- **Ind**: Independent random assignment (No regard for network structure)
- **Eps-Net**: Epsilon net (Knows about network structure but does not know about communities, standard GCR) (Eckles et al., 2016)
- **Community Eps-Net with True U**: Perform an epsilon net on each community sub-graph (Knows about network structure and existence of communities, community informed design)
- **Community Eps-Net with Estimated U**: Perform an epsilon net on each community sub-graph by estimated community label (Knows about network structure and existence of communities, community informed design)

While the main goal of this work is bias reduction, we also consider the potential for bias/variance trade-off. Results comparing the root mean squared error (RMSE) are given in the supplement. Consistently our methods maintain lower RMSE than standard GCR and independent randomization when the interference is community driven. In the main text of this section, we report the bias of the different approaches *relative* to the bias under independent assignment. This is defined as the average absolute bias for the difference in means estimator under a particular method divided by the average absolute bias obtained from independent randomization.

In the first set of simulations (Sections 5.1-5.3) we concentrate on networks that have a true underlying community structure (whether or not that community structure is informative of the true interference). In Sections 5.4 and 5.5 we explore the behavior of our approach when there are no ground truth communities or when there might be a slight mismatch between the communities driving interference and the communities observed in the network.

As a note, for each SBM parameterization, we simulate multiple networks and multiple clusterings for each network. As a result, the fraction of nodes with incorrectly estimated labels acts as a proxy for difficulty of community detection thus only community aware designs are impacted by this. Other methods are agnostic to this fact and are shown to perform fairly consistently across all regimes.

5.1 Two Communities

Throughout the simulation in this subsection we investigate networks with two ground truth communities. There are $N = 1600$ individuals in the SBM, split equally between the two communities. The probability of within community edges is set to 0.04, and we vary the probability of cross-community edges to demonstrate the behavior of our randomization as the community detection problem becomes more difficult. Outcomes are generated following Eqn (7). We consider three scenarios:

1. Community interference: The true underlying interference mechanism for outcomes is within community level interference. That is, a node is only

influenced by the treatment of within community connections, and we let $\boldsymbol{\gamma} = (20/50, 40/50)$.

2. Community agnostic interference: The true underlying interference mechanism for outcomes is full neighborhood level interference. That is, all neighbors influence a node equally, and we let $\boldsymbol{\gamma} = (20/50, 20/50)$.
3. Anti community interference: The true underlying interference mechanism for outcomes is that individuals are only influenced by neighbors that do not share their community and thus $\boldsymbol{\Gamma}$ has a dis-assortative structure:

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0 & 20/50 \\ 40/50 & 0 \end{bmatrix}$$

.

The results are presented in Figures 4-6. Unsurprisingly, our proposed approach performs exceptionally well in the first scenario (Figure 4) when either the true labels are known or the community detection task is easy. Importantly, as the community detection problem becomes more difficult, our proposed approach reduces to the standard GCR method.

When there is community agnostic interference, Figure 5 shows that there are not many gains from the community detection step of CID. The difference between GCR and CID can likely be explained by the fact that there are more within community edges and so the clusters used in CID are tighter and more representative then the ones in GCR.

Lastly, we see that our approach suffers substantially when the interference mechanism is completely misspecified (Figure 6). In this setting, the community interference assumption is false and thus there is no reason to expect our design implementation to be successful. If this phenomenon is known, our algorithm could be amended to only consider cross-community ties for selecting clusters. However, if CID is implemented and the community detection problem is challenging, then we see that there is not a substantial loss in bias estimation.

5.2 Community Level Average Treatment Effect

Since communities are likely influenced by treatment in different ways, consider community level treatment effects:

$$\tau_k = \frac{1}{N_k} \sum_{i=1}^{N} [Y_i(\boldsymbol{Z} = 1) - Y_i(\boldsymbol{Z} = 0)] \mathbb{1}[U_{ik} = 1] \tag{10}$$

where $N_k = \sum_{i=1}^{N} \mathbb{1}[U_{ik} = 1]$. This quantity can be estimated using a difference in means estimator that only uses nodes in community k (whether those labels are learned or estimated). Through community informed design, these quantities are easily obtained. However, for the independent and standard GCR epsilon net methods, community information can only be leveraged after the experiment by partitioning outcomes by community labels. Figure 7 shows

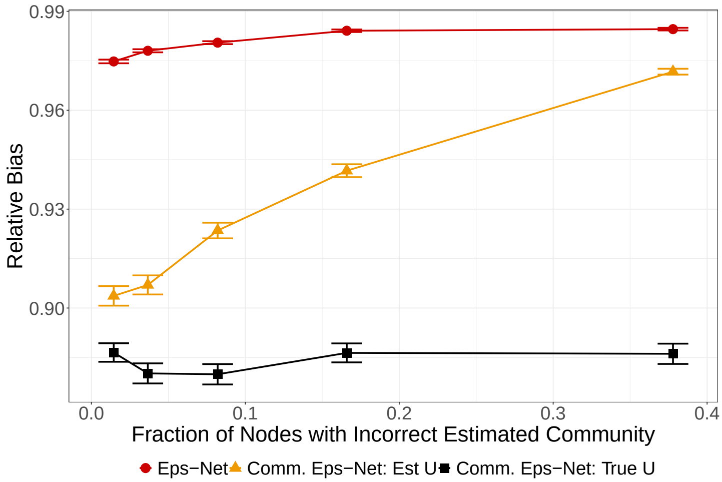

Fig. 4 Results for community interference simulation. The x-axis is the average fraction of nodes with incorrectly estimated labels for a given regime. The y-axis is relative bias for the difference in means estimator. Standard error bars are over 2000 simulations.

relative bias for estimating the community level treatment effects from the simulation with community interference in Section 5.1. From this experiment, again the benefits of community informed design are apparent.

5.3 Varying number of communities

For the following simulation, consider performance for different values of K. Let N scale with the number of communities such that there are 250 nodes in each community ($N = 250 \times K$) and let $\boldsymbol{\gamma}$ depend on K such that a sequence of length K is generated with values between 10/50 and 1. For generating the networks, define the SBM parameters to be $a = .04$ and $b = .03/(K-1)$. Note that N and b are dependent on K to maintain equal expected degree for the network, however slight variability in outcomes persists due to $\boldsymbol{\gamma}$.

Figure 8 shows that conditioning on communities is beneficial across values of K. The community detection problem becomes easier as N increases thus the method using estimated $\boldsymbol{U}$ matches that of true $\boldsymbol{U}$ for high K.

5.4 No Community Structure or Interference

Unlike the first three simulations, in this section we consider a network generative model that does not exhibit explicit community structure. Similarly, the outcome model relies on all edges in the network and so the community neighborhood interference assumption is violated. We follow the data generation procedure described in Ugander and Yin (2020) which uses small world networks and a multiplicative response model. Details of parameter specification

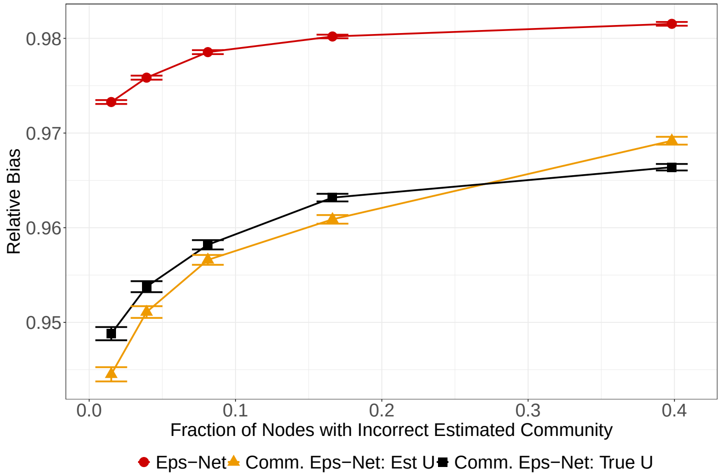

Fig. 5 Results for community agnostic interference. The x-axis is the average The x-axis is the average fraction of nodes with incorrectly estimated labels for a given regime. The y-axis is relative bias for the difference in means estimator. Standard error bars are over 2000 simulations

are provided in the supplement. Again, since the underlying network model does not exhibit meaningful community structure, the estimated communities used in CID are themselves not relevant and so we do not expect any performance gains from using CID as opposed to standard GCR. Figure 9 shows the output of the two approaches relative to independent design: the performance of community informed design with epsilon net nearly matches that of the non community informed version.

5.5 Potentially mismatched communities

Much of the work on network analysis has been driven by the study of student-student networks (Hoff, Fosdick, Volfovsky, & Stovel, 2013; Mayer & Puller, 2008; Rienties & Nolan, 2014; Sentse, Kiuru, Veenstra, & Salmivalli, 2014). While grade levels and genders represent natural communities within a network, it is not necessarily the case that these community labels are to be recoverable from observed in-school networks (Mathews & Volfovsky, 2021). Motivated by a recent collection of network-driven experiments on the impact of anti-bullying interventions (Paluck et al., 2016, 2020) we leverage the observed networks and observed grade and gender labels for individuals in those networks to study the performance of our experimental design.

As part of the original data collection in Paluck et al. (2016), students were asked to record up to 10 friends (for the purposes of this simulation we consider undirected versions of these networks, making the observed degree slightly larger), making the problem of community detection substantially harder since the networks are censored. We study four schools with different sample sizes

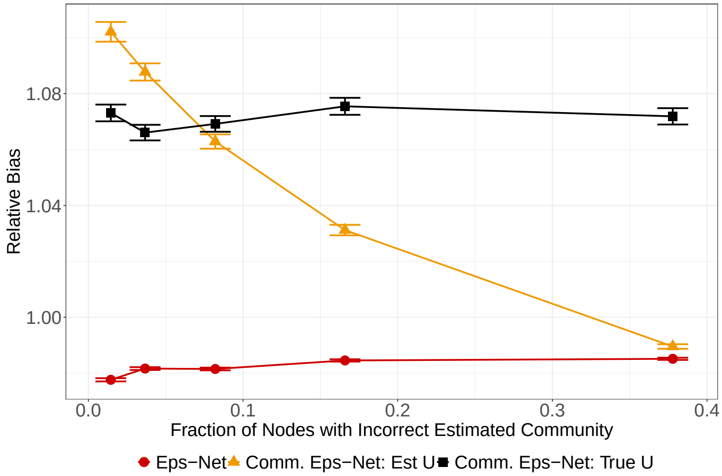

Fig. 6 Results for anti community interference. This figure shows the average relative bias for GATE estimation over 2000 simulations when response is now influenced by between community neighbors

($N = 126, 104, 370, 530$) and so for larger N the censoring makes the network appear sparser, again potentially affecting the performance of community detection and our proposed approach.

For each of the schools, outcomes are generated according to Eqn (7). Under this model, ground truth communities are defined as unique grade and gender pairs. That is, if two students share grade and gender, they share a community. Under this definition, we assume that interference in outcomes of students are only driven by the treatment of neighbors who share the same grade and gender. K is the count of unique gender-grade pairs in a particular school (ranging from 4 to 6 depending on the grades within the school). Students with NA values for grade and/or gender are dropped from the data.

For each school the within community effect, γ, is a vector of length K with values ranging from 0.4 to 1.6, and the direct effect is $\beta = 1$. As shown in Figure 10, leveraging community information reduces relative bias across all of the schools. However, we also note that the differences between standard GCR and community informed methods are not always the same. For example, both Schools A and B are relatively small and we see that CID with estimated community labels nearly matches CID with true labels. On the other hand, we see the smallest improvement between GCR and CID with estimated communities in School D which has the sparsest network, making the community detection problem the hardest.

6 DISCUSSION

This paper has proposed a new experimental design that leads to a reduction of bias of the naive difference-in-means estimator in the estimation of the

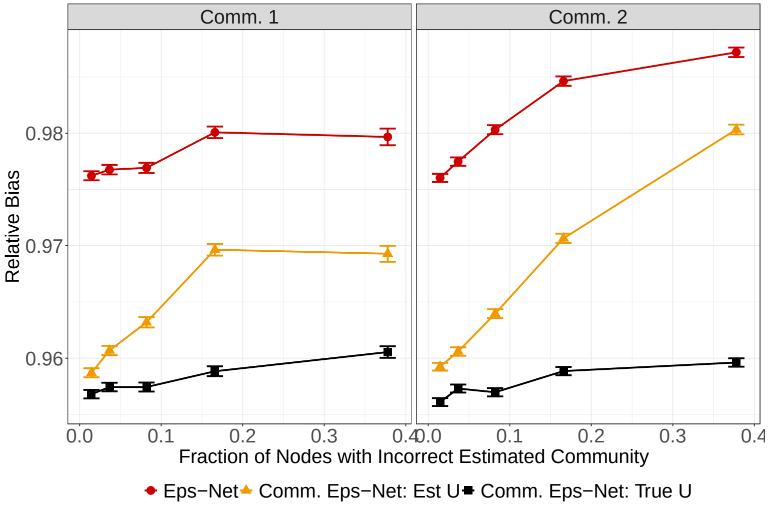

Fig. 7 The average relative bias for community level treatment effect estimation over 2000 simulations. These community level effects are from the simulation generated in Figure 4

global average treatment effect when interference is community driven. The approach improves on graph cluster randomization techniques by conditioning on the community structure of the graph (estimated or known) and provably reducing the fraction of low-quality randomizations. Importantly, the community interference assumption is meaningful in many applied settings, while realistic violations of it do not lead to a substantive reduction in the quality of the proposed design. In settings where the community structure must be estimated from an observed network, we demonstrate both analytically and empirically that the improvement due to CID decreases as the community detection problem becomes more difficult, but this step does not lead to performance that is worse than naive graph cluster randomization. Further, we demonstrate that when within and cross community ties are influential, our method still improves estimation when community structure is present.

Network interference has been recognized as an important obstacle in experimental design, leading to many recent advances. For example, Zhou, Liu, Li, and Hu (2020) propose a cluster adaptive network testing procedure with a sequential cluster adaptive randomization and a cluster adjusted estimator for the average treatment effect. The proposed approach requires observing additional covariate information on each of the individuals in the network. While this can improve the underlying clustering of the network, such data may be unavailable at times (such as when networks are elicited prior to experimentation) or latent community structure might not be correlated with observed covariates. Another recent approach proposes a cluster based regression adjustment that improves estimation of the GATE as well as testing for interference

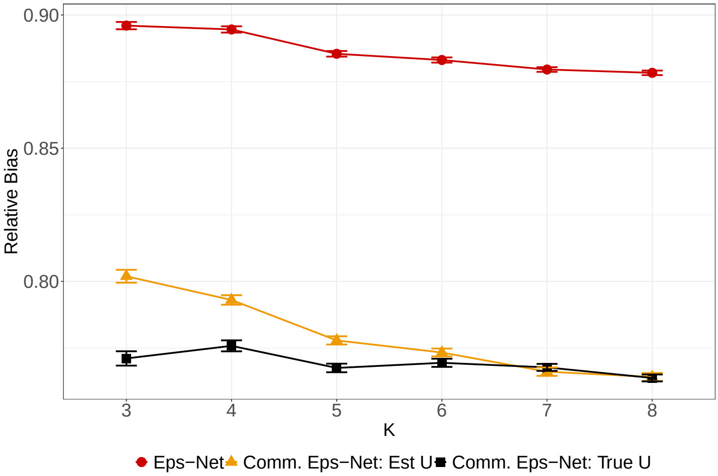

Fig. 8 The y axis is relative bias for varying K equally sized communities with standard error bars over 2000 simulations per value of K

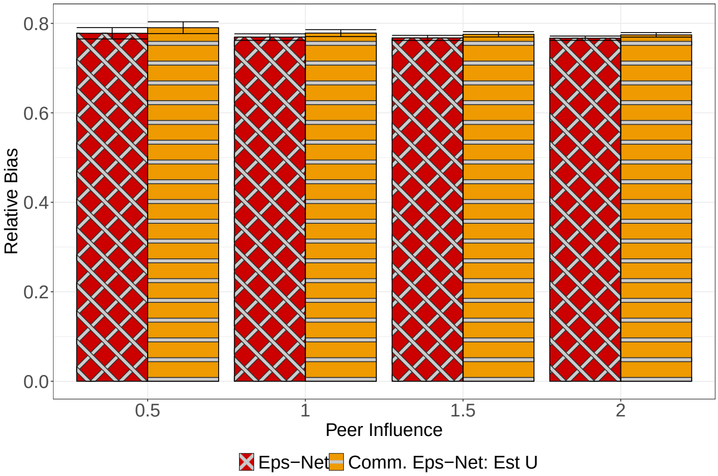

Fig. 9 Using response model from Ugander and Yin (2020) in their section 6.2. Standard error bars are over 100 simulations. The x-axis represents their peer influence parameter that is given in the supplement

between individuals (Karrer et al., 2021). They show how tracking exposure to treatment can be used to further reduce variance in estimating the GATE but again rely heavily on the availability of additional side information about the individuals in the study.

Given the recent focus on covariates in the literature, it is a natural future direction to incorporate covariate information into the community informed

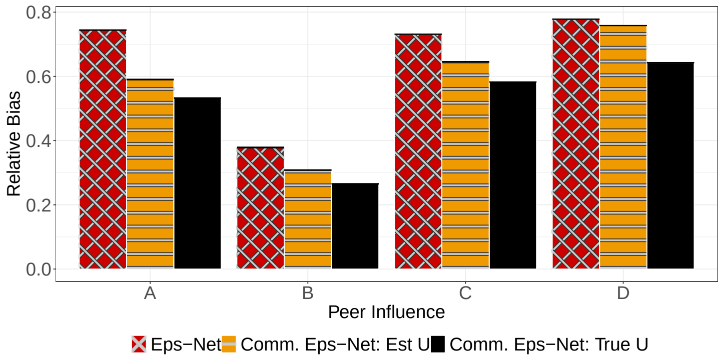

Fig. 10 Relative bias for various schools from the middle school data set where outcomes are simulated based off of the true observed networks with standard error bars over 750 simulations

design procedure. This can be done by incorporating covariates into the community detection problem directly (Binkiewicz et al., 2017; Shen, Amini, Josephs, & Lin, 2022; Yan & Sarkar, 2021) or by considering community detection as a component of a network regression problem (Hoff, 2008; Mathews & Volfovsky, 2021). These approaches can further be coupled with post-hoc estimators that adjust for the potential community structure that may not have been accounted for during the design phase of an experiment.

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, *18*(1), 6446–6531.

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

Aukett, R., Ritchie, J., Mill, K. (1988). Gender differences in friendship patterns. *Sex roles*, *19*(1-2), 57–66.

Awan, U., Morucci, M., Orlandi, V., Roy, S., Rudin, C., Volfovsky, A. (2020). Almost-matching-exactly for treatment effect estimation under network interference. *International conference on artificial intelligence and statistics* (pp. 3252–3262).

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.F., ... Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

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

Bhattacharyya, S., & Bickel, P.J. (2014). Community detection in networks using graph distance. *arXiv preprint arXiv:1401.3915*.

Binkiewicz, N., Vogelstein, J.T., Rohe, K. (2017). Covariate-assisted spectral clustering. *Biometrika*, *104*(2), 361–377.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Bruna, J., & Li, X. (2017). Community detection with graph neural networks. *stat*, *1050*, 27.

Chamberlain, B., Kasair, C., Rotheram-Fuller, E. (2007). Involvement or isolation? the social networks of children with autism in regular classrooms. *J Autism Dev Disord*, *37*(2).

Eckles, D., Karrer, B., Ugander, J. (2016). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, *5*(1).

Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75–174.

Girvan, M., & Newman, M.E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, *99*(12), 7821–7826.

Hoff, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems 20* (pp. 657–664). Cambridge, MA: MIT Press.

Hoff, P., Fosdick, B., Volfovsky, A., Stovel, K. (2013). Likelihoods for fixed rank nomination networks. *Network Science*, *1*(3), 253–277.

Holland, P., Laskey, K., Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*, 109-137.

Igarashi, T., Takai, J., Yoshida, T. (2005). Gender differences in social network development via mobile phone text messages: A longitudinal study. *Journal of Social and Personal Relationships*, *22*, 691-713.

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.

Karrer, B., Shi, L., Bhole, M., Goldman, M., Palmer, T., Gelman, C., … Sun, F. (2021). Network experimentation at scale. *Proceedings of the 27th acm sigkdd conference on knowledge discovery & data mining* (pp.

3106–3116).

Karwa, V., & Airoldi, E.M. (2018). A systematic investigation of classical causal inference strategies under mis-specification due to network interference. *arXiv preprint arXiv:1810.08259*.

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., Pohlmann, N. (2013). Online controlled experiments at large scale. *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1168–1176).

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., Zhang, P. (2013). Spectral redemption in clustering sparse networks. , *110*(52), 20935–20940.

10.1073/pnas.1312486110

Manski, C.F. (1995). *Identification problems in the social sciences.* Harvard University Press.

Mathews, H., Mayya, V., Volfovsky, A., Reeves, G. (2019). Gaussian mixture models for stochastic block models with non-vanishing noise. *2019 ieee 8th international workshop on computational advances in multi-sensor adaptive processing (camsap)* (pp. 699–703).

Mathews, H., & Volfovsky, A. (2021). Latent community adaptive network regression. *arXiv preprint arXiv:2112.06097*.

Mayer, A., & Puller, S.L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of public economics*, *92*(1-2), 329–347.

Mayya, V., & Reeves, G. (2019). Mutual information in community detection with covariate information and correlated networks. *2019 57th annual allerton conference on communication, control, and computing (allerton)* (pp. 602–607).

McPherson, M., Smith-Lovin, L., Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.

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

Paluck, E.L., Shepherd, H., Aronow, P.M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, *113*(3), 566–571. Retrieved from https://www.pnas.org/content/113/3/566 https://arxiv.org/abs/https://www.pnas.org/content/113/3/566.full.pdf
10.1073/pnas.1514483113

Paluck, E.L., Shepherd, H.R., Aronow, P. (2020). Changing climates of conflict: A social network experiment in 56 schools, new jersey, 2012-2013.

10.3886/ICPSR37070.v2

Reeves, G., Mayya, V., Volfovsky, A. (2019). The geometry of community detection via the mmse matrix. *2019 ieee international symposium on information theory (isit)* (pp. 400–404).

Rienties, B., & Nolan, E.-M. (2014). Understanding friendship and learning networks of international and host students using longitudinal social network analysis. *International Journal of Intercultural Relations*, *41*, 165–180.

Rohe, K., Chatterjee, S., Yu, B., et al. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, *39*(4), 1878–1915.

Rubin, D.B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, *25*(3), 279-292. Retrieved from https://www.sciencedirect.com/science/article/pii/0378375890900778

https://doi.org/10.1016/0378-3758(90)90077-8

Särndal, C.-E., Swensson, B., Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Sävje, F. (2021). Causal inference with misspecified exposure mappings. *arXiv preprint arXiv:2103.06471*.

Sävje, F., Aronow, P.M., Hudgens, M.G. (2021). Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, *49*(2), 673–701.

Sentse, M., Kiuru, N., Veenstra, R., Salmivalli, C. (2014). A social network approach to the interplay between adolescents' bullying and likeability over time. *Journal of youth and adolescence*, *43*(9), 1409–1420.

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.

Shen, L., Amini, A., Josephs, N., Lin, L. (2022). Bayesian community detection for networks with covariates. *arXiv preprint arXiv:2203.02090*.

Shrum, W., Cheek Jr., N., Hunter, S. (1988). Friendship in school: gender and racial homophily. *Sociology of Education*, *61*, 227-239.

Staber, U. (1993). Friends, acquaintances, strangers: gender differences in the structure of enterpreneurial networks. *Journal of Small Business and Entrepreneurship*, *11*, 73-82.

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

Toulis, P., & Kao, E. (2013). Estimation of causal peer influence effects. *International conference on machine learning* (pp. 1489–1497).

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

Ugander, J., & Yin, H. (2020). Randomized graph cluster randomization. *arXiv preprint arXiv:2009.02297*.

Xu, Y., Chen, N., Fernandez, A., Sinno, O., Bhasin, A. (2015). From infrastructure to culture: A/b testing challenges in large scale social networks. *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (pp. 2227–2236).

Yan, B., & Sarkar, P. (2021). Covariate regularized community detection in sparse graphs. *Journal of the American Statistical Association*, *116*(534), 734–745.

Zhou, Y., Liu, Y., Li, P., Hu, F. (2020). Cluster-adaptive network a/b testing: From randomization to estimation. *arXiv preprint arXiv:2008.08648*.