Statistical Methods & Applications
https://doi.org/10.1007/s10260-022-00679-6

ORIGINAL PAPER

Community informed experimental design

Heather Mathews[1] · **Alexander Volfovsky**[1]

Accepted: 5 November 2022

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 et al. 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.

✉ Alexander Volfovsky
av136@duke.edu

Heather Mathews
heather.mathews@duke.edu

[1] Department of Statistical Science, Duke University, Durham, North Carolina, USA

Published online: 23 January 2023

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)*:

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

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 estimand 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 of the network. For example, it is well established that not all connections between individuals in a network are created equal and some ties are stronger or more influential than others. For example, same gender ties among college students are more likely and yield more meaningful connections than cross-gender ties in social media and face-to-face contact (Igarashi et al. 2005). In what follows we formalize a new neighborhood interference assumption that makes it clear that only those important connections within a network lead to causal interference — this in turn suggests a novel design that allows for better estimation of the GATE in this setting.

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 ($A_{ij} = A_{ji}$). Importantly, the development in this article only requires knowledge of the network A and does not assume access to any additional information about the individuals in the network. As such, identifying "important" connections or edges in a network must be done by simply looking at the adjacency matrix itself. To do this *we concentrate on networks that exhibit community structure and argue that within community connections are more likely to lead to interference than cross-community connections*. We discuss the potential downsides of such assumptions in Section 1.2.

Below we discuss several common simplifying causal 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.

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, (Rubin 1990; Manski 1995), 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 (Toulis and Kao 2013; Aronow and Samii 2017; Jagadeesan et al. 2020; Sussman and Airoldi 2017; Sävje et al. 2021).

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 $g_i(\boldsymbol{Z}) = \big(Z_i, \{Z_j : A_{ij} = 1\}\big)$ or the simplified neighborhood exposure assumption takes $g_i(\boldsymbol{Z}) = \Big(Z_i, \sum_{\{j:A_{ij}=1\}} Z_j > 0\Big)$. 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 and 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 effects 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 Ugander et al. (2013) and Eckles et al. (2016) 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 et al. 1988; Staber 1993; Chamberlain et al. 2007; Bail et al. 2018), which should be reflected in the design. Specifically, we consider the setting where individuals belong to distinct communities and the probability of connections between two individuals is only a function of their (likely unobserved) community membership. To formalize notation, let individual i belong to one of K 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 $\sum_{l=1}^{K} U_{il} = 1$. When the probability that individuals i and j have a connection or an edge between them is only a function of U_i and U_j, this fully specifies the stochastic block model (SBM), a type of popular network model that has received a lot of attention due to its tractability and success at describing real world networks (Holland et al. 1983; Abbe 2017; Adamic and Glance 2005).

An important aspect of such networks is that they encode homophilous relationships — individuals within the same community behave similarly towards others both in terms of how they form connections (Lorrain and White 1971; White et al. 1976; Holland et al. 1983; Faust and Wasserman 1992) and in terms of other social processes. For example, non-romantic same sex and opposite sex friendships exhibit community structure (Aukett et al. 1988). More importantly, edges within

communities are special: same-sex edges are associated with greater sharing and intimate activities. This would suggest that deploying a treatment (e.g. a pamphlet for health screenings) within such communities would likely lead to interference within the communities but not outside the communities. Similar differences have been observed in an entrepreneurial setting (Staber 1993). Beyond that, it has been observed that online (e-mail) and offline (phone, in person) communication are correlated among small social circles which again suggests that similarity within a network may be associated with similar behaviors in other domains (Kossinets and Watts 2006).

This framing motivates the definition of **community network interference** where only individuals connected by within community edges can interfere with one another. Formally, let

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

be the community network interference function — for unit i it extracts the entries of the treatment allocation vector belonging to i's friends from within the same community. We say that the community network inteference assumption holds if for treatment allocation vectors $\boldsymbol{Z}$, $\boldsymbol{Z}'$ we have $Y_i(\boldsymbol{Z}) = Y_i(\boldsymbol{Z}')$ if $g_i(\boldsymbol{Z}) = g_i(\boldsymbol{Z}')$.

This assumption is stronger than the one made in the general neighborhood interference literature since it further restricts interference, but it is justifiable when treatment is personal and local (e.g health or finances related) and community structure is likely. For example, in Section 5, we base our simulations on network data from an anti-bullying experiment in middle schools (Paluck et al. 2020, 2016), where there are natural friendship communities (gender and grade) as well as possible latent communities.

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 randomizing treatment. 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.

Potential limitations and criticisms of the assumption. It is important to note that we are particularly concerned with interference (the treatment of my friend affects my outcome) rather than contagion (the outcome of my friend affects my outcome). In contagious settings, such as when studying product adoption (Aral et al. 2009), it has been documented that cross-community ties are crucial for broad adoption of a product. Beyond this, the strength of weak ties has been touted as an important avenue for economic mobility (Granovetter 1973; Aldrich and Dubini 1991), which contrasts with our notion of important edges in a network. Importantly, such "weak" edges are usually defined as cross-community edges with few mutual connections and low interaction frequency (Rajkumar et al. 2022). A clear example of this is a network with directed edges such as Twitter: low-status (regular) individuals may follow high-status (influencer) individuals thus forming essentially two communities where only one type of cross-community edges are important for interference. In such a setting, treating a high status individual may lead to changes in the outcomes of lower status individuals, but not the other way around. Evaluating or developing trials on such networks using our proposed assumption would be inappropriate.

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.

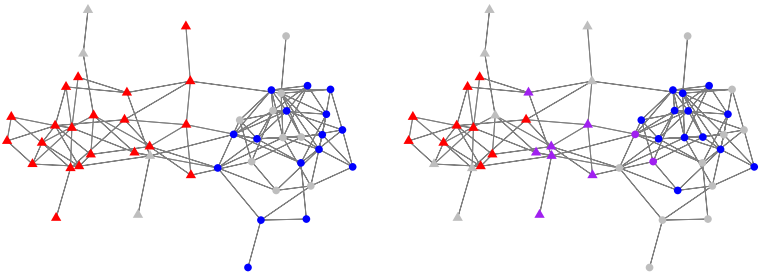

Fig. 1 This figure shows two versions of the same network where each has a clustering design implemented. The left plot shows example cluster assignments under community informed design while the right shows standard graph cluster randomization. This network is generated from a 2 community stochastic 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

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

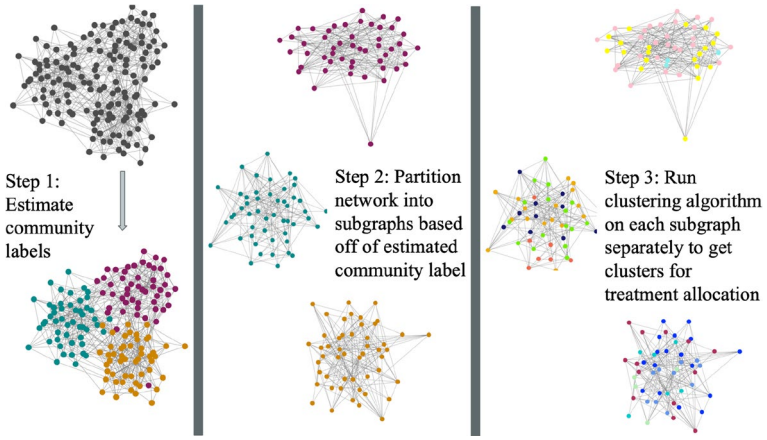

Fig. 2 This figure is a visualization representation for steps 1-3 in Algorithm 1

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.

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

1. 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
2. Create a community induced sub-graph for each community based off of $\hat{\boldsymbol{U}}$ that contains only within community edges
3. 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, $C_{\hat{U}}$. This ensures that only nodes belonging to the same community can be in the same cluster.
4. Assign treatment according to some design, $s(C_{\hat{U}})$. This could denote a pairing design, randomization designs with covariate information, etc.

2.1 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 and Bickel 2014; Bruna and Li 2017; Rohe et al. 2011; Mathews et al. 2019; Blondel et al. 2008), 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; Reeves et al. 2019; Mayya and Reeves 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 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 $\hat{\boldsymbol{U}}$.

Throughout we assume that the number of communities K is known, but note that this is not a limitation of the proposed approach, as the number of communities can be learned directly from the adjacency matrix. To address this we consider simulations in Section 5 where the community detection problem with K known is easier or harder — when it is easy then there is significant separation of the eigenvalues of the Graph Laplacian and so K can be easily identified from the data; when it is harder then it would be similarly hard to estimate K from the adjacency matrix and so performance of the procedure would be expected to suffer (Rohe et al. 2011; Newman and Reinert 2016; Budel and Van Mieghem 2020; Geng et al. 2019). This is addressed in the simulation section.

2.2 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 used again, the goals of the previous and current step differ, and, as such, different algorithms are better suited for each task. For example, in Eckles et al. (2016) the authors report that performing this step with the Louvain algorithm (Blondel et al. 2008) simulataneously results in large variance increases and large bias decreases as compared to the 3-net algorithm.

We use 3-net due to its positive performance in the literature as demonstrated in Ugander and Yin (2020). Further, this algorithm is equipped with useful theoretical properties for estimating network exposure probabilities that are 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.

2.3 Randomization of clusters

Finally, for this work, $s(C_{\hat{U}})$ 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

For the theoretical development in this section we will consider a further restriction on the community network interference assumption by assuming that each within community friend contributes equally and so the amount of interference is summarized by the total number of within community friends of i who are treated. As such, let $g_i = \sum_{j \in N_i^c} Z_j$ (where N_i^c is the neighborhood of within community neighbors of i) and so in a slight abuse of notation we have $g_i(\mathbf{Z}) = (Z_i, g_i)$, allowing us to write the potential outcome as $Y_i(Z_i, g_i)$. Under this assumption, the potential outcomes can be decomposed as $Y_i(Z_i, g_i) = C_i(Z_i) + B_i(g_i)$ (this decomposition has been used when studying interference in e.g. Karwa and Airoldi (2018), Sussman and Airoldi (2017), and Jagadeesan et al. (2020)). Here $Y_i(1, g_i) - Y_i(0, g_i) = C_i(1) - C_i(0)$ can be interpreted as the direct effect when keeping g_i constant. Similarly, $B_i(g_i)$ relates to the difference in potential outcomes of the indirect effect when holding the treatment of individual i constant. The true GATE, τ, is then equal to:

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

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:

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

The bias can then be written as:

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

where we define the following weights as:

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

and

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

The values of the above weights are determined by the chosen design and accompanying interference assumptions. Note that for ease of exposition we treat N_T and N_C as fixed quantities. Treating them as fixed is not an overly unrealistic or stringent assumption: in sufficiently large networks where within community degree is concentrated, the clusters will largely have the same number of individuals, meaning that knowing the probability of assignment will also provide explicit knowledge of N_T and N_C under complete or paired randomizations. We further note that it is reasonable to expect N_T to be approximately the same under the different designs we consider (CID and GCR) and so the calculations below would still reduce to comparisons of inclusion probabilities. Below we consider $N_T = N_C = 1/2$, though the choice of any $N_T = Nq, N_C = N(1-q)$ would be appropriate.

Following the calculations for general interference in Karwa and Airoldi (2018), the bias can be rewritten as:

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

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

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

Above, $g_i \neq \{0, d_i^c\}$ is equivalent to $g_i \in \{1, \ldots, (d_i^c - 1)\}$ 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_{max}^c 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, $\mathbb{P}(Z_i = 1, G_i = d_i^c) \geq \frac{0.5}{d_{max}^c \times (1+d_{max}^c)}$. By assumption, to obtain $\pi_i^{CID}(1, d_i^c)$, we simply divide $\mathbb{P}(Z_i = 1, G_i = d_i^c)$ by $N/2$:

$$\frac{1}{Nd_{max}^c \times (1+d_{max}^c)} \leq \pi_i^{CID}(1,d_i^c) \leq \frac{1}{2N} \leq \frac{1}{N}, \tag{6}$$

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, $\pi_i^{CID}(0, d_i^c) = 0$ and $\pi_i^{CID}(1, d_i^c) < 1/N$. However, we are not guaranteed that $\pi_i^{GCR}(0, d_i^c) = 0$. Further, under GCR, $\mathbb{P}(Z_i = 1, G_i = d_i^c) \geq \frac{0.5}{d_{max} \times (1+d_{max})}$. Since $d_{max} \geq d_{max}^c$, 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 $\mathbb{P}(Z_i = 1, G_i = d_i^c)$ 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 $g_i \approx d_i^c$ 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 $\mathbb{P}(1, d_i^c)$ 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

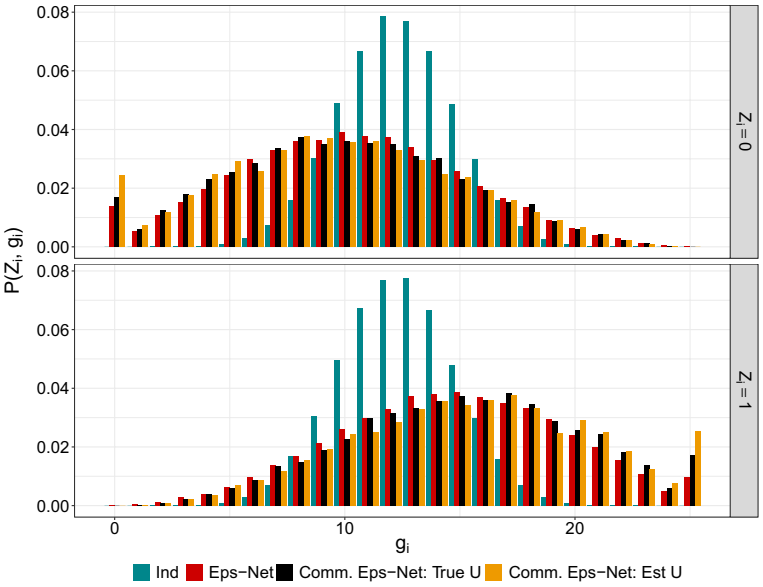

Fig. 3 Simulated $\mathbb{P}(Z_i, g_i)$ over 4000 simulations. This figure demonstrates how community informed design puts higher probability on more desirable quantities

a model assisted approach (Basse and Airoldi 2018; Särndal et al. 2003). Consider the following outcome model:

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

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

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 et al. 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 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:

$$\begin{aligned} &E[Y_i(Z=1) - Y_i|Z_i = 1] \\ &\quad = E[\gamma U_i \sum_j A_{ij} \times 1[U_i = U_j] - \gamma U_i \sum_j A_{ij} \times 1[U_i = U_j] \times Z_j|\hat{U}] \end{aligned}$$

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:

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

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 $K = 2$:

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

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 without regard for network structure.
- **Eps-Net**: Epsilon net that knows about network structure but does not know about communities (standard GCR as in Eckles et al. 2016).
- **Community Eps-Net with True** U : Performs epsilon net on each community sub-graph using ground truth community labels.
- **Community Eps-Net with Estimated** U : Performs epsilon net on each community sub-graph using estimated community labels.

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, 5.2, 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.

For each distinct SBM specification, we simulate multiple networks and multiple randomization clusterings for each network. As a result, the fraction of nodes with incorrectly estimated labels acts as a proxy for the difficulty of community detection. To ensure that community aware designs that *know* the community structure remain consistent across the board (when interference is community driven), we keep the within community probabilities the same and only vary the across community probabilities to make the community detection problem more difficult. Throughout we assume that the number of communities K is known apriori and so it does not need to be estimated. Since we include a spectrum of easy to difficult community detection problems in our simulations, these can also serve as a proxy for the ease of estimating the number of communities explicitly: when the problem is easy then K is easy to estimate and so does not warrant its own study; when the problem is hard then K is also hard to estimate (and would likely be under estimated due to the nature of community detection) and so CID would look more similar to GCR.

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, 5, 6. The $x-$axis in these figures represents the difficulty of the community detection problem, measured in terms of the fraction

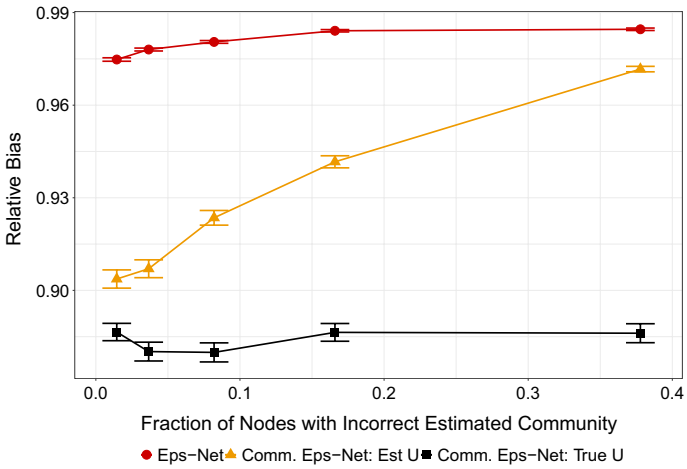

Fig. 4 Results for community interference simulation. The x-axis is the average fraction of nodes with incorrectly estimated labels for a given regime. The y-axis is relative bias for the difference in means estimator for different designs compared to independent random assignment. Standard error bars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

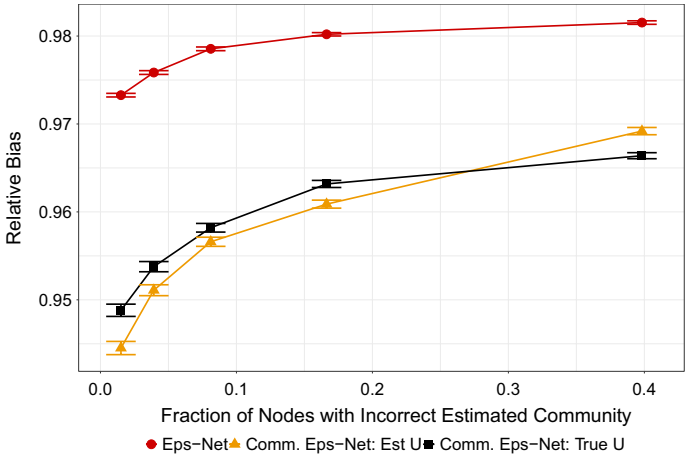

Fig. 5 Results for community agnostic interference. The x-axis is the average fraction of nodes with incorrectly estimated labels for a given regime. The y-axis is relative bias for the difference in means estimator for different designs compared to independent random assignment. Standard error bars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

of nodes assigned to the wrong community during the community detection step. 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. We also note that GCR appears to be not very sensitive to the changes in the network in this

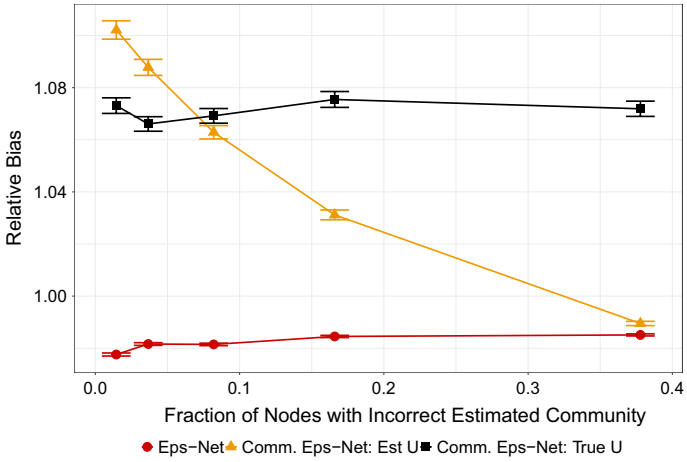

Fig. 6 Results for anti-community interference. The x-axis is the average fraction of nodes with incorrectly estimated labels for a given regime. The y-axis is relative bias for the difference in means estimator for different designs compared to independent random assignment. Standard error bars are over 2000 simulations

simulation; this is likely because the vast majority of edges in the network are within community (even when the community detection problem is more difficult), and so the changing network structures likely do not induce substantively different biases in estimation.

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 substantially more within community edges than across community edges across the networks we study in this section and so most of the interference happens within a community.

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. Similarly, we observe standard GCR performing similarly to the setting of community informed interference since it suffers from a symmetric misspecification (that some edges are less important than others).If the anti-community interference 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 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 $a = 0.04$ and $b = 0.03/(K - 1)$. 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.

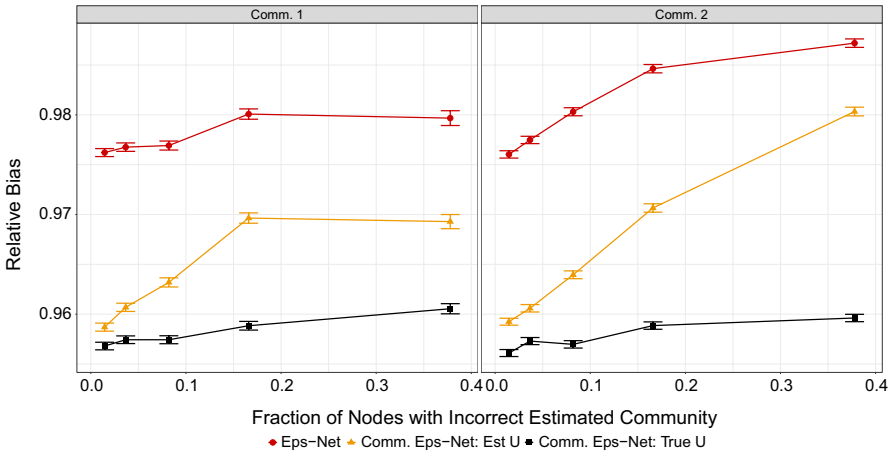

Fig. 7 Relative bias for the difference in means estimator for different designs compared to independent random assignment for community level treatment effect estimation over 2000 simulations reported in Section 5.2

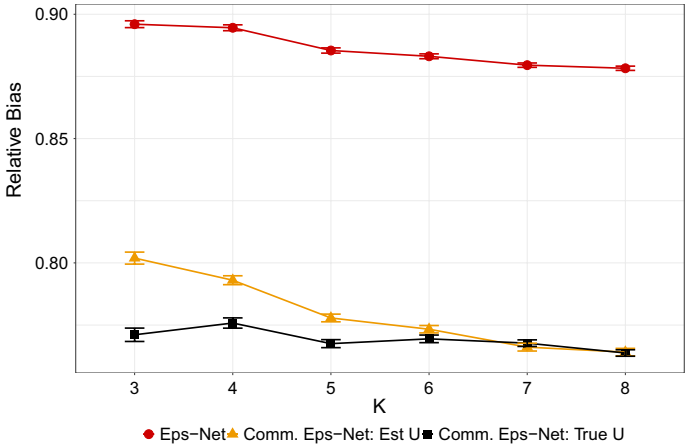

Fig. 8 The y-axis is relative bias for the difference in means estimator for different designs compared to independent random assignment for varying K equally sized communities with standard error bars over 2000 simulations per value of K

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 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. Table 1 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.

Table 1 Results for the relative bias of epsilon-net and community informed design to independent assignment in the setting where there is no community structure or interference. Simulations are based on the response model from Section 6.2 of Ugander and Yin (2020). Standard errors (in parenthesis) are over 100 simulations. Parameter details are provided in the supplement

Peer influence	Eps-Net	Comm. Eps-Net: Est U
0.5	0.78 (0.01)	0.79 (0.01)
1	0.77 (0.01)	0.78 (0.01)
1.5	0.77 (0.01)	0.78 (0.01)
2	0.77 (0.00)	0.77 (0.01)

5.5 Potentially mismatched communities

Much of the work on network analysis has been driven by the study of student-student networks (Hoff et al. 2013; Rienties and Nolan 2014; Mayer and Puller 2008; Sentse et al. 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 and Volfovsky 2021). Motivated by a recent collection of network-driven experiments on the impact of anti-bullying interventions (Paluck et al. 2020, 2016) 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 ($N = 126, 104, 370, 530$) 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, $\boldsymbol{\gamma}$, 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 Table 2, 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

Table 2 Relative bias for four schools from the middle school data set where outcomes are simulated based off of the true observed networks with standard error over 750 simulations. All standard errors were $< 10^{-2}$ and are omitted for clarity of notation

	A	B	C	D
Ind	6.57	8.65	7.21	7.96
Eps-Net	4.89	3.12	5.26	6.19
Comm. Eps-Net: Est U	3.82	1.83	4.65	6.03
Comm. Eps-Net: True U	3.49	1.88	4.19	5.11

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 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 (such as non-within-community edges *also* leading to interference) 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 et al. (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 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 on community detection, it is a natural future direction to incorporate covariate information into the community informed design procedure. This can be done by incorporating covariates into the community detection problem directly (Binkiewicz et al. 2017; Yan and Sarkar 2021; Shen et al. 2022) or by considering community detection as a component of a network regression problem (Mathews and Volfovsky 2021; Hoff 2008). 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.

In addition to incorporating covariate information, alternative estimators such as the Horvitz-Thompson, Hajek and other inverse probability weighted estimators can be used to address some bias concerns. This strategy has been leveraged in several causal inference with networks settings (Aronow and Samii 2017), including the works studying graph cluster randomization (Ugander et al. 2013; Eckles et al. 2016; Ugander and Yin 2020). We eschew these estimators in favor of the difference in means estimator because of a desire for an exceedingly simple to explain estimator, concerns of very small exposure probabilities (leading to large variances of estimators such as the Horvitz-Thompson) and the general difficulty in computing these exposure probabilities (with exact computation often being NP-hard). For example, Ugander and Yin (2020) use an order of magnitude more Monte Carlo samples as there are nodes in the network to compute the probabilities and still sometimes end up with relative errors that are order 0.1. Nonetheless, inverse weighting plays an important role in causal inference and may provide some robustness properties against misspecification of exposures (Sävje et al. 2021).

As we mention in the introduction and discuss in the simulation section, there are settings where the within-community interference assumption is inappropriate. This suggests the need for data-driven adaptive designs that can first identify the type of interference and then implement an appropriate high quality experimental design that targets that interference pattern specifically. One potential approach for settings where the population of interest is sufficiently large is to first perform a small pilot experiment to identify the type of interference within the network (e.g. whether community informed or anti-community informed) followed by a large scale, properly calibrated experiment on the larger population. Identifying the type of interference can be achieved using exact testing or by imposing parametric assumptions on the potential outcomes (Athey et al. 2018; Puelz et al. 2019). These tests are have been shown to be powered against certain alternatives, but much work is still needed to properly identify the alternatives of interest in our settings and to fully analyze such adaptive data-driven designs.

Beyond studying the GATE, there is a great interest in the interference literature in estimands relating to a fraction of individuals being treated versus none at all or to estimating direct or indirect effects separately. This is usually addressed in the context of developing estimators but there is a dearth of literature on novel designs for such contexts. One of the complicating factors is that the same design rarely allows for high quality estimation of multiple estimands (Jagadeesan et al. 2020). Our design is unfortunately no different since it targets the GATE specifically and, for example, would not be effective at estimating direct effects. Some strategies, such as potentially combining two stage randomization (Hudgens and Halloran 2008) with CID or GCR or studying marginalized estimands such as in Sävje et al. (2021) appear promising in allowing us to expand the class of estimands we can study using simple estimators as those used throughout this paper.

Acknowledgements The authors gratefully acknowledge financial support from the Statistical and Applied Mathematical Sciences Institute, the National Science Foundation (DMS 2046880) and the Army Research Institute. (W911NF1810233).

Declarations

Competing interests The authors do not have any competing interests.

Code availability Code will be made available for all simulation studies and no additional data was generated for this manuscript.

References

Abbe E (2017) Community detection and stochastic block models: recent developments. J Mach Learn Res 18(1):6446–6531

Adamic LA, Glance N (2005) The political blogosphere and the 2004 us election: divided they blog. Proceedings of the 3rd international workshop on link discovery (pp. 36–43)

Aldrich H, Dubini P (1991) Personal and extended networks are central to the entrepreneurial process. J Bus Ventur 6(5):305–313

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influencebased contagion from homophily-driven diffusion in dynamic networks. Proc Nat Acad Sci 106(51):21544–21549

Aronow PM, Samii C (2017) Estimating average causal effects under general interference, with application to a social network experiment. Ann Appl Stat 11(4):1912–1947

Athey S, Eckles D, Imbens GW (2018) Exact p values for network interference. J Am Stat Assoc 113(521):230–240

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 CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MF, Volfovsky A (2018) Exposure to opposing views on social media can increase political polarization. Proc Nat. Acad. Sci. 115(37):9216–9221

Basse GW, Airoldi EM (2018) Model-assisted design of experiments in the presence of network-correlated outcomes. Biometrika 105(4):849–858

Bhattacharyya S, Bickel PJ (2014) Community detection in networks using graph distance. arXiv preprint arXiv:1401.3915

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

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):P10008

Bruna J, Li X (2017) Community detection with graph neural networks. Stat 1050:27

Budel G, Van Mieghem P (2020) Detecting the number of clusters in a network. J Complex Netw 8(6):047

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):230–242

Eckles D, Karrer B, Ugander J (2016) Design and analysis of experiments in networks: reducing bias from interference. J Causal Inference 5(1):7530

Faust K, Wasserman S (1992) Blockmodels: interpretation and evaluation. Soc Netw 14(1–2):5–61

Geng J, Bhattacharya A, Pati D (2019) Probabilistic community detection with unknown number of communities. J Am Stat Assoc 114(526):893–905

Granovetter MS (1973) The strength of weak ties. Am J Soc 78(6):1360–1380

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

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

Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. Soc Netw 5(2):109–137

Hudgens MG, Halloran ME (2008) Toward causal inference with interference. J Am Stat Assoc 103(482):832–842
Igarashi T, Takai J, Yoshida T (2005) Gender differences in social network development via mobile phone text messages: A longitudinal study. J Soc Pers Relatsh 22:691–713
Jagadeesan R, Pillai NS, Volfovsky A (2020) Designs for estimating the treatment effect in networks with interference. Ann Stat 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 EM (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)
Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. Science 311(5757):88–90
Krzakala F, Moore C, Mossel E, Neeman J, Sly A, Zdeborová L, Zhang P (2013) Spectral redemption in clustering sparse networks. Proc Nat Acad Sci 110(52):20935–20940. https://doi.org/10.1073/pnas.1312486110
Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. J Math Soc 1(1):49–80
Manski CF (1995) Identification problems in the social sciences. Harvard University Press, Cambridge
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 SL (2008) The old boy (and girl) network: social network formation on university campuses. J Pub Econ 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
Newman ME, Reinert G (2016) Estimating the number of communities in a network. Phys Rev Lett 117(7):078301
Paluck EL, Shepherd H, Aronow PM (2016). Changing climates of conflict: A social network experiment in 56 schools. Proc Nat Acad Sci, 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 EL, Shepherd HR, Aronow P (2020) Changing climates of conflict: a social network experiment in 56 schools. Proceedings of the National Academy of Sciences. NJ 10.3886/ICPSR37070.v2
Puelz D, Basse G, Feller A, Toulis P (2019). A graph-theoretic approach to randomization tests of causal effects under general interference. arXiv preprint arXiv:1910.10862
Rajkumar K, Saint-Jacques G, Bojinov I, Brynjolfsson E, Aral S (2022) A causal test of the strength of weak ties. Science 377(6612):1304–1310
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. Int J Intercult Relat 41:165–180
Rohe K, Chatterjee S, Yu B et al (2011) Spectral clustering and the highdimensional stochastic blockmodel. Ann Stat 39(4):1878–1915
Rubin DB (1990). Formal mode of statistical inference for causal effects. J Stat Plann 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 and Business Media, Berlin
Sävje F (2021). Causal inference with misspecified exposure mappings. arXiv preprint arXiv:2103.06471
Sävje F, Aronow PM, Hudgens MG (2021) Average treatment effects in the presence of unknown interference. Ann Stat 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. J Youth Aadolesc 43(9):1409–1420

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

Staber U (1993) Friends, acquaintances, strangers: gender differences in the structure of enterpreneurial networks. J Small Bus Entrep 11:73–82

Sussman DL, Airoldi EM (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. In International conference on machine learning. PMLR, NY, 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

White HC, Boorman SA, Breiger RL (1976) Social structure from multiple networks. i. blockmodels of roles and positions. Am J Soc 81(4):730–780

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. J Am Stat Assoc 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