The Annals of Statistics
2020, Vol. 48, No. 2, 679–712
https://doi.org/10.1214/18-AOS1807

DESIGNS FOR ESTIMATING THE TREATMENT EFFECT IN NETWORKS WITH INTERFERENCE

BY RAVI JAGADEESAN[1], NATESH S. PILLAI[2] AND ALEXANDER VOLFOVSKY[3]

[1]*Harvard Business School, ravi.jagadeesan@gmail.com*
[2]*Department of Statistics, Harvard University, pillai@fas.harvard.edu*
[3]*Department of Statistical Science, Duke University, alexander.volfovsky@duke.edu*

In this paper, we introduce new, easily implementable designs for drawing causal inference from randomized experiments on networks with interference. Inspired by the idea of matching in observational studies, we introduce the notion of considering a treatment assignment as a "quasi-coloring" on a graph. Our idea of a perfect quasi-coloring strives to match every treated unit on a given network with a distinct control unit that has identical number of treated and control neighbors. For a wide range of interference functions encountered in applications, we show both by theory and simulations that the classical Neymanian estimator for the direct effect has desirable properties for our designs.

1. Introduction. In this paper, we construct and analyze new designs for estimating treatment effects from randomized experiments in networks with interference. With the proliferation of network data and the steady increase in the number of experiments conducted on networks, understanding the behavior of individuals in a network has become an important issue in many scientific fields. Epidemiologists study the transmission of disease over social networks [1], computer scientists are interested in information diffusion in large computer networks [9, 26] and sociologists study the effects of school integration on friendship networks [17]. While much of the early statistical work on networks focused on models for understanding network formation [10, 12], there has been a recent surge in drawing causal inference from experiments on networks [7, 21–24].

A time-honored approach to performing causal inference from randomized experiments entails the following steps [11, 18, 19]: (i) define the population of units, (ii) define the treatment assignment and (iii) define the quantity (or estimand) of interest. When an experiment is conducted on a network, we must revisit each of these elements. First, the object of inference can be the network, the edges of the network or the nodes of the network. We focus on the case where the nodes are the experimental units and our population is just the observed units. Next, the treatment assignment mechanisms proposed in this paper are conditional on a given network and thus the events that any two units receive treatment are not independent. This choice is in stark contrast to usual Bernoulli-type randomization mechanisms where treatment is assigned to units independently or with very weak dependence. Finally, our estimand of interest is the *direct* treatment effect—that is, the effect of treatment on the treated unit irrespective of the treatment status of the rest of the network—discussed below.

Much of the recent work on causal inference on networks studies generic Bernoulli-type randomization schemes and construct various estimators for minimizing their Mean-Squared Error (MSE); notable exceptions are the recent papers [7, 8]. In contrast, we fix an estimator of interest and focus on the *design* of treatment assignments. We study the classical Neymanian estimator that takes the difference between the means of the outcome for treated nodes

Received August 2017; revised December 2018.
MSC2010 subject classifications. 62-02, 62K05, 91D30.
Key words and phrases. Experimental design, network interference, Neyman estimator, symmetric interference model, homophily.

and the control nodes. Our approach is motivated by two key observations. First, the Neymanian estimator is ubiquitously used. It is a natural estimator for the direct effect and improves on reweighted versions of it (such as Horvitz–Thompson, Hajek, etc.) due to its prima facie interpretability. Second, it has been emphasized by many researchers that for objective causal inference, "design trufmps analysis" [20]. It is known that the Neymannian estimator is biased under standard designs such as Bernoulli trials (every unit has probability of treatment p) and a completely randomized design (a fraction p of the units is assigned to treatment). We consider a more natural randomization scheme that works to remove the effects of interference by balancing relevant distributions of interference-relevant parameters between treated and untreated nodes.

Conceptually, our main contribution is the idea of considering a treatment assignment as a "quasi-coloring" of a graph (see Definition 5.1). Roughly speaking, a treatment assignment is a *perfect quasi-coloring*,[1] if for every treated vertex v (represented by black dots, say), there is a nontreated vertex v' (represented by white dots) that has the same number of treated and nontreated neighbors as that of v. Thus having a perfect quasi-coloring on a graph G ensures that one can color the graph in such a way that for every black vertex, there exists a *distinct* white vertex with identically colored neighbors. Figure 1 shows two ways to color the nodes of a square, where one coloring is a perfect quasi-coloring and the other is not.

Our notion of perfect quasi-colorings is inspired by the notion of covariate balance in the context of matching in observational studies. For any given network, if a treatment assignment mechanism satisfies our notion of quasi-coloring, we prove that the Neymanian estimator for the direct treatment effect is unbiased for a wide range of families of interference effects encountered in practice. This result replicates the behavior of the Neymanian estimator in classical randomized experiments.

It turns out that, for many graphs, perfect quasi-colorings are not available or may be very difficult to construct. To circumvent this issue, we develop treatment assignment mechanisms that correspond to "approximately perfect quasi-colorings." The closer an approximately perfect quasi-coloring is to a perfect quasi coloring, the smaller its bias. Based on this notion, we develop a new *restricted randomization* design that reduces bias and variance. In networks in which a perfect quasi-coloring is not possible, we provide easily implementable algorithms to construct designs with desirable properties—see the "partitioning by degree" design in Definition 6.1. This design implements a stratified sampling method—where vertices are stratified by degree—to ensure similarity between the treated and control groups. In settings with additional covariates, vertices can be stratified by this additional information as well to ensure balancedness in covariates between the treated and control groups—see Section 7.

We derive upper bounds on the bias and variance of our estimator under a few different settings of approximate quasi-colorings. These results are then used to prove the asymptotic

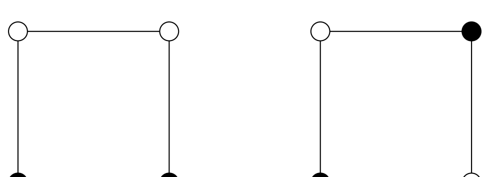


FIG. 1. *The coloring of G on the left is a perfect quasi-coloring, since both the black vertices and the white vertices have exactly two neighbors of opposite colors. The coloring of G on the right is not a perfect quasi-coloring because both the black vertices have two white neighbors where as both the white vertices have two black neighbors.*

[1] The word coloring is reserved for something specific in graph theory; thus we use the phrase "quasi-coloring."

consistency of the Neymannian estimator for our proposed randomization schemes in both dense and sparse asymptotic regimes for network growth. We further show that our results persist in settings in which interference effects depend on covariates for suitable stratified sampling schemes. We demonstrate the efficacy of our proposed randomization scheme in a series of simulations, varying both the type of interference and the network. Our proofs are different in the cases of dense and sparse graphs and are thus of independent interest in the two cases.

1.1. *Background and literature.* In situations in which the experimental units are connected in a network, some of the usual assumptions used in other settings are not likely to hold. For example, the stable unit treatment value assumption [18] requires that the outcome for a unit only depends on its own treatment, and in particular is independent of the treatment assignment mechanism. For networks this condition can be violated in several ways: it is likely that either the outcomes of units are associated with the treatment of their network neighbors (*interference*) or that the treatment effect passes temporally across the network (*contagion*). Further complications arise due to the likely similarity in behavior of connected units (*homophily*). It has been previously demonstrated that while interference and contagion can affect causal inference on a network differently, they are difficult, if not impossible, to distinguish [21]. These complications lead to a difficulty in specifying an estimand of interest [13]. The four main estimands in the presence of a network are (i) the effect of treatment were it applied to the whole network versus no one in the network (*total network treatment effect*), (ii) the direct effect of treatment on the treated unit irrespective of the treatment status of the rest of the network (*direct treatment effect*), (iii) the spillover effect of treatment of the network on a single unit irrespective of its treatment (*indirect treatment effect*) and (iv) the sum of the direct and indirect effect (*total nodal treatment effect*).

Different estimands lead to different inference procedures—both from a design and an analysis point of view. We focus on the design of experiments targeting the direct treatment effect. Other recent work has targeted different estimands. In [2], the authors consider different reweighting and post-stratification estimators for settings where the interference is characterized by an exposure mapping. In particular, they require the probability of all possible exposures (the combination of a unit's treatment with the interference the unit experiences) of a unit to be greater than zero. This positivity assumption may not be tenable as the number of exposures grows, such as if the true interference is linear in the number of treated neighbors. In [6], the author studies estimators for monotone treatment effects and constructs asymptotically consistent bounds for such estimates. The paper [7] studies total network effects by considering a cluster-randomized-design in conjunction with Horvitz–Thompson and Hajek estimators. In [22], the authors construct unbiased estimators for direct and indirect treatment effects for a fixed design. Of particular interest here is Theorem 6.5 of [22] that demonstrates that the post-stratified estimator has minimal integrated variance under strong symmetry conditions (on the graph and on the design). We explore the performance of this estimator in a completely randomized design in Section 9.

The aforementioned papers study the effects of interference on estimation and many make the common assumption that the interference is limited to the immediate neighborhood of a node, or at least employ functions of the adjacency matrix in specifying the procedure for estimating effects. When interference can be arbitrary, the performance of these procedures depends on the deviation of the true interference structure from multi-hop neighborhood interfere. We will also make this assumption, but our work can be easily generalized to different patterns of interference (see Section 10). Another simplifying assumption that is frequently made requires the interference effect to be symmetric—that is, that each interfering unit contributes the same indirect effect. We demonstrate results under several classes of interference patterns in which this assumption fails.

1.2. *Paper guide.* In Section 2, we introduce a basic model for interference and the Neymanian estimator. In Section 3, we discuss some restricted randomizations. In Section 4, we describe a symmetric interference model. In Section 5, we define our notion of quasi-coloring. We derive the bounds for the MSE of the Neymanian estimator in Section 6. In Section 7, we relax the assumption that interference is symmetric. In Section 8, we study the effects of (possibly latent) heterogeneity in baseline covariates on the treatment effect estimation. The results from a simulation study are given in Section 9. We close with a short discussion. The proofs are given in the appendices.

1.3. *Notation.* Fix $n \in \mathbb{N}$ and let G be a graph with $|V(G)| = rn$. The treatment units are nodes of G. We treat pn nodes of G and leave qn nodes as controls—where $p + q = r$—so $\frac{p}{r}$ fraction of the nodes are treated.

Throughout the paper, we will assume that G has no isolated vertices. Let $\mathcal{N}(v)$ denote the set of neighbors of a vertex $v \in V(G)$ and let $d(v) = |\mathcal{N}(v)|$ denote the degree of v. Also define the minimum and maximum degrees

$$d_{\min} = \min_{v\in V(G)} d(v), \qquad d_{\max} = \max_{v\in V(G)} d(v). \tag{1.1}$$

We will denote by $\binom{V(G)}{k}$ the family of all k-element subsets of $V(G)$. Similarly, for $1 \leq m_i \leq rn$ with $\sum m_i = rn$, define

$$\binom{V(G)}{m_1, \dots, m_k} = \{(A_1, \dots, A_k), A_k \subset V(G), |A_k| = m_k, A_k \cap A_\ell = \varnothing, \forall k \neq \ell\}.$$

In particular, $\binom{V(G)}{r,\dots,r}$ denotes the set of all partitions of $V(G)$ into sets of size r. For $r \in \mathbb{N}$, the set $\{1, 2, \dots, r\}$ is denoted by $[r]$. For sets $A, B \subset V(G)$, $A \Delta B$ denotes the symmetric set difference.

For $\mathrm{T} \subset V(G)$, let $1_{\mathrm{T}}(\cdot)$ denote the indicator function

$$1_{\mathrm{T}}(v) = \begin{cases} 1 & \text{if } v \in \mathrm{T}, \\ 0 & \text{if } v \notin T. \end{cases}$$

For $\mathrm{T} \subseteq V(G)$ and $v \in V(G)$, let

$$\chi_v^{\mathrm{T}} = \begin{cases} q & \text{if } v \in \mathrm{T}, \\ -p & \text{if } v \notin \mathrm{T}. \end{cases}$$

For the reader's convenience, we finish this section with a table (Table 1) of notation for the key quantities introduced in future sections.

2. The model and the estimator. For each vertex $v \in V(G)$, let $x_v, t_v \in \mathbb{R}$ be constants and let $f_v : 2^{\mathcal{N}(v)} \to \mathbb{R}$ be a function such that $f_v(\varnothing) = 0$ for all $v \in V(G)$. We study the linear model

$$y_v = x_v + 1_{\mathrm{T}}(v)t_v + f_v(\mathrm{T} \cap \mathcal{N}(v)), \qquad v \in V(G), \tag{2.1}$$

where $\mathrm{T} \subset V(G)$ denotes the treatment group. The quantity x_v reflects the outcome for node v under control. The function f_v denotes the interference effect. For every vertex v, it is only a function of its treated neighbors $\mathrm{T} \cap \mathcal{N}(v)$.

This model (without observed covariates) is a member of the class of neighborhood interference models introduced by [22]. In particular, they demonstrate that this parametrization is equivalent to the potential outcomes notation of [18] under specific assumptions on the additivity and symmetry of the effects. In particular, equation (2.1) corresponds to the additivity

TABLE 1
Table of notation

x_v:	covariate of v
t_v:	direct treatment effect on v
y_v:	observed outcome for v
f_v:	interference function for v
$\widehat{t}_{\mathrm{Neyman}}$:	Neymannian estimator
t_{ideal}:	Neymannian estimator ignoring interference
$\mathcal{B}$:	set of possible bidegrees
f:	interference function in a symmetric model
$\mathbf{d}$:	metric on $\mathcal{B}$
$\vec{d}_{\mathrm{T}}$:	T-bidegree
$\mathcal{B}_\pi$:	set of possible bidegrees of vertices in π
f_π:	interference function in a symmetric model with types
$\mathbf{d}_\pi$:	metric on $\mathcal{B}_\pi$

of main effects assumption (ANIA), the second-most-general model in [22]. That is, x_v is the baseline, t_v is the direct treatment effect (defined as the effect of treatment on node v when no one else is treated) and f_v is the interference effect. While new estimators for the average treatment effect were constructed in [22], we focus on better designs for the Neymanian estimator defined below.

Formally, we define the average direct treatment effect as

$$\bar{t} = \frac{1}{rn} \sum_{v \in V(G)} t_v, \tag{2.2}$$

which is our estimand of interest. When $|\mathrm{T}| = pn$, define the Neymanian estimator

$$\widehat{t}_{\mathrm{Neyman}} = \frac{1}{pqn}\left(q \sum_{v \in \mathrm{T}} y_v - p \sum_{v \in V(G)\setminus \mathrm{T}} y_v\right) = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} y_v. \tag{2.3}$$

When $p = q = 1$ and $r = 2$, the estimator $\widehat{t}_{\mathrm{Neyman}}$ has the usual form

$$\widehat{t}_{\mathrm{Neyman}} = \frac{1}{n}\left(\sum_{v \in \mathrm{T}} y_v - \sum_{v \in V(G)\setminus \mathrm{T}} y_v\right).$$

Define the quantity

$$t_{\mathrm{ideal}} = \frac{1}{pqn}\left(q \sum_{v \in \mathrm{T}} (x_v + t_v) - p \sum_{v \in V(G)\setminus \mathrm{T}} x_v\right). \tag{2.4}$$

The difference[2]

$$\xi = \widehat{t}_{\mathrm{Neyman}} - t_{\mathrm{ideal}} = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} f_v\big(\mathrm{T} \cap \mathcal{N}(v)\big) \tag{2.5}$$

is the "net interference effect" on the Neymannian estimator. We first show that bounds on $|\mathbb{E}_{\mathrm{T}}(\xi)|$ lead to bounds on the bias of $\widehat{t}_{\mathrm{Neyman}}$. Here, $\mathbb{E}_{\mathrm{T}}$ denotes that the expectation is taken over the treatment assignment mechanism.

LEMMA 2.1. *Suppose that* $\mathrm{T} \subset V(G)$ *is selected in a fashion so that* $\mathbb{P}(v \in \mathrm{T}) = \frac{p}{r}$ *for all* $v \in V(G)$. *Then*, $\mathbb{E}_{\mathrm{T}}(\widehat{t}_{\mathrm{Neyman}}) - \bar{t} = \mathbb{E}_{\mathrm{T}}(\xi)$.

[2]The quantity ξ is a function of the treatment T, but we suppress this dependence for notational convenience.

PROOF. The quantity t_{ideal} in (2.4) can be written as

$$t_{\text{ideal}} = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}}\bigl(x_v + 1_{\mathrm{T}}(v) t_v\bigr).$$

Since $\mathbb{P}(v \in \mathrm{T}) = \frac{p}{r}$ for all $v \in V(G)$, we obtain that

$$\mathbb{E}_{\mathrm{T}}(t_{\text{ideal}}) = \frac{1}{pn} \sum_{v \in V(G)} \mathbb{P}(v \in T) t_v + \frac{1}{pqn} \sum_{v \in V(G)} x_v \mathbb{E}_{\mathrm{T}}\bigl(\chi_v^{\mathrm{T}}\bigr) = \bar{t}.$$

Thus, we have

$$\mathbb{E}_{\mathrm{T}}(\widehat{t}_{\text{Neyman}}) - \bar{t} = \mathbb{E}_{\mathrm{T}}(t_{\text{ideal}} + \xi) - \bar{t} = \mathbb{E}_{\mathrm{T}}(\xi),$$

proving the lemma. □

3. Restricted randomizations. Throughout this paper, we consider groups of r experimental units; in each group, p units are assigned to treatment. Fix a partition $\mathcal{P} = (S_1, \ldots, S_n) \in \binom{V}{r,\ldots,r}$ of the vertices V into sets $S_i = \{w_i^1, \ldots, w_i^r\}$. Define a random vector $\vec{B} = (B_1, \ldots, B_n) \in \binom{[r]}{p}^n$ with B_i i.i.d. uniform on $\binom{[r]}{p}$. Conditional on the partition $\mathcal{P}$, we define our treatment assignment mechanism to be

$$\mathrm{T}_{\vec{B},\mathcal{P}} = \bigl\{w_i^j \mid j \in B_i\bigr\}. \tag{3.1}$$

Thus, we give treatment to the vertex w_i^j when $j \in B_i$.

The standard completely randomized design (CRD) for treatment assignments is recovered when $\mathcal{P}$ is sampled uniformly from the set $\binom{V(G)}{r,\ldots,r}$. In this section, we obtain bounds for the bias of $\widehat{t}_{\text{Neyman}}$ with treatment assignment $\mathrm{T}_{\vec{B},\mathcal{P}}$—for a fixed partition $\mathcal{P} \in \binom{V(G)}{r,\ldots,r}$.

3.1. *General upper bound on bias.* The following definition introduces a useful framework for quantifying the variability of the interference effect f_v across the units.

DEFINITION 3.1. For $v \in V(G)$, the function f_v is called *K_v-Lipschitz* if

$$\bigl|f_v(A) - f_v(B)\bigr| \le \frac{K_v |A \Delta B|}{d(v)} \tag{3.2}$$

for $K_v > 0$ and all $A, B \subset \mathcal{N}(v)$.

Thus, the Lipschitz constant K_v provides an upper bound on the amount that treating a proportion of the neighbors of v can affect y_v.

EXAMPLE 3.2. The linear interference function $f_v(A) = \gamma|A|$ is $|\gamma| d(v)$-Lipschitz. Moreover, the normalized linear interference function $f_v(A) = \gamma \frac{|A|}{d(v)}$ is $|\gamma|$-Lipschitz.

The following lemma bounds the bias of $\widehat{t}_{\text{Neyman}}$ with treatment $\mathrm{T}_{\vec{B},\mathcal{P}}$ when f_v is Lipschitz. The strategy of the proof is to apply Lemma 2.1 to reduce to bounding the expectation of ξ. The Lipschitz condition yields a termwise bound on ξ in (2.5). Given $v \in V$, we let

$$\mathcal{P}_v = S_i \quad \text{where } v \in S_i \tag{3.3}$$

denote the element of $\mathcal{P}$ to which v belongs.

LEMMA 3.3. *Suppose the function f_v is K_v-Lipschitz. Then for the partition $\mathcal{P}$ and the treatment assignment in* (3.1), *we have*

$$\left|\mathbb{E}_{\mathrm{T}}(\xi)\right| \leq \frac{1}{nr(r-1)} \sum_{v \in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)} K_v. \tag{3.4}$$

Lemma 3.3 yields the following important observation.

LEMMA 3.4. *If every element of $\mathcal{P}$ is an independent set in G—that is, if $\{v, v'\} \notin E(G)$ whenever $\{v, v'\} \subseteq S_i$—and $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$, then $\mathbb{E}_{\mathrm{T}}(\widehat{t}_{\text{Neyman}}) = \bar{t}$.*

PROOF. Indeed, in this case, the right-hand side of Lemma 3.3 has no terms in this case. Thus, we have $|\mathbb{E}_{\mathrm{T}}(\xi)| = 0$ and the proof follows from Lemma 2.1. $\square$

Lemma 3.4 implies that if we choose clusters (S_i) of independent sets and then randomize within those clusters for the treatment assignment, then $\widehat{t}_{\text{Neyman}}$ will be unbiased. Thus, a design principle will be to ensure that elements of $\mathcal{P}$ do not contain too many edges of G using appropriate randomizations.

3.2. *Random choices of $\mathcal{P}$.* In this section, we assume that the function f_v is K_v-Lipschitz. Define the *average Lipschitz constant*

$$\bar{K} = \frac{1}{rn} \sum_{v \in V(G)} K_v. \tag{3.5}$$

EXAMPLE 3.5. Let $f_v(A) = \gamma|A|$ for some γ. Then f_v is K_v-Lipschitz with $K_v = |\gamma| d(v)$. When the underlying graph G has average degree m, we have that $\bar{K} = |\gamma| m$.

Choosing $\mathcal{P}$ randomly can help reduce the bias, as the following proposition shows. As will be seen in the sequel, it will be helpful to restrict the randomization of $\mathcal{P}$ to reduce the MSE.

PROPOSITION 3.6. *When $\mathcal{P}$ is sampled uniformly from $\binom{V(G)}{r,\dots,r}$, we have*

$$\mathbb{E}_{\mathcal{P}}\left|\mathbb{E}_{\vec{B}}(\xi \mid \mathcal{P})\right| \leq \frac{\bar{K}}{rn-1},$$

where $\bar{K}$ is as in (3.5).

The following result is immediate from Proposition 3.6.

COROLLARY 3.7. *When T is sampled uniformly from $\binom{V(G)}{pn}$, we have*

$$\left|\mathbb{E}_{\mathrm{T}}(\xi)\right| \leq \frac{\bar{K}}{rn-1}.$$

Corollary 3.7 generalizes a result in [14] for the case of $f_v(S) = \gamma|S|$ and $p = q = 1$. As mentioned in Example 3.5, when G has average degree m, we have that $\bar{K} = \gamma m$. Thus, by Corollary 3.7, we obtain that $|\mathbb{E}_{\mathrm{T}}(\xi)| \leq \frac{\gamma m}{2n-1}$.

4. Symmetric interference model. In this section, we introduce a simple—but natural—type of interference function where the interference effect on a vertex depends only on the numbers of its neighbors that are not treated. Let

$$d\big(V(G)\big) = \big\{d(v), v \in V(G)\big\}$$

denote the set of degrees of vertices in G and let

$$\mathcal{B} = \big\{(a,b) \in \mathbb{Z}^2_{\geq 0} \mid a + b \in d\big(V(G)\big)\big\} \tag{4.1}$$

denote the set of pairs of natural numbers that sum to elements of $d(V(G))$. Hence, $\mathcal{B}$ is the set of possible bi-degrees of vertices in G (for a partition of the vertices of G into two subsets).

DEFINITION 4.1. The collection of functions $\{f_v : v \in G\}$ is called a *symmetric interference model without types* if there is a function $f : \mathcal{B} \to \mathbb{R}$ such that

$$f_v(S) = f\big(|S|, \big|\mathcal{N}(v) \setminus S\big|\big) \tag{4.2}$$

for all $v \in V(G)$.

In Definition 4.1, all vertices share the same interference function. Moreover, the interference effect on a vertex depends only on the numbers of its neighbors that are (not) treated—not the identities of its treated neighbors. In the next section, we will allow different *types* of vertices to have different interference functions.

EXAMPLE 4.2. The family of interference functions $f_v(S) = \gamma|S|$ is achieved in a symmetric interference model without types when $f(a,b) = \gamma a$ in Definition 4.1. A similar related example is that $f_v(S) = \gamma\frac{|S|}{d(v)}$ is achieved in a symmetric inference model when $f(a,b) = \frac{\gamma a}{a+b}$ in Definition 4.1.

EXAMPLE 4.3. In many natural examples, treating neighbors beyond a certain threshold number of treated neighbors does not change the interference effect. This interference pattern can be captured in our model by setting $f_v(S) = \gamma \min\{|S|, k\}$—corresponding to interference only due to the first k treated neighbors—and $f_v(S) = \gamma \min\{\frac{|S|}{d(v)}, p/r\}$—corresponding to interference by only the first p/r proportion of treated neighbors. Both of these cases are examples of symmetric interference models.

For $\mathrm{T} \subseteq V(G)$ and $v \in V(G)$, let

$$\vec{d}_{\mathrm{T}}(v) = \big(\big|\mathrm{T} \cap \mathcal{N}(v)\big|, \big|\mathcal{N}(v) \setminus \mathrm{T}\big|\big)$$

denote the T-*bidegree of* v, which is the pair of the number of treated and number of untreated neighbors of v. Let $\Delta^0(\mathcal{B})$ denote the space of finite, signed measures on $\mathcal{B}$ of total mass 0. When $|\mathrm{T}| = pn$, define the measure $D_{\mathrm{T}} \in \Delta^0(\mathcal{B})$ by

$$D_{\mathrm{T}}(u) = \frac{1}{pqn}\sum_{v \in V} \chi_v^{\mathrm{T}} \delta_{\vec{d}_{\mathrm{T}}(v)}(u) \quad \text{for } u \in \mathcal{B}, \tag{4.3}$$

where $\mathcal{B}$ is as defined in (4.1). Here, we write $\delta_{\vec{d}_{\mathrm{T}}(v)}$ for a Dirac mass at $\vec{d}_{\mathrm{T}}(v) \in \mathcal{B}$. Clearly, $D_{\mathrm{T}}(\mathcal{B}) = \sum_{u \in \mathcal{B}} D_{\mathrm{T}}(u) = 0$. In symmetric interference models without types, the quantity ξ in equation (2.5) can be expressed compactly as

$$\xi = \int_{\mathcal{B}} f\, dD_{\mathrm{T}} = \sum_{u \in \mathcal{B}} f(u) D_{\mathrm{T}}(u). \tag{4.4}$$

Hence, D_{T} is the kernel that yields the net interference effect ξ when integrated against f.

5. Perfect quasi-colorings and designs for symmetric interference model. In this section, we introduce our idea of perfect quasi-colorings, which we use to construct designs for the symmetric interference model. Throughout this subsection, we will assume that $r = 2$ and $p = q = 1$, so that the target treatment fraction is $\frac{1}{2}$.

The following notion of *perfect quasi-coloring* lets us identify the treatment groups so that the interference effect ξ is identically zero.

DEFINITION 5.1. A *perfect quasi-coloring* is a set $Q \in \binom{V(G)}{n}$ that satisfies $D_Q = 0$.

The following result implies that $\xi = 0$ for the treatment groups $\mathrm{T} = Q$ and $\mathrm{T} = V(G) \setminus Q$ if and only if Q is a perfect quasi-coloring.

PROPOSITION 5.2. *Let $Q \in \binom{V(G)}{n}$. The following are equivalent in a symmetric model.*

- *Q is a perfect quasi-coloring.*
- *$V(G) \setminus Q$ is a perfect quasi-coloring.*
- *If $\mathrm{T} = Q$, for every function f_v of the form (4.2), we have $\xi = 0$.*
- *If $\mathrm{T} = V(G) \setminus Q$, for every function f_v of the form (4.2), we have $\xi = 0$.*
- *If the treatment T is chosen uniformly and randomly between Q and $V(G) \setminus Q$, for every function f_v of the form (4.2), we have $\xi = 0$.*

REMARK 5.3. Intuitively, randomizing between $\mathrm{T} = Q$ and $\mathrm{T} = V(G) \setminus Q$ when Q is a perfect quasi-coloring makes $\widehat{t}_{\mathrm{Neyman}}$ unbiased because (1) interference effects cancel and (2) each vertex is treated with probability $\frac{1}{2}$—so each treatment effect enters the estimate with probability $\frac{1}{2}$.

PROOF OF PROPOSITION 5.2. First, we show that Q is a perfect quasi-coloring if and only if $V(G) \setminus Q$ is. Define $\tau : \mathcal{B} \to \mathcal{B}$ by $\tau(a, b) = (b, a)$. Let $\tau_* D_Q$ be the push forward measure of D_Q by the function τ. By construction, we have $\tau_* D_Q = -D_{V(G)\setminus Q}$. Thus we conclude that $D_Q = 0$ if and only if $D_{V(G)\setminus Q} = 0$.

Next, we prove that Q is a perfect quasi-coloring if and only if $\xi = 0$ for all f when $\mathrm{T} = Q$. Since, $\xi = \int_{\mathcal{B}} f \, dD$ by equation (4.4), this assertion is immediate. The lemma follows because the distribution of ξ with T chosen uniformly at random between Q and $V(G) \setminus Q$ is a $\frac{1}{2}$–$\frac{1}{2}$ mixture of point masses at the values of ξ with $\mathrm{T} = Q$ and $\mathrm{T} = V(G) \setminus Q$. □

The following example shows that highly homogeneous graphs admit perfect quasi-colorings.

EXAMPLE 5.4 (Perfect quasi-colorings exist in graphs consisting of many disjiont copies of a smaller graph). Let H be an arbitrary graph with $|V(H)| > 1$. Let $G = H \times \{0, 1\}^{V(H)}$ denote the disjoint union of $2^{|V(H)|}$-many copies of H. To be precise, the set of vertices of G is

$$V(G) = \{(v, (\epsilon_w)_{w\in V(H)}) \mid v \in H \text{ and } \epsilon_w \in \{0, 1\} \text{ for all } w \in V(H)\},$$

and there is an edge between $(v, (\epsilon_w)_{w\in V(H)})$ and $(v', (\epsilon'_w)_{w\in V(H)})$ in G if and only if $\{v, v'\} \in E(H)$ and $\epsilon_w = \epsilon'_w$ for all $w \in V(H)$. We claim that

$$Q = \{(v, (\epsilon_w)_{w\in V(H)}) \mid \epsilon_v = 1\}$$

is a perfect quasi-coloring of G. To see this, define an involution $\psi : V(G) \to V(G)$ by

$$\psi(v, \epsilon) = (v, (\epsilon_{V(H)\setminus\{v\}}, 1 - \epsilon_v)).$$

Note that, for all $w \in V(G)$, exactly one of w and $\psi(w)$ is in Q and w and $\psi(w)$ have the same number of neighbors in Q (resp. $V(G) \setminus Q$). It follows that $D_Q = 0$.

The class of graphs considered by Example 5.4 is quite specific. Unfortunately, not even $2k$-regular graphs need to admit a perfect quasi-coloring, as the following example shows.

EXAMPLE 5.5 (A hexagon does not have a perfect quasi-coloring). Let G be a hexagon. Thus $V(G) = \{1, 2, \dots, 6\}$ with an edge drawn between i and $i+1$ modulo 6 for all i. Let $B \in \binom{V(G)}{3}$.

We claim that B is not perfect. Indeed, if B contains three consecutive elements of $V(G)$, then the support of D_B contains $(2, 0)$. If B does not contain any three consecutive elements of $V(G)$, then the support of D_B contains $(0, 2)$. In either case, we have $D_B \neq 0$. This example motivates studying other estimators in addition to $\widehat{t}_{\mathrm{Neyman}}$; see Section 10 for more on this point.

Example 5.5 suggests that it might not be fruitful to search for perfect quasi-colorings in arbitrary graphs. In general, we can only hope to control the size of ξ. Proposition 5.2 yields that $\xi = 0$ for a perfect quasi-coloring. It is then natural to ask whether an "almost perfect quasi-coloring" will imply that the corresponding ξ is close to zero. In the next section, we show that this intuition indeed holds, quantify it, and use it constructing new designs.

5.1. *Quantifying the notion of perfect quasi-coloring.* Our strategy is to use the Wasserstein norm to quantify the approximation $D_{\mathrm{T}} \approx 0$. Let $\mathbf{d}$ be a pseudo-metric on $\mathcal{B}$. For $f : \mathcal{B} \to \mathbb{R}$, define the Lipschitz norm

$$\|f\|_{\mathbf{d}} = \sup_{u_1, u_2 \in \mathcal{B}, u_1 \neq u_2} \frac{|f(u_1) - f(u_2)|}{\mathbf{d}(u_1, u_2)}.$$

For a measure $D \in \Delta^0(\mathcal{B})$, define the Wasserstein norm

$$\|D\|_{\mathbf{d}_{\mathrm{w}}} = \sup_{\|f\|_{\mathbf{d}} \leq 1} \left\| \int_{\mathcal{B}} f \, dD \right\|.$$

Since the total mass is 0 for any $D \in \Delta^0(\mathcal{B})$, we have that

$$\|D\|_{\mathbf{d}_{\mathrm{w}}} \leq \frac{1}{2} \operatorname{diam}(\mathcal{B}) \|D\|_{\mathrm{TV}},$$

where $\| - \|_{\mathrm{TV}}$ denotes the total variation norm. From equation (4.4), we can deduce that if the interference function $f : \mathcal{B} \to \mathbb{R}$ is Lipschitz with respect to a metric $\mathbf{d}$, then

$$|\xi| \leq \|f\|_{\mathbf{d}} \cdot \|D_{\mathrm{T}}\|_{\mathbf{d}_{\mathrm{w}}}. \tag{5.1}$$

For a treatment assignment T that is a perfect quasi-coloring, we have $D_{\mathrm{T}} = 0$ and thus $\xi = 0$. Equation (5.1) shows that ξ is continuous in $\|D_{\mathrm{T}}\|_{\mathbf{d}_{\mathrm{w}}}$.

While (5.1) holds for any pseudo-metric $\mathbf{d}$, we use the following pseudo-metric $\mathbf{d} = \mathbf{d}_K$.

DEFINITION 5.6. Fix $K = (K_1, K_2)$ with $K_1 \geq 0$ and $K_2 > 0$. Define a pseudo-metric $\mathbf{d}_K$ on $\mathcal{B}$ by

$$\mathbf{d}_K\big((a, b), (c, d)\big) = K_1 \frac{|a + b - c - d|}{d_{\max}} + K_2 \left| \frac{a}{a+b} - \frac{c}{c+d} \right| \tag{5.2}$$

for all $(a, b), (c, d) \in \mathcal{B}$, where $d_{\max}$ is as in (1.1).

REMARK 5.7. Since we assume that G does not have any isolated vertices, $\mathbf{d}_K$ is indeed a well-defined pseudo-metric on $\mathcal{B}$.

REMARK 5.8. The choice of a metric is crucial for our estimates. The main point here is that the chosen metric must capture the key features of the interference model. To measure the similarity of two vertices, the metric $\mathbf{d}_K$ in (5.2) just takes the differences in the fraction of the treated neighbors and the differences of the degrees between the vertices. This choice is justified in our setting, because the symmetric interference model by definition depends only on these quantities. Different metrics could be used for other settings.

For $\mathcal{P} = (S_1, \dots, S_n) \in \binom{V(G)}{r,\dots,r}$, define a constant

$$C_{\mathcal{P}} = \frac{2}{d_{\max}(r-1)} \sum_{i=1}^{n} \sum_{\{v,v'\}\subseteq S_i} |d(v) - d(v')|. \tag{5.3}$$

The following proposition bounds the L^2 norm of $\|D_{\mathrm{T}}\|_{\mathbf{d}_{\mathrm{w}}}$.

PROPOSITION 5.9. *Fix $\mathcal{P} \in \binom{V(G)}{r,\dots,r}$ and let $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$ as in (3.1). We have*

$$\sqrt{\mathbb{E}_{\vec{B}} \|D_{\mathrm{T}}\|^2_{\mathbf{d}_{\mathrm{w}}}} \le \frac{K_1}{\sqrt{pq}n} C_{\mathcal{P}} + \frac{4K_2}{rn} \sum_{v\in V(G)} \frac{1}{\sqrt{d(v)}} + \frac{K_2}{pqn} \sum_{v\in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)},$$

where $\mathcal{P}_v$ is as in (3.3).

The idea behind the proof of Proposition 5.9 is to bound the contributions of each vertex to the left-hand side, and use the fact that $\mathrm{T} \cap S_i$ and $\mathrm{T} \cap S_j$ are independent for $i \neq j$, where $\mathcal{P} = (S_1, \dots, S_n)$.

Proposition 5.9 and equation (5.1) imply the following upper bound on the L^2 norm of ξ for the randomization scheme $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

COROLLARY 5.10. *Let the interference function $f : \mathcal{B} \to \mathbb{R}$ be such that $\|f\|_{\mathbf{d}_K} \le 1$. Then*

$$\sqrt{\mathbb{E}_{\vec{B}} |\xi|^2} \le \frac{K_1}{\sqrt{pq}n} C_{\mathcal{P}} + \frac{1}{rn} \sum_{v\in V(G)} \frac{4K_2}{\sqrt{d(v)}} + \frac{K_2}{pqn} \sum_{v\in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}$$

for all $\mathcal{P}$ when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

REMARK 5.11. In the case of a complete graph on rn vertices, we have $C_{\mathcal{P}} = 0$ and hence

$$\sqrt{\mathbb{E}_{\vec{B}} |\xi|^2} \le \frac{4K_2}{\sqrt{rn-1}} + \frac{K_2 r(r-1)}{pq(rn-1)} \le \frac{4K_2}{\sqrt{rn-1}} + \frac{K_2 r}{pqn},$$

where the second inequality holds because $\frac{r-1}{rn-1} \le \frac{1}{n}$. Thus, for fixed p, q, r, we have $\sqrt{\mathbb{E}_{\vec{B}} |\xi|^2} = O(n^{-1/2})$.

6. New designs and MSE for $\hat{t}_{\mathrm{Neyman}}$. In this section, we use the idea of perfect quasi-coloring and Proposition 5.9 to construct new designs and derive bounds for the MSE of $\hat{t}_{\mathrm{Neyman}}$. We study the dense ($d_{\min} \to \infty$ as $n \to \infty$) and sparse ($d_{\max} = o(\sqrt{n})$ as $n \to \infty$) cases separately, as our methods and assumptions are different for dense vs. sparse graphs. We note that sparsity is essentially an asymptotic property and so for a single observed network it is not immediately apparent which regime to consider. In cases where the network is collected prior to an experiment, the sampling procedure can inform which regime is more appropriate by providing evidence for the rate of growth of $d_{\min}$ and $d_{\max}$.

6.1. *Dense graphs.* A key term appearing in the right-hand side of Proposition 5.9 is the constant $C_{\mathcal{P}}$, which is solely a function of the partition $\mathcal{P}$. Thus, we seek designs that lead to smaller values of $C_{\mathcal{P}}$. To this end, we introduce the following new design, which we call "partitioning by degree."

DEFINITION 6.1. Let $\{w_i^*\}$, $1 \le i \le rn$ be an enumeration of the vertices of G such that $d(w_i^*) \ge d(w_{i'}^*)$ whenever $i > i'$. Choose

$$S_i = \{w_j^*, (i-1)r + 1 \le j \le ir\}$$

for $1 \le i \le n$. Finally set

$$\mathcal{P}^* = (S_1, \dots, S_n). \tag{6.1}$$

Thus, the partition $\mathcal{P}^*$ is chosen by first ordering the vertices by degree and then grouping vertices of similar degree. The randomization scheme $T_{\vec{B},\mathcal{P}^*}$ is therefore obtained by approximately stratifying vertices by degree. The following is a key observation.

LEMMA 6.2. *For $\mathcal{P}^*$ chosen according to partitioning by degree as in Definition 6.1, we have*

$$C_{\mathcal{P}^*} \le r, \tag{6.2}$$

where $C_{\mathcal{P}^*}$ in the corresponding constant in Corollary 5.10.

PROOF. Breaking each appearance of $d(v) - d(v')$ in (5.3) into a sum of terms of the form $d(w_k^*) - d(w_{k+1}^*)$, we have

$$\begin{aligned}
C_{\mathcal{P}^*} &= \frac{2}{d_{\max}(r-1)} \sum_{i=1}^{n} \sum_{1 \le j < j' \le r} \bigl(d(w^*_{r(i-1)+j}) - d(w^*_{r(i-1)+j'})\bigr) \\
&\le \frac{2}{d_{\max}(r-1)} \sum_{i=1}^{n} \sum_{1 \le j < j' \le r} \bigl(d(w^*_{r(i-1)+1}) - d(w^*_{ri})\bigr) \\
&= \frac{2}{d_{\max}(r-1)} \sum_{i=1}^{n} \left[\frac{r(r-1)}{2}\bigl(d(w^*_{r(i-1)+1}) - d(w^*_{ri})\bigr)\right] \\
&= \frac{r}{d_{\max}} \sum_{i=1}^{n} \bigl(d(w^*_{r(i-1)+1}) - d(w^*_{ri})\bigr)] \\
&\le \frac{r}{d_{\max}} \bigl(d(w_1^*) - d(w_{rn}^*)\bigr) \le \frac{r}{d_{\max}} \cdot d_{\max} = r,
\end{aligned}$$

as desired. □

As an immediate consequence of Lemmata 3.3 and 6.2, Proposition 5.9, and Corollary 5.10, we have the following bound.

THEOREM 6.3. *When* $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}^*}$,

$$\sqrt{\mathbb{E}_{\mathrm{T}} \|D\|^2_{\mathbf{d}_{\mathrm{w}}}} \le \frac{K_1 r}{\sqrt{pqn}} + \frac{4K_2}{\sqrt{d_{\min}}} + \frac{r K_2 \min\{r-1, d_{\min}\}}{pq \cdot d_{\min}},$$

where $d_{\min}$ *is as in* (1.1). *If in addition, the interference function* f *satisfies* $\|f\|_{\mathbf{d}_K} \le 1$, *then*

$$|\mathbb{E}_{\mathrm{T}}\xi| \le \frac{\min\{r-1, d_{\min}\}(K_1+K_2)}{(r-1)d_{\min}},$$

$$\sqrt{\mathbb{E}_{\mathrm{T}}\xi^2} \le \frac{K_1 r}{\sqrt{pqn}} + \frac{4K_2}{\sqrt{d_{\min}}} + \frac{rK_2\min\{r-1, d_{\min}\}}{pq\cdot d_{\min}}.$$

Theorem 6.3 immediately yields that interference does not affect the consistency of the estimator $\widehat{t}_{\mathrm{Neyman}}$ for our randomized design when G grows to be sufficiently large and dense.

COROLLARY 6.4. *Let* $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}^*}$ *and fix* $p, q, r \in \mathbb{N}$. *If* $\|f\|_{\mathbf{d}} \le 1$, *then*

$$|\mathbb{E}_{\mathrm{T}}\xi| = O\left(\frac{K_1+K_2}{d_{\min}}\right),$$

$$\mathbb{E}_{\mathrm{T}}\xi^2 = O\left(\frac{K_1^2}{n^2} + \frac{K_2^2}{d_{\min}}\right).$$

In particular, if $d_{\min} \to \infty$ *and* $n \to \infty$, *then* $\mathbb{E}_{\mathrm{T}}\xi^2 \to 0$.

The following example shows that the restricted randomization $\mathrm{T}_{\vec{B},\mathcal{P}^*}$, where $\mathcal{P}^*$ is obtained by partitioning by degree as in (6.1), can substantially outperform the CRD in terms of reducing the mean squared error of $\widehat{t}_{\mathrm{Neyman}}$.

EXAMPLE 6.5. Let $p = q = 1$. Let $V(G) = \{v_1, \dots, v_{2k}, w_1, \dots, w_{2k}\}$, and let the edges of G be $\{v_i, v_j\}$. Thus, G is the disjoint union of a complete graph on $2k$ vertices $V = \{v_1, \dots, v_{2k}\}$ with $2k$ additional vertices $W = \{w_1, \dots, w_{2k}\}$. Consider a symmetric linear interference model $f(a,b) = \gamma a$.

Fix $\mathrm{T} \in \binom{V(G)}{n}$ and let $\alpha = \alpha(T) = |T \cap V|$. It is straightforward to verify that

$$\xi = \frac{\gamma(\alpha(\alpha-1) - (2k-\alpha)\alpha)}{2k} = \frac{\gamma}{2k}\alpha(2\alpha - 2k - 1).$$

When T is chosen uniformly and randomly from $\binom{V(G)}{2k}$, by the Central Limit Theorem, we have

$$\frac{\alpha - k}{\sqrt{k}} \to_D \mathcal{N}(0, 1/2)$$

as $k \to \infty$. While $\mathbb{E}_{\mathrm{T}}\xi \to 0$ as $n \to \infty$, it can be verified using the formulae for higher moments of normal distributions that

$$\left(\mathbb{E}_{\mathrm{T}}|\xi|^2\right)^{1/2} \sim \gamma\sqrt{k}$$

as $n \to \infty$.

On the other hand, note that any $\mathcal{P}^*$ according to (6.1) consists of a partition of V into pairs and a partition of W into pairs. Therefore, when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}^*}$, we have $\alpha = k$ and hence $\xi = -\frac{\gamma}{2}$.

Of course in the above example, the graph G contains isolated vertices $\{w_1, \dots, w_{2k}\}$. The conclusions noted above are qualitatively the same if we add some small number of edges among $\{w_1, \dots, w_{2k}\}$ and between $\{v_1, \dots, v_{2k}\}$ and $\{w_1, \dots, w_{2k}\}$, with $d_{\min} \to \infty$ at a sufficiently slow rate.

Example 6.5 illustrates that our treatment design can improve on the completely random design when there is a high degree of heterogeneity in the degrees of vertices.

6.2. *Sparse graphs*. For sparse graphs, the bias and MSE bounds implied by Theorem 6.3 are a bit weak. In this setting, it is helpful to randomize over all choices of $\mathcal{P}^*$ in order to reduce bias. To this end, we introduce the following randomized version of the design introduced in Definition 6.1.

DEFINITION 6.6. Let $S \subseteq V(G)$ be such that no r vertices in S have the same degree and the number of vertices in $V(G) \setminus S$ of each degree is divisible by r. Let $\mathcal{P}_0^{**}$ be sampled uniformly from the set of partitions of $V(G) \setminus S$ into sets of r vertices of the same degree. Let $S = \{w_1, \dots, w_{rk}\}$ with

$$d(w_1) \geq d(w_2) \geq \cdots \geq d(w_{rk}).$$

Let

$$\mathcal{P}^{**} = \bigl(\{w_1, \dots, w_r\}, \dots, \{w_{rk-r+1}, \dots, w_{rk}\}, \mathcal{P}_0^{**}\bigr).$$

Thus the main difference between designs in Definitions 6.1 and 6.6 is that in the latter, we randomize over all vertices with the same degree instead of merely fixing an ordering. We now give the L^2 bounds on ξ for the randomization scheme $\mathrm{T}_{\vec{B},\mathcal{P}^{**}}$.

PROPOSITION 6.7. *If* $\|f\|_{\mathbf{d}} \leq 1$, *then*

$$|\mathbb{E}_{\mathrm{T}}\xi| \leq \frac{K_1 r}{pqn} + \frac{3K_2(d_{\max} - d_{\min})}{n}\left(\frac{r}{pqd_{\min}} + \frac{3}{r}\right),$$

$$\sqrt{\mathbb{E}_{\mathrm{T}}\xi^2} \leq \frac{K_1 r}{pqn} + \frac{K_2(d_{\max} - d_{\min})r}{pqnd_{\min}} + \frac{6K_2\sqrt{r^2 d_{\max}^2 + 1}}{\sqrt{n} \cdot d_{\min}}$$
$$+ \frac{rK_2 \min\{r-1, d_{\min}\}\sqrt{r^2 d_{\max}^2 + 1}}{pq\sqrt{n} \cdot d_{\min}},$$

where $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}^{**}}$ *and* $\mathcal{P}^{**}$ *is as in Definition 6.6.*

To prove Proposition 6.7, we decompose ξ into effects from each of the parts of $\mathcal{P}^{**}$. Sparsity ensures that there is not too much dependence between the contributions of the various parts. The following two simple lemmata in probability then imply a bound on the variance of ξ, which, when coupled with Lemma 3.3, yields an L^2 bound.

LEMMA 6.8. *For a sequence of random variables* $X_1, \dots, X_k$,

$$\operatorname{Var}\left(\sum_{i=1}^k X_i\right) \leq \sum_{i,j=1}^k \operatorname{Var}(X_i)\bigl|\operatorname{Corr}(X_i, X_j)\bigr|.$$

PROOF. For all $x, y \geq 0$, we have $2\sqrt{xy} \leq x + y$. It follows that

$$2\operatorname{Cov}(X_i, X_j) = 2\operatorname{Corr}(X_i, X_j)\sqrt{\operatorname{Var}(X_i)}\sqrt{\operatorname{Var}(X_j)}$$
$$\leq \bigl|\operatorname{Corr}(X_i, X_j)\bigr|\bigl(\operatorname{Var}(X_i) + \operatorname{Var}(X_j)\bigr)$$

for all i, j. Thus, we have

$$\operatorname{Var}\left(\sum_{i=1}^k X_i\right) = \sum_{i,j} \operatorname{Cov}(X_i, X_j)$$

$$\leq \frac{1}{2} \sum_{i,j} \left|\mathrm{Corr}(X_i, X_j)\right| \left(\mathrm{Var}(X_i) + \mathrm{Var}(X_j)\right)$$

$$= \sum_{i,j=1}^{k} \mathrm{Var}(X_i) \left|\mathrm{Corr}(X_i, X_j)\right|,$$

as desired. □

LEMMA 6.9. *Let $X_1, \dots, X_n$ be real valued random variables. Suppose that for each i, there exist at most κ indices j such that X_i and X_j are not independent. Then, we have*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \leq \kappa \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

PROOF. Since independent random variables are uncorrelated, for each index i, there exist at most κ indices j such that $\mathrm{Corr}(X_i, X_j) \neq 0$. It follows that $\sum_{j=1}^{k} |\mathrm{Corr}(X_i, X_j)| \leq \kappa$ for all i. The lemma thus follows from Lemma 6.8. □

Proposition 6.7 immediately yields the following MSE bounds for $\widehat{t}_{\mathrm{Neyman}}$ in the sparse regime.

COROLLARY 6.10. *Let $\mathrm{T} = \mathrm{T}_{\vec{B}, \mathcal{P}^{**}}$ and fix $p, q, r \in \mathbb{N}$. If $\|f\|_{\mathbf{d}} \leq 1$, then*

$$|\mathbb{E}_T \xi| = O\left(\frac{K_1}{n} + \frac{K_2(d_{\max} - d_{\min})}{n}\right),$$

$$\mathbb{E}_{\mathrm{T}} \xi^2 = O\left(\frac{K_1^2}{n^2} + \frac{K_2^2 d_{\max}^2}{n d_{\min}}\right).$$

In particular, if $n \to \infty$ with $d_{\max} = o(\sqrt{n})$, then the MSE *of $\widehat{t}_{\mathrm{Neyman}}$ is $o(1)$.*

7. Interference with types. In this section, we generalize the symmetric interference model from Section 4 to incorporate heterogeneity in susceptibility to interference, and derive the MSE bounds for $\widehat{t}_{\mathrm{Neyman}}$ under this extension. More precisely, we allow the interference function f to depend on an exogenously specified "type" of a vertex.

DEFINITION 7.1. The function f_v is a *symmetric interference model with types* if there exists a partition Π of $V(G)$ into sets of size divisible by r such that there are functions $(f_\pi)_{\pi \in \Pi}$ with

$$f_v(S) = f_{\Pi_v}\left(|S|, \left|\mathcal{N}(v) \setminus S\right|\right)$$

for all $v \in V(G)$. Here, f_π is real-valued with domain

$$\mathcal{B}_\pi = \left\{(a, b) \in \mathbb{Z}_{\geq 0}^2 \mid a + b \in d(\pi)\right\}.$$

REMARK 7.2. The case of $\Pi = \{V(G)\}$ recovers the symmetric interference model (without types) in Definition 4.1.

Let $\Delta^0(\mathcal{B}_\pi)$ denote the space of finite, signed measures on $\mathcal{B}_\pi$ of total mass 0. When $\pi \subseteq V(G)$ is such that $|\mathrm{T} \cap \pi| = \frac{p|\pi|}{r}$, let

$$\Delta^0(\mathcal{B}_\pi) \ni D^\pi = D_{\mathrm{T}}^\pi = \frac{1}{pqn} \sum_{v \in \pi} \chi_v^{\mathrm{T}} \delta_{\vec{d}_{\mathrm{T}}(v)}.$$

Intuitively, D^π measures the contribution of π to D. Indeed, we have

$$\xi = \sum_{\pi \in \Pi} \int_{\mathcal{B}_\pi} f_\pi \, dD^\pi.$$

7.1. *Perfect quasi-colorings for interference with types.* The structure of perfect quasi-colorings extends to the setting of interference models with types. For this subsection, we assume that $p = q = 1$, so the target treatment fraction is $\frac{1}{2}$. The analogue of Definition 5.1 is the following definition.

DEFINITION 7.3. A *perfect quasi-coloring of G with respect to the type partition* Π is a set $B \in \binom{V(G)}{n}$ that satisfies $D_B^\pi = 0$ and $|B \cap \pi| = |\pi|/2$ for all $\pi \in \Pi$.

Definition 7.3 recovers Definition 5.1 by taking $\Pi = \{V(G)\}$. The analogue of Proposition 5.2 is the following result.

PROPOSITION 7.4. *Let $B \in \binom{V(G)}{n}$ be such that $|B \cap \pi| = |\pi|/2$ for all $\pi \in \Pi$. The following are equivalent in a symmetric model with types.*

- *B is a perfect quasi-coloring.*
- *$V(G) \setminus B$ is a perfect quasi-coloring.*
- *$\xi = 0$ for all $(f_\pi)_{\pi \in \Pi}$ with treatment $T = B$.*
- *$\xi = 0$ for all $(f_\pi)_{\pi \in \Pi}$ with treatment $T = V(G) \setminus B$.*
- *$\xi = 0$ for all $(f_\pi)_{\pi \in \Pi}$ with treatment chosen uniformly and randomly between B and $V(G) \setminus B$.*

The proof of Proposition 7.4 is similar to the proof of Proposition 5.2. Example 5.5 shows that perfect quasi-colorings need not exist in general, while the following example generalizes Example 5.4 to exhibit a class of graphs and type partitions in which perfect quasi-colorings exist.

EXAMPLE 7.5 (Perfect quasi-colorings exist in the graph consisting of copies of a smaller graph). Let H be an arbitrary graph with $|V(H)| > 1$, and let Π_0 be a partition of the vertices of H. Let $G = H \times \{0, 1\}^{V(H)}$, and define a partition Π of $V(G)$ by

$$\Pi = \{\pi \times \{0, 1\}^{V(H)} \mid \pi \in \Pi_0\}.$$

It is straightforward to verify that

$$B = \{v, (\epsilon_w)_{w \in V(H)} \mid u_v = 1\}$$

is a perfect quasi-coloring of G with respect to the type partition Π.

7.2. *Semi-restricted randomization.* We consider a stratified sampling scheme. For each $\pi \in \Pi$, let T_π be drawn uniformly and randomly from $\binom{\pi}{p|\pi|/r}$, with $(\mathrm{T}_\pi)_{\pi \in \Pi}$ independent. Define $T_\Pi = \bigcup_{\pi \in \Pi} \mathrm{T}_\pi$. Intuitively, T_Π is a stratified sampler based on a stratification of vertices by type (i.e., the partition Π).

We can represent the stratified sampler treatment group in terms of a restricted randomization treatment group as follows. To fix notation, given a partition Π of $V(G)$, let $\binom{\Pi}{r, \ldots, r}$ denote the set of partitions $\mathfrak{P} = (S_1, \ldots, S_n)$ of $V(G)$ into sets of size r such that S_i lies in an element of Π for every i. That is, $\binom{\Pi}{r, \ldots, r}$ is the set of partitions of $V(G)$ into sets of size r that refine Π. Let $\mathcal{P}$ be sampled uniformly from $\binom{\Pi}{r, \ldots, r}$ (independently of $\vec{B}$). Then, $\mathrm{T}_{\vec{B}, \mathcal{P}}$ has the same distribution as T_Π.

7.3. *MSE bounds*. In this section, we prove bounds on the L^2 norm of ξ for the randomization scheme $\mathrm{T} = \mathrm{T}_\Pi$. We will use the following metric $\mathbf{d} = \mathbf{d}_K$.

DEFINITION 7.6. Fix $K > 0$, define the metric $\mathbf{d}$ on $\mathcal{B}_\pi$ by

$$\mathbf{d}_K\big((a,b),(c,d)\big) = K\left|\frac{a}{a+b} - \frac{c}{c+d}\right|$$

for all $(a,b),(c,d) \in \mathcal{B}_\pi$.

The analogue of Proposition 5.9 in this setting is the following result.

PROPOSITION 7.7. *For all $\mathcal{P} \in \binom{\Pi}{r,\dots,r}$, we have*

$$\sqrt{\mathbb{E}_{\vec{B}}\Big[\Big(\sum_{\pi\in\Pi}\|D^\pi\|_{\mathbf{d}_{\mathrm{w}}}\Big)^2\Big]} \le \frac{1}{rn}\sum_{v\in V(G)}\frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn}\sum_{v\in V(G)}\frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}$$

when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$. *If* $\|f_\pi\|_{\mathbf{d}} \le 1$ *for all* $\pi \in \Pi$, *then*

$$\sqrt{\mathbb{E}_{\mathrm{T}}\xi^2} \le \frac{1}{rn}\sum_{v\in V(G)}\frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn}\sum_{v\in V(G)}\frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}$$

when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

Proposition 7.7 also implies a bound on the L^2 norm of ξ when $\mathrm{T} = \mathrm{T}_\Pi$.

COROLLARY 7.8. *If* $\|f_\pi\|_{\mathbf{d}} \le 1$ *for all* $\pi \in \Pi$, *then*

$$\sqrt{\mathbb{E}_{\mathrm{T}}\xi^2} \le \frac{1}{rn}\sum_{v\in V(G)}\frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn}\sum_{v\in V(G)}\frac{(r-1)|\Pi_v \cap \mathcal{N}(v)|}{(|\Pi_v|-1)d(v)}$$

when $\mathrm{T} = \mathrm{T}_\Pi$.

PROOF. Recall that the distribution of T_Π is the distribution of $T_{\mathcal{P}}$ when $\mathcal{P}$ is chosen uniformly at random from the set $\binom{\Pi}{r,\dots,r}$ of partitions. In this case, we have that

$$\begin{aligned}\mathbb{E}_{\mathcal{P}}|\mathcal{P}_v \cap \mathcal{N}(v)| &= \sum_{w\in\mathcal{N}(v)}\mathbb{P}_{\mathcal{P}}[w \in \mathcal{P}_v] = \sum_{w\in\mathcal{N}(v)} 1_{\Pi_v}(w)\frac{r-1}{|\Pi_v|-1} \\ &= \frac{(r-1)|\Pi_v \cap \mathcal{N}(v)|}{|\Pi_v|-1}.\end{aligned}$$

The result follows. $\square$

As in Section 6, Corollary 7.8 implies a consistency result for the Neymannian estimator.

COROLLARY 7.9. *Let* $\mathrm{T} = \mathrm{T}_\Pi$ *and fix* $p, q, r \in \mathbb{R}$. *If* $\|f_\pi\|_{\mathbf{d}} \le 1$ *for all* $\pi \in \Pi$, *then*

$$\mathbb{E}_{\mathrm{T}}\xi^2 \le O\Big(\frac{K^2}{d_{\min}}\Big).$$

In particular, if $d_{\min} \to \infty$, *then* $\mathbb{E}_{\mathrm{T}}\xi^2 \to 0$.

The analogue of Proposition 6.7 is the following result.

PROPOSITION 7.10. *If* $\| f_\pi \|_{\mathbf{d}} \le 1$ *for all* $\pi \in \Pi$, *then*

$$|\mathbb{E}_{\mathrm{T}}\xi| \le \frac{K|\Pi|}{rn},$$

$$\frac{\sqrt{\mathbb{E}_{\mathrm{T}}\xi^2}}{K} \le \frac{\sqrt{2|\Pi|}}{\sqrt{nr \cdot d_{\min}}} + \frac{4\sqrt{r^2 d_{\max}^2 + 1}}{\sqrt{n \cdot d_{\min}}} + \frac{r \min\{r-1, d_{\min}\}\sqrt{r^2 d_{\max}^2 + 1}}{pq\sqrt{n} \cdot d_{\min}}$$

when $\mathrm{T} = \mathrm{T}_\Pi$.

As in Section 6, Proposition 7.10 implies a consistency result for the Neymannian estimator.

COROLLARY 7.11. *Let* $\mathrm{T} = \mathrm{T}_\Pi$ *and fix* $p, q, r \in \mathbb{R}$. *If* $\| f_\pi \|_{\mathbf{d}} \le 1$ *for all* $\pi \in \Pi$, *then*

$$|\mathbb{E}_{\mathrm{T}}\xi| = O\left(\frac{K|\Pi|}{n}\right),$$

$$\mathbb{E}_{\mathrm{T}}\xi^2 = O\left(\frac{K^2|\Pi|}{nd_{\min}} + \frac{K^2 d_{\max}^2}{nd_{\min}}\right).$$

In particular, if $n \to \infty$ *with* $d_{\max} = o(\sqrt{n})$ *and* $|\Pi| = o(\sqrt{n})$, *then* $\mathbb{E}_{\mathrm{T}}\xi^2 \to 0$.

Thus, in the sparse setting, it is important that Π is not too large, that is, that degree heterogeneity is not too large. The analogous condition in Corollary 6.10 is that there are not too many different degrees in the graph, which follows from the requirement that $d_{\max} = o(\sqrt{n})$.

8. Nodal similarity and types. In this section, we directly bound the MSE of $\widehat{t}_{\mathrm{Neyman}}$ in a model that allows covariate similarity between vertices within elements of Π. We state results solely assuming that individuals of similar types have similar outcomes while conditioning on the graph. This allows us to control for (possibly latent) heterogeneity among the outcomes of individuals in the graph. As a specific example, if the graph is homophilous and the elements of Π include individuals who are more likely to connect to each other, then our approach provides an avenue for considering both homophily and interference.

For $\pi \in \Pi$, let

$$x_\pi = \frac{1}{|\pi|}\sum_{v \in \pi} x_v, \qquad t_\pi = \frac{1}{|\pi|}\sum_{v \in \pi} t_v,$$

be the average covariate effect and average treatment effect respectively within a type. For $v \in V(G)$, let

$$\epsilon_v = x_v - x_{\Pi_v} + \frac{q}{r}(t_v - t_{\Pi_v})$$

be the discrepancy between an individual node's behavior and their type average. Then

$$\sigma^2 = \frac{1}{rn}\sum_{v \in V(G)} \epsilon_v^2$$

captures the sum of squared differences between nodes and their type averages within a graph. Thus σ^2 has an inverse relationship with the similarity of individuals within elements of Π. The following result, which generalizes Lemma 2.1, bounds the MSE of t_{ideal}.

PROPOSITION 8.1. *For all partitions Π of $V(G)$ into sets of size divisible by r, we have*

$$\mathbb{E}_{\mathrm{T}} t_{\text{ideal}} = \bar{t},$$

$$\operatorname{Var}_{\mathrm{T}}(t_{\text{ideal}}) \leq \frac{2r\sigma^2}{pqn}$$

when $\mathrm{T} = \mathrm{T}_{\Pi}$.

Coupling Proposition 8.1 with bounds on ξ yields the following bias and MSE bounds for $\widehat{t}_{\text{Neyman}}$.

COROLLARY 8.2. *If* $\| f_\pi \|_{\mathbf{d}} \leq 1$ *for all* $\pi \in \Pi$, *then*

$$|\mathbb{E}_{\mathrm{T}} \widehat{t}_{\text{Neyman}} - \bar{t}| \leq \frac{K|\Pi|}{n},$$

$$\sqrt{\mathbb{E}_{\mathrm{T}} (\widehat{t}_{\text{Neyman}} - \bar{t})^2} \leq \frac{1}{rn} \sum_{v \in V(G)} \frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn} \sum_{v \in V(G)} \frac{(r-1)|\Pi_v \cap \mathcal{N}(v)|}{(|\Pi_v| - 1) \cdot d(v)} + \frac{\sigma\sqrt{2r}}{\sqrt{pqn}}$$

when $\mathrm{T} = \mathrm{T}_{\Pi}$.

PROOF. Follows from Corollary 7.8 and Proposition 8.1. $\square$

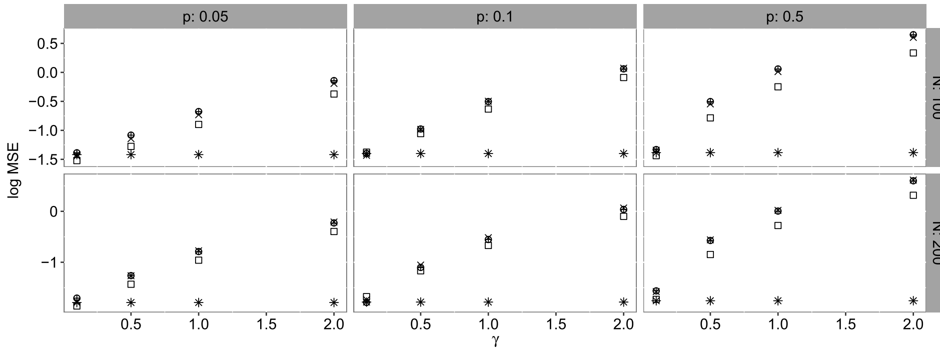
The results of this section are closely related to the work of Basse and Airoldi [4] on optimal design with network correlated outcomes that are induced by homophily but no interference. Furthermore, if the types are unknown, homophily is suspected (i.e., individuals with similar covariates and outcomes are more likely to form connections), and it is suspected that individuals with similar baseline characteristics have similar interference functions then the types can be estimated from the graph. This will incur an additional penalty due to the estimation error of the types (we do not pursue this here).

9. Simulations. In this section we conduct a series of simulations to demonstrate the efficacy of our approach. We vary the type of the network and strength and method of the interference. For each of the simulations we consider the model

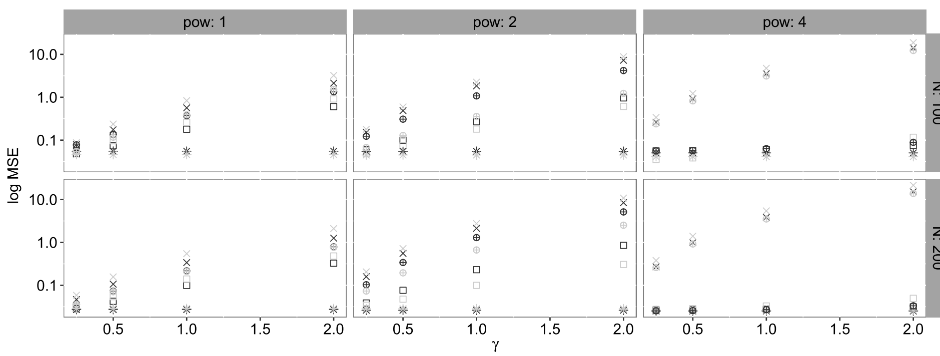
$$y_v = x_v + t_v 1_{v \in T} + f_v(\mathrm{T} \cap \mathcal{N}(v)),$$

where $x_v \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1)$ and $t_v \overset{\text{i.i.d.}}{\sim} \mathrm{N}(2, 0.25)$. That is, the baseline outcome for all of the individuals in the graph is centered at 0 with a variance of 1, while the treatment effect for everyone is centered at 2 with a variance of 0.25. We consider three treatment regimes: our approach (described in Section 6), a variant on our approach that further ensures that individuals within the same partition are not connected by an edge[3] and the completely randomized design (CRD) where exactly half of all units are treated randomly. We report log mean

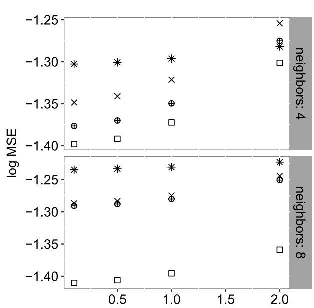
[3]This approach minimizes two terms in the bound in Corollary 5.10. To minimize the term $\sum_{v \in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}$, we note that the only nonzero contribution is from individuals who are both neighbors and are within the same partition. In the simulations, we first construct a partition and then check if any nodes within a partition are connected by an edge. If such nodes exist we swap one of them with a node of the same degree from a different partition. This process is repeated in a greedy fashion until either all partitions are neighbor-free or no further improvements can be made. By only swapping nodes with equal degree, we maintain all of the results of Section 6.



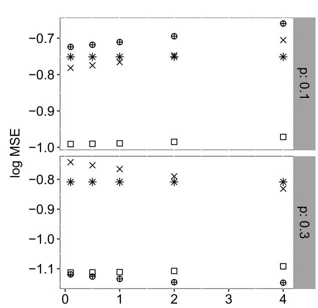
(a) Erdos–Renyi graphs with linear interference



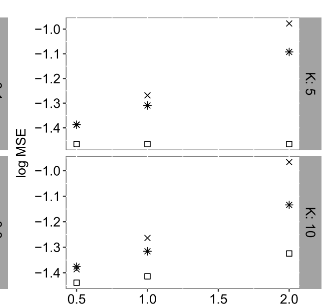
(b) Preferential attachment graphs with linear interference. Darker colors: $m = 4$, lighter colors: $m = 6$



(c) Small world graphs with fractional interference



(d) Stochastic blockmodel with homophily and linear interference



(e) k-star graphs with influencer interference

FIG. 2. *Simulation results for different randomization schemes and estimators.* $\times$ *refers to the Neymanian estimator coupled with a completely randomized deisign.* $*$ *refers to the stratified estimator after a completely randomized design.* $\oplus$ *refers to the Neymanian estimator after ordering by degree only and* $\square$ *refers to the Neymanian estimator after ordering by degree and ensuring that partitions do not overlap with the edges of the graph.*

squared errors (log MSE) for the Neymanian estimator for the three randomizations as well as log MSE for the estimator the post-stratifies on treated degree after a CRD randomization in Figure 2. The MSEs are calculated over 10,000 simulated randomizations for each approach.

9.1. *Erdős–Renyi graphs.* In the first simulation, we generate a graph $G \sim \mathrm{ER}(N, p)$ with N nodes and overall density p. The Erdős–Renyi model $\mathrm{ER}(N, p)$ is an independent-edge random graph model where an edge between node v and v' exists with probability p. We consider two graph sizes, 100 and 200, and three graph densities, 0.05, 0.1 and 0.5.

An important property of Erdős–Renyi graphs is that they are extremely dense (expected degree is Np) and the degrees of their nodes concentrate [16]. Because of this trait, a randomization scheme based on the degree distribution of the graph is unlikely to perform well. Our proposed procedure behaves similarly to the standard completely randomized design. We consider the symmetric linear interference function $f_v(A) = \gamma|A|$. The parameter γ controls the amount of linear interference experienced by any unit with larger values meaning more interference. It varies from 0.1 to 2. Figure 2(a) demonstrates that the estimators based on degrees only and CRD randomization have approximately the same log MSE, with the CRD even exhibiting better behavior for denser graphs and higher levels of interference (such as $p = 0.5, \gamma = 2$). The partition scheme based on degrees that further tries to reduce the number of edges within partition sets outperforms both. For symmetric linear interference the post-stratified estimator performs very well since it correctly identifies the exposure function. We also considered the symmetric fractional interference functions

$$f_v(A, B) = \gamma|A|/\big(|A| + |B|\big), \qquad f_v(A, B) = p \times N \times \gamma|A|/\big(|A| + |B|\big)$$

where under the latter, the expected interference is approximately the same as in the linear interference setting. In both settings, we observe behavior similar to the linear interference—this conclusion is easily explained by the concentration of degrees in an Erdős–Renyi graph, which entails that the fractional and linear interference functions lead to equivalent performance for the different estimators.

9.2. *Preferential attachment graphs.* In a second simulation, we generate a graph $G \sim \mathrm{PA}(N, \mathrm{pow}, m)$ with N nodes, pow power of the preferential attachment (PA) and m new edges at each step of the graph growth [3]. These graphs are constructed by staring with a single vertex and adding one new vertex at a time. The new vertex forms an edge with an existing vertex v with probability proportional to $d(v)^{\mathrm{pow}}$. Each new vertex forms m new edges. This process continues until there are N vertices in the graph. These graphs have power law degree distributions and hence are sparse with many small degree nodes and a few large hubs.

As the power parameter grows, the density of the graph and the number of nodes with the same degree grow with it. Because of this property, we expect an increase in log MSE as a function of power. It is clear that log MSE increases with power since it produces denser graphs that are more likely to have too many nodes with the same degree. The difference between the performance of estimators based on the CRD and the restricted randomization is likely explained by the special behavior of super-linear preferential attachment [15]. The results of these simulations are presented in Figure 2(b). When $\mathrm{pow} = 4$ and $m = 4$ most graphs have four central nodes that are connected to everyone else. As such, only these central nodes induce any form of interference on the other nodes and so the restricted randomization ideally allocated treatment. The CRD does not take this structure into account and so frequently is likely to allocate all of the central nodes to treatment or control, leading to increased bias and variance. When $m = 6$ there are enough perturbations in the system to lead to poorer performance by the restricted randomization. On the other hand, when $\mathrm{pow} \in (1, 2]$, small m frequently lead to the creation of an odd (not even) number of central nodes (in which case ordering by degree requires pairing at least one very high degree node with a low degree node), while a large m produces a large amount of heterogeneity in the degrees. In this setting, the restricted randomization approach prefers more heterogeneity as it balances the interference among nodes. In all of these settings, the CRD performs worse than the restricted randomization. We note that as in Section 9.1, the interference function is linear so the post-stratified estimator correctly identifies the exposure function and so is able to perform very well.

9.3. *Small world graphs*. In a third simulation (whose results are depicted in Figure 2(c)), we generate a graph $G \sim \mathrm{SW}(N, n, p)$ with N nodes, n neighbors on the original lattice and $p = 0.2$ the probability of rewiring [25]. These graphs have both a small diameter and high local connectivity. We consider a fractional interference function $f_v(A, B) = \gamma|A|/(|A| + |B|)$ where A are the treated neighbors and B are the untreated neighbors of a unit. Our design that only considers degrees performs well, though as γ increases it's improvement over the other approaches decays. On the other hand, our design that both treats degrees and eliminates connected edges from the partitions does not suffer as much as γ increases. Of note here is that the post-stratified estimator performs very poorly as it is not robust to different interference functions. This behavior is consistent across different values of n.

9.4. *Stochastic blockmodel graphs and homophily*. In a fourth simulation (whose results are depicted in Figure 2(d)), we generate a graph $G \sim \mathrm{SBM}(N, 2, p)$ with N nodes, 2 groups, with the probability of an edge between members of group one being $0.5 + p$ and the probability of an edge between members of group two being $0.5 - p$. The probability of an edge between members of the two groups is 0.1. The first group includes 25% of individuals while the second group includes 75%. The interference function in this setting is taken to be $f_v(A) = \gamma|A|/d_{\max}$ where this normalization keeps the interference on the order of the treatment effect. Lastly, for this simulation only, we shift x_v to be centered at -2 for group one and 2 for group two. That is, individuals with similar baseline information are more likely to be connected (and hence interfere with each other). Our randomization approach that accounts for both degree and reduces the number of edges within partitions performs substantially better than the naive and post-stratified estimators after a CRD. This demonstrates that the presence of homophily reduces the efficacy of the post-stratification estimator even when the interference function has the same form as the estimator. This behavior is consistent across values of p.[4] Lastly, we note that when $p = 0.2$ (Figure 3 in Appendix E) the two groups have equal average degree (this is an often studied regime in the stochastic blockmodel literature [5]); in this case the stratified estimator conflates people in group 1 with people in group 2 even though they have extremely different baselines, leading to extremely poor performance. The other approaches behave similarly to the Erdős–Renyi setup.

9.5. *Ego networks and individualized interference*. In a fifth simulation (whose results are depicted in Figure 2(e)), we generate graphs of size $N = 100$ made up of separate K-stars (for $K = 5, 10$). This structure is frequently observed in social networks, in which the central node in each K-star can represent an influencer. As such, we consider an "influencer" interference function—if a central node is treated, he interferes with those connected to him, but he is never interfered with. This interference function is further individualized by letting each influencer interfere at a random rate of (sampled from Gamma(2, 2)). In this setting, a CRD completely ignores the structure of the network and so is extremely likely to end up with randomizations in which all of the central nodes are treated (this is not corrected by post-stratifying on treated degree). We see that our proposed randomization appropriately accounts

[4]Our design that does not eliminate connected edges from the partitions appears to be more volatile (additional figures in Appendix E): when $p = 0.1$ there are more edges within groups and the groups have substantively different average degrees, making the behavior of the naive estimator under this randomization similar to its behavior in the Erdős–Renyi setting (as similar degree people are likely to be connected). On the other hand, when $p = 0.3$, the larger group has almost the same probability of within group connections as it does out-of-group connections, meaning that similar degree individuals have a large probability of not being connected. In Figure 3 in Appendix E we see that for other graphs generated from the same SBM regime this approach can have substantively different results. On the other hand the randomization that accounts for degrees and eliminates edges within partitions has consistent results (in terms of log MSE).

for this structure and while the theory of Section 4 does not apply to this type of interference function, we still see good performance in this simulation. This behavior is consistent across values of K.

10. Discussion. This article provided a new approach to bounding the bias and mean squared error of the Neymanian estimator of the average treatment effect under interference. It introduced the notion of quasi-coloring to better understand the balance needed in the randomization scheme to account for interference. Based on quasi-colorings, we developed a restricted randomization scheme that has good theoretical properties and performs well in simulations.

There are a number of directions for future research. The general notion of perfect quasi-colorings provides intuition for the construction of other linear unbiased estimators. For example, we can construct a partial-perfect-quasi-coloring by only treating one node. This treatment produces the unbiased estimator $Y_{\text{treated}} - \bar{Y}_c$, where $\bar{Y}_c$ is the average outcome of all the control units that are not neighbors of the treated unit. The weights associated with treated and control units are still interpretable.

It is also possible to extend the machinery of this paper to other estimands and estimators of interest. For example, one estimand is the interference effect of having exactly one treated neighbor. Following the strategy taken in this paper, Neyman-type estimators, such as the difference in outcomes of control (resp. treated) nodes with one treated neighbor and no treated neighbors, can be developed. In turn, restricting the randomization scheme appropriately may reduce bias and variance of the estimate. Our methodology can be further extended to more general effects.

APPENDIX A: BOUNDS ON BIAS

For $v \in V(G)$ and $\mathrm{T} \subseteq \binom{V(G)}{n}$, let

$$\xi_v = \chi_v^{\mathrm{T}} f_v\big(\mathrm{T} \cap \mathcal{N}(V)\big)$$

denote the interference effect on v, so that ξ in (2.5) can be expressed as

$$\xi = \frac{1}{pqn} \sum_{v \in V(G)} \xi_v.$$

Given $v, w \in V(G)$, define the *weight of w on v* as

$$\begin{aligned} W_v(w) &= \sup_{A \subseteq \mathcal{N}(v) \setminus \{w\}} \big| f_v(A) - f_v\big(A \cup \{w\}\big)\big| && \text{if } w \in \mathcal{N}(v) \\ &= 0 && \text{otherwise.} \end{aligned} \tag{A.1}$$

LEMMA A.1. *For a partition $\mathcal{P} = \{S_1, \ldots, S_n\} \in \binom{V(G)}{r, \ldots, r}$ and the treatment assignment mechanism $\mathrm{T}_{\vec{B},\mathcal{P}}$ given in (3.1), we have*

$$\big|\mathbb{E}_{\vec{B}}(\xi)\big| \le \frac{1}{nr(r-1)} \sum_{i=1}^{n} \sum_{\{w,w'\} \in \binom{S_i}{2}} \big(W_w(w') + W_{w'}(w)\big).$$

The full generality of Lemma A.1 may be of use in a weighted interference model, as the formalism of weights allows one to capture the fact that different connections may have different strengths. Including a weak connection (low weight edge) in $\mathcal{P}$ will affect the bias less than including a strong connection. The following result will imply Lemma A.1.

LEMMA A.2. *For all* $j = 1, \dots, n$ *and* $v \in S_j$, *we have*

$$\left|\mathbb{E}_{B_j}[\xi_v \mid B_1, \dots, B_{j-1}, B_{j+1}, \dots, B_n]\right| \le \frac{pq}{r(r-1)} \sum_{w \in S_j \setminus \{v\}} W_v(w)$$

when $\mathrm{T} = \mathrm{T}_{\vec{B}, \mathcal{P}}$.

PROOF. Without loss of generality, assume that $j = 1$ and $v = w_1^1$. When $1 \in B_1$, define a random variable B_1' with values in $\binom{[r]}{p}$ by choosing B_1' uniformly from

$$\left\{A \in \binom{[r]}{p} \;\middle|\; B_1 \setminus A = \{1\}\right\}.$$

Let $B_i' = B_i$ for $i \neq 1$. Denote by ξ' (resp. ξ_v') the interference effect ξ (resp. the interference effect ξ_v' on v) for the treatment group $\mathrm{T}' = \mathrm{T}_{\vec{B}', \mathcal{P}}$.

When $1 \in B_1$, we have

$$p\xi_v + q\xi_v' = pq\big(f_v(\mathrm{T} \cap \mathcal{N}(v)) - f_v(\mathrm{T}' \cap \mathcal{N}(v))\big),$$

so that

$$\left|p\xi_v + q\xi_v'\right| \le pq\left|f_v(\mathrm{T} \cap \mathcal{N}(v)) - f_v(\mathrm{T}' \cap \mathcal{N}(v))\right| \le W_v(w),$$

where $\mathrm{T} \Delta \mathrm{T}' = \{v, w\}$. Taking expectations with respect to B_1', it follows that, when $1 \in B_1$, we have

$$\mathbb{E}_{B_1'}\left[\left|p\xi_v + q\xi_v'\right| \mid \vec{B}\right] \le \frac{pq}{r-1} \sum_{w \in S_1 \setminus \{v\}} W_v(w)$$

so that

$$\left|\mathbb{E}_{B_1}\left[\frac{p\xi_v + q\xi_v'}{r} \mid \{1 \notin B_1\}, B_2, \dots, B_n\right]\right| \le \frac{pq}{r(r-1)} \sum_{w \in S_1 \setminus \{v\}} W_v(w) \tag{A.2}$$

by the triangle inequality.

Note that $\mathcal{L}(B_1 \mid 1 \notin B_1) = \mathcal{L}(B_1' \mid 1 \in B_1)$, where $\mathcal{L}$ denotes the law of a random variable. It follows that

$$\mathbb{E}_{B_j}[\xi_v \mid 1 \notin B_1, B_2, \dots, B_n] = \mathbb{E}_{B_1'}\left[\xi_v' \mid 1 \in B_1, B_2, \dots, B_n\right]. \tag{A.3}$$

Combining (A.2) and (A.3) and using the fact that

$$\Pr[1 \in B_1 \mid B_2, \dots, B_n] = \frac{p}{r},$$

we obtain the lemma. $\square$

PROOF OF LEMMA A.1. It follows from Lemma A.2 that

$$|\mathbb{E}_{\vec{B}} \xi_v| \le \frac{pq}{r(r-1)} \sum_{w \in S_j \setminus \{v\}} W_v(w).$$

Summing over $v \in V(G)$, we have

$$|\mathbb{E}_{\vec{B}} \xi| \le \frac{1}{pqn} \sum_{v \in V(G)} |\mathbb{E}_{\vec{B}} \xi_v| \le \frac{1}{nr(r-1)} \sum_{i=1}^{n} \sum_{\{w, w'\} \in \binom{S_i}{2}} \big(W_w(w') + W_{w'}(w)\big),$$

as desired. $\square$

PROOF OF LEMMA 3.3. From (A.1) and (3.2), it follows that

$$W_v(w) \le \frac{K_v}{d(v)},$$

with $W_v(w) = 0$ if $\{v, w\} \notin E(G)$. The lemma therefore follows from Lemma A.1. □

PROOF OF PROPOSITION 3.6. By the linearity of expectation, we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}}\biggl[\sum_{v\in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)} K_v\biggr] \\
&= \sum_{v\in V(G)} \frac{K_v}{d(v)} \sum_{v\in e\in E(G)} \mathbb{P}(e \subseteq S_i \text{ for some } i) \\
&= \sum_{v\in V(G)} \frac{K_v}{d(v)} \sum_{v\in e\in E(G)} \frac{r-1}{rn-1} = \sum_{v\in V} \frac{K_v}{d(v)} \cdot d(v) \cdot \frac{r-1}{rn-1} \\
&= \frac{r-1}{rn-1} \sum_{v\in V(G)} K_v.
\end{aligned}$$

The proposition follows, by Lemma 3.3. □

APPENDIX B: BOUNDS ON MSE: DENSE CASE

The following L^2 bound is the key to the proofs of all of the MSE bounds.

LEMMA B.1. *For all* $\mathcal{P} = (S_1, \dots, S_n) \in \binom{V(G)}{r,\dots,r}$ *and all* $v, v' \in S_j$, *we have*

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\vec{B}}\bigl[\mathbf{d}\bigl(\vec{d}(v), \vec{d}(v')\bigr)^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\bigr]} \\
&\quad \le \frac{K_1}{d_{\max}} |d(v) - d(v')| \\
&\quad\quad + \frac{K_2}{r}\biggl(\frac{\sqrt{2pq}}{\sqrt{d(v)}} + \frac{\sqrt{2pq}}{\sqrt{d(v')}} + \frac{q \cdot 1_{E(G)}(\{v, v'\})}{d(v)} + \frac{p \cdot 1_{E(G)}(\{v, v'\})}{d(v')}\biggr)
\end{aligned}$$

when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

PROOF. Note that

$$\mathbf{d}_K\bigl(\vec{d}_{\mathrm{T}}(v), \vec{d}_{\mathrm{T}}(v')\bigr) = \frac{K_1}{d_{\max}} |d(v) - d(v')| + \frac{K_2}{r} |F|,$$

where

$$F = \frac{r|\mathrm{T} \cap \mathcal{N}(v)|}{d(v)} - \frac{r|\mathrm{T} \cap \mathcal{N}(v')|}{d(v')}.$$

Thus, it suffices to prove that

$$\begin{aligned}
\sqrt{\mathbb{E}_{\vec{B}}\bigl[F^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\bigr]} &\le \frac{\sqrt{2pq}}{\sqrt{d(v)}} + \frac{\sqrt{2pq}}{\sqrt{d(v')}} \\
&\quad + \frac{q \cdot 1_{E(G)}(\{v, v'\})}{d(v)} + \frac{p \cdot 1_{E(G)}(\{v, v'\})}{d(v')}.
\end{aligned}$$

For $w \in V(G)$, let

$$F_w = \chi_w^T \left(\frac{1_{\mathcal{N}(v)}(w)}{d(v)} - \frac{1_{\mathcal{N}(v')}(w)}{d(v')} \right).$$

Note that

$$\mathbb{E}_{\vec{B}}\left[(\chi_w^T)^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T} \right] = \begin{cases} pq & \text{if } w \notin S_j, \\ \dfrac{pqr - p^2 - q^2}{r-2} & \text{if } w \in S_j. \end{cases}$$

In particular, we have

$$\mathbb{E}_{\vec{B}}\left[(\chi_w^T)^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T} \right] \le pq.$$

It follows that

$$\begin{aligned} \mathbb{E}_{\vec{B}}\left[F_w^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T} \right] &\le pq \left(\frac{1_{\mathcal{N}(v)}(w)}{d(v)} - \frac{1_{\mathcal{N}(v')}(w)}{d(v')} \right)^2 \\ &\le pq \frac{1_{\mathcal{N}(v)}(w)}{d(v)^2} + pq \frac{1_{\mathcal{N}(v')}(w)}{d(v')^2}. \end{aligned}$$

For all i, let

$$F_i = \frac{r|\mathrm{T} \cap S_i \cap \mathcal{N}(v)| - p|S_i \cap \mathcal{N}(v)|}{d(v)} - \frac{r|\mathrm{T} \cap S_i \cap \mathcal{N}(v')| - p|S_i \cap \mathcal{N}(v')|}{d(v')}.$$

For $i \neq j$ and $w, w' \in S_i$ with $w \neq w'$, we have

$$\mathrm{Corr}\left(\chi_w^{\mathrm{T}}, \chi_{w'}^{\mathrm{T}} \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right) = -\frac{1}{r-1}.$$

By Lemma 6.8, we have

$$\begin{aligned} &\mathrm{Var}_{B_i}\left(F_i \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right) \\ &\quad \le \sum_{w, w' \in S_i} \mathrm{Var}_{B_i}\left(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right) \left|\mathrm{Corr}_{B_i}\left(F_w, F_{w'} \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right)\right| \\ &\quad \le \left(\sum_{w \in S_i} \mathrm{Var}_{B_i}\left(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right) \right) \left(1 + \frac{r-1}{r-1} \right) \\ &\quad = 2 \sum_{w \in S_i} \mathrm{Var}_{B_i}\left(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right). \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}_{B_i}\left[F_i^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T} \right] &\le 2 \sum_{w \in S_i} \mathbb{E}_{\vec{B}}\left[F_w^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T} \right] \\ &\le \frac{2pq|S_i \cap \mathcal{N}(v)|}{d(v)^2} + \frac{2pq|S_i \cap \mathcal{N}(v')|}{d(v')^2}. \end{aligned}$$

Similarly, for $w, w' \in S_i \setminus \{v, v'\}$ with $w \neq w'$ we have

$$\mathrm{Corr}\left(\chi_w^{\mathrm{T}}, \chi_{w'}^{\mathrm{T}} \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\right) = -\frac{1}{r-3}$$

for $w, w' \in S_j \setminus \{v, v'\}$. By Lemma 6.8, we have

$$\begin{aligned}
&\operatorname{Var}_{B_j}\big(F_j - F_v - F_{v'} \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big) \\
&\quad \le \sum_{w,w' \in S_j \setminus \{v,v'\}} \operatorname{Var}_{B_j}\big(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big) \\
&\qquad \times \big|\operatorname{Corr}_{B_j}\big(F_w, F_{w'} \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big)\big| \\
&\quad \le \Big(\sum_{w \in S_j \setminus \{v,v'\}} \operatorname{Var}_{B_j}\big(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big)\Big)\Big(1 + \frac{r-3}{r-3}\Big) \\
&\quad = 2 \sum_{w \in S_j \setminus \{v,v'\}} \operatorname{Var}_{B_j}\big(F_w \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}_{\vec{B}}\big[(F_j - F_v - F_{v'})^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big] \\
&\quad \le 2 \sum_{w \in S_j \setminus \{v,v'\}} \mathbb{E}_{\vec{B}}\big[F_w^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big] \\
&\quad \le \frac{2pq|S_j \cap \mathcal{N}(v) \setminus \{v'\}|}{d(v)^2} + \frac{2pq|S_j \cap \mathcal{N}(v') \setminus \{v'\}|}{d(v')^2}.
\end{aligned}$$

As $B_1, \dots, B_n$ are independent (even conditioned on the event that $v \in \mathrm{T}$ and $v' \notin \mathrm{T}$), it follows that

$$\mathbb{E}_{\vec{B}}\Big[\Big(\sum_i F_i - F_v - F_{v'}\Big)^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\Big] \le \frac{4pq}{d(v)} + \frac{4pq}{d(v')},$$

so that

$$\sqrt{\mathbb{E}_{\vec{B}}\Big[\Big(\sum_i F_i - F_v - F_{v'}\Big)^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\Big]} \le \frac{\sqrt{2pq}}{\sqrt{d(v)}} + \frac{\sqrt{2pq}}{\sqrt{d(v')}}.$$

Noting that $F = \sum_i F_i$ and using the fact that $|F_v| \le \frac{q \cdot 1_{E(G)}(\{v,v'\})}{d(v)}$ and $|F_{v'}| \le \frac{p \cdot 1_{E(G)}(\{v,v'\})}{d(v')}$, it follows that

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\vec{B}}\big[F^2 \mid v \in \mathrm{T} \text{ and } v' \notin \mathrm{T}\big]} \\
&\quad \le \frac{\sqrt{2pq}}{\sqrt{d(v)}} + \frac{\sqrt{2pq}}{\sqrt{d(v')}} + \frac{q \cdot 1_{E(G)}(\{v,v'\})}{d(v)} + \frac{p \cdot 1_{E(G)}(\{v,v'\})}{d(v')}
\end{aligned}$$

and the proof is finished. □

For $v \in S_i$, define

$$D_{\mathrm{T}}^v = 1_{\mathrm{T}}(v)\Big(q\delta_{\vec{d}(v)} - \sum_{w \in S_i \setminus \mathrm{T}} \delta_{\vec{d}(w)}\Big).$$

Intuitively, D_{T}^v is the contribution of v to D_{T} if v is treated and is 0 if v is not treated. Note that

$$D_{\mathrm{T}}^v = \sum_{w \in S_i \setminus \{v\}} 1_{\mathrm{T}}(v) 1_{V(G) \setminus \mathrm{T}}(w)(\delta_{\vec{d}(v)} - \delta_{\vec{d}(w)}).$$

By the triangle inequality, we have

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{v}\|_{\mathbf{d}}^{2}]} \\
&\quad\le \sqrt{\frac{pq}{r(r-1)}} \sum_{w \in S_i \setminus \{v\}} \sqrt{\mathbb{E}[\mathbf{d}(\vec{d}(v), \vec{d}(w))^2 \mid v \in \mathrm{T} \text{ and } w \notin \mathrm{T}]} \\
&\quad\le \frac{\sqrt{pq}}{r-1} \sum_{w \in S_i \setminus \{v\}} \sqrt{\mathbb{E}[\mathbf{d}(\vec{d}(v), \vec{d}(w))^2 \mid v \in \mathrm{T} \text{ and } w \notin \mathrm{T}]}.
\end{aligned} \tag{B.1}$$

PROOF OF PROPOSITION 5.9. By Lemma B.1 and (B.1), we have

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{v}\|_{\mathbf{d}}^{2}]} \\
&\quad\le \frac{K_1\sqrt{pq}}{(r-1)d_{\max}} \sum_{w \in S_i \setminus \{v\}} |d(v) - d(w)| + \frac{2K_2 pq}{r\sqrt{d(v)}} \\
&\qquad + \sum_{w \in S_i \setminus \{v\}} \frac{2K_2 pq}{r(r-1)\sqrt{d(w)}} + \frac{K_2\sqrt{pq}}{r(r-1)} \sum_{w \in S_i \cap \mathcal{N}(v)} \left(\frac{q}{r \cdot d(v)} + \frac{p}{r \cdot d(w)}\right) \\
&\quad\le \frac{K_1\sqrt{pq}}{(r-1)d_{\max}} \sum_{w \in S_i \setminus \{v\}} |d(v) - d(w)| + \frac{2K_2 pq}{r\sqrt{d(v)}} \\
&\qquad + \sum_{w \in S_i \setminus \{v\}} \frac{2K_2 pq}{r(r-1)\sqrt{d(w)}} + K_2 \sum_{w \in S_i \cap \mathcal{N}(v)} \left(\frac{q}{r \cdot d(v)} + \frac{p}{r \cdot d(w)}\right).
\end{aligned}$$

Noting that $D_{\mathrm{T}} = \frac{1}{pqn} \sum_{v \in V(G)} D_{\mathrm{T}}^{v}$, the triangle inequality implies that

$$\begin{aligned}
\sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}\|_{\mathbf{d}}^{2}]} &\le \frac{1}{pqn} \sum_{v \in V(G)} \sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{v}\|_{\mathbf{d}}^{2}]} \\
&\le \frac{K_1}{\sqrt{pq}n} C_{\mathcal{P}} + \frac{1}{rn} \sum_{v \in V(G)} \frac{4K_2}{\sqrt{d(v)}} + \frac{K_2}{pqn} \sum_{v \in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)},
\end{aligned}$$

as claimed. □

PROOF OF PROPOSITION 7.7. By Lemma B.1 and (B.1), we have

$$\begin{aligned}
\sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{v}\|_{\mathbf{d}}^{2}]} &\le \frac{2Kpq}{r\sqrt{d(v)}} + \sum_{w \in S_i \setminus \{v\}} \frac{2Kpq}{r(r-1)\sqrt{d(w)}} \\
&\qquad + K \sum_{w \in S_i \cap \mathcal{N}(v)} \left(\frac{q}{r \cdot d(v)} + \frac{p}{r \cdot d(w)}\right).
\end{aligned}$$

Noting that $D_{\mathrm{T}}^{\pi} = \frac{1}{pqn} \sum_{v \in \pi} D_{\mathrm{T}}^{v}$, we have

$$\sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{\pi}\|_{\mathbf{d}}^{2}]} \le \frac{1}{rn} \sum_{v \in \pi} \frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn} \sum_{v \in \pi} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}. \tag{B.2}$$

Summing over $\pi \in \Pi$, it follows that

$$\sqrt{\mathbb{E}_{\vec{B}}\Big[\Big(\sum_{\pi \in \Pi} \|D_{\mathrm{T}}^{\pi}\|_{\mathbf{d}}\Big)\Big]^2} \le \sum_{\pi \in \Pi} \sqrt{\mathbb{E}_{\vec{B}}[\|D_{\mathrm{T}}^{\pi}\|_{\mathbf{d}}^2]}$$
$$\le \frac{1}{rn} \sum_{v \in V(G)} \frac{4K}{\sqrt{d(v)}} + \frac{K}{pqn} \sum_{v \in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)},$$

as claimed. □

APPENDIX C: BOUNDS ON MSE: SPARSE CASE

As Proposition 3.6 shows, introducing randomness can help reduce bias. We will first need a generalization of Proposition 3.6 to a class of semi-restricted randomizations.

Assume that the function f_v is K_v-Lipshitz and define the quantity

$$K_{\max} = \max_{v \in V(G)} K_v. \tag{C.1}$$

PROPOSITION C.1. *Fix a partition Π of $V(G)$ into sets of size divisible by r. When $\mathcal{P}$ is sampled uniformly from $\binom{\Pi}{r,\dots,r}$, we have*

$$\mathbb{E}_{\mathcal{P}}\big|\mathbb{E}_{\vec{B}}(\xi|\mathcal{P})\big| \le \frac{2K_{\max}|\Pi|}{rn}$$

when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

PROOF. By the linearity of expectation, we have

$$\mathbb{E}_{\mathcal{P}}\Big[\sum_{v \in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)} K_v\Big]$$
$$= \sum_{v \in V(G)} \frac{K_v}{d(v)} \sum_{v \in e \in E(G)} \mathbb{P}(e \subseteq S_i \text{ for some } i)$$
$$\le \sum_{\pi \in \Pi} \sum_{v \in \pi} \frac{K_v}{d(v)} \sum_{v \in e \in E(g)} \frac{r-1}{|\pi|-1} = \sum_{\pi \in \Pi} \sum_{v \in \pi} \frac{K_v}{d(v)} \cdot d(v) \cdot \frac{r-1}{|\pi|-1}$$
$$\le (r-1)K_{\max} \sum_{\pi \in \Pi} \frac{|\pi|}{|\pi|-1} \le 2(r-1)K_{\max}|\Pi|.$$

The proposition follows, by Lemma 3.3. □

C.1. With types. The following lemma will be used in the proof of Proposition 7.10. Recall $K_{\max}$ and $d_{\min}$ from equations (C.1) and (1.1) respectively.

LEMMA C.2. *When $\mathcal{P}$ is sampled uniformly from $\binom{\Pi}{r,\dots,r}$, we have*

$$\mathbb{E}_{\mathcal{P}}\big(\mathbb{E}_{\vec{B}}(\xi \mid \mathcal{P})\big)^2 \le \frac{2K_{\max}^2|\Pi|}{nr \cdot d_{\min}}$$

when $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$.

PROOF. The proof of Proposition C.1 shows that

$$\mathbb{E}_{\mathcal{P}}\left[\sum_{v\in V(G)} \frac{K_v \cdot |\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}\right] \le 2(r-1)K_{\max}|\Pi|.$$

We have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}}\left[\left\{\sum_{v\in V(G)} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)} K_v\right\}^2\right] \\
&\quad \le \mathbb{E}_{\mathcal{P}}\left[\sum_{v\in V(G)} \frac{K_v \cdot |\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}\right] \cdot \max_{\mathcal{P}\in\binom{\Pi}{r,\dots,r}}\left[\sum_{v\in V(G)} \frac{K_v \cdot |\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}\right] \\
&\quad \le 2(r-1)K_{\max}|\Pi| \cdot \frac{nr(r-1)K_{\max}}{d_{\min}} \\
&\quad = \frac{2nr(r-1)^2K_{\max}^2|\Pi|}{d_{\min}}.
\end{aligned}$$

The lemma follows, by Lemma 3.3. □

PROOF OF PROPOSITION 7.10. Let $\mathcal{P} = (S_1, \dots, S_n)$ be sampled uniformly from $\binom{\Pi}{r,\dots,r}$ and let $\mathrm{T} = \mathrm{T}_{\vec{B},\mathcal{P}}$. For $i = 1, 2, \dots, n$, define

$$\xi_i = \sum_{v\in S_i} \xi_v.$$

Note that if ξ_i and ξ_j are dependent given $\mathcal{P}$ and $i \ne j$, then either there is an edge between S_i and S_j or there exists k such that there are edges between S_i and S_k and between S_k and S_j. In particular, for fixed i, there are at most $r^2d_{\max}^2 + 1$ values of j such that ξ_i and ξ_j are dependent given $\mathcal{P}$. By Lemmata B.1 and 6.9, it follows that

$$\begin{aligned}
\mathrm{Var}_{\vec{B}}(\xi \mid \mathcal{P}) &\le (r^2d_{\max}^2 + 1)\sum_{i=1}^n \mathrm{Var}_{\vec{B}}(\xi_i \mid \mathcal{P}) \\
&\le (r^2d_{\max}^2 + 1)\sum_{i=1}^n \mathbb{E}_{\vec{B}}\left[\|D_{\mathrm{T}}^{S_i}\|_{\mathbf{d}}^2 \mid \mathcal{P}\right].
\end{aligned}$$

By (B.2) in the proof of Proposition 7.7, it follows that

$$\begin{aligned}
\frac{\mathrm{Var}_{\vec{B}}(\xi \mid \mathcal{P})}{K^2} &\le (r^2d_{\max}^2 + 1)\sum_{i=1}^n \left(\frac{1}{rn}\sum_{v\in S_i} \frac{4}{\sqrt{d(v)}} + \frac{1}{pqn}\sum_{v\in S_i} \frac{|\mathcal{P}_v \cap \mathcal{N}(v)|}{d(v)}\right)^2 \\
&\le \frac{r^2d_{\max}^2 + 1}{n^2}\sum_{i=1}^n \left(\frac{4}{\sqrt{d_{\min}}} + \frac{r\min\{r-1, d_{\min}\}}{pq \cdot d_{\min}}\right)^2 \\
&= \frac{r^2d_{\max}^2 + 1}{n}\left(\frac{4}{\sqrt{d_{\min}}} + \frac{r\min\{r-1, d_{\min}\}}{pq \cdot d_{\min}}\right)^2.
\end{aligned}$$

Taking square roots yields that

$$\sqrt{\mathrm{Var}_{\vec{B}}(\xi \mid \mathcal{P})} \le \frac{4K\sqrt{r^2d_{\max}^2 + 1}}{\sqrt{n \cdot d_{\min}}} + \frac{rK\min\{r-1, d_{\min}\}\sqrt{r^2d_{\max}^2 + 1}}{pq\sqrt{n} \cdot d_{\min}}.$$

The bound on $|\mathbb{E}_T\xi|$ is given by Proposition C.1. It remains to prove the bound on $\mathbb{E}_\mathrm{T}\xi^2$. Lemma C.2 and the previous paragraph together imply that

$$\sqrt{\mathbb{E}_\mathrm{T}\xi^2} = \sqrt{\mathbb{E}_{\mathcal{P}}\mathbb{E}_{\vec{B}}(\xi \mid \mathcal{P})^2 + \mathbb{E}_{\mathcal{P}}\operatorname{Var}(\xi \mid \mathcal{P})}$$
$$\leq \frac{K\sqrt{2|\Pi|}}{\sqrt{nr}\cdot d_{\min}} + \frac{4K\sqrt{r^2d_{\max}^2+1}}{\sqrt{n}\cdot d_{\min}} + \frac{rK\min\{r-1, d_{\min}\}\sqrt{r^2d_{\max}^2+1}}{pq\sqrt{n}\cdot d_{\min}},$$

as desired. $\square$

C.2. Without types. The key to the proof of Proposition 6.7 is to note that the treatment group $T_{\vec{B},\mathcal{P}^{**}}$ has the same distribution as the treatment group T_Π for a suitably chosen Π.

PROOF OF PROPOSITION 6.7. Let $D = d(V(G)\setminus S)$. For $d \in D$, let $V_d = \{v \in V(G)\setminus S \mid d(v) = d\}$. Define

$$\Pi = \bigl(S_1, \dots, S_k, (V_d)_{d\in D}\bigr).$$

By construction, we have that $|\Pi| \leq 2(d_{\max} - d_{\min}) + 1 \leq 2d_{\max}$ and that $k \leq d_{\max} - d_{\min}$.

For $d \in D$, define $g_{V_d} = f|_{V_d}$. For $1 \leq i \leq k$, define

$$g_{S_i}(a,b) = f\left(\left\lfloor \frac{a\cdot \max_{v\in S_i} d(v)}{a+b}\right\rfloor, \left\lceil \frac{b\cdot \max_{v\in S_i} d(v)}{a+b}\right\rceil\right).$$

It is straightforward to verify that f and $g_{\Pi(v)}$ agree on $\{(a,b) \in \mathbb{Z}_{\geq 0}^2 \mid a+b = d(v)\}$ for all $v \notin S$ and

$$\bigl|f(a,b) - g_{\Pi(v)}(a,b)\bigr| \leq K_1\frac{|a+b-\max_{v\in S_i} d(v)|}{d_{\max}} + \frac{K_2}{d_{\min}}.$$

Define an auxiliary random variable

$$\zeta_v^\mathrm{T} = \chi_v^\mathrm{T} g_{\Pi(v)}(\vec{d}_T(v))$$

and let

$$\zeta = \frac{1}{pqn}\sum_{v\in V(G)}\zeta_v^\mathrm{T}.$$

The discussion of the previous paragraph shows that $\zeta_v^\mathrm{T} = \xi_v^\mathrm{T}$ for all $v \in V(G)\setminus S$. It follows that

$$|\zeta - \xi| \leq \sum_{v\in S}\bigl|\zeta_v^\mathrm{T} - \xi_v^\mathrm{T}\bigr|$$
$$\leq \frac{K_1}{d_{\max}}\sum_{i=1}^k\sum_{u\in S_i}\Bigl|d(u) - \max_{v\in S_i} d(v)\Bigr| + \frac{K_2}{d_{\min}}\sum_{i=1}^n |S_i|$$
$$\leq \frac{K_1 r}{d_{\max}}\sum_{i=1}^k\Bigl|\max_{v\in S_i} d(v) - \min_{v\in S_i} d(v)\Bigr| + \frac{K_2 rk}{d_{\min}}$$
$$\leq K_1 r + \frac{K_2 r(d_{\max} - d_{\min})}{d_{\min}}.$$

Because $|\Pi| \leq 2d_{\max}$, the proposition then follows by bounding ζ for the treatment $\mathrm{T} = \mathrm{T}_\Pi$ using Proposition 7.10. $\square$

APPENDIX D: HOMOPHILY

PROOF OF PROPOSITION 8.1. The first assertion follows from Lemma 2.1 because $\mathbb{P}(v \in \mathrm{T}) = \frac{p}{r}$ for all $v \in V(G)$.

Note that

$$t_{\text{ideal}} = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} y_v = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} (x_v + 1_{\mathrm{T}}(v) t_v).$$

As $|\mathrm{T} \cap \pi| = \frac{p|\pi|}{r}$ for all $\pi \in \Pi$, we have that $\sum_{v \in \pi} \chi_v^{\mathrm{T}} = 0$ for all $\pi \in \Pi$. It follows that

$$\sum_{v \in V(G)} \chi_v^{\mathrm{T}} \left(x_{\Pi_v} + \frac{q}{r} t_{\Pi_v} \right) = \sum_{\pi \in \Pi} \left(x_\pi + \frac{q}{r} t_\pi \right) \sum_{v \in \pi} \chi_v^{\mathrm{T}} = 0.$$

Hence, we have that

$$\begin{aligned} t_{\text{ideal}} &= \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} \left(x_v - x_{\Pi_v} - \frac{q}{r} t_{\pi_v} + 1_{\mathrm{T}}(v) t_v \right) \\ &= \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} \left(\epsilon_v + 1_{\mathrm{T}}(v) t_v - \frac{q}{r} t_v \right). \end{aligned}$$

If $v \in \mathrm{T}$, then we have that

$$\chi_v^{\mathrm{T}} \left(1_{\mathrm{T}}(v) t_v - \frac{q}{r} t_v \right) = q t_v - \frac{q^2}{r} t_v = \frac{pq}{r} t_v,$$

while if $v \notin \mathrm{T}$, then we have that

$$\chi_v^{\mathrm{T}} \left(1_{\mathrm{T}}(v) t_v - \frac{q}{r} t_v \right) = -p \left(-\frac{q}{r} t_v \right) = \frac{pq}{r} t_v.$$

It follows that

$$t_{\text{ideal}} = \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} \epsilon_v + \frac{1}{pqn} \sum_{v \in V(G)} \frac{pq}{r} t_v = \bar{t} + \frac{1}{pqn} \sum_{v \in V(G)} \chi_v^{\mathrm{T}} \epsilon_v.$$

Note that $\mathrm{Corr}(\chi_v^{\mathrm{T}}, \chi_w^{\mathrm{T}}) = -\frac{1}{r-1}$ if $v \neq w$ lie in a single part of Π and $\mathrm{Corr}(\chi_v^{\mathrm{T}}, \chi_w^{\mathrm{T}}) = 0$ if v and w lie in different parts of Π. Thus, we have

$$\sum_{w \in V(G)} \left| \mathrm{Corr}(\chi_v^{\mathrm{T}}, \chi_w^{\mathrm{T}}) \right| = 2$$

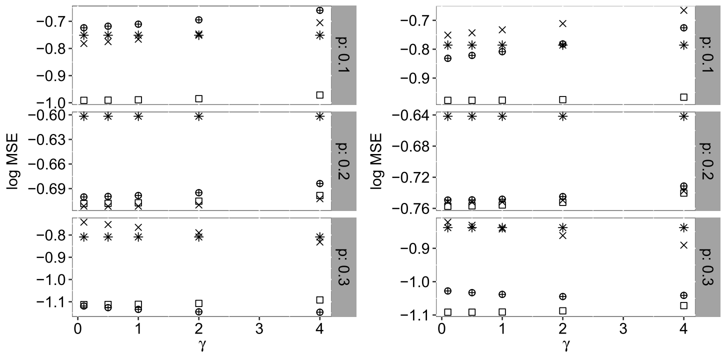
for all $v \in V(G)$. By Lemma 6.8, it follows that

$$\mathrm{Var}(t_{\text{ideal}}) \leq \frac{1}{p^2 q^2 n^2} \sum_{v \in V(G)} 2 \, \mathrm{Var}(\chi_v^{\mathrm{T}}) \epsilon_v^2 = \frac{2 r \sigma^2}{pqn},$$

as desired. □

APPENDIX E: ADDITIONAL SIMULATION RESULTS

Figure 3 provides further results for the stochastic blockmodel simulation of Section 9.4. We note that the behavior of the Neymanian estimator after ordering by degree and ensuring that partitions do not overlap with the edges of the graph maintains its behavior between the two graphs, while the approach that only accounts for degrees exhibits different behavior conditional on the graph.



(a) Graph 1 (b) Graph 2

FIG. 3. *Simulation results for different randomization schemes and estimators across two different baseline graphs.* $\times$ *refers to the Neymanian estimator coupled with a completely randomized deisign.* $*$ *refers to the stratified estimator after a completely randomized design.* $\oplus$ *refers to the Neymanian estimator after ordering by degree only and* $\square$ *refers to the Neymanian estimator after ordering by degree and ensuring that partitions do not overlap with the edges of the graph.*

Acknowledgments. The authors thank Edoardo Airoldi, Dean Eckles, Vishesh Karwa and Daniel Sussman for helpful conversations as well as two anonymous reviewers, an associate editor and the editor for detailed and helpful comments. Part of this research was conducted while RJ was an Economic Design Fellow at the Harvard Center of Mathematical Sciences and Applications.

The first author was supported in part by an NSF Graduate Research Fellowship under Grant No. DGE-1745303. Part of this research was conducted while he was an Economic Design Fellow at the Harvard Center of Mathematical Sciences and Applications.

The second author was supported in part by an ONR grant.

The third author was supported in part by a NSF MSPRF and an NIH grant.

REFERENCES

[1] AIELLO, A. E., SIMANEK, A. M., EISENBERG, M. C., WALSH, A. R., DAVIS, B., VOLZ, E., CHENG, C., RAINEY, J. J., UZICANIN, A. et al. (2016). Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemics* **15** 38–55.

[2] ARONOW, P. M. and SAMII, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* **11** 1912–1947. MR3743283 https://doi.org/10.1214/16-AOAS1005

[3] BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

[4] BASSE, G. W. and AIROLDI, E. M. (2015). Optimal design of experiments in the presence of network-correlated outcomes. ArXiv e-prints.

[5] CALTAGIRONE, F., LELARGE, M. and MIOLANE, L. (2018). Recovering asymmetric communities in the stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **5** 237–246. MR3859911 https://doi.org/10.1109/TNSE.2017.2758201

[6] CHOI, D. (2017). Estimation of monotone treatment effects in network experiments. *J. Amer. Statist. Assoc.* **112** 1147–1155. MR3735366 https://doi.org/10.1080/01621459.2016.1194845

[7] ECKLES, D., KARRER, B. and UGANDER, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference* **5**.

[8] FORASTIERE, L., AIROLDI, E. M. and MEALLI, F. (2016). Identification and estimation of treatment and interference effects in observational studies on networks. ArXiv Preprint arXiv:1609.06245.

[9] GRUHL, D., GUHA, R., LIBEN-NOWELL, D. and TOMKINS, A. (2004). Information diffusion through blogspace. In *Proceedings of the* 13*th International Conference on World Wide Web* 491–501. ACM.

[10] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 https://doi.org/10.1198/016214502388618906

[11] HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. With discussion and a reply by the author. MR0867618

[12] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[13] HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

[14] KARWA, V. and AIROLDI, E. (2016). Bias of classic estimates under interference. In Preparation.

[15] KRAPIVSKY, P. L., REDNER, S. and LEYVRAZ, F. (2000). Connectivity of growing random networks. *Phys. Rev. Lett.* **85** 4629.

[16] LU, L. and PENG, X. (2013). Spectra of edge-independent random graphs. *Electron. J. Combin.* **20** Paper 27, 18. MR3158266

[17] MOODY, J. (2001). Race, school integration, and friendship segregation in America. *Amer. J. Sociol.* **107** 679–716.

[18] RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

[19] RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

[20] RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Ann. Appl. Stat.* **2** 808–804. MR2516795 https://doi.org/10.1214/08-AOAS187

[21] SHALIZI, C. R. and THOMAS, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* **40** 211–239. MR2767833 https://doi.org/10.1177/0049124111404820

[22] SUSSMAN, D. L. and AIROLDI, E. M. (2017). Elements of estimation theory for causal effects in the presence of network interference. ArXiv Preprint arXiv:1702.03578.

[23] TOULIS, P. and KAO, E. (2013). Estimation of causal peer influence effects. In *ICML* 1489–1497.

[24] TOULIS, P., VOLFOVSKY, A. and AIROLDI, E. M. (2018). Propensity score methodology in the presence of network entanglement between treatments. arXiv preprint arXiv:1801.07310.

[25] WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature* **393** 440.

[26] YANG, J. and COUNTS, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *ICWSM* **10** 355–358.