Biometrika (2018), **105**, 4, *pp.* 849–858
Printed in Great Britain
doi: 10.1093/biomet/asy036
Advance Access publication 6 August 2018

Model-assisted design of experiments in the presence of network-correlated outcomes

By GUILLAUME W. BASSE

Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02138, U.S.A.
gbasse@fas.harvard.edu

and EDOARDO M. AIROLDI

Department of Statistical Science, Fox School of Business, Temple University, 1810 Liacouras Walk, Philadelphia, Pennsylvania 19122, U.S.A.
airoldi@alumni.harvard.edu

Summary

In this paper we consider how to assign treatment in a randomized experiment in which the correlation among the outcomes is informed by a network available pre-intervention. Working within the potential outcome causal framework, we develop a class of models that posit such a correlation structure among the outcomes. We use these models to develop restricted randomization strategies for allocating treatment optimally, by minimizing the mean squared error of the estimated average treatment effect. Analytical decompositions of the mean squared error, due both to the model and to the randomization distribution, provide insights into aspects of the optimal designs. In particular, the analysis suggests new notions of balance based on specific network quantities, in addition to classical covariate balance. The resulting balanced optimal restricted randomization strategies are still design-unbiased when the model used to derive them does not hold. We illustrate how the proposed treatment allocation strategies improve on allocations that ignore the network structure.

Some key words: Causal inference; Degree distribution; Network balance; Network data; Optimal treatment allocation; Randomized experiment; Rerandomization.

1. Introduction

The past decade has witnessed a surge of interest in causal analyses in the context of social networks, social media platforms and online advertising (Christakis & Fowler, 2007; Aral et al., 2009; Bakshy et al., 2011, 2012; Bond et al., 2012; Gui et al., 2015; Kim et al., 2015; Phan & Airoldi, 2015; Cavusoglu et al., 2016). From a statistical perspective, the challenging aspect of these applications is how to account for the presence of connections, or network data, observed pre-intervention, possibly with uncertainty. While there is a well-developed literature on several aspects of the statistical analysis of network data (Wasserman & Faust, 1994; Bickel & Chen, 2009; Goldenberg et al., 2010; Kolaczyk & Csárdi, 2014), the literature on methods for experimentation and causal analyses that use observed connections is still nascent (Rosenbaum, 2007; Hudgens & Halloran, 2008; Toulis & Kao, 2013; Ogburn & VanderWeele, 2017).

The need to account for network connections in causal analyses has led scholars to focus on two specific problem settings: network interference (Toulis & Kao, 2013; Ugander et al., 2013;

Aronow & Samii, 2017; Eckles et al., 2017), where the potential outcomes of unit i are functions of the treatment assigned to unit i and of the treatments assigned to other units that are related to unit i through the network; and network-correlated outcomes (McPherson et al., 2001; Shalizi & Thomas, 2011; Manski, 2013), where the network informs the correlation among the potential outcomes because the potential outcomes of unit i depend on its covariates, and the covariates of units that are connected are more similar than the covariates of those that are not. In this paper, we focus on the second setting, which has received less attention.

Restricted randomization as a way to increase the precision of estimates has a long tradition (Yates, 1948; Youden, 1972; Simon, 1979; Bailey, 1983; Higham et al., 2015). The basic idea is that some assignments are considered problematic and should be excluded; stratified randomization, for instance, implicitly excludes assignments for which certain covariates of interest are unbalanced between the treatment arms. In networks, the challenge is to identify the features that must be balanced, which makes it difficult to know how to restrict the randomization. Drawing inspiration from the model-assisted survey sampling literature (e.g., Särndal et al., 2003), we propose a model-assisted approach to design. We posit a working model for the potential outcomes specified conditionally on a network observed pre-intervention, and then restrict the randomization to assignments for which the estimator of interest achieves a low mean squared error. The class of models we propose leads to analytical expressions for the mean squared error that suggest new notions of balance in terms of network statistics related to the degree distribution. We also develop theoretical results showing that our model-assisted restricted randomization approach maintains the design-unbiasedness of the difference-in-means estimator even when the model is misspecified, and reduces its expected variance when the model holds.

2. Analytical insights for evaluating allocations

2·1. *Causal inference set-up*

We work within the potential outcomes framework (Rubin, 1974; Holland, 1986; Imbens & Rubin, 2015). We consider a population of N units, a binary treatment, denoted by $Z_i = 1$ if unit i is assigned to treatment, and real-valued outcomes, denoted by Y_i. The corresponding vectors are denoted by Y and Z. We make the stable unit-treatment-value assumption, so the outcome of unit i is a function only of the treatment assigned to it, $Y_i(Z) = Y_i(Z_i)$, thus excluding interference (Rubin, 1974). We consider a finite population setting, where the potential outcomes $Y(Z)$ are unknown constant quantities, given Z. The only source of variation is how treatment is allocated to units, which is done according to a distribution on the space of all binary vectors of length N, called the randomization distribution.

To illustrate model-assisted restricted randomization, we consider the average treatment effect as the inferential target of interest, $\tau^* = N^{-1}\sum_{i=1}^{N}\{Y_i(1) - Y_i(0)\}$, and the difference-in-means estimator of the average treatment effect,

$$\hat{\tau}(Y \mid Z) = \frac{\sum_{i=1}^{N} Z_i Y_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N}(1 - Z_i) Y_i}{\sum_{i=1}^{N}(1 - Z_i)}. \tag{1}$$

2·2. *The normal-sum model*

The model-assisted approach to experimental design requires a model, which is used to improve the inferential properties of the difference-in-means estimator when the model holds. We posit a model that depends on a network, which is available at the design stage.

Consider N units and an undirected network $\mathcal{G}$ among them or, equivalently, a binary adjacency matrix A of size $N \times N$ with the added constraint that $A_{ii} = 1$ for all i; we call A the extended adjacency matrix. The neighbourhood of a unit i is defined as the index set $\mathcal{N}_i = \{j : A_{ij} = 1 \text{ or } A_{ji} = 1\}$. Consider the model

$$X_j \sim N(\mu, \sigma^2), \tag{2}$$

$$Y_i(0) \mid X \sim N\big(\textstyle\sum_{j \in \mathcal{N}_i} X_j, \gamma^2\big), \tag{3}$$

$$Y_i(1) = Y_i(0) + \tau, \tag{4}$$

where $N(m, v)$ denotes the normal distribution with mean m and variance v. The network induces correlation among the control potential outcomes because the mean of each $Y_i(0)$ is the sum of the covariate values, X_j, of units j in a neighbourhood of unit i. The effect of treatment is additive. Equations (2)–(4) define the normal-sum model. The implied model for the observed outcomes, Y^{obs}, is given in the Supplementary Material. We generalize this model in § 2·5 but will otherwise focus on the normal-sum model for clarity in presenting the restricted randomization approach.

The normal-sum model provides a useful abstraction for exploring the problem of optimal design of experiments in the presence of network-correlated outcomes. An illustration will help to anchor the intuition. The normal-sum model arises naturally, for example, when considering the time users spend on a social media platform. Consider the binary treatment Z_i to be the exposure to a new feature of the website designed to increase engagement and time spent online, and let $Y_i(Z_i)$ be the time spent online by user i when assigned to treatment Z_i. The causal effect of interest τ is then the effect of the new feature on the time spent online. Let us assume a constant, additive treatment effect for simplicity. In the absence of network connections and in the absence of treatment, X_i is the expected value of $Y_i(0)$ conditional on X_i. So X_i can be thought of as the intrinsic propensity of user i to spend time on the website. The model then captures the fact that the time spent on the website by user i increases with the number of his or her neighbours, with their propensities to spend time on the website, and with the exposure to the new feature if the treatment has an effect. The rest of the paper explores the implications of the normal-sum model for designing optimal treatment allocation strategies.

2·3. *Interpretation of the mean squared error for a fixed treatment allocation vector*

We compute the mean squared error of the difference-in-means estimator according to the normal-sum model for Y^{obs}, defined as $\text{MSE}(\hat{\tau} \mid Z) = E\{(\hat{\tau} - \tau^*)^2 \mid Z\}$ for a fixed treatment allocation vector Z; we call this quantity the conditional mean squared error. We have

$$\text{MSE}(\hat{\tau} \mid Z) = \underbrace{\mu^2\{\delta_{\mathcal{N}}(Z)\}^2}_{\text{bias}^2} + \underbrace{\gamma^2\omega(Z)^{\mathrm{T}}\omega(Z) + \sigma^2\omega(Z)^{\mathrm{T}}A^{\mathrm{T}}A\omega(Z)}_{\text{variance}}.$$

We can identify desirable assignments by evaluating their conditional mean squared error. This idea is the basis for the model-assisted restricted randomization strategies in § 3·2.

In the absence of specific constraints on the number of treated units, different treatment allocation vectors will generally have a different number of treated and untreated units, defined as $N_1 = \sum_i Z_i$ and $N_0 = \sum_i (1 - Z_i)$, respectively, both functions of Z. Then the bias term,

$$\mu\delta_{\mathcal{N}} = \mu\big(N_1^{-1} \textstyle\sum_{\{i:Z_i=1\}} |\mathcal{N}_i| - N_0^{-1} \sum_{\{i:Z_i=0\}} |\mathcal{N}_i|\big), \tag{5}$$

is proportional to the difference in the average neighbourhood sizes of treated and untreated units, and measures a lack of balance between the two groups in terms of the average degree. A larger

value of the mean μ amplifies the contribution of this imbalance to the mean squared error. Since the designer does not control μ, desirable treatment assignments minimize bias by balancing the average neighbourhood size between treated and untreated units. The first variance term is

$$\gamma^2 \omega^{\mathrm{T}} \omega = \gamma^2 \big(N_1^{-1} + N_0^{-1}\big), \tag{6}$$

which is minimized when $N_1 = N_0$. This term penalizes the difference between the number of treated and untreated units. A larger value of γ amplifies the contribution of this imbalance to the mean squared error. This result is consistent with classical results on the optimality of balanced randomization for estimating the average treatment effect in the absence of network-correlated outcomes (e.g., Imbens & Rubin, 2015, Ch. 6). The second variance term involves features of the network; it is

$$\sigma^2 \omega^{\mathrm{T}} A^{\mathrm{T}} A \omega = (\sigma^2/N_1^2) \sum_{\{i,j:\, Z_i = Z_j = 1\}} |\mathcal{N}_i \cap \mathcal{N}_j| \tag{7}$$

$$+ (\sigma^2/N_0^2) \sum_{\{i,j:\, Z_i = Z_j = 0\}} |\mathcal{N}_i \cap \mathcal{N}_j| \tag{8}$$

$$- \{2\sigma^2/(N_1 N_0)\} \sum_{\{i,j:\, Z_i = 1, Z_j = 0\}} |\mathcal{N}_i \cap \mathcal{N}_j|. \tag{9}$$

The term on the right-hand side of (7) is proportional to the average number of shared neighbours among pairs of units both assigned to the treatment group. The term (8) is proportional to the average number of shared neighbours among pairs of units both assigned to the control group. The term (9) is proportional to the average number of shared neighbours among pairs of units where one is assigned to treatment and the other to control. Considering the signs of these three factors, the second variance term may be minimized by assigning units with shared neighbours to different groups, and by avoiding the assignment of entire clusters of units that are densely connected to either treatment or control.

2·4. *Interpretation of the mean squared error averaged over allocation vectors*

Next, we compute the mean squared error of the difference-in-means estimator according to the normal-sum model and the distribution on the allocation vectors implied by a complete randomization strategy, which assigns equal probability to all of the treatment allocation vectors Z for which the numbers of units in treatment and control are fixed at (N_0, N_1). We refer to this quantity, $\mathrm{MSE}(\hat{\tau}) = E[E\{(\hat{\tau} - \tau^*)^2 \mid Z\}]$, as the marginal mean squared error. It is

$$\mathrm{MSE}(\hat{\tau}) = \big(N_1^{-1} + N_0^{-1}\big)\big(\gamma^2 + \sigma^2\big) \tag{10}$$

$$+ \big(N_1^{-1} + N_0^{-1}\big)\Bigg\{\underbrace{\sigma^2(|\bar{\mathcal{N}}| - 1)}_{C_1} \underbrace{- \frac{2\sigma^2}{N(N-1)} \sum_{i<j} |\mathcal{N}_i \cap \mathcal{N}_j|}_{C_2} + \underbrace{\frac{\mu^2 N}{N-1}(|\bar{\mathcal{N}}|^2 - |\bar{\mathcal{N}}|^{\langle 2 \rangle})}_{C_3}\Bigg\}.$$

The right-hand side of (10) is the mean squared error of the difference-in-means estimator due to a complete randomization strategy in the absence of a network, since $(\gamma^2 + \sigma^2)$ is the total variance implied by the network-sum model. The three terms C_1, C_2 and C_3 can be seen as contributions to the variance due to the presence of network-correlated outcomes. The term C_1

is proportional to the average degree of the nodes; thus networks with higher average degrees will tend to yield higher mean squared errors, ceteris paribus. The term C_2 is proportional to the average number of shared neighbours among all pairs of nodes; thus networks that are locally denser will tend to have lower mean squared error, ceteris paribus. The term C_3 is proportional to the variance of observed degrees; thus low variability in the degree of the nodes will lead to lower mean squared error, ceteris paribus. This contribution need not be positive, because of term C_2, which summarizes average local density.

2·5. *More general models of network-correlated potential outcomes*

The normal-sum model introduced in § 2·2 is a special case of a more general model that replaces (3) with the more general formulation

$$Y_i(0) \mid X \sim N\big[g\{(X_j)_{j\in\mathcal{N}_i}\}, \gamma^2\big],$$

with regularity conditions on the function g which essentially ensure that for any subset of nodes $\mathcal{S} \subset \mathcal{N}_i$, the conditional expectation $E[g\{(X_j)_{j\in\mathcal{N}_i}\} \mid (X_j)_{j\in\mathcal{S}}]$ is well behaved. We detail the positivity, symmetry and monotonicity properties as well as the general form of the mean squared error for this model in the Supplementary Material, and we show that the general form of the mean squared error suggests that good designs seek to decrease the number of neighbours shared within treatment groups and increase the number of units shared between treatment groups, while balancing the sizes of the groups and the distribution of neighbourhood sizes. These derivations indicate that the network balance criteria the proposed restricted randomizations are based upon extend well beyond the normal-sum model. Moreover, model-assisted strategies come with theoretical guarantees that hold regardless of the validity of the model, as we show next.

3. Methodology and theory

3·1. *Classical randomization and restricted randomization strategies*

Randomization strategies are probability distributions on the set of binary vectors $\mathcal{Z}$. Restricted randomization strategies are probability distributions implied by discarding allocation vectors $Z \in \mathcal{Z}$ according to a set of rules. According to a Bernoulli randomization strategy with parameter $p \in (0, 1)$, each treatment allocation vector $Z \in \mathcal{Z}$ has individual treatments Z_i drawn as independent Bernoulli random variables with probability of success p. A completely randomized design with parameters (N_0, N_1), where $N_0 + N_1 = N$, considers only treatment allocation vectors $Z \in \mathcal{Z}$ such that $\sum_{i=1}^{N} Z_i = N_1$, and assigns equal probability to them. If $N_0 = N_1 = N/2$, we refer to this as a balanced completely randomized design.

Restricted randomization strategies stem from the observation that when designing an experiment, it is often clear how to evaluate whether a treatment allocation vector is undesirable. For instance, when an allocation vector Z leads to statistical imbalance for one or more key covariates, it leaves the door open to confounding even in the presence of randomization (Gosset, 1938). Indeed, the most common form of restricted randomization is to discard treatment allocations that lead to covariate imbalances (Lock-Morgan & Rubin, 2012).

3·2. *Model-assisted restricted randomization strategies*

We introduce four model-assisted designs, which differ in the degree of reliance on the model. First, we consider balanced restricted randomization strategies, which discard treatment allocation vectors where the number of treated units N_1 differs from the number of untreated units N_0, or

differs by more than 1 when N is odd. This strategy aims at minimizing the contribution of the total variance to the conditional mean squared error, according to (6).

Second, we introduce unbiased restricted randomization strategies, which discard treatment allocation vectors where the average number of neighbours for treated units differs from the average number of neighbours for untreated units. This strategy aims at minimizing the contribution of the bias to the conditional mean squared error, as suggested by the discussion of (5).

Third, we introduce optimal restricted randomization strategies, which favour treatment allocation vectors that minimize the average number of shared neighbours among pairs of treated units, according to (7), minimize the average number of shared neighbours among pairs of untreated units, according to (8), and maximize the average number of shared neighbours among pairs of units one of which is treated and the other untreated, according to (9).

Let $\mathcal{Z} = \{0, 1\}^N$ be the set of all possible treatment allocation vectors on N units. Formally, we can define sets of allocations corresponding to the restricted randomizations defined above:

$$\mathcal{Z}^{\mathrm{b}} = \{Z \in \mathcal{Z} : N_1 - N_0 = 0\}, \tag{11}$$

$$\mathcal{Z}^{\mathrm{u}} = \{Z \in \mathcal{Z} : N_1^{-1} \textstyle\sum_{\{i:Z_i=1\}} |\mathcal{N}_i| - N_0^{-1} \sum_{\{i:Z_i=0\}} |\mathcal{N}_i| = 0\}, \tag{12}$$

$$\mathcal{Z}^{\mathrm{o}} = \{Z \in \mathcal{Z} : \mathrm{MSE}(\hat{\tau} \mid Z) \leqslant q_\alpha^{\mathrm{MSE}}\}, \tag{13}$$

where q_α^{MSE} is the α quantile of the distribution of the conditional mean squared error. These subsets of assignments depend on network statistics that the normal-sum model suggests as relevant for computing the conditional mean squared error, discussed in § 2·3.

The rest of the paper focuses on the first three model-assisted strategies: balanced restricted randomization, which assigns equal probability to all $Z \in \mathcal{Z}^{\mathrm{b}}$; balanced unbiased restricted randomization, which assigns equal probability to all $Z \in \mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{u}}$; and balanced unbiased optimal restricted randomization, which assigns equal probability to all $Z \in \mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{u}} \cap \mathcal{Z}^{\mathrm{o}}$.

The fourth model-assisted strategy, which we refer to as unconstrained optimal restricted randomization, aims to trade off small increases in bias for significant reductions in variance. It assigns equal probability to all $Z \in \mathcal{Z}^{\min}$, defined as

$$\mathcal{Z}^{\min} = \{Z \in \mathcal{Z} : \arg\min \mathrm{MSE}(\hat{\tau} \mid Z)\}. \tag{14}$$

The set $\mathcal{Z}^{\min}$ is usually either a singleton or a set of small cardinality. This makes it challenging to perform randomization-based inference using this design; in particular, the approach proposed in § 3·5 is often unfeasible in practice. This design is largely of theoretical interest.

3·3. *Model-based optimal treatment allocation strategies*

The model-assisted strategies in § 3·2 use a model for the outcomes to select allocations that improve properties of the difference-in-means estimator. The natural next step is to use the model to derive a better estimator for the average treatment effect, replacing for instance the difference-in-means estimator with the maximum likelihood estimator of τ under the normal-sum model. The estimator $\hat{\tau}_{\mathrm{MLE}}$ and its conditional mean squared error are derived in the Supplementary Material. The optimal maximum likelihood design is then the model-based restricted randomization strategy that assigns equal probability to all $Z \in \mathcal{Z}^{\mathrm{MLE}}$, defined as

$$\mathcal{Z}^{\mathrm{MLE}} = \{Z \in \mathcal{Z} : \arg\min \mathrm{MSE}(\hat{\tau}_{\mathrm{MLE}} \mid Z)\}. \tag{15}$$

In the Supplementary Material, we show that the maximum likelihood estimator for τ is not robust with respect to misspecification of the model or the network, unlike the model-assisted

restricted randomization designs. When evaluating the performance of model-based strategies, we fix parameters μ, σ and γ at their true values and treat τ as the only unknown parameter.

3·4. *Restricted randomizations via rerandomization*

A general approach to sampling from arbitrary restricted randomization designs, referred to as rerandomization, has recently been formalized by Lock-Morgan & Rubin (2012). Let ϕ be a binary function such that assignment Z belongs to the restricted randomization set if and only if $\phi(Z) = 1$. A simple way to sample from the restricted randomization design is via rejection sampling: draw an assignment Z from the original design, and then keep the assignment if $\phi(Z) = 1$, or reject it if $\phi(Z) = 0$. In our setting, the restricted sets in (11)–(13) can be defined in terms of different functions ϕ. Denote the indicator function by $I(\cdot)$; then

$$\phi^{\mathrm{b}}(Z) = I\Big\{\textstyle\sum_{i=1}^{N} Z_i = \sum_{i=1}^{N}(1 - Z_i)\Big\},$$
$$\phi^{\mathrm{u}}(Z) = I\big\{\mu\delta_{\mathcal{N}}(Z) = 0\big\},$$
$$\phi^{\mathrm{o}}(Z) = I\Big\{\mathrm{MSE}(\hat{\tau} \mid Z) \leqslant q_{\alpha}^{\mathrm{MSE}}\Big\}.$$

Thus rerandomization can be used to sample from the restricted randomization designs we proposed. It is particularly useful when performing exact tests and computing confidence intervals, as we show next.

3·5. *Inference via inversion of a sequence of exact Fisher tests*

There are traditionally three types of confidence interval in randomization-based inference: Neyman intervals, bootstrap intervals, and Fisher intervals. Neyman intervals are usually obtained using an asymptotic normal approximation to the distribution of the difference-in-means estimator (Imbens & Rubin, 2015, Ch. 6). This approach works well for simple designs for which the asymptotic variance can be estimated, but is challenging with more complicated designs. Li et al. (2017) proposed an asymptotic theory of rerandomization. Unfortunately, the asymptotic regime considered in that paper is not compatible with our setting: it requires the number of covariates to be fixed in the asymptotic regime, whereas in our case the quantities that are analogous to covariates include the number of neighbours shared by each pair of units, which grows with the number of units in the asymptotic regimes of interest; it also requires the constraints to be a function only of the vector of differences in means between treated and control units for the observed covariates, and of the covariance matrix of that vector, which does not hold in our case. Bootstrap intervals are difficult to implement since the correlation structure of the outcomes may be complex.

Instead, we propose using Fisher intervals, which are obtained by inverting a sequence of Fisher exact tests (e.g., Rosenbaum, 2002). This can be accomplished by rerandomization (Lock-Morgan & Rubin, 2012, § 2.2) but with the proposed restricted randomization distributions as the permutation distributions.

We illustrate by simulation the potential gains from Fisher intervals based on restricted randomization. For a fixed network of 500 nodes, we generated 200 realizations of the potential outcomes according to the normal-sum model, as well as 200 observed assignments. For each realization, we computed Fisher confidence intervals based on balanced optimal restricted randomization, balanced unbiased restricted randomization, and balanced complete randomization, with a nominal test size of $\alpha = 5\%$. The intervals based on balanced optimal restricted randomization have a median length of 5·5 with 90% interquantile range [5·1, 5·7]; those based on balanced unbiased restricted randomization have a median length of 5·7 with 90% interquantile

range [5·4, 5·9]. In contrast, the intervals based on balanced complete randomization have median length 6·5 with 90% interquantile range [6·0, 6·6]. The coverage for all three methods is 95%, as expected. These results suggest that restricted randomization inference reduces the length of the intervals while maintaining nominal coverage. More details are given in the Supplementary Material.

3·6. *Theory*

Model-assisted designs have desirable inferential properties even when the model they rely on for evaluating treatment allocations is wrong. We show that the difference-in-means estimator is design-unbiased (Särndal et al., 2003) for the restricted randomization strategies developed in § 3·2.

Definition 1 (Design unbiasedness). *An estimator $\hat{\tau}$ is unbiased with respect to a distribution on $\mathcal{Z}$, typically referred to as a design on $\mathcal{Z}$, if $E_{\mathcal{Z}}(\hat{\tau} - \tau) = 0$.*

The main result is the following.

Theorem 1. *The difference-in-means estimator $\hat{\tau}$ defined in* (1) *is an unbiased estimator of the average treatment effect with respect to the following distributions:*

(i) *the uniform distribution on $\mathcal{Z}^{\mathrm{b}}$, which defines the balanced design;*
(ii) *the uniform distribution on $\mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{u}}$, which defines the balanced unbiased design;*
(iii) *the uniform distribution on $\mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{o}}$, which defines the balanced optimal design;*
(iv) *the uniform distribution on $\mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{u}} \cap \mathcal{Z}^{\mathrm{o}}$, which defines the balanced unbiased optimal design.*

As a consequence of design-unbiasedness and of the increasingly nested supports, we can compare variances of $\hat{\tau}$ implied by the designs in Theorem 1, in expectation.

Corollary 1. *The estimator $\hat{\tau}$ defined in* (1) *satisfies*

$$E\{\mathrm{var}_{\mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{o}}}(\hat{\tau} \mid Y)\} \leqslant E\{\mathrm{var}_{\mathcal{Z}^{\mathrm{b}}}(\hat{\tau} \mid Y)\}.$$

Similar inequalities can be derived for any pair of nested designs in Theorem 1. These results are based on symmetry arguments, which is why $\mathcal{Z}^{\mathrm{b}}$ is always part of the support of designs that make the difference-in-means estimator unbiased. This notion of symmetry is made precise in the following lemma.

Lemma 1. *For Z in $\mathcal{Z}^{\mathrm{b}}$, $\hat{\tau}(1 - Z) = 2\tau - \hat{\tau}(Z)$.*

As a consequence, if we required the unconstrained optimal design to be balanced by restricting its support to $\mathcal{Z}^{\mathrm{b}} \cap \mathcal{Z}^{\mathrm{min}}$, we would recover design-unbiasedness for the difference-in-means estimator. However, we do not consider balanced unconstrained optimal designs.

4. Discussion

The idea behind model-assisted design is fairly general, two key elements being the estimator and the model. The theoretical guarantees in § 3·6 are limited to estimators satisfying the symmetry condition of Lemma 1, and to the model family introduced in § 2·5. Extending the theory to a larger class of estimators and models is conceptually feasible, although it would often lead to complex expressions for the mean squared error and hard-to-interpret balance criteria.

The designs presented in (14) and (15) can be seen as extreme versions of the model-assisted approach, perhaps closer in spirit to the optimal model-based design literature (Kiefer, 1959). However, the mean squared errors minimized in the model-based and model-assisted design of experiments are associated with different estimators. Randomization inference based on the restricted distributions that the designs in (14) and (15) imply is often impractical, since the sets $\mathcal{Z}^{\text{MLE}}$ and $\mathcal{Z}^{\min}$ are generally too small. Moreover, even when feasible, inference based on these designs relies heavily on model-induced constraints, by requiring stringent balance of terms appearing in the conditional mean squared error, and is in general not robust with respect to model misspecification.

In practice, there are often additional issues to consider, which we have ignored for simplicity of exposition. Covariates will have to be taken into account, and the parameters μ, σ^2 and γ will need to be specified or estimated. One option is to specify point priors (Box & Lucas, 1959); another option is to specify full priors and work with the integrated mean squared error. In both situations, historical data and pilot studies could be used to calibrate these priors, and are recommended for optimal design in practice (Kim et al., 2015). Our theory for a more general model of network-correlated outcomes, as well as simulation studies, both detailed in the Supplementary Material, show that the efficiency gains one can expect to achieve with model-assisted design of experiments are robust with respect to misspecification.

This paper has introduced model-assisted design of network experiments and illustrated its use in a simple setting. Although the model in § 2·5 is fairly general, it focuses on estimating the average treatment effect in the presence of network-correlated outcomes and homophily, ignoring other phenomena of interest, including network interference, peer influence and contagion. Developing model-assisted design strategies for estimating other causal effects in more complicated settings, such as the presence of network interference and confounding due to homophily, is one of the directions we are currently pursuing.

Acknowledgement

The authors thank Iavor Bojinov for helpful comments. We are grateful to the editor, associate editor, and two reviewers for constructive feedback. This work was partially supported by the U.S. National Science Foundation, the Office of Naval Research, a Shutzer Fellowship and a Sloan Research Fellowship. The work was performed while the authors were at Harvard University.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of all the theoretical results, simulations illustrating the robustness of the proposed strategies with respect to misspecification, and additional technical details.

References

Aral, S., Muchnik, L. & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Nat. Acad. Sci.* **106**, 21544–9.

Aronow, P. M. & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Statist.* **11**, 1912–47.

Bailey, R. A. (1983). Restricted randomization. *Biometrika* **70**, 183–98.

Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on Twitter. In *Proc. 4th ACM Int. Conf. Web Search and Data Mining*. New York: Association for Computing Machinery, pp. 65–74.

Bakshy, E., Rosenn, I., Marlow, C. & Adamic, L. (2012). The role of social networks in information diffusion. In *Proc. 21st Int. Conf. World Wide Web*. New York: Association for Computing Machinery, pp. 519–28.
Bickel, P. J. & Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Nat. Acad. Sci.* **106**, 21068–73.
Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E. & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–8.
Box, G. E. P. & Lucas, H. (1959). Design of experiments in non-linear situations. *Biometrika* **46**, 77–90.
Cavusoglu, H., Phan, T. Q., Cavusoglu, H. & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Info. Syst. Res.* **27**, 848–79.
Christakis, N. A. & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New Engl. J. Med.* **357**, 370–9.
Eckles, D., Karrer, B. & Ugander, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *J. Causal Infer.* **5**, DOI: 10.1515/jci-2015-0021.
Goldenberg, A., Zheng, A. X., Fienberg, S. E. & Airoldi, E. M. (2010). A survey of statistical network models. *Foundat. Trends Mach. Learn.* **2**, 129–233.
Gosset, W. J. (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29**, 363–78.
Gui, H., Xu, Y., Bhasin, A. & Han, J. (2015). Network A/B testing: From sampling to estimation. In *Proc. 24th Int. Conf. World Wide Web*. New York: Association for Computing Machinery, 399–409.
Higham, R., Tharmanathan, P. & Birks, Y. (2015). Use and reporting of restricted randomization: A review. *J. Eval. Clin. Practice* **21**, 1205–11.
Holland, P. W. (1986). Statistics and causal inference. *J. Am. Statist. Assoc.* **81**, 945–60.
Hudgens, M. G. & Halloran, M. E. (2008). Toward causal inference with interference. *J. Am. Statist. Assoc.* **103**, 832–42.
Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.
Kiefer, J. (1959). Optimum experimental designs. *J. R. Statist. Soc.* B **21**, 272–319.
Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H. & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *Lancet* **386**, 145–53.
Kolaczyk, E. D. & Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. New York: Springer.
Li, X., Ding, P. & Rubin, D. B. (2017). Asymptotic theory of rerandomization in treatment-control experiments. *arXiv:* 1604.00698v4.
Lock-Morgan, K. & Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40**, 1263–82.
Manski, C. F. (2013). Identification of treatment response with social interactions. *Economet. J.* **16**, S1–23.
McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Ann. Rev. Sociol.* **27**, 415–44.
Ogburn, E. L. & VanderWeele, T. J. (2017). Vaccines, contagion, and social networks. *Ann. Appl. Statist.* **11**, 919–48.
Phan, T. Q. & Airoldi, E. M. (2015). A natural experiment of social network formation and dynamics. *Proc. Nat. Acad. Sci.* **112**, 6595–600.
Rosenbaum, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statist. Sci.* **17**, 286–327.
Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *J. Am. Statist. Assoc.* **102**, 191–200.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688–701.
Särndal, C.-E., Swensson, B. & Wretman, J. (2003). *Model Assisted Survey Sampling*. New York: Springer.
Shalizi, C. R. & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociol. Meth. Res.* **40**, 211–39.
Simon, R. (1979). Restricted randomization designs in clinical trials. *Biometrics* **35**, 503–12.
Toulis, P. & Kao, E. (2013). Estimation of causal peer influence effects. *J. Mach. Learn. Res.* **28**, 1489–97.
Ugander, J., Karrer, B., Backstrom, L. & Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proc. 19th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, pp. 329–37.
Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
Yates, F. (1948). Contribution to the discussion of "The validity of comparative experiments" by F. J. Anscombe. *J. R. Statist. Soc.* A **111**, 204–5.
Youden, W. J. (1972). Randomization and experimentation. *Technometrics* **14**, 13–22.

[*Received on* 2 *July* 2015. *Editorial decision on* 20 *August* 2017]

Biometrika (2017), **xx**, x, *pp*. 1–25

Printed in Great Britain

Supplementary material for 'Model-assisted design of experiments in the presence of network correlated outcomes'

BY G. W. BASSE AND E. M. AIROLDI

Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

gbasse@fas.harvard.edu airoldi@fas.harvard.edu

1. MORE GENERAL MODELS OF NETWORK-CORRELATED OUTCOMES

1·1. *Elements of experimental design for more general models*

Although the exact network balance criteria will depend on the model, some broad experimental design guidelines are available for a relatively large class of models. Consider the family,

$$\begin{aligned} Y_i(0) \mid X &\sim \mathcal{N}\Big[g\{(X_j)_{j\in\mathcal{N}_i}\}, \gamma^2\Big], \\ X_i &\sim \mathcal{N}(\mu, \sigma^2), \\ Y_i(1) &= Y_i(0) + \tau, \end{aligned}$$

where g satisfies the following regularity conditions. There exists a real-valued set function ϕ and a real-valued function of three variables $h(\cdot,\cdot,\cdot)$ such that for any collection $\{X_k\}_{k\in\chi}$ indexed by a finite set χ and any subset of indices $\mathcal{S}\subset\chi$, the following hold: (i) $E[g\{(X_k)_{k\in\chi}\} \mid \{X_k\}_{k\in\mathcal{S}}] = h[|\chi|, |\mathcal{S}|, \phi\{(X_k)_{k\in\mathcal{S}}\}]$; (ii) $h(n,s,\cdot)$ is a monotone function of its third argument, for n and s fixed. That is, $h(n,s,\cdot)$ is either non-increasing for every n and s, or non-decreasing for every n and s; and (iii) If $s_0 \in \chi\cap\bar{\mathcal{S}}$, the quantity $\phi[\{X_{s_0}\cup(X_k)_{k\in\mathcal{S}}\}]$ is a non-decreasing function of X_{s_0}, for $\{X_k\}_{k\in\mathcal{S}}$ fixed.

We now state our main new theorem. We will give examples of functions g satisfying the constraints immediately after:

THEOREM 1. *If the regularity conditions above hold, then*

$$\text{MSE}(\hat{\tau} \mid Z) = \{\delta(Z)\}^2 + \gamma^2\omega(Z)^T\omega(Z) + \omega(Z)^T\Sigma\,\omega(Z),$$

where

$$\delta(Z) = N_1^{-1}\sum_{i:Z_i=1} q(|\mathcal{N}_i|) - N_0^{-1}\sum_{i:Z_i=0} q(|\mathcal{N}_i|), \qquad \gamma^2\omega(Z)^T\omega(Z) = \gamma^2(N_1^{-1}+N_0^{-1}),$$

and

$$\begin{aligned}\omega(Z)^T\Sigma\,\omega(Z) = (1/N_1^2)&\sum_{i,j:Z_i=Z_j=1} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &+ (1/N_0^2)\sum_{i:j:Z_i=Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &- \{2/(N_1N_0)\}\sum_{i,j:Z_i=1,Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &+ (1/N_1^2)\sum_{i:Z_i=1} v(|\mathcal{N}_i|) + (1/N_0^2)\sum_{i:Z_i=0} v(|\mathcal{N}_i|).\end{aligned}$$

The functions q, v and m all depend on g, μ, and σ^2 and are defined in Lemma 2, below. The function m has the following properties:

(1) $m(a,b,c) \geq 0$ *for all* $a,b \geq 0$ *and* $c \leq min(a,b)$.

(2) m *is symmetric in its first two arguments:* $m(a,b,c) = m(b,a,c)$

(3) m *is a non-decreasing function of its third argument, when the first two arguments are held constant.*

The functions q and h in the theorem depend on the specific choice of g, and on the parameters. However when looking at the variance one can see that everything else being equal, the mean square error is minimized by decreasing within-group overlap, and by increasing between-group overlap, precisely as in the basic normal-sum model. A general heuristic for experimental design would then be: minimize shared neighbors within treatment and control groups, maximize shared neighbors between these groups, while keeping the distribution of neighborhood sizes similar in both treatment and control groups.

1·2. *Example models*

We now give examples of functions g satisfying the three regularity conditions stated above.

Example 1 *(normal-sum model).* Consider $g\{(X_j)_{j\in\mathcal{N}_i}\} = \sum_{j\in\mathcal{N}_i} X_j$. Let $\mathcal{S}\subset\mathcal{N}_i$. Then

$$E[g\{(X_j)_{j\in\mathcal{N}_i}\} \mid \{X_k\}_{k\in\mathcal{S}}] = (|\mathcal{N}_i| - |\mathcal{S}|)\mu + \sum_{k\in\mathcal{S}} X_k = h[|\mathcal{N}_i|,|\mathcal{S}|,\phi\{(X_k)_{k\in\mathcal{S}}\}],$$

with $\phi\{(X_k)_{k\in\mathcal{S}}\} = \sum_{k\in\mathcal{S}} X_k$. It is clear that $h(n,s,x)$ is an increasing function of x for fixed n and s, and that $\phi\{(X_k)_{k\in\mathcal{S}}\}$ is a non-decreasing function of any element of the set, after fixing the others. So $g\{(X_j)_{j\in\mathcal{N}_i}\} = \sum_{j\in\mathcal{N}_i} X_i$ satisfies all three regularity conditions.

Example 2 *(normal-mean model).* Consider $g\{(X_j)_{j\in\mathcal{N}_i}\} = |\mathcal{N}_i|^{-1}\sum_{j\in\mathcal{N}_i} X_j$. Let $\mathcal{S}\subset\mathcal{N}_i$. Then

$$\begin{aligned}E[g\{(X_j)_{j\in\mathcal{N}_i}\} \mid \{X_k\}_{k\in\mathcal{S}}] &= |\mathcal{N}_i|^{-1}(|\mathcal{N}_i| - |\mathcal{S}|)\mu + |\mathcal{N}_i|^{-1}\sum_{k\in\mathcal{S}} X_k\\ &= h[|\mathcal{N}_i|,|\mathcal{S}|,\phi\{(X_k)_{k\in\mathcal{S}}\}],\end{aligned}$$

with $\phi\{(X_k)_{k\in\mathcal{S}}\} = \sum_{k\in\mathcal{S}} X_k$. Again, $h(n,s,x)$ is an increasing function of x for fixed n and s. The other regularity conditions can be verified since the function ϕ is the same as in the previous example. Additional details on the normal-mean model are provided in Section 4·4 of this Supplement.

Example 3 *(normal-threshold model).* Consider $g\{(X_j)_{j\in\mathcal{N}_i}\} = I(\sum_{j\in\mathcal{N}_i} X_j > c)$, where $I(\cdot)$ is the indicator function. Let $\mathcal{S} \subset \mathcal{N}_i$. With regards to the first regularity condition we have:

$$\begin{aligned} E[g\{(X_j)_{j\in\mathcal{N}_i}\} \mid \{X_k\}_{k\in\mathcal{S}}] &= P\Big(\sum_{j\in\mathcal{N}\setminus\mathcal{S}} X_j > c - \sum_{k\in\mathcal{S}} X_k \mid \{X_k\}_{k\in\mathcal{S}}\Big) \\ &= 1 - \Phi\Big\{\frac{(c - \sum_{j\in\mathcal{S}} X_j) - (|\mathcal{N}_i| - |\mathcal{S}|)\mu}{(|\mathcal{N}_i| - |\mathcal{S}|)\sigma^2}\Big\} \\ &= h[|\mathcal{N}_i|, |\mathcal{S}|, \phi\{(X_k)_{k\in\mathcal{S}}\}], \end{aligned}$$

with $\phi\{(X_k)_{k\in\mathcal{S}}\} = \sum_{k\in\mathcal{S}} X_k$. The first regularity follows from the properties of the normal cumulative density function Φ. The other two regularity conditions follow once again from the function ϕ being the same as in the previous examples.

Example 4 *(normal-max model).* Consider $g\{(X_j)_{j\in\mathcal{N}_i}\} = \max\{(X_j)_{j\in\mathcal{N}_i}\}$. As before, let $\mathcal{S} \subset \mathcal{N}_i$. Also, let $\bar{\mathcal{S}} = \mathcal{N}_i - \mathcal{S}$.

$$\begin{aligned} E[g\{(X_j)_{\mathcal{N}_i}\} \mid \{X_k\}_{k\in\mathcal{S}}] &= E\Big(I\Big[\max\{(X_k)_{k\in\bar{\mathcal{S}}}\} > \max\{(X_k)_{k\in\mathcal{S}}\}\Big] \max\{(X_k)_{k\in\bar{\mathcal{S}}}\} \\ &\quad + I\Big[\max\{(X_k)_{k\in\bar{\mathcal{S}}}\} < \max\{(X_k)_{k\in\mathcal{S}}\}\Big] \max\{(X_k)_{k\in\mathcal{S}}\} \mid \{X_k\}_{k\in\mathcal{S}}\Big) \\ &= P\Big[\max\{(X_k)_{k\in\bar{\mathcal{S}}}\} > \max\{(X_k)_{k\in\mathcal{S}}\} \mid (X_k)_{k\in\mathcal{S}}\Big] \\ &\quad \times E\Big[\max\{(X_k)_{k\in\bar{\mathcal{S}}}\} \mid \max\{(X_k)_{k\in\bar{\mathcal{S}}}\} > \max\{(X_k)_{k\in\mathcal{S}}\}\Big] \\ &\quad + P\Big[\max\{(X_k)_{k\in\bar{\mathcal{S}}}\} < \max\{(X_k)_{k\in\mathcal{S}}\} \mid \{X_k\}_{k\in\mathcal{S}}\Big] \max\{(X_k)_{k\in\mathcal{S}}\} \\ &= h[|\mathcal{N}_i|, |\mathcal{S}|, \phi\{(X_k)_{k\in\mathcal{S}}\}] \end{aligned}$$

where $\phi\{(X_k)_{\mathcal{S}}\} = \max\{(X_k)_{\mathcal{S}}\}$. We have used the fact that the distribution of $\max\{(X_k)_{k\in\bar{\mathcal{S}}}\}$ depends on $\mathcal{S}$ and $\mathcal{N}_i$ only through $|\mathcal{N}_i| - |\mathcal{S}|$, since the X_j's are iid. This means that $P[\max\{(X_k)_{\bar{\mathcal{S}}}\} > \max\{(X_k)_{\mathcal{S}}\} \mid \{X_k\}_{\mathcal{S}}]$ is a function of $|\mathcal{N}_i|$ and $|\mathcal{S}|$ and $\max\{(X_k)_{\mathcal{S}}\}$. It is easy to verify the last two regularity conditions are also satisfied.

1·3. *Proofs*

Where there is no confusion possible we will use the shortcut $g(\mathcal{A})$ to denote $g\{(X_k)_{k\in\mathcal{A}}\}$ and similarly, we will use $\phi(\mathcal{A})$ in place of $\phi\{(X_k)_{k\in\mathcal{A}}\}$. We will write $E(\cdot \mid \mathcal{A})$ as a shortcut for $E\{\cdot \mid (X_k)_{k\in\mathcal{A}}\}$. This will lighten the notation considerably.

LEMMA 1. *Let $\{X_k\}_{k\in\chi}$ a set of independent and identically distributed random variables, were χ is any set of integers. If the regularity conditions hold, then there exists a function $m(\cdot,\cdot,\cdot)$ such that for any $\chi_1 \subset \chi$ and $\chi_2 \subset \chi$ two sets of indices such that $\chi_1 \neq \chi_2$, and such that $\mathcal{S} = \chi_1 \cap \chi_2 \neq \emptyset$, we have the following:*

$$cov\Big\{g(\chi_1), g(\chi_2)\Big\} = m(|\chi_1|, |\chi_2|, |\mathcal{S}|),$$

and the function m satisfies:

1. *(positivity)* $m(a,b,c) \geq 0$ *for all* $a, b \geq 0$ *and* $c \leq min(a,b)$
2. *(partial symmetry)* $m(a,b,c) = m(b,a,c)$
3. *(monotonicity) For all* a *and* b*, the function* $m : c \to m(a,b,c)$ *is non-decreasing*

Proof. Let $\{X_k\}_{k\in\chi}$ a set of independent and identically distributed random variables, were χ is any set of integers. Let $\chi_1 \subset \chi$ and $\chi_2 \subset \chi$ such that $\chi_1 \neq \chi_2$ and $S = \chi_1 \cap \chi_2 \neq \emptyset$. Then

$$\begin{aligned} cov\Big\{g(\chi_1), g(\chi_2)\Big\} &= E\Big[cov\{g(\chi_1), g(\chi_2)\} \mid \mathcal{S}\Big] \\ &\quad + cov\Big[E\{g(\chi_1) \mid \mathcal{S}\}, E\{g(\chi_2) \mid \mathcal{S}\}\Big] \\ &= 0 + cov\Big[E\{g(\chi_1) \mid \mathcal{S}\}, E\{g(\chi_2) \mid \mathcal{S}\}\Big] \\ &= cov\Big[h\{|\chi_1|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2|, |\mathcal{S}|, \phi(\mathcal{S})\}\Big], \end{aligned}$$

where the last equality uses the first regularity condition. Since the X's are independent and identically distributed, the covariance depends on $\{X_k\}_{k\in\mathcal{S}}$ only through $|\mathcal{S}|$, the number of random variables in the set, and so we can write

$$\begin{aligned} cov\Big\{g(\chi_1), g(\chi_2)\Big\} &= cov\Big[h\{|\chi_1|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2|, |\mathcal{S}|, \phi(\mathcal{S})\}\Big] \\ &= m(|\chi_1|, |\chi_2|, |\mathcal{S}|), \end{aligned}$$

where we emphasize once again that implicitly, m depends on g and on the model. We now study the properties of m:

Partial Symmetry: Since the covariance is symmetric, it is clear that:

$$m(|\chi_1|, |\chi_2|, |\mathcal{S}|) = m(|\chi_2|, |\chi_1|, |\mathcal{S}|)$$

Positivity: By the second regularity condition, $h(n, s, \cdot)$ is either a non-decreasing function of its third argument for all n, or a non-increasing function of its third argument for all n and s. But it is known (see e.g, Thorisson (1995), section 2) that the covariance of two monotone functions of random variables is positive. Thus

$$cov\Big[h\{|\chi_1|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2|, |\mathcal{S}|, \phi(\mathcal{S})\}\Big] \geq 0,$$

and we have:

$$m(|\chi_1|, |\chi_2|, |\mathcal{S}|) \geq 0.$$

Monotonicity: Let χ_1' and χ_2' such that $|\chi_1'| = |\chi_1|$ and $|\chi_2'| = |\chi_2|$ and $\chi_1' \cap \chi_2' = \{s_0\} \cup \mathcal{S} \equiv \mathcal{S}'$. On the one hand

$$cov\{g(\chi_1'), g(\chi_2')\} = cov[h\{|\chi_1'|, |\mathcal{S}'|, \phi(\mathcal{S}')\}, h\{|\chi_2'|, |\mathcal{S}'|, \phi(\mathcal{S}')\}], \tag{1}$$

and on the other hand

$$\begin{aligned} cov\{g(\chi_1'), g(\chi_2')\} &= E[cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}\}] \\ &\quad + cov[E\{g(\chi_1') \mid \mathcal{S}\}, E\{g(\chi_2') \mid \mathcal{S}\}] \\ &= E[cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}\}] \\ &\quad + cov[h\{|\chi_1'|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2'|, |\mathcal{S}|, \phi(\mathcal{S})\}]. \end{aligned}$$

But

$$\begin{aligned} cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}\} &= E\Big[Cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}'\} \mid \mathcal{S}\Big] \\ &\quad + cov\Big[E\{g(\chi_1') \mid \mathcal{S}'\}, E\{g(\chi_2') \mid \mathcal{S}'\} \mid \mathcal{S}\Big] \\ &= cov\Big[h\{|\chi_1'|, |\mathcal{S}'|, \phi(\mathcal{S}')\}, h\{|\chi_2'|, |\mathcal{S}'|, \phi(\mathcal{S}')\} \mid \mathcal{S}\Big]. \end{aligned}$$

The only random element in the covariance of the last line is X_{s_0}, since we condition on all the other random variables. But by the third regularity condition, the function

$$X_{s_0} \to \phi\{(X_{s_0}) \cup (X_k)_{k\in\mathcal{S}}\}$$

for fixed $\{X_k\}_{k\in\mathcal{S}}$ is non-decreasing. Thus as above the covariance will be positive (Thorisson (1995)). That is

$$cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}\} \geq 0,$$

and so we also have:

$$E\Big[cov\{g(\chi_1'), g(\chi_2') \mid \mathcal{S}\}\Big] \geq 0.$$

Putting it all together,

$$cov\{g(\chi_1'), g(\chi_2')\} \geq cov[h\{|\chi_1'|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2'|, |\mathcal{S}|, \phi(\mathcal{S})\}]$$

and combining with Equation 1, the following holds

$$cov[h\{|\chi_1'|, |\mathcal{S}'|, \phi(\mathcal{S}')\}, h\{|\chi_2'|, |\mathcal{S}'|, \phi(\mathcal{S}')\}] \geq cov[h\{|\chi_1'|, |\mathcal{S}|, \phi(\mathcal{S})\}, h\{|\chi_2'|, |\mathcal{S}|, \phi(\mathcal{S})\}].$$

That is,

$$m(|\chi_1|, |\chi_2|, |\mathcal{S}| + 1) \geq m(|\chi_1|, |\chi_2|, |\mathcal{S}|),$$

and so by induction, m is a non-decreasing function of its third argument. □

LEMMA 2. *If g satisfies the regularity conditions, we have:*

$$\begin{aligned} E\{Y_i(Z_i) \mid Z\} &= \tau Z_i + q(|\mathcal{N}_i|), \\ var\{Y_i(Z_i) \mid Z\} &= \gamma^2 + v(|\mathcal{N}_i|), \\ cov\{Y_i(Z_i), Y_j(Z_j) \mid Z\} &= m(|\mathcal{N}_i|, |\mathcal{N}_j|, |\mathcal{N}_i \cap \mathcal{N}_j|) \quad i \neq j, \end{aligned}$$

where:

$$q(|\mathcal{N}_i|) = E\{g(\mathcal{N}_i) \mid Z\},$$
$$v(|\mathcal{N}_i|) = var\{g(\mathcal{N}_i) \mid |Z\},$$

and

$$m(|\mathcal{N}_i|, |\mathcal{N}_j|, |\mathcal{N}_i \cap \mathcal{N}_j|) = cov\Big[h\{|\mathcal{N}_i|, |\mathcal{N}_i \cap \mathcal{N}_j|, \phi(\mathcal{N}_i \cap \mathcal{N}_j)\}, h\{|\mathcal{N}_j|, |\mathcal{N}_i \cap \mathcal{N}_j|, \phi(\mathcal{N}_i \cap \mathcal{N}_j)\} \mid Z\Big],$$

if $|\mathcal{N}_i \cap \mathcal{N}_j| \neq 0$ *and* $m(|\mathcal{N}_i|, |\mathcal{N}_j|, 0) \equiv 0$. *The functions* q, v *and* m *depend implicitly on* g *and on the model for the* X*'s. Moreover,* m *satisfies the properties of positivity, partial symmetry and monotonicity defined in Lemma* 1.

Proof. We proceed in order:

Expectation:

$$\begin{aligned} E\{Y_i(Z_i) \mid Z\} &= E\{Y_i(0) + Z_i\tau \mid Z\} \\ &= E[E\{Y_i(0) \mid X, Z\} \mid Z] + \tau Z_i \\ &= E\{g(\mathcal{N}_i) \mid Z\} + \tau Z_i. \end{aligned}$$

Variance:

$$\begin{aligned} var\{Y_i(Z_i) \mid Z\} &= var\{Z_i\tau + Y_i(0) \mid Z\} \\ &= E[var\{Y_i(0) \mid X, Z\} \mid Z] + var[E\{Y_i(0) \mid X, Z\} \mid Z] \\ &= E(\gamma^2 \mid Z) + var\{g(\mathcal{N}_i) \mid Z\} \\ &= \gamma^2 + var\{g(\mathcal{N}_i) \mid Z\}. \end{aligned}$$

Covariance:

$$\begin{aligned} cov\{Y_i(Z_i), Y_j(Z_j) \mid Z\} &= cov\{Z_i\tau + Y_i(0), Z_j\tau + Y_j(0) \mid Z\} \\ &= E[cov\{Y_i(0), Y_j(0) \mid X, Z\} \mid Z] \\ &\quad + cov[E\{Y_i(0) \mid X, Z\}, E\{Y_j(0) \mid X, Z\} \mid Z] \\ &= cov\{g(\mathcal{N}_i), g(\mathcal{N}_j) \mid Z\}. \end{aligned}$$

We now apply Lemma 1 to the right-hand side of the last equality, with $\chi_1 = \mathcal{N}_i$, $\chi_2 = \mathcal{N}_j$ and $\mathcal{S} = \mathcal{N}_i \cap \mathcal{N}_j$, and we immediately obtain

$$cov\{g(\mathcal{N}_i), g(\mathcal{N}_j) \mid Z\} = \begin{cases} m(|\mathcal{N}_i|, |\mathcal{N}_j|, |\mathcal{N}_i \cap \mathcal{N}_j|) & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

The properties of m are also obtained from Lemma 1. □

Proof of Theorem 1 (of the Supplementary Material). Recall that

$$\hat{\tau} = N_1^{-1} \sum_{i:Z_i=1} Y_i(1) - N_0^{-1} \sum_{i:Z_i=0} Y_i(0).$$

We have

$$\begin{aligned}\text{bias}(\hat{\tau}, \tau \mid Z) &= E(\hat{\tau} \mid Z) - \tau \\ &= N_1^{-1} \sum_{i:Z_i=1} E\{Y_i(1) \mid Z\} - N_0^{-1} \sum_{i:Z_i=0} E\{Y_i(0) \mid Z\} - \tau \\ &= N_1^{-1} \sum_{i:Z_i=1} \{\tau + q(|\mathcal{N}_i|)\} - N_0^{-1} \sum_{i:Z_i=0} q(|\mathcal{N}_i|) - \tau \\ &= N_1^{-1} \sum_{i:Z_i=1} q(|\mathcal{N}_i|) - N_0^{-1} \sum_{i:Z_i=0} q(|\mathcal{N}_i|),\end{aligned}$$

where we have used Lemma 2. We now turn to the variance. Let $\omega(Z_i) = Z_i/N_1 - (1 - Z_i)/N_0$. We have

$$\begin{aligned}var(\hat{\tau} \mid Z) &= var\Big\{ \sum_i \omega(Z_i) Y_i(0) + \tau \mid Z \Big\} \\ &= \sum_{i=1}^{N} \omega(Z_i)^2 var\{Y_i(0) \mid Z\} + \sum_{i \neq j}^{N} \omega(Z_i)\omega(Z_j) cov\{Y_i(0), Y_j(0) \mid Z\},\end{aligned}$$

with

$$\omega(Z_i)^2 = \begin{cases} 1/N_1^2 & \text{if } Z_i = 1, \\ 1/N_0^2 & \text{if } Z_i = 0. \end{cases}$$

Applying Lemma 2 to the first term of the right-hand side of the last line gives

$$\begin{aligned}\sum_{i=1}^{N} \omega(Z_i)^2 \cdot var\{Y_i(0) \mid Z\} &= \sum_{i=1}^{N} \omega(Z_i)^2 \{\gamma^2 + v(|\mathcal{N}_i|)\} \\ &= \sum_{i:Z_i=1} \omega(1)^2 \gamma^2 + \sum_{i:Z_i=0} \omega(0)^2 \gamma^2 + \sum_{i:Z_i=1} \omega(1)^2 v(|\mathcal{N}_i|) + \sum_{i:Z_i=0} \omega(0)^2 v(|\mathcal{N}_i|) \\ &= \gamma^2 (N_1^{-1} + N_0^{-1}) + (1/N_1^2) \sum_{i:Z_i=1} v(|\mathcal{N}_i|) + (1/N_0^2) \sum_{i:Z_i=0} v(|\mathcal{N}_i|).\end{aligned}$$

Now notice that

$$\omega(Z_i)\omega(Z_j) = \begin{cases} 1/N_1^2 & \text{if } Z_i = Z_j = 1, \\ 1/N_0^2 & \text{if } Z_i = Z_j = 0, \\ -1/(N_1 N_0) & \text{otherwise,} \end{cases}$$

so applying Lemma 2 to the second term of the right-hand side of the last line gives

$$\begin{aligned}\sum_{i\neq j}^{N}\omega(Z_i)\omega(Z_j)cov\{Y_i(0),Y_j(0)\mid Z\} &= \sum_{i\neq j}^{N}\omega(Z_i)\cdot\omega(Z_j)\cdot m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &= (1/N_1^2)\sum_{i\neq j:Z_i=Z_j=1} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &\quad+(1/N_0^2)\sum_{i\neq j:Z_i=Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &\quad-\{2/(N_1N_0)\}\sum_{i\neq j:Z_i=1,Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|),\end{aligned}$$

where the $-2/(N_1N_0)$ term is obtained by symmetry of m with respect to its first two argument.

Now using the fact that $\text{MSE}(\hat{\tau},\tau\mid Z)=\text{bias}(\hat{\tau},\tau\mid Z)^2+var(\hat{\tau}\mid Z)$,

$$\begin{aligned}\text{MSE}(\hat{\tau},\tau|Z) &= \left\{N_1^{-1}\sum_{i:Z_i=1}q(|\mathcal{N}_i|)-N_0^{-1}\sum_{i:Z_i=0}q(|\mathcal{N}_i|)\right\}^2\\ &\quad+\gamma^2(N_1^{-1}+N_0^{-1})\\ &\quad+(1/N_1^2)\sum_{i\neq j:Z_i=Z_j=1} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &\quad+(1/N_0^2)\sum_{i\neq j:Z_i=Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &\quad-\{2/(N_1N_0)\}\sum_{i\neq j:Z_i=1,Z_j=0} m(|\mathcal{N}_i|,|\mathcal{N}_j|,|\mathcal{N}_i\cap\mathcal{N}_j|)\\ &\quad+(1/N_1^2)\sum_{i:Z_i=1}v(|\mathcal{N}_i|)+(1/N_0^2)\sum_{i:Z_i=0}v(|\mathcal{N}_i|),\end{aligned}$$

which completes the proof. □

2. FISHER INTERVALS BASED ON RESTRICTED RANDOMIZATION

2·1. *Inferential procedure*

This section simply combines together classic results on inverting Fisher tests (Rosenbaum et al. 2002) and on exact tests with restricted randomization (Lock Morgan & Rubin 2012). The inferential procedure assumes that we have imposed balance, so $Z\in\mathcal{Z}^b$, which is the case considered in the article. We begin by describing how to obtain a p-value for the following type of sharp null hypothesis,

$$H_{\tau^*}: Y_i(1)=Y_i(0)+\tau^* \quad \forall\, i,$$

then we describe how to invert a sequence of such tests to obtain a confidence interval. Let $Z\sim\mathcal{R}$ be any of the restricted randomization schemes we proposed that imposes exact balance on the size of the treatment groups. Define

$$T(Z,a,b)=\frac{2}{N}\left\{\sum_{i=1}^{N}Z_ia_i-\sum_{i=0}^{N}(1-Z_i)b_i\right\}.$$

A p-value for H_{τ^*} is obtained as follows:

1. let $T^{obs} = \hat{\tau}(Z^{obs}, Y^{obs}, Y^{obs})$.
2. Let $Y^*(1) = Y^{obs} + (1 - Z^{obs})\tau^*$ and $Y^*(0) = Y^{obs} - Z^{obs}\tau^*$.
3. For $k = 1, \ldots, M$, let $Z^{(k)} \sim \mathcal{R}$, compute $T_k = T\{Z_k, Y^*(0), Y^*(1)\}$.
4. Compute the two-sided p-value $p_{\tau^*}^{(M)} = M^{-1} \sum_{k=1}^{M} I(|T_k| > |T^{obs}|)$.

The p-value $p_{\tau^*}^{(M)}$ thus obtained is a monte-carlo approximation of the true p-value p_{τ^*} for the test H_{τ^*}. It can be made arbitrarily precise by increasing M. The difference with the traditional Fisher test here occurs in Step 3, when we sample $Z \sim \mathcal{R}$, where $\mathcal{R}$ is a restricted randomization design. This is usually performed using rerandomization as described in the main article. Confidence intervals are then obtained by inverting a sequence of such Fisher exact tests. Specifically, let $\tau^{min} < \tau^{max}$ be such that $p_{\tau^{min}}^{M} < \alpha$, $p_{\tau^{max}}^{M} < \alpha$, and such that there exists $\tau^{min} < \tau^* < \tau^{max}$ such that $p_{\tau}^{M} > \alpha$. These can always be found, by construction. Let $\delta_K = (\tau^{max} - \tau^{min})/K$. Then a $100(1 - \alpha)$ interval for τ^* can be constructed as follows:

1. For $k = 1, \ldots, K$, compute $p_{\tau^{min}+k\delta_K}^{M}$.
2. Define $k^{(low)} = \min\{k : p_{\tau^{min}+k\delta_K}^{M} > \alpha\}$ and $k^{(high)} = \max\{k : p_{\tau^{min}+k\delta_K}^{M} > \alpha\}$.
3. The interval $[\tau^{min} + k^{(low)} \cdot \delta_K, \tau^{min} + k^{(high)} \cdot \delta_K]$ is a $100 \times (1 - \alpha)$ interval for τ^*.

2·2. *Simulation study*

Figure 1 shows the percentage reduction in the size of confidence intervals obtained with balanced optimal restricted randomization to that of the confidence intervals obtained with complete randomization . Figure 2 show the same information but comparing the balanced unbiased restricted randomization to balanced complete randomization. Each panel shows the percentage reduction in the size of confidence intervals obtained by using our method, for a single network. Each row correspond to a different degree of misspecification. The columns indicate how we chose the network displayed in the panels: for the panel in the first column and first row, we chose the network for which the mean percentage reduction was smallest, in the correctly specified case. The other panels can be interpreted similarly. The coverage for all methods was 95%, as expected.

The key message of these figures is that our methods lead to smaller confidence intervals, even under strong misspecification, and that this reduction is consistent across different networks.

3. NUMERICAL RESULTS AND ROBUSTNESS TO MISSPECIFICATION

3·1. *Design of simulation experiments*

We consider four families of networks: Erdös-Renyi, power law, stochastic blockmodel, and small world on a ring lattice. We do this for convenience, but without loss of generality, since the formulas for the mean square error in Sections 2·3–2·4 and the theory and methods in Section 3 depend on observed network statistics. We generate 100 networks, each with 500 nodes, from these families. These networks all have comparable edge densities of roughly 0·08 with stadard errors of 0·02 by design. The outcomes are generated according to the model in Equations 2–4 of the article, with parameters $\mu = 1$, $\sigma = 2$ and $\gamma = 1$. We note that several allocation strategies

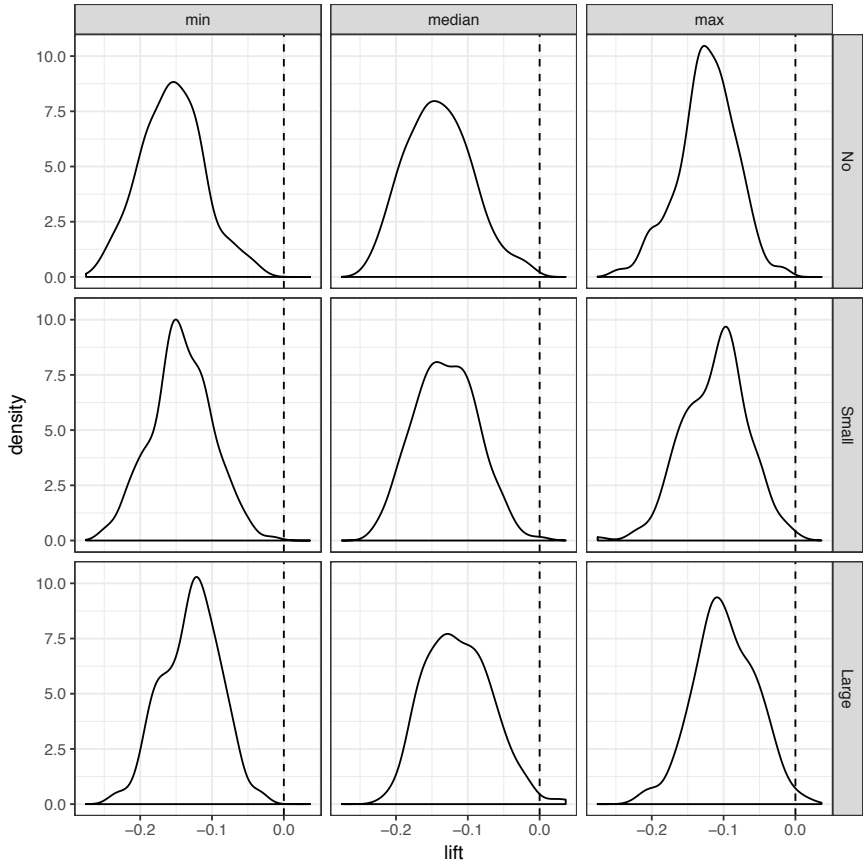

Fig. 1: Reduction is the size of Fisher intervals obtained using our balanced optimal restricted randomization strategy, with $\alpha = 0{\cdot}05$, compared to the size of Fisher intervals obtained using a balanced randomization design.

described in Section 3·2 require solving optimization problems, for which we can only provide approximate solutions. All optimizations are carried out via stochastic optimization (Goldberg & Holland 1988). We discuss the variability in the results due to this approximation when appropriate.

3·2. *Comparative performance analysis*

The goal of this set of simulations is to quantify the order of magnitude of improvements in mean square error an analyst can expect, under controlled conditions. In these simulations, we compare the performance of the different estimators when the data are simulated from the model in Equations 2–4. For each of the 400 networks described in Section 3·1, we generate 300 assignments for each of the methods described in Section 3. For each assignment we compute the mean square error. Thus the results here compare performance of the randomization strategies coupled with the simple difference-in-means estimator. We postpone the discussion of the maximum likelihood estimator to the following section.

Figure 3 shows the mean square error densities for seven randomization strategies, estimated from 30,000 replicated experiments for each network family. Figure 4 shows the mean square error densities for seven randomization strategies, estimated from 300 replicated experiments for

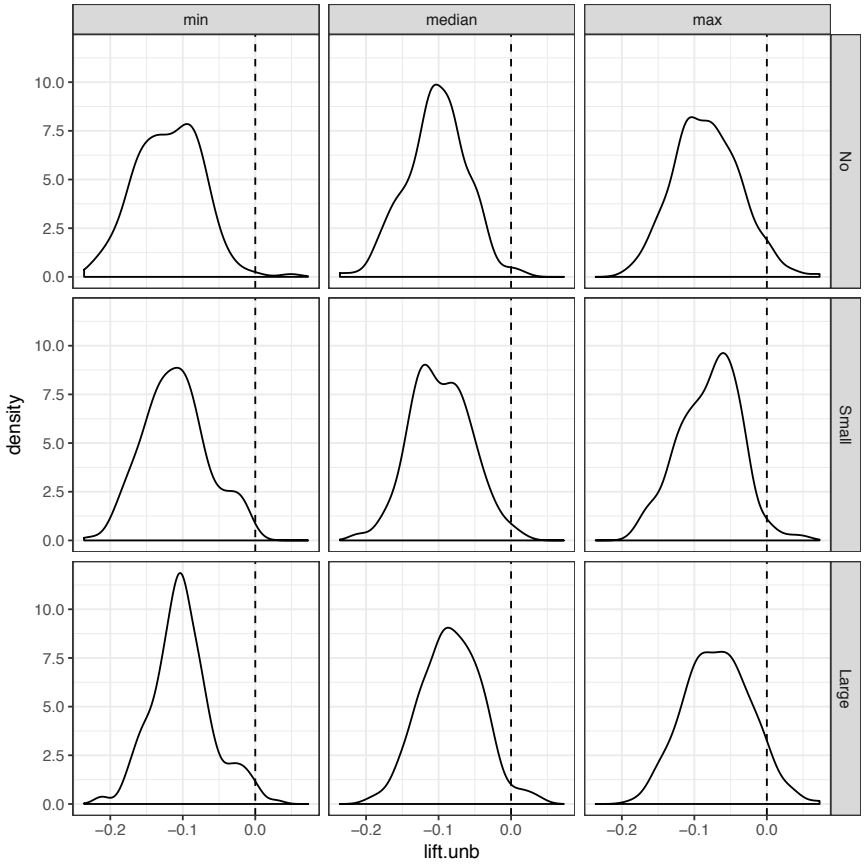

Fig. 2: Reduction is the size of Fisher intervals obtained using our balanced unbiased restricted randomization strategy, with $\alpha = 0{\cdot}05$, compared to the size of Fisher intervals obtained using a balanced randomization design.

the first simulated network in each family. In both figures, we truncated the x-axis at 5, however the mean square errors for the Bernoulli and balanced randomizations take values as high as 10.

The results suggest a number of observations. Balancing the number of treated and control units generally improves the mean squared error over Bernoulli randomizations. Removing bias through restricted randomization, by balancing the average degree of treated and untreated units as suggested by Equation 5 of the article, generally improves the mean squared error over balanced randomizations. Restricted randomization designs that effectively discard balanced randomizations with high mean square error often outperform balanced unbiased restricted randomizations, with the exception of power law networks. In these networks, the balanced unbiased restricted randomization still outperforms the restricted randomization that keeps only allocation vectors with the 20% highest mean square error. This happens because the degree distribution in power law networks is very skewed and even the top 20% allocations in terms of mean square error display quite a lack of balance in terms of average degree between treated and non treated units.

We note that different restricted randomization strategies explore the treatment allocation vectors with different criteria, thus these improvements are simply a consequence of the difficulty in exploring a vast space; we sampled 300 allocation vectors in a space that has roughly 2^{100}

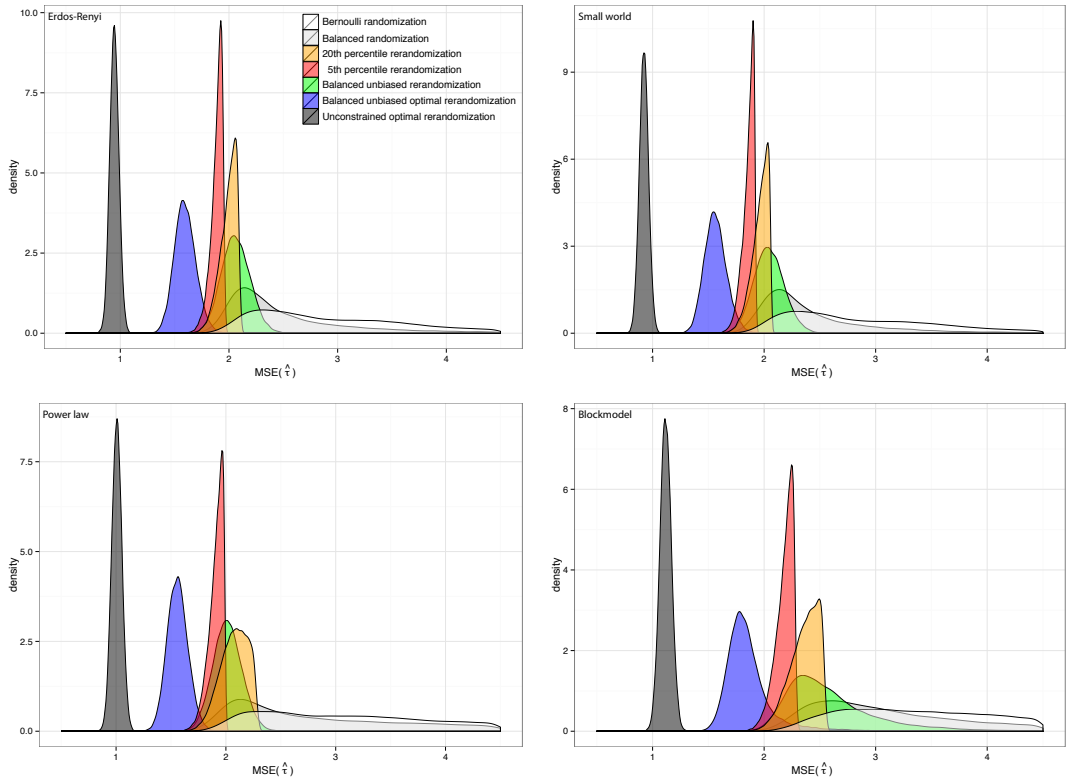

Fig. 3: Distributions of the marginal mean square error for seven randomization strategies, each estimated from 30,000 replicated experiments for each network family.

elements. In our experience in designing large experiments practitioners typically generate tens of thousands of allocation vectors, but only look closely at hundreds of them, so our simulation is realistic in this respect (Kim et al. 2015). The other three network families, on the other hand, have much more symmetric degree distributions, and so explicitly disregarding balance on average degree does not lead to heavy bias and higher mean square errors.

Interestingly, balanced unbiased restricted randomization that explicitly control the variance terms in the conditional mean square error as suggested in Equations 7–9 substantially improve the mean squared error over balanced restricted randomization based on the overall mean square error. This suggest that is is unlikely to find allocation vectors with good variance control in a small set of allocations. Unconstrained restricted randomization that directly control the sum of bias and variance terms substantially improve the mean squared error over balanced unbiased optimal restricted randomization. This is consistent with the findings in classical estimation tasks, where a small increase in bias may lead to larger reductions in variance, and thus to lower mean square error.

Overall, model-assisted restricted randomization designs perform better, under ideal conditions. The theory in Section 3·6 provides assurances in terms of unbiasedness when the model does not hold.

3·3. *Robustness to network misspecification*

The goal of this set of simulations is to quantify the loss in performance of the randomization strategies we are considering when the network the model conditions on is misspecified.

We perturbed each of the 400 networks simulated for Section 3·1 by randomly rewiring different proportions of the edges. For each perturbed network we generated 100 allocation vectors using six randomization strategies: those considered in the previous section with the exclusion of the Bernoulli randomizations. Perturbations in the network only affect the proposed model-assisted restricted randomization strategies, which rely on explicit bias and variance terms that now depend on the perturbed network, while the outcomes are generated according to a model that relies on the unperturbed network. In addition, we consider the randomization strategy that minimizes the analytical expression for the mean square error of the maximum likelihood estimator, in Equation 4·5, which also leverages the perturbed network for estimating the parameter τ. We evaluated assignments in terms of marginal mean squared error, computed using the unperturbed networks.

Figure 5 displays the resulting mean square errors for the seven randomization strategies described above, and for the maximum likelihood estimator under a balanced complete randomization as a baseline for the maximum likelihood estimator. This baseline allows us to quantify the effects of model misspecification on the mean square error because of failures in the estimation task only, and to contrast it with the effects model misspecification on the mean square error because of failures in both estimation and optimal treatment allocation tasks. The four panels in

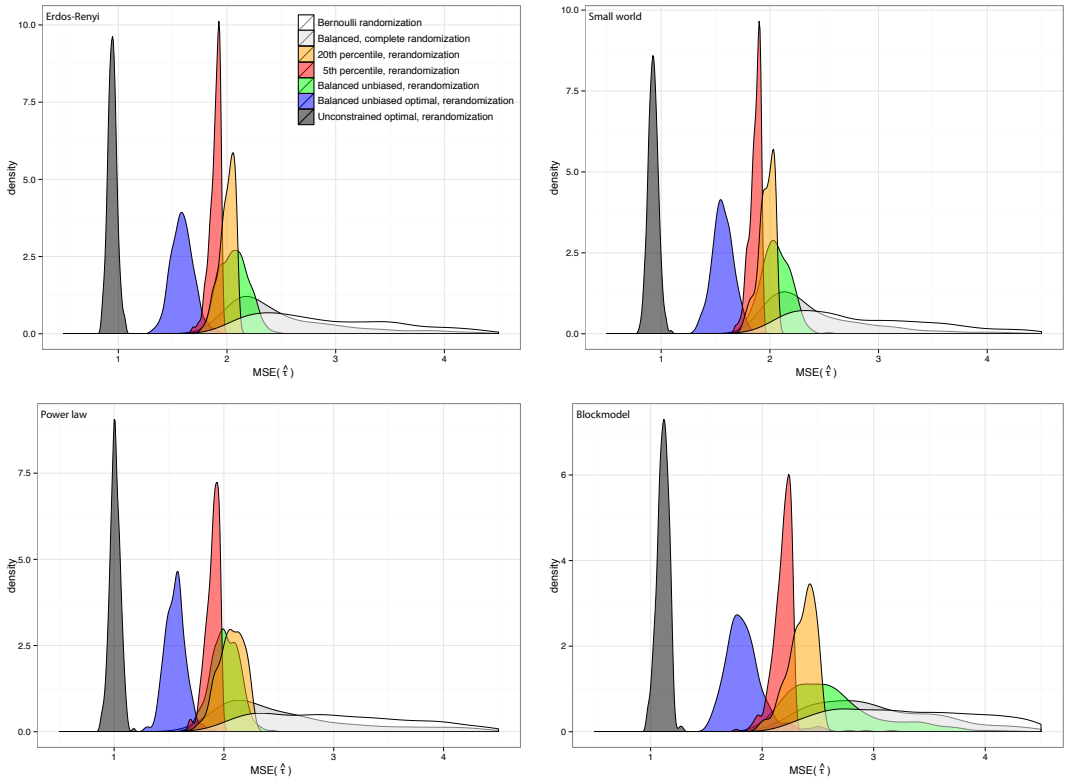

Fig. 4: Distributions of the marginal mean square error for seven randomization strategies, each estimated from 300 replicated experiments on a single simulated network for each family.

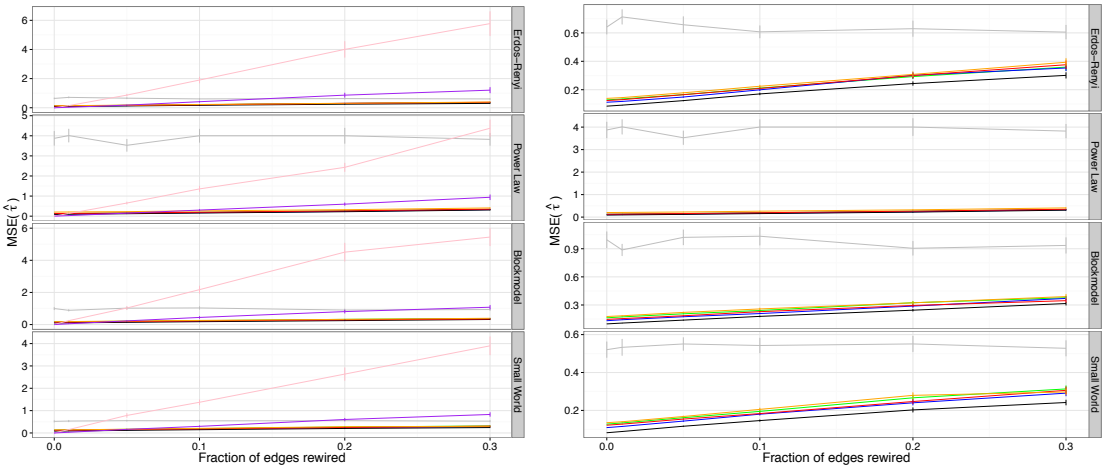

Fig. 5: Left: Robustness of eight randomization strategies with colors as in previous Figures, maximum likelihood estimator optimal in purple, and maximum likelihood optimal baseline in pink, to increasing perturbations of the network underlying the model for the outcomes in Equations 2–4. Right: Focus on the robustness to model misspecification of the six randomization strategies for the difference-in-means estimator.

Figure 5 show the mean square error for the four different network families. The x-axis measures the fraction of edges rewired that defines the severity of the network perturbation. For instance, at 0·01 we rewire 1% of the edges; in networks with 500 nodes and density 0·15, this corresponds to 188 edges on average. At zero, mean square errors correspond to unperturbed networks.

The results suggest a few observations. There is a clear contrast between randomization strategies based on the maximum likelihood estimator and those based on the difference-in-means estimator. While strategies targeting the maximum likelihood estimator outperform the other strategies in the absence of model misspefication, even for modest misspecification, such as 5% rewiring, their mean square error increases substantially and exceeds that of strategies based on the difference-in-means estimator. Perhaps surprisingly, the balanced complete randomization for the maximum likelihood estimator performs worst than the optimizing treatment assignment for the maximum likelihood estimator for the range of misspecification explored. This oversensitivity to model misspecification makes approaches that depend on the maximum likelihood estimator both for the design and estimation unattractive options in practice.

In contrast, randomization strategies based on the difference-in-means estimator are generally insensitive to increasing amounts of misspecification, which is plausible since this estimator does not depend on the network. No amount of misspecification in the range we consider alters the ordering the proposed restricted randomization strategies suggested in Section 3·2, in terms of average marginal mean square error over the simulated networks.

3·4. *Robustness to prior misspecification*

The goal of this set of simulations is to quantify the loss in performance of the randomization strategies we are considering when parameters in the model for the outcomes are misspecified.

For each of the 400 networks used in the previous simulations, we generated 100 assignments from each of the six randomization strategies based on the difference-in-means estimator, for each of the nine sets of parameters specified in Table 1. Recall that γ^2 is the variance of the

outcomes, and (μ, σ^2) are mean and variance of the individual features that induce correlation among the outcomes along a given network. The parameter sets are listed in increasing order or average marginal mean square error expected under unconstrained optimal rerandomizations. The first set of parameter values gives the values used to generate the outcomes.

Figure 4 shows the resulting mean square errors for the six randomization strategies. The results suggest a few observations. The balanced complete randomization strategy, in gray, is insensitive to the changes in the model since it does not rely on any aspect of it for assigning treatment. The two balanced rerandomization strategies, 5% in red and 20% in orange, select allocations based on their conditional mean square error, which is computed using misspecified parameters; these strategies suffer in settings where both the parameter that controls the bias μ is wrongly assumed to be negligible and the parameter that controls the variability in the outcomes γ is wrongly assumed to be much bigger than its real value. The balanced unbiased rerandomization strategy, in green, is insensitive to misspecification; it disregards the variance components of the mean square error, thus eliminating sensitivity to the misspecification of γ and σ, and it selects allocations that zero out the term $\delta_{\mathcal{N}}$ in Equation 5, thus eliminating any potential adverse effects due to the misspecification of μ. The balanced unbiased optimal rerandomization strategy, in blue, is generally robust to parameter misspecification, while achieving low mean square error. Interestingly, the optimal unconstrained rerandomization strategy, in black, which despite parameter misspecifications achieves the lowest mean square error, in settings 8–9 suffers from trading too little bias for variance, since it wrongly assumes a high value for μ, and thus looses its advantage over the balanced unbiased optimal restricted randomization strategy.

4. Analytical derivations

4·1. *Model for the observed data for the normal-sum model*

We first derive the correlation between control potential outcomes

$$cov\{Y_i(0), Y_j(0)\} = \begin{cases} |\mathcal{N}_i \cap \mathcal{N}_j|\sigma^2, & \text{if } i \neq j, \\ |\mathcal{N}_i \cap \mathcal{N}_j|\sigma^2 + \gamma^2, & \text{if } i = j. \end{cases}$$

Set no.	μ	γ	σ
1	1	1	2
2	20	20	0.1
3	0.1	0.1	0.1
4	20	20	20
5	0.1	20	0.1
6	0.1	20	20
7	0.1	0.1	20
8	20	0.1	0.1
9	20	0.1	20

Tab. 1: Parameters values for Figure 4.

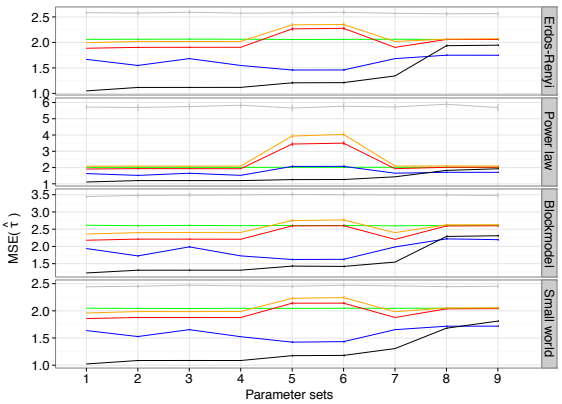

Fig. 4: Robustness to prior misspecifications.

It follows that

$$\begin{aligned} cov\{Y(Z), Y(Z) \mid Z\} &= cov\{Y(0) + Z\tau, Y(0) + Z\tau \mid Z\} \\ &= cov\{Y(0), Y(0)\} \\ &= A^T A\sigma^2 + I\gamma^2. \end{aligned}$$

It then follows that the observed model can be written

$$Y^{\text{obs}} = Y(Z^{\text{obs}}) \mid Z^{\text{obs}} \sim \text{MVN}(A\mu + \tau Z^{obs}, \gamma^2 I + \sigma^2 AA^T). \tag{2}$$

where MVN denotes the multivariate-normal distribution.

4·2. *Derivation of the conditional mean square error for the normal-sum model*

It follows from the calculations in Appendix 4·1 that:

$$\begin{aligned} var(\hat{\tau} \mid Z) &= var\{\omega(Z)^T Y(0)\} \\ &= \omega(Z)^T cov\{Y(0), Y(0)\}\omega(Z) \\ &= \omega(Z)^T \bigg(A^T A\sigma^2 + I\gamma^2 \bigg)\omega(Z), \end{aligned}$$

where $\omega(Z) = Z/N_1 - (1-Z)/N_0$. So we have shown that

$$var(\hat{\tau} \mid Z) = \omega(Z)^T \bigg(A^T A\sigma^2 + I\gamma^2 \bigg)\omega(Z).$$

It is quick to derive

$$\begin{aligned} E(\hat{\tau} \mid Z) &= E\bigg\{ N_1^{-1} \sum_{i=1}^{N} Z_i Y_i(1) - N_0^{-1} \sum_{i=1}^{N} (1 - Z_i) Y_i(0) \mid Z \bigg\} \\ &= \tau + \sum_{i=1}^{N} \{Z_i/N_1 - (1-Z_i)/N_0\} E\{Y_i(0)\} \\ &= \tau + \sum_{i=1}^{N} \{Z_i/N_1 - (1-Z_i)/N_0\} E(\sum_{j \in \mathcal{N}_i} X_j) \\ &= \tau + \sum_{i=1}^{N} \{Z_i/N_1 - (1-Z_i)/N_0\} |\mathcal{N}_i| \mu \\ &= \tau + \delta_{\mathcal{N}} \mu, \end{aligned}$$

where $\delta_{\mathcal{N}} = (\sum_{Z_i=1} |\mathcal{N}_i|)/N_1 - (\sum_{Z_i=0} |\mathcal{N}_i|)/N_0$. Thus

$$\text{bias}(\hat{\tau} \mid Z)^2 = E(\hat{\tau} - \tau \mid Z)^2 = \mu^2 (\delta_{\mathcal{N}})^2,$$

which gives the following conditional mean square error:

$$\text{MSE}(\hat{\tau} \mid Z) = \mu^2 (\delta_{\mathcal{N}})^2 + \omega(Z)^T \bigg(A^T A\sigma^2 + I\gamma^2 \bigg)\omega(Z).$$

4·3. *Derivation of the marginal mean square error for the normal-sum model*

The purpose of this section is to derive the analytical expression for:

$$E\{(\hat{\tau}-\tau)^2\} = E[E\{(\hat{\tau}-\tau)^2 \mid Y\}].$$

We start by noticing that

$$E(\hat{\tau}) = E\{E(\hat{\tau}) \mid Y\} = E(\tau) = \tau,$$

so $\text{bias}(\hat{\tau}) = 0$. Also, we have

$$\begin{aligned} var(\hat{\tau}) &= E\{var(\hat{\tau} \mid Y)\} + var\{E(\hat{\tau} \mid Y)\} \\ &= E\{var(\hat{\tau} \mid Y)\} + var(\tau) \\ &= E\{var(\hat{\tau} \mid Y)\}, \end{aligned}$$

so

$$\text{MSE}(\hat{\tau}) = E\{var(\hat{\tau} \mid Y)\}.$$

We focus on this quantity, we have

$$E\{var(\hat{\tau} \mid Y)\} = E(S_1^2)/N_1 + E(S_0^2)/N_0.$$

We start with $E(S_1^2)$ and we will show later that it's actually equal to $E(S_0^2)$. Also, we will use the shorthand $Y_i = Y_i(1)$ and $\bar{Y} = \bar{Y}(1)$ in order to simplify the notation. Let

$$\begin{aligned} D_i &= (Y_i - \bar{Y}_i)^2 \\ &= Y_i^2 + \bar{Y}^2 - 2Y_i\bar{Y}, \end{aligned}$$

and remember that

$$\begin{aligned} E(Y_i) &= |\mathcal{N}_i|\mu + \tau, \\ var(Y_i) &= \gamma^2 + |\mathcal{N}|\sigma^2, \\ cov(Y_i, Y_j) &= |\mathcal{N}_i \cap \mathcal{N}_j|\sigma^2 \quad i \neq j, \end{aligned}$$

so

$$\begin{aligned} S_1^2 &= (N-1)^{-1} \sum_i D_i \\ &= (N-1)^{-1} \sum_{i=1}^N Y_i^2 + (N-1)^{-1} \sum_i \bar{Y}^2 - \{2/(N-1)\} \sum_{i=1}^N Y_i\bar{Y} \\ &= (N-1)^{-1} \sum_{i=1}^N Y_i^2 + \{N/(N-1)\}\bar{Y}^2 - \{2N/(N-1)\}\bar{Y}^2 \\ &= (N-1)^{-1} \sum_{i=1}^N Y_i^2 - \{N/(N-1)\}\bar{Y}^2 \\ &= B_1 - B_2, \end{aligned}$$

where $B_1 = (N-1)^{-1} \sum_{i=1}^N Y_i^2$ and $B_2 = \{N/(N-1)\}\bar{Y}^2$. We have

$$\begin{aligned} E(Y_i^2) &= var(Y_i) + E(Y_i)^2 \\ &= \gamma^2 + |\mathcal{N}_i|\sigma^2 + (|\mathcal{N}_i|\mu + \tau)^2, \end{aligned}$$

so

$$E(B_1) = (N-1)^{-1}\left\{N\gamma^2 + N\sigma^2\overline{|\mathcal{N}|} + \sum_i (\overline{|\mathcal{N}_i|}\mu + \tau)^2\right\}$$

and

$$\begin{aligned} E(\bar{Y}^2) &= var(\bar{Y}) + E(\bar{Y})^2 \\ &= (1/N^2)1^t\Sigma 1 + (\tau + \mu\overline{\mathcal{N}})^2. \end{aligned}$$

Notice that

$$\begin{aligned} 1^T\Sigma 1 &= N\gamma^2 + \sigma^2\sum|\mathcal{N}_i| + 2\sum_{i<j}|\mathcal{N}_i \cap \mathcal{N}_j|\sigma^2 \\ &= N\gamma^2 + N\sigma^2\overline{|\mathcal{N}|} + 2\sigma^2\sum_{i<j}|\mathcal{N}_i \cap \mathcal{N}_j|, \end{aligned}$$

so

$$\begin{aligned} E[B_2] &= \{N(N-1)\}^{-1}1^T\Sigma 1 + \{N/(N-1)\}(\tau + \mu\overline{|\mathcal{N}|})^2 \\ &= (N-1)^{-1}\gamma^2 + (N-1)^{-1}\sigma^2\overline{|\mathcal{N}|} + 2\{N(N-1)\}^{-1}\sigma^2\sum_{i<j}|\mathcal{N}_i \cap \mathcal{N}_j| + \{N/(N-1)\}(\tau + \mu\overline{|\mathcal{N}|})^2 \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^N(|\mathcal{N}_i|\mu + \tau)^2 - N(\tau + \mu\overline{|\mathcal{N}|})^2 &= N\tau^2 + 2\tau\mu N\overline{|\mathcal{N}|} \\ &\quad + \mu^2\sum_{i=1}^N|\mathcal{N}_i|^2 - \tau^2 N - N\mu^2\overline{|\mathcal{N}|}^2 - 2N\tau\mu\overline{|\mathcal{N}|} \\ &= \mu^2 N(\overline{|\mathcal{N}|^2} - \overline{|\mathcal{N}|}^2), \end{aligned}$$

which leads to:

$$E(S_1^2) = \gamma^2 + \sigma^2\overline{|\mathcal{N}|} - 2\sigma^2\{N(N-1)\}^{-1}\sum_{i<j}|\mathcal{N}_i \cap \mathcal{N}_j| + \mu^2\{N/(N-1)\}(\overline{|\mathcal{N}|^2} - \overline{|\mathcal{N}|}^2).$$

Clearly, none of the above would change for S_0^2 since the τ's cancel out. So finally,

$$\begin{aligned} E\{var(\hat{\tau} \mid Y)\} = (N_1^{-1} + N_0^{-1})\Big[&\gamma^2 + \sigma^2\overline{|\mathcal{N}|} \\ &- 2\sigma^2\{N(N-1)\}^{-1}\sum_{i<j}|\mathcal{N}_i \cap \mathcal{N}_j| \\ &+ \mu^2\{N/(N-1)\}(\overline{|\mathcal{N}|^2} - \overline{|\mathcal{N}|}^2)\Big]. \end{aligned}$$

A simple sanity check is to look at what would happen if there was no network. That is, if $|\mathcal{N}_i| = 1$ for all i, and $|\mathcal{N}_i \cap \mathcal{N}_j| = 0$ for all $i \neq j$. The above formula then reduces to $(N_0^{-1} + N_1^{-1})(\gamma^2 + \sigma^2)$, which is correct. This suggests a refactorization of the equation above:

$$E\{var(\hat{\tau} \mid Y)\} = V_1 + V_2,$$

where

$$V_1 = (N_1^{-1} + N_0^{-1})(\gamma^2 + \sigma^2)$$

is the variance term in the absence of a network, and

$$\begin{aligned} V_2 = (N_1^{-1} + N_0^{-1}) \Big[& \sigma^2(\overline{|\mathcal{N}|} - 1) \\ & - 2\sigma^2\{N(N-1)\}^{-1} \sum_{i<j} |\mathcal{N}_i \cap \mathcal{N}_j| \\ & + \mu^2\{N/(N-1)\}(\overline{|\mathcal{N}|^2} - \overline{|\mathcal{N}|}^2 \Big] \end{aligned}$$

is the variance term correction when a network structure is present. So in conclusion, we have:

$$\text{MSE}(\hat{\tau}) = V_1 + V_2,$$

which completes the proof.

4·4. *Analysis of the difference-in-means estimator under the normal-mean model*

We consider the normal-means model, as an alternative:

$$\begin{aligned} X_j &\sim \text{ Normal } (\mu, \sigma^2), \\ Y_i(0) \mid X &\sim \text{ Normal } (|\mathcal{N}_i|^{-1} \textstyle\sum_{j \in \mathcal{N}_i} X_j, \gamma^2), \\ Y_i(1) &= Y_i(0) + \tau. \end{aligned}$$

It easy to verify that for all Z, we have: $E(\hat{\tau} \mid Z) = 0$. Then, as in the sum model, the variance can be expressed as

$$var(\hat{\tau} \mid Z) = \omega(Z)^T var\{Y(0) \mid Z\}\omega(Z),$$

where

$$\begin{aligned} var\{Y(0) \mid Z\} &= E\{var(Y \mid X, Z) \mid Z\} + var\{E(Y \mid X, Z) \mid Z\} \\ &= E(\gamma^2 I \mid Z) + var(\tilde{A}X \mid Z) \\ &= \gamma^2 I + \sigma^2 \tilde{A}\tilde{A}^T, \end{aligned}$$

and $\tilde{A}$ is the matrix such that $\tilde{A}_{ij} = A_{ij}/|\mathcal{N}_i|$. And so finally,

$$\text{MSE}(\hat{\tau} \mid Z) = \gamma^2 \omega(Z)^T \omega(Z) + \omega(Z)^T \tilde{A}\tilde{A}^T \omega(Z),$$

which we can write in longer form as:

$$\begin{aligned} \text{MSE}(\hat{\tau}|Z) = \gamma^2(N_1^{-1} + N_0^{-1}) & \\ + (\sigma^2/N_1^2) \sum_{i,j:Z_i=Z_j=1} & \frac{|\mathcal{N}_i \cap \mathcal{N}_j|}{|\mathcal{N}_i||\mathcal{N}_j|} \\ + (\sigma^2/N_0^2) \sum_{i,j:Z_i=Z_j=0} & \frac{|\mathcal{N}_i \cap \mathcal{N}_j|}{|\mathcal{N}_i||\mathcal{N}_j|} \\ - \{2\sigma^2/(N_1 N_0)\} \sum_{i,j:Z_i=1,Z_i=0} & \frac{|\mathcal{N}_i \cap \mathcal{N}_j|}{|\mathcal{N}_i||\mathcal{N}_j|}. \end{aligned}$$

The first term penalizes, as before, imbalance in the sizes of the treatment groups. The last three terms look a lot like what we had with the sum-model, except that we now have weighted averages.

4·5. *Analysis of the maximum likelihood estimator under the normal-sum model*

The naive estimator does not make any reference to the network. This is not the case for the maximum likelihood estimator, and we need to introduce a distinction between true and observed network. Let A_0 be the adjacency matrix associated with the true unobserved network, and A be the adjacency matrix associated with the noisy observed network. The model we use will be based on the observed network, while the evaluation will be with respect to the true network. We have shown in the observed model of (2) that the observed outcomes are jointly multivariate normal. Let $v = A\mu + \tau Z^{obs}$ be the mean, and let $\Sigma = \gamma^2 I + \sigma^2 AA^T$ be the variance. We also denote by $\Sigma_0 = \gamma^2 I + \sigma^2 A_0 A_0^T$ the variance based on the true covariance matrix. Finally, define $\mu^* = A\mu$ and $\mu_0 = A_0\mu$. With this, standard results show that

$$\begin{aligned}\frac{d}{d\tau}\log P\{Y^{obs}(Z) \mid \mu,\sigma,\gamma\} = 0 &\Leftrightarrow \{Y(Z) - \mu^* - \tau Z\}^T\Sigma^{-1}Z = 0\\ &\Leftrightarrow \hat{\tau}^{\text{MLE}}(Z) = \frac{\{Y(Z)-\mu^*\}\Sigma^{-1}Z}{Z^T\Sigma^{-1}Z}.\end{aligned}$$

Under the true model,

$$\begin{aligned}E(\hat{\tau}^{\text{MLE}} \mid Z) &= \frac{(\mu_0^* + \tau Z - \mu^*)^T\Sigma^{-1}Z}{Z^T\Sigma^{-1}Z}\\ &= \frac{(\mu_0^* - \mu^*)^T\Sigma^{-1}Z}{Z^T\Sigma^{-1}Z} + \tau\end{aligned}$$

and the bias is

$$\text{bias}(\hat{\tau}^{\text{MLE}} \mid Z) = \frac{(\mu_0^* - \mu^*)^T\Sigma^{-1}Z}{Z^T\Sigma^{-1}Z}.$$

The variance is

$$\begin{aligned}var(\hat{\tau}^{\text{MLE}} \mid Z) &= \{1/(Z^T\Sigma^{-1}Z)^2\}var\{Y(Z)\Sigma^{-1}Z \mid Z\}\\ &= \{1/(Z^T\Sigma^{-1}Z)^2\}(\Sigma^{-1}Z)^T var\{Y(Z) \mid Z\}\Sigma^{-1}Z\\ &= \{1/(Z^T\Sigma^{-1}Z)^2\}(\Sigma^{-1}Z)^T\Sigma_0\Sigma^{-1}Z\\ &= \frac{Z^T\Sigma^{-1}\Sigma_0\Sigma^{-1}Z}{(Z^T\Sigma^{-1}Z)^2},\end{aligned}$$

and so finally,

$$\text{MSE}(\hat{\tau}^{\text{MLE}}|Z) = \left\{\frac{(\mu_0^* - \mu^*)^T\Sigma^{-1}Z}{Z^T\Sigma^{-1}Z}\right\}^2 + \frac{Z^T\Sigma^{-1}\Sigma_0\Sigma^{-1}Z}{(Z^T\Sigma^{-1}Z)^2}.$$

5. PROOFS

5·1. *Proof of Corollary* 1

Proof of Corollary 1. The key intuition for the proof is that $\mathcal{Z}^b \cap \mathcal{Z}^o \subset \mathcal{Z}^b$ and that the assignments that are in $\mathcal{Z}^b$ and not in $\mathcal{Z}^b \cap \mathcal{Z}^o$ have large model mean square error. In all the proofs below, we will write $E_{\mathcal{Z}^b}(\cdot)$ as a shorthand for $E(\cdot \mid Z \in \mathcal{Z}^b)$, and similarly for the variance,

covariance bias and mean square error. Since $\text{bias}_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y) = \text{bias}_{\mathcal{Z}^b}(\hat{\tau} \mid Y) = 0$, we have

$$\begin{aligned} E\{var_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y)\} &= E\Big\{\text{bias}_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y)^2 + var_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y)\Big\} \\ &= E\{\text{MSE}_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y)\} \\ &= E[E_{\mathcal{Z}^b \cap \mathcal{Z}^o}\{(\hat{\tau} - \tau)^2 \mid Y\}] \\ &= E_{\mathcal{Z}^b \cap \mathcal{Z}^o}[E\{(\hat{\tau} - \tau)^2 \mid Z\}] \end{aligned}$$

and similarly,

$$\begin{aligned} E\{var_{\mathcal{Z}^b}(\hat{\tau} \mid Y)\} &= E_{\mathcal{Z}^b}[E\{(\hat{\tau} - \tau)^2 \mid Z\}] \\ &= E_{\mathcal{Z}^b}[I(Z \in \mathcal{Z}^b \cap \mathcal{Z}^o)E\{(\hat{\tau} - \tau)^2 \mid Z\} + I(Z \in \mathcal{Z}^b \backslash \mathcal{Z}^b \cap \mathcal{Z}^o)E\{(\hat{\tau} - \tau)^2 \mid Z\}] \\ &= P(Z \in \mathcal{Z}^b \cap \mathcal{Z}^o)E_{\mathcal{Z}^b}[E\{(\hat{\tau} - \tau)^2 \mid Z\} \mid Z \in \mathcal{Z}^b \cap \mathcal{Z}^o] \\ &\quad + \{1 - P(Z \in \mathcal{Z}^b \cap \mathcal{Z}^o)\}E_{\mathcal{Z}^b}[E\{(\hat{\tau} - \tau)^2 \mid Z\} \mid Z \in \mathcal{Z}^b \backslash \mathcal{Z}^b \cap \mathcal{Z}^o] \\ &= pE_{\mathcal{Z}^b \cap \mathcal{Z}^o}[E\{(\hat{\tau} - \tau)^2 \mid Z\}] + (1-p)E_{\mathcal{Z}^b}[E\{(\hat{\tau} - \tau)^2 \mid Z\} \mid Z \in \mathcal{Z}^b \backslash \mathcal{Z}^b \cap \mathcal{Z}^o] \\ &\geq pE_{\mathcal{Z}^b \cap \mathcal{Z}^o}[E\{(\hat{\tau} - \tau)^2 \mid Z\}] + (1-p)q_\alpha \\ &\geq E_{\mathcal{Z}^b \cap \mathcal{Z}^o}[E\{(\hat{\tau} - \tau)^2 \mid Z\}] \\ &= E\{var_{\mathcal{Z}^b \cap \mathcal{Z}^o}(\hat{\tau} \mid Y)\}, \end{aligned}$$

which concludes the proof.

5·2. *Proof of Lemma* 1

Proof of Lemma 1. Let $Z \in \mathcal{Z}^b$, and let $Z^* = 1 - Z$. Clearly, $N_1(1-Z) = N_1(Z) = N/2$ and $N_0(Z) = N_0(1-Z) = N/2$. Then

$$\begin{aligned} \hat{\tau}(Z) + \hat{\tau}(1-Z) &= \left\{(N/2)^{-1}\sum_{Z_i=1} Y_i(1) - (N/2)^{-1}\sum_{Z_i=0} Y_i(0)\right\} \\ &\quad + \left\{(N/2)^{-1}\sum_{1-Z_i=1} Y_i(1) - (N/2)^{-1}\sum_{1-Z_i=0} Y_i(0)\right\} \\ &= \left\{(N/2)^{-1}\sum_{Z_i=1} Y_i(1) - (N/2)^{-1}\sum_{Z_i=0} Y_i(0)\right\} + \left\{\frac{1}{N/2}\sum_{Z_i=0} Y_i(1) - \frac{1}{N/2}\sum_{Z_i=1} Y_i(0)\right\} \\ &= (N/2)^{-1}\sum_{i}^{N} Y_i(1) - (N/2)^{-1}\sum_{i}^{N} Y_i(0) \\ &= 2\tau, \end{aligned}$$

which completes the proof. □

5·3. *Proof of Theorem* 1

There are four parts to this theorem: we must show that the estimator is unbiased under the uniform designs on $\mathcal{Z}^b$, $\mathcal{Z}^b \cap \mathcal{Z}^u$, $\mathcal{Z}^b \cap \mathcal{Z}^o$, and $\mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o$. Th proofs follow the same general ideas, so we will skip the details whenever the proofs are similar.

Proof of Theorem 1*, part (i): unbiasedness under the uniform distribution on* $\mathcal{Z}^b$. This proof could be carried exactly as above. The longer proof that we use introduces concepts that will be reused in most of the following proofs, but in a simple scenario.
By definition,

$$Z \in \mathcal{Z}^b \Rightarrow Z^* = 1 - Z \in \mathcal{Z}^b.$$

Now, introduce for all i the sets:

$$\mathcal{Z}^b_{i=1} = \{Z \in \mathcal{Z}^b : Z_i = 1\}, \quad \mathcal{Z}^b_{i=0} = \{Z \in \mathcal{Z}^b : Z_i = 0\}$$

Notice that $Z \in \mathcal{Z}^b_{i=1} \Leftrightarrow Z^* \in \mathcal{Z}^b_{i=0}$, which implies $|\mathcal{Z}^b_{i=1}| = |\mathcal{Z}^b_{i=0}|$. And since

$$\mathcal{Z}^b_{i=1} \cup \mathcal{Z}^b_{i=0} = \mathcal{Z}^b \quad \text{and} \quad \mathcal{Z}^b_{i=1} \cap \mathcal{Z}^b_{i=0} = \emptyset,$$

we conclude that

$$|\mathcal{Z}^b_{i=1}| = |\mathcal{Z}^b_{i=0}| = |\mathcal{Z}^b|/2,$$

which implies that

$$P_{\mathcal{Z}^b}(Z_i = 1) = \frac{|\mathcal{Z}^b_{i=1}|}{|\mathcal{Z}^b|} = \frac{1}{2} = \frac{|\mathcal{Z}^b_{i=0}|}{|\mathcal{Z}^b|} = P_{\mathcal{Z}^b}(Z_i = 0),$$

for all i. We can then write

$$\begin{aligned} E_{\mathcal{Z}^b}[\hat{\tau}] &= E_{\mathcal{Z}^b}\left[(2/N)\sum_{i=1}^{N}\{I(Z_i = 1)Y_i(1) - I(Z_i = 0)Y_i(0)\}\right] \\ &= (2/N)\sum_{i=1}^{N}\{P_{\mathcal{Z}^b}(Z_i = 1)Y_i(1) - P_{\mathcal{Z}^b}(Z_i = 0)Y_i(0)\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\{Y_i(1) - Y_i(0)\} \\ &= \tau, \end{aligned}$$

which completes the proof. □

Proof of Theorem 1*, part (iii): unbiasedness under the uniform distribution on* $\mathcal{Z}^b \cap \mathcal{Z}^o$. It is clear from the previous proof, that the key element of the proof is to show that

$$P_{\mathcal{Z}^b \cap \mathcal{Z}^o}(Z_i = 1) = P_{\mathcal{Z}^b \cap \mathcal{Z}^o}(Z_i = 0) = \frac{1}{2},$$

for all i. For this, we by start proving that

$$Z \in \mathcal{Z}^b \cap \mathcal{Z}^o \Rightarrow Z^* = 1 - Z \in \mathcal{Z}^b \cap \mathcal{Z}^o.$$

By Lemma 1 of the article, $\hat{\tau}(Z^*) = 2\tau - \hat{\tau}(Z)$. Now, let $Z \in \mathcal{Z}^b \cap \mathcal{Z}^o$, we then have

$$\begin{aligned}
\text{MSE}(\hat{\tau}|Z^*) &= E[\{\hat{\tau}(Z^*) - \tau\}^2 \mid Z] \\
&= E[\{2\tau - \hat{\tau}(Z) - \tau\}^2 \mid Z] \\
&= E[\{\tau - \hat{\tau}(Z)\}^2 \mid Z] \\
&= \text{MSE}(\hat{\tau} \mid Z) \\
&\leq q_\alpha^{\text{MSE}},
\end{aligned}$$

which means that $Z^* \in \mathcal{Z}^b \cap \mathcal{Z}^o$. So we proved that

$$Z \in \mathcal{Z}^b \cap \mathcal{Z}^o \Rightarrow Z^* = 1 - Z \in \mathcal{Z}^b \cap \mathcal{Z}^o.$$

The rest of the proof unfolds exactly as in the proof of (i). □

Proof of Theorem 1*, part (ii): unbiasedness under the uniform distribution on* $\mathcal{Z}^b \cap \mathcal{Z}^u$.
Here again, the key is to show that

$$P_{\mathcal{Z}^b \cap \mathcal{Z}^u}(Z_i = 1) = P_{\mathcal{Z}^b \cap \mathcal{Z}^u}(Z_i = 0) = \frac{1}{2}.$$

Let $Z \in \mathcal{Z}^b \cap \mathcal{Z}^u$, then by definition,

$$\text{bias}(\hat{\tau}, \tau \mid Z) = E\{\hat{\tau}(Z) - \tau \mid Z\} = 0,$$

but then, since $\mathcal{Z}^b \cap \mathcal{Z}^u \subset \mathcal{Z}^b$, by the Lemma 1 , we have

$$\begin{aligned}
\text{Bias}(\hat{\tau}|Z^*) &= E[\hat{\tau}(Z^*) - \tau \mid Z] \\
&= E\{2\tau - \hat{\tau}(Z) - \tau \mid Z\} \\
&= -E\{\hat{\tau}(Z) - \tau\} \\
&= -\text{bias}(\hat{\tau} \mid Z) \\
&= 0,
\end{aligned}$$

which implies, that $Z^* \in \mathcal{Z}^b \cap \mathcal{Z}^u$. So we proved that

$$Z \in \mathcal{Z}^b \cap \mathcal{Z}^u \Rightarrow Z^* = 1 - Z \in \mathcal{Z}^b \cap \mathcal{Z}^u.$$

The rest of the proof follows as in the previous two proofs.

Proof of Theorem 1*, part (iv): unbiasedness under the uniform distribution on* $\mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o$.
This proof is exactly the same as the previous one. We will just prove the fact that if $\mathcal{Z}^b \cap \mathcal{Z}^u \neq \emptyset$, then $\mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o$ contains at least two elements. Let $Z \in \mathcal{Z}^b \cap \mathcal{Z}^u$. There exists a Z_0 that minimizes the mean square error on the set $\mathcal{Z}^b \cap \mathcal{Z}^u$ and so $Z_0 \in \mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o$. But we have shown in the proof of (i) that for $Z \in \mathcal{Z}^b$, we have $\text{MSE}(\hat{\tau}, \tau \mid Z) = \text{MSE}(\hat{\tau}, \tau \mid Z^*)$. And so since $Z_0 \in \mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o \subset \mathcal{Z}^b$, we have

$$\text{MSE}(\hat{\tau} \mid Z_0) = \text{MSE}(\hat{\tau} \mid Z_0^*) = \min_{Z \in \mathcal{Z}^b \cap \mathcal{Z}^u} \{\text{MSE}(\hat{\tau} \mid Z)\}$$

which means that $Z^* \in \mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o$, and so $|\mathcal{Z}^b \cap \mathcal{Z}^u \cap \mathcal{Z}^o| \geq 2$. □

REFERENCES

GOLDBERG, D. E. & HOLLAND, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning* **3**, 95–99.

KIM, D. A., HWONG, A. R., STAFFORD, D., HUGHES, D. A., O'MALLEY, A. J., FOWLER, J. H. & CHRISTAKIS, N. A. (2015). Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *The Lancet* **386**, 145–153.

LOCK MORGAN, K. & RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics* **40**, 1263–1282.

ROSENBAUM, P. R. et al. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science* **17**, 286–327.

THORISSON, H. (1995). Coupling methods in probability theory. *Scandinavian Journal of Statistics* **22**, 159–182.