Near-Optimal Experimental Design for Networks: Independent Block Randomization

Ozan Candogan, Chen Chen, Rad Niazadeh

The University of Chicago Booth School of Business
{ozan.candogan, cc459, rad.niazadeh}@chicagobooth.edu

September 29, 2021

Abstract

Motivated by the prevalence of experimentation in online platforms and social networks, we consider the problem of designing randomized experiments to assess the effectiveness of a new market intervention for a network of users. An experiment assigns each user to either the treatment or the control group. The outcome of each user depends on her assignment as well as the assignments of her neighbors. Given the experiment, the unbiased Horvitz-Thompson estimator is used to estimate the total market effect of the treatment. The decision maker chooses randomized assignments of users to treatment and control, in order to minimize the worst-case variance of this estimator. We focus on networks that can be partitioned into communities, where the users in the same community are densely connected, and users from different communities are only loosely connected. In such settings, it is almost without loss to assign all users in the same community to the same variant (treatment or control). The problem of designing the optimal randomized assignments of communities can be formulated as a linear program with an exponential number of decision variables and constraints in the number of communities—and hence, is generally computationally intractable.

We develop a family of practical experiments that we refer to as *independent block randomization (IBR)* experiments. Such an experiment partitions communities into blocks so that each block contains communities of similar sizes. It then treats half of the communities in each block (chosen uniformly at random) and does so independently across blocks. The optimal community partition can be obtained in a tractable way using dynamic programming. We show that these policies are asymptotically optimal when the number of communities grows large and no community size dominates the rest. In the special case where community sizes take values in a finite set and the number of communities of each size is a fixed proportion of the total number of communities, the loss is only a constant that is independent of the number of communities. Beyond the asymptotic regime, we show that the IBR experiment is a $\frac{7}{3}$-approximation for any problem instance. We also examine the performance of the IBR experiments on data-driven numerical examples, including examples based on Facebook subnetworks.

Subject classifications: Experimental design, cluster-based randomization, social networks, interference, approximation algorithms, asymptotic optimality.

1 Introduction

Experimental design is a celebrated branch of statistics—rooted in the pioneering work of Fisher in the 1920s and 1930s (Fisher 1935). In recent years, thanks to the rapid decrease in the cost of conducting experiments in online platforms, experimental design has become a prevalent tool for improving the operations of online marketplaces and social networks. Such platforms often conduct binary experiments, also known as A/B testing, before launching a new feature or introducing a market intervention, as they strive to make data-driven product decisions. The experiment exposes a (randomized) group of targeted users to the new feature, or equivalently assigns each user to either the *treatment* or the *control* group. The platform then uses the resulting outcomes to estimate the new feature's *total market effect*, i.e., the difference in total user outcomes if this feature is introduced to the entire market. Accurately estimating this quantity enables the platform to informatively decide on whether to deploy this new feature.[1]

The aforementioned platforms exhibit complex network effects. Consequently, unless designed carefully, the experiments could suffer from *interference*, where one user's assignment to the treatment or control affects another user's outcome (or behaviour). For example, passengers in a ride-sharing platform share the same supply of drivers; hence, enabling prime time subsidies for passengers in one neighborhood can impact the service experienced by passengers in nearby neighborhoods (Chamandy 2016). Similarly, advertisers in an ad-exchange platform might compete in the same publisher's auction (Barajas et al. 2016), connected users of a social network might be involved in the same daily activities (Eckles et al. 2016), and hosts in an online hospitality platform (such as Airbnb) might share the same pool of guests (Cui et al. 2017, Holtz et al. 2020). In all of these examples, a user's response to treatment may contaminate the outcomes of other users, thereby resulting in bias or inaccuracy when one estimates the total market effect.

To alleviate such interference, a common practice (e.g., see Chapter 22 of Kohavi et al. 2020) is to partition the network of users into almost-disconnected communities (i.e., groups of users) that only exhibit a minor amount of interference among each other (Eckles et al. 2016, Koutra 2017). For example, riders or hosts in different cities are unlikely to interfere with each other; advertisers of totally different products may have different potential publishers; and finally, users in a social network usually form communities based on their geography, interests, and beliefs. The platform then assigns all users in a community to the same variant in the experiment to (hopefully) remove

[1]When a new feature is rolled out, it often impacts all market participants (see, e.g., Chamandy (2016)), which motivates our focus on the total market effect. A complementary research direction, which is outside the scope of the present work, involves personalized feature deployments and adapting the design of experiments accordingly.

much of the interference, and hence have a relatively unbiased or accurate estimation for the total market effect.[2] While such community-level assignment is helpful, it is not clear how the platform can obtain the "best" (correlated) randomized assignment.

1.1 Our Contributions

We focus on a setting with a fixed set of experimental units that interfere with each other. We represent them with a network, whose nodes correspond to the experimental units and edges capture the interference patterns. Specifically, the outcome of a node is the treatment (control) outcome if the node and all her neighbors are assigned to the treatment (control) variant. This is motivated by the applications mentioned earlier. For instance, in the context of our social networks application, the outcome of a user depends on her assignment as well as those individuals she interacts with (i.e., her neighbors). Hence, the social network coincides with the aforementioned network representing the interference pattern. For ease of exposition, and motivated by this application, in the remainder of the paper we refer to our experimental units/nodes as users.[3]

The decision maker chooses an experiment which is a randomized assignment of each user to either treatment or control. Then, he uses the Horvitz-Thompson unbiased estimator (Horvitz and Thompson 1952) to estimate the total market effect.[4] The objective is to design an optimal joint distribution to assign treatments, so as to minimize the variance of this estimator. As the potential outcomes are uncertain, we formulate the problem as a robust optimization problem against the worst-case values (also known as the adversarial values) of the unknown potential outcomes.

We focus on the common scenario where the network of users can be partitioned into several disjoint communities by ignoring a relatively small number of connections—see, e.g., Eckles et al. (2016), Easley et al. (2010), Jackson (2010) for more context. We initially design our experiments and derive our results, assuming that such connections are absent. Later in Section 5, we show that similar results to those we obtain in this setting readily carry over to settings where a small number of cross-community connections are present.[5]

When cross-community connections are absent, it is natural to focus on *community-level assign-*

[2] See Section 1.1 of Wager and Xu (2021) for a related discussion.

[3] Note that this modeling approach is useful in the context of the other motivating applications discussed earlier. For instance, in the context of the ride-sharing application, nodes correspond to different locations of a city. The outcome in a location depends on whether this location and the neighboring ones are treated or not.

[4] Throughout, we use the pronoun "he" to refer to the decision maker, and "she" to refer to a user.

[5] In fact, this structural property motivated the growing literature on community detection. For instance, Newman (2006) writes "One issue that has received a considerable amount of attention is the detection and characterization of community structure in networks, meaning the appearance of densely connected groups of vertices, with only sparser connections between groups."

ments, i.e., randomized assignments of *all* users in each community to either treatment or control. In this case, our problem boils down to *designing (negative) correlation* between communities in order to minimize the variance of the Horvitz-Thompson estimator against adversarial choices of the potential outcomes for each community.

In the special case of the above problem where community sizes are identical, we illustrate the *price of independence*: We show that randomly assigning half of the communities to treatment is optimal, whereas simply assigning treatment to each community independently is only a 4-approximation (as we illustrate through a simple example). The problem of obtaining the optimal (negative) correlation under the worst-case vector of outcomes can be formulated as a linear program. However, this problem appears intractable in general, as both the number of decision variables and the number of constraints are exponential in the number of communities. Even if we could solve for the optimal experiment, it turns out that it has another subtle drawback: it often involves complicated correlation structures in the assignments of treatments, making it possibly difficult to implement and interpret in practice. Motivated by this, we ask the following natural research question:

> *Can we design simple, computationally efficient, and interpretable randomized experiments that perform approximately well compared to the optimal experiment (or even are near or asymptotically optimal)?*

As our main technical contribution, we answer the above question in the affirmative. In particular, inspired by the special case of the optimal community-level assignment problem with identical community sizes, we develop a family of practical experiments that we refer to as *independent block randomization (IBR)* experiments.[6] Specifically, we partition communities into blocks so that each block contains communities of (approximately) similar sizes. We then assign the treatment variant to half of the communities in each block that are selected uniformly at random, and do so independently across blocks. Recall that doing so yields the optimal experiment in the aforementioned special case, but this is not necessarily the case in general. In fact, the suboptimality of the IBR experiments originates from two sources: (a) the loss due to the independence of assignments in different blocks, and (b) the loss from ignoring the community size differences within a block. The key idea behind our policies is to partition communities into blocks in a way that makes these losses as small as possible, and our analysis relies on showing that these losses can indeed be substantially

[6] We often drop the word "randomization" and simply refer to them as "independent block experiments" or "independent block policies".

reduced through careful choices of the partitions.

Since the blocks are treated independently, the worst-case variance of an IBR experiment is the sum of the worst-case variances for each block. We provide a full characterization of the worst-case potential outcomes (and hence the worst-case variance) within a block. This characterization gives us a handle to analyze the variance of this family of experiments. In addition, it enables us to solve for the optimal partitioning of the communities into blocks in a way that minimizes the variance, through the solution of a simple dynamic program. We then focus on analyzing the performance gap between the optimal IBR experiment and the optimal experiment for the community-level assignment problem. Specifically, we demonstrate and prove the following results:

1. We first (in Section 4.1) show that the optimal IBR experiment is a $\frac{7}{3}$-approximation for any problem instance. To this end, we consider a simple surrogate IBR experiment that groups every *fixed* k communities together first, and then solves for the optimal partition for each group separately to find the final blocks. By (i) bounding the variance of the optimal experiment from below in two separate ways—one in terms of the largest community and one in terms of the squared sum of all community sizes, and (ii) bounding from above the variance of the surrogate experiment in terms of similar terms; we obtain an approximation factor for this surrogate experiment as a function of k. The optimal choice of k in our analysis (which happens to be $k = 4$) provides the desired approximation ratio.

2. We then (in Section 4.2) focus on asymptotic analysis, and show that the IBR experiment with optimal partition is asymptotically optimal (among all experiments) in the regime with many communities and when no community dominates the rest (asymptotically) in terms of size. We also provide an example which shows that our "no dominating community" condition is necessary for any IBR experiment to be asymptotically optimal.

3. In Section 4.2.1, for some special cases of the asymptotic regime, we also obtain stronger results with a more careful analysis of the performance loss. Specifically, when the community sizes take values in a finite set that does not change as the problem scales, we show that the simple IBR experiment that places all the communities of exactly same size in the same block only increases the worst-case variance by an additive $O\left(\sqrt{n}\right)$ term compared to the optimal experiment.[7] If, in addition, the number of communities of each size is a fixed proportion of the total number of communities, this simple IBR experiment only increases the worst-case

[7]On the other hand, the worst-case variance of the optimal experiment scales linear in n.

variance by an additive $O(1)$ term, i.e., a constant amount.

4. Next, in Section 4.2.2, we introduce a simple member of the IBR family, which we refer to as the *logarithmic partition IBR* experiment. This simple experiment groups the communities into blocks by making sure that the ratio of the largest to the smallest community sizes in each block is upper-bounded by a given fixed constant $\eta > 1$. We then show that setting η appropriately is sufficient for such an IBR experiment to be asymptotically optimal, albeit at a slower convergence rate compared to the optimal IBR experiment. Notably, this logarithmic partitioning IBR experiment requires no explicit optimization to find its blocks, which is in contrast to the DP-based optimal IBR experiment.

5. In Section 5, we show that community-level experiments are essentially without loss of optimality when the network of users can be partitioned into disjoint communities by ignoring only a small number of connections, and users in the same community are densely connected. More formally, we show that by restricting the Horvitz-Thompson estimator to only *uncontaminated* users—i.e., those that have no cross-community connections—we obtain an approximate estimator to the true total market effect. Doing so only incurs a small bias, while the optimal IBR experiment for the uncontaminated users and their communities still is approximately and asymptotically optimal.

6. In Section 6, motivated by applications related to two-sided online marketplaces in which the experimental units (i.e., units that the decision maker assigns with treatment and control) and the analysis units (i.e., units generating the outcomes) are different, we consider a bipartite network model to capture the interference.[8] As it turns out, our results and analyses readily extend to this more general bipartite network setting.

7. Finally, in Section 7, we examine the performance of our IBR experiments on both synthetic examples and data-driven examples based on Facebook subnetworks. We demonstrate the following numerical performance guarantees for our (optimal) IBR experiment:

   (a) It improves the variance substantially relative to independent cluster/community-level randomization and improves upon other heuristic designs—both on average and in the worst case.

[8]Examples include e-commerce platforms that connect buyers and sellers (e.g., Amazon, eBay, and Facebook Marketplace), online ad-exchange platforms that connect publishers and advertisers (e.g., Google AdX), and online hospitality services that connect hosts and guests (e.g., Airbnb).

(b) In realistic instances, it performs substantially better than the performance guarantees provided by our theoretical results.

Furthermore, while the optimal experiment (without restriction to the IBR experiments) has a fairly complicated and counter-intuitive structure, our near-optimal IBR experiment admits a remarkably simple structure, which in turn makes our experiment easier to interpret and implement than the optimal experiment.

1.2 Practical Insights and Implications

The data-science pipeline in online marketplaces and social networks often deals with interference via clustering—which enables reducing bias and controlling variance. Our IBR experiment naturally complements the prevalent practice of using independent cluster-based randomization and other heuristics, and naturally integrates to this pipeline. Specifically, by introducing engineered negative correlation through our block structure, we manage to decrease the variance of the Horvitz-Thompson estimator substantially. While our robust design framework theoretically establishes this insight, our numerical simulations highlight that we achieve this goal in practical instances of our problem with community sizes and numbers extracted from data (see Section 7.2.2). Moreover, under less pessimistic average-case models of potential outcomes, we establish the same qualitative takeaways continue to hold (see Appendix D). We view the combination of theoretical results and data-driven numerical simulations as evidence for the practicality of our proposed design in applications of cluster-based experimental design over networks. Implementing our proposed design in practice would likely necessitate addressing additional practical challenges. We discuss some practically relevant considerations and associated challenges in Section 8.1.

1.3 Related Literature

Foundations of Experimental Design Experimental design has found far-reaching applications to guide decision making from medical trials to social sciences, and recently in online marketplaces and social networks. It is grounded in causal inference, but instead of inferring causality from purely observational data, a decision maker can choose how to gather data to gain more statistical efficiency and more convincing empirical evidence of causality. The process often involves randomization, and a more careful design of randomization based on optimization helps to maximize the statistical power. Great expositions of related topics include Owen (2020), Kohavi et al. (2020), and Imbens and Rubin (2015).

Applications in Networks A number of recent papers have studied experimental design with interference in social network settings; e.g., Eckles et al. (2016), Aronow et al. (2017), and Ugander and Yin (2020). Many of these proposed a graph-cluster-based randomization. We consider a similar problem and focus on networks that can be partitioned into several densely connected communities after ignoring a small number of connections. Such a network structure occurs in many applications, as discussed earlier in the introduction. For these networks, we develop a practical independent block experiment, and we formally analyze the performance and bound the gap relative to the optimal experiment of our model. Several other papers also considered similar experimental design problems on bipartite networks; e.g., Zigler and Papadogeorgou (2021), Pouget-Abadie et al. (2019), Doudchenko et al. (2020), and Harshaw et al. (2021). See Section 6 for the extension of our results and our approach to bipartite networks. Lastly, from a more practical viewpoint, Karrer et al. (2020), Rolnick et al. (2019), and Gui et al. (2015) demonstrated the implication of network experimental design for Facebook Stories and Jobs on Facebook, Google Search, and LinkedIn Feed recommendations, respectively, and shared valuable hands-on insights from practice.

Applications in Online Platforms Recent experimental design work has also considered interference in the context of markets. Johari et al. (2020) uses a mean-field model to investigate the biases of different experiments in the steady state of system dynamics, and suggests two-sided randomization to reduce bias. Wager and Xu (2021) uses a related mean-field analysis to study the effect of small changes of experimental interventions on marketplace equilibrium. Bojinov et al. (2020) and Glynn et al. (2020) study the optimal design of switchback experiments, in which a decision maker can alternate between two variants over time. Xiong et al. (2019) studies the experimental design with a staggered rollout constraint, i.e., the treatment can not be removed once implemented. There are also plentiful relevant empirical works: e.g., see Ostrovsky and Schwarz (2011) for internet advertising auctions, Blake and Coey (2014) for online auctions at eBay, Zhang et al. (2020) for price promotions in online retail platforms, Yu et al. (2020) for revealing delay information in ride-sharing platforms, and Holtz et al. (2020) and Holtz and Aral (2020) for price experiments in online hospitality platforms. Many of these have considered cluster/community level randomization or experiments that are similar in spirit.[9]

[9]For example, Ostrovsky and Schwarz (2011) and Blake and Coey (2014) considered keyword and auction level assignments, Yu et al. (2020) adopted queuing level assignments, and Holtz et al. (2020) and Holtz and Aral (2020) considered randomization over Airbnb listing clusters, to mitigate the interference among users.

Applications in Other Fields The clustering/community level randomization also has broad applications in many other fields; e.g., in political science, public health, medicine, and education. We refer to Imai et al. (2009) and the literature therein for a detailed discussion. Interestingly, Imai et al. (2009) recommends a "pair matching experiment" if the community-level randomization is used. This experiment pairs similar (in terms of size and/or related covariates) communities and randomly assigns one community from each pair to treatment. While the authors do not provide theoretical justification for this recommendation, they offer empirical evidence for the usefulness of this design based on comparison with other heuristic experiments. Our work, on the other hand, provides a theoretical foundation for the aforementioned design. Specifically, we show that a simple family of IBR experiments, which subsumes the aforementioned design, can achieve much of the benefit from the optimal (correlated) randomized assignment. Moreover, our optimal IBR experiment provably improves over the pair matching experiments. We also provide numerical evidence which further supports this claim and illustrates that the improvement is often quite large.

Robust Design Framework We adopt a robust design approach to experimental design (Berger 2013, Chapter 5) and in particular, study the problem of minimizing the variance of the estimator against the worst-case value of potential outcomes. Several recent works also use a similar approach. For example, Bojinov et al. (2020) studies a switchback experimental design problem. The authors restrict attention to experiments that partition a finite time horizon into slots, and assign treatment or control variants independently to each slot. With this restriction, the worst-case potential outcomes take the same extreme point of the uncertainty set, regardless of the experiment. Our problem allows for general joint assignment distributions. The worst-case potential outcomes depend on the specific correlation of the assignments, and thus are experiment-dependent. Harshaw et al. (2019) considers a similar robust design problem against potential outcomes. In their setting, each unit has a covariate that can potentially predict the potential outcomes. A decision maker solves for an optimal experiment to trade off covariate balancing and robustness. The authors assume that potential outcomes belong to an ℓ_2-ball, and show that the problem is equivalent to aligning eigenvectors of the resulting correlation matrix in desired directions. They further develop a randomized experiment based on the Gram-Schmidt walk algorithm.

Alternative Approaches to Experimental Design Other recent papers have considered models in which a covariate of a unit is correlated with the potential outcome in a certain way. The optimal experiment usually involves covariate balancing, such that the treatment and control groups are similar in terms of the covariates (Bertsimas et al. 2015, Bertsimas et al. 2019, Kallus 2018, Bhat et al. 2020 and Harshaw et al. 2019). We instead work with a model of potential outcomes with minimal assumptions.

The rich literature on online learning and multi-armed bandits (e.g., see Lattimore and Szepesvári 2020 and Slivkins 2019 for surveys) can also be viewed as a form of adaptive and sequential experimental design. There, the decision maker is allowed to switch between variants (i.e., arms), the system is assumed to be stationary and have rapid feedback (i.e., no carryover effect), and the objective is to find the best variant with minimum cumulative regret or number of trials. For example, see Hadad et al. (2019) and Bibaut et al. (2021). Finally, we point out that the experimental design problem of minimizing the variance of an (unbiased) estimator is also related to various variance-reduction methods in the simulation literature (e.g., Asmussen and Glynn 2007).

2 Optimal Experimental Design Problem

We first describe our interference model in Section 2.1. We then formalize the problem of designing the (robust) optimal experiment against worst-case outcomes in its general form. In Section 2.2, we introduce a closely related combinatorial optimization problem termed as the *optimal correlation design* problem, which itself might be of independent interest. This problem captures the essence of our framework and suggests *community-level (randomized) assignments* for running experiments over a network of users with loosely connected or disjoint communities. Later in Section 3 and Section 4, we provide algorithmic solutions to obtain (constant or near-optimal) approximations for this problem.

2.1 Problem Formulation

Notation For any positive integer n, we let $[n] = \{1, 2, \dots, n\}$ denote a sequence of integers starting from 1 and ending with n. For a subset $S \subseteq [n]$, we let $S^c = [n] \setminus S$ denote its complement. Finally, we let $\Re$ be the set of correlation matrices, i.e., matrices that are positive semi-definite with diagonal entries equal to one. The size of the correlation matrices will be clear from the context.

Model We consider the setting where a decision maker aims to design a randomized binary experiment over a network $G = (\mathcal{V}, \mathcal{E})$ of users, i.e., picking a randomized assignment of each user to either of the two possible variants: *treatment* (i.e., the variant "1") or *control* (i.e., the variant "0"). Each node $j \in \mathcal{V}$ of the network represents a user. Suppose the network has $m \in \mathbb{N}$ users.

Without loss of generality, let $\mathcal{V} = [m]$. There is an undirected edge between a pair $\{i, j\}$ of users if they interact with each other, i.e., assigning one user to either treatment or control can possibly impact the outcome of the other user.

Given an assignment of the users to treatment/control in the network G, the outcome of each user j will be determined based on the assignments of both the user and all of her neighbors $\mathcal{N}_j \triangleq \{i \in [m] : \{i, j\} \in \mathcal{E}\}$. In particular, for a user $j \in [m]$, we let $y_{j1} \in \mathbb{R}$ (and $y_{j0} \in \mathbb{R}$ respectively) be the potential outcome when both the user and all of her neighbors receive the treatment variant (the control variant respectively). Clearly, at most one of the potential outcomes y_{j1} and y_{j0} is observed for any user j under any assignment. Note that if for some user j, the nodes in $\{j\} \cup \mathcal{N}_j$ are not all assigned to the same variant, then no meaningful outcome will be associated with that user.[10] We consider the setting where these potential outcomes $\{y_{j1}\}_{j\in[m]}$ and $\{y_{j0}\}_{j\in[m]}$ are unknown to the decision maker before he designs the randomized experiment. Throughout the paper, we impose the following assumption:

Assumption 2.1 (*Uncertainty Sets of Potential Outcomes*)**.** The potential outcomes are deterministic, non-negative and bounded from above; without loss of generality, we let $y_{j1}, y_{j0} \in [0, 1]$ for all users $j \in [m]$.

In many applications, it is reasonable to assume that the potential outcomes are non-negative;[11] however, our methodology in this paper is general and can be applied to other outcome ranges by proper modifications. We comment further on this in Remark 2.2.

Objective In applications of binary experimental design (or A/B testing), the decision maker conducts randomized experiments to eventually select one of the two possible options for the entire network of users—each corresponding to either treatment or control.[12] In order to make an informed choice, the decision maker would like to estimate the *total market effect* τ, which is the difference between the sum of the outcomes when all users receive the treatment and when all users receive

[10] As mentioned earlier, and formalized below, we are interested in the estimation of the difference in total user outcomes when all users are exposed to the same variant. We make no assumptions on the user outcomes when the assignment of a user is partial in the sense described here. Thus, such assignments are not meaningful for the aforementioned estimation problem.

[11] For example, the number of completed rides during a certain time period by a ride-sharing platform, the number of clicks in an online advertisement setting, and the amount of revenue created from a certain marketplace intervention, etc., all need to be non-negative.

[12] This scenario is in particularly relevant when a platform plans not to personalize among its users, and hence everyone should either be eventually assigned to the treatment option or the control one.

the control. Formally, we have

$$\tau \triangleq \sum_{j\in[m]} y_{j1} - \sum_{j\in[m]} y_{j0}.$$

In presence of network interference, even designing an unbiased estimator for τ given a randomized experiment can be challenging. To overcome this challenge, we focus on the celebrated Horvitz-Thompson unbiased estimator (Horvitz and Thompson 1952) to estimate the total market effect. Specifically, for a user $j \in [m]$, let z_{j1} be a binary random variable such that $z_{j1} = 1$ if user j's outcome is the treatment outcome (i.e., user j and all her neighbors receive the treatment variant) and $z_{j1} = 0$ otherwise. Analogously, let z_{j0} be a binary random variable such that $z_{j0} = 1$ if user j's outcome is the control outcome and $z_{j0} = 0$ otherwise. Using these indicator variables, the Horvitz-Thompson estimator $\hat{\tau}$ of the total market effect τ is

$$\hat{\tau} \triangleq \sum_{j\in[m]} y_{j1} \frac{z_{j1}}{\mathbb{P}\big[z_{j1} = 1\big]} - \sum_{j\in[m]} y_{j0} \frac{z_{j0}}{\mathbb{P}\big[z_{j0} = 1\big]}. \tag{1}$$

By the linearity of expectations, $\mathbb{E}\big[\hat{\tau}\big] = \tau$; thus, the Horvitz-Thompson estimator is indeed an unbiased estimator.

Given the Horvitz-Thompson unbiased estimator, the objective of the decision maker is to design an experiment, i.e., a joint distribution for assigning treatment and control to users, that minimizes the variance of the above estimator, i.e,

$$\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right] = \mathbb{E}\left[\left(\hat{\tau} - \mathbb{E}[\hat{\tau}]\right)^2\right] = \mathbb{E}\left[(\hat{\tau} - \tau)^2\right].$$

Note that since the Horvitz-Thompson estimator is unbiased, minimizing the variance is equivalent to minimizing the mean square error of the estimator.

Worst-Case Outcomes and Min-Max Optimization The variance of the Horvitz-Thompson estimator τ depends both on the assignment distribution (i.e., the experiment) and the value of unknown potential outcomes. To minimize the variance given the uncertainty in outcomes, we follow a robust design approach. More specifically, we aim to design an experiment that minimizes the variance $\mathbb{V}\mathrm{ar}\left[\hat{\tau}\right]$ of the above estimator against an adversarial selection of the potential outcomes—which basically corresponds to the worst-case possible outcomes $y_{j0,}, y_{j1} \in [0, 1]$ for all $j \in [m]$ for a given experiment.

Since the treatment and control variants have the same range for potential outcomes (due to Assumption 2.1) and have symmetric roles in the expression for variance (i.e., swapping the assignment of treatment and control in any experiment does not change the variance under the worst-case outcome), we only focus on experiments that are indifferent between treatment and control throughout the paper. We state this property formally in Assumption 2.2. This mild assumption implies that users have an equal chance to receive the treatment or control. Moreover, as will be seen later, it simplifies the analysis and exposition.

Assumption 2.2 (*Treatment/Control Symmetry*)**.** For a given experiment P, for any subset $S \subseteq [m]$ of users, let $P(S)$ denote the probability that users in set S receive the treatment, whereas users not in set S receive the control.[13] We assume $P(S) = P(S^c)$ for any subset $S \subseteq [m]$.

Each experiment P can be described by a probability distribution $\{P(S)\}_{S \subseteq [m]}$ for randomized assignment of treatments over subsets of $[m]$ that satisfies Assumption 2.2. We denote the set of experiments that satisfy Assumption 2.2 by[14] $\mathcal{P} \subseteq \Delta\left(2^{[m]}\right)$. This set can be explicitly expressed as follows:

$$\mathcal{P} = \left\{ P \in [0,1]^{2^{[m]}} : \text{(i) } P(S) \geq 0,\, \forall\, S \subseteq [m]; \quad \text{(ii) } \sum_{S \subseteq [m]} P(S) = 1; \quad \text{(iii) } P(S) = P(S^c),\ \forall\, S \subseteq [m] \right\}.$$

The problem of finding the optimal experiment for minimizing the variance under the adversarial selection of potential outcomes is formulated as the following min-max optimization problem:

$$V^{\mathrm{OPT}} = \min_{P \in \mathcal{P}} \max_{y_{j0}, y_{j1} \in [0,1],\, j \in [m]} \mathbb{V}\mathrm{ar}\big[\hat{\tau}\big], \qquad \text{(Min-Max Optimization)}$$

where the variance $\mathbb{V}\mathrm{ar}\big[\hat{\tau}\big]$ of the estimator depends both on the treatment assignment distribution P and the value of unknown potential outcomes. Specifically, let $y_1 = (y_{j1})_{j \in [m]}$ and $y_0 = (y_{j0})_{j \in [m]}$ denote the vectors of potential treatment and control outcomes, and $\tilde{z}_1 = \big(z_{j1}/\mathbb{P}\big[z_{j1} = 1\big]\big)_{j \in [m]}$ and $\tilde{z}_0 = \big(z_{j0}/\mathbb{P}\big[z_{j0} = 1\big]\big)_{j \in [m]}$ denote the vectors of normalized indicators. We have

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\hat{\tau}\right] &= \mathbb{V}\mathrm{ar}\Big[\tilde{z}_1^{\mathrm{T}} y_1 - \tilde{z}_0^{\mathrm{T}} y_0\Big] = \mathbb{V}\mathrm{ar}\big[\tilde{z}_1^{\mathrm{T}} y_1\big] + \mathbb{V}\mathrm{ar}\big[\tilde{z}_0^{\mathrm{T}} y_0\big] - 2 \cdot \mathbb{C}\mathrm{ov}\big[\tilde{z}_1^{\mathrm{T}} y_1, \tilde{z}_0^{\mathrm{T}} y_0\big] \\ &= y_1^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_1 \tilde{z}_1^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big) y_1 + y_0^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_0 \tilde{z}_0^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big) y_0 - 2 \cdot y_1^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_1 \tilde{z}_0^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big) y_0. \end{aligned} \tag{2}$$

Note that for a given experiment P, the expression (2) for variance is a convex quadratic function of the outcomes $\{y_{j0}\}_{j \in [m]}$ and $\{y_{j1}\}_{j \in [m]}$. This (and other structural properties of this problem)

[13]While we use the shorthand notation $P(S)$ to denote the probability of treatment assignments to users in S, we reserve the notation $\mathbb{P}[\cdot]$ to denote the induced probability distribution.

[14]Throughout the paper, we use 2^S to denote the power set of S and $\Delta(S)$ to denote the probability simplex over S.

makes it computationally challenging to use the formulation in (Min-Max Optimization) to obtain the optimal experiment. Before we detail this point, we first introduce another closely related formulation, which also shares similar challenges.

2.2 Optimal Correlation Design for Community-Level Assignments

In many applications, the network of users can be decomposed into disjoint communities after ignoring only a small number of edges, and the users in the same community are densely connected. As mentioned earlier, a common practice is to partition the network into such communities and then assign all the users in a community to the same variant (where each variant has a marginal probability of $\frac{1}{2}$ due to Assumption 2.2). We first assume that the network consists of several disjoint communities and we study the optimal joint distribution of assignments at the community level. In Section 5, we show that such community-level assignments are nearly optimal when the network consists of densely-connected communities plus a few cross-community connections.

Let n be the number of communities. For each community $i \in [n]$, we let $S_i \subseteq [m]$ be the set of users and $w_i = |S_i|$ be the number of users in the community. For each community i, let z_i be a random binary variable such that $z_i = 1$ if all users in community i are assigned to treatment, and $z_i = 0$ if all users are assigned to control. A community-level assignment is specified by a joint distribution $\mathbb{P}[\cdot]$ of these Bernoulli random variables z_i with marginal probabilities being $\mathbb{P}[z_i = 1] = \mathbb{P}[z_i = 0] = \frac{1}{2}$ for each community.

We first show in Lemma 2.1 that for community-level experiments, the worst-case potential outcome is such that all users in the same community always take the same value.

Lemma 2.1 (*Structure of the Worst-Case Potential Outcomes*). *With any community-level experiment, the worst-case outcome is such that $y_{j1} = y_{j0} = \tilde{y}_i \in \{0, 1\}$ for all users $j \in S_i$. The corresponding variance of the Horvitz-Thompson estimator is*

$$\mathbb{V}\mathrm{ar}[\hat{\tau}] = 4 \sum_{i \in [n]} y_i^2 + 4 \sum_{i \in [n]} \sum_{[n] \ni k \neq i} \sigma_{ik} y_i y_k = 4 y^T \Sigma y,$$

where $y_i = w_i \tilde{y}_i \in \{0, w_i\}$ is the sum of potential outcomes over community i, $y = (y_i)_{i \in [n]} \in \mathbb{R}^n$ is the concatenation of these potential outcomes, σ_{ik} is the correlation coefficient of assignments z_i and z_k, and $\Sigma = (\sigma_{ik})_{i,k \in [n]} \in \Re$ is the correlation matrix of the assignments $\{z_i\}_{i \in [n]}$.

We prove Lemma 2.1 in Appendix A.1. Similar to before, a community-level experiment can be described by a probability distribution $\{P(S)\}_{S \subseteq [n]}$, where $P(S)$ is the probability of assigning all users in the set of communities $S \subseteq [n]$ to the treatment. The set of all joint community-level assignment distributions that satisfy Assumption 2.2, hereafter $\mathcal{P}_c \subseteq \Delta\left(2^{[n]}\right)$, is now given by:

$$\mathcal{P}_c = \left\{ P \in [0,1]^{2^{[n]}} : \text{(i) } P(S) \geq 0,\ \forall\, S \subseteq [n]; \quad \text{(ii) } \sum_{S \subseteq [n]} P(S) = 1; \quad \text{(iii) } P(S) = P(S^c),\ \forall\, S \subseteq [n] \right\}.$$

From Lemma 2.1, the optimization problem of obtaining the optimal community-level assignment can be written as

$$V_c^{\mathrm{OPT}} = \min_{P \in \mathcal{P}_c} \max_{y \in \times_{i \in [n]} [0, w_i]} 4 y^T \Sigma\big(P\big) y. \tag{3}$$

Here, $\Sigma\big(P\big)$ is the correlation matrix of community assignments under experiment P, and V_c^{OPT} denotes the worst-case variance of an optimal community-level experiment.

Since all potential outcomes are non-negative by Assumption 2.1, achieving a small objective value in (3) necessitates having a correlation matrix $\Sigma\big(P\big)$ with large (in absolute terms) negative off-diagonal entries. Thus, intuitively, (3) can be viewed as a problem of designing *optimal negative correlation* among communities' assignments. Given that the problem of achieving large negative correlations between pairs of random variables (through the appropriate choice of some decision variables) is quite natural, we suspect that our formulation and approach could be of interest in other settings as well.

Note that (Min-Max Optimization) and (3) are very similar in terms of their objectives and feasible regions. In general, these problems are challenging to solve. To see this, let us focus on (3):

(i) First, note that the inner problem is to maximize a quadratic convex function. Since the objective is not concave, off-the-shelf optimization algorithms do not guarantee achieving an optimal solution. Moreover, due to convexity, the maximum is always achieved at an extreme point of the feasible region satisfying $y_i \in \{0, w_i\}$ for all $i \in [n]$. One way to solve this problem is to evaluate the objective at the extreme points, which is computationally difficult due to the exponential number of extreme points (in the number of communities).

(ii) Second, the outer problem involves minimization over the joint distribution of binary assignments, and the number of decision variables is exponential in the number of communities as well. It may be possible to develop approximation algorithms that, e.g., rely on a semi-definite programming relaxation of the inner problem, and a relaxation of the outer problem to allow for any correlation matrix—not necessarily only to those achievable by a random binary assignment. But it remains unclear how to efficiently compute a joint binary assignment distribution from such a correlation matrix, even though the correlation matrix is indeed feasible.[15]

(iii) Third, (3) can be formulated as a linear program (see (9) in Appendix A.3). However, this program has an exponential number of decision variables and constraints in the number of communities: We have the constraints defining $\mathcal{P}_c$, plus we have one constraint for each extreme point of the potential outcomes' uncertainty set (to encode the worst-case objective). Hence, solving this linear program directly is not computationally tractable.[16]

Motivated by these challenges, in Section 3, we consider a family of practical experiments that we refer to as independent block experiments, and we show that (i) they are easy to compute and (ii) they admit provable performance guarantees. Before introducing this family formally, we conclude this section with two remarks. The first one is on the independence of the optimal community-level assignment from the specific interference model, and the second one is on the non-negative potential outcomes assumption.

Remark 2.1 (*Optimal Community-Level Assignment and the Specific Interference Model*)**.** We derived formulation (3) for finding the optimal community-level assignment for the specific interference model introduced in Section 2.1. On the other hand, once we focus on community-level assignments, the Horvitz-Thompson estimator's variance depends only on community sizes and the correlation matrix of community-level assignments. This is because, by fixing the community-level assignment, the joint distribution of the binary random variables $\{z_{j1}\}$ and $\{z_{j0}\}$ are determined.[17] Thus, by (1), the distribution (and hence variance) of the Horvitz-Thompson estimator is determined as well—regardless of the choice of the interference model. Since the specific interference pattern can be hard to infer in practice, the above fact is valuable and highlights the fact that our results (on approximately optimal community-level experiments) are not restricted to a particular interference

[15]Given a correlation matrix $\Sigma \in \Re$, one can obtain a heuristic random assignment based on the standard random hyperplane rounding ideas (Chapter 6 of Williamson and Shmoys 2011). This random assignment has a correlation matrix $\frac{2}{\pi}\arcsin(\Sigma) \succeq \frac{2}{\pi}\Sigma$, and can increase the worst-case variance considerably compared to the worst-case variance with the correlation matrix Σ.

[16]Similar comments also apply to (Min-Max Optimization). Note that this problem faces another challenge: it does not readily admit a linear programming formulation. This is because, unlike (3), in (Min-Max Optimization) some users can be assigned to treatment while their neighbors are assigned to control (or vice versa). Thus, one needs to also optimize over the marginal probabilities $q_j \triangleq \mathbb{P}\big[z_{j1} = 1\big] = \mathbb{P}\big[z_{j0} = 1\big]$ (where the latter equality holds by Assumption 2.2). If we fix the marginal probabilities, then the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ of the estimator is a linear function of the joint probabilities P. Hence, similar to (3), one can re-formulate (Min-Max Optimization) as a linear program with exponentially many decision variables and constraints. On the other hand, the optimal value of this linear program is in general not a convex function of the marginal probabilities q_j, and it is not clear how to tractably search for the best marginal probabilities. This makes (Min-Max Optimization) computationally even harder to solve.

[17]Recall that users in the same community always receive the same variant, and users of different communities do not interfere with each other.

model.[18]

Remark 2.2 (*Alternative Uncertainty Sets*). In this section, we restricted our attention to settings with non-negative potential outcomes (as captured by Assumption 2.1). Importantly, our approach is general and can be applied to other ranges of potential outcomes as well. It turns out that the choice of the range of potential outcomes qualitatively impacts the structure of the optimal experimental design. For instance, for the special case when the uncertainty set of each user j's potential outcomes is $y_{j1}, y_{j0} \in [-1, 1]$ (or more generally, when the uncertainty set is symmetric around zero), an optimal experiment solving (3) simply assigns each community independently to either treatment or control with probability 1/2 (see Appendix A.2). Interestingly, this simple experimental design seems to be optimal in very restrictive settings. When the potential outcomes do not belong to an interval symmetric around zero, e.g., as in the setting considered in this section, more intricate designs are needed to ensure low variance in the worst-case. We revisit this point and illustrate the sub-optimality of independent assignments in the absence of symmetry in the next section as well as Section 7.

3 Independent Block Randomization

In this section, we introduce our main family of experiments for community-level assignments, which attain approximate (or near-optimal) solutions to (3). To provide insights on the design of the experiments in this family, we first (in Section 3.1) discuss the *price of independence* in community-level assignments, i.e., the variance gap between the optimal solution to (3) and the community-level assignment that treats communities independently. Subsequently, in Section 3.2, we formally define our proposed family of experiments, which are defined through a partition of communities. We also provide an approach for efficiently computing the optimal partition. Finally, in Appendix E, we discuss a way to construct confidence intervals for an IBR experiment.

3.1 Price of Independence: Optimal Experiment with Equal-Size Communities

Before we introduce the family of IBR experiments, as a warm-up, we study the case where community sizes are equal. In this case, Proposition 3.1 shows that the optimal community-level experiment randomly assigns half of the communities to treatment.

[18] We still need the specific interference model of Section 2.1 to obtain our findings in Section 5 and establish that community-level experiments remain near-optimal within the set of all experiments.

Proposition 3.1 (*Optimal Experiments with Equal-Size Communities*). *Suppose there are n communities with equal sizes $\{w_i\}_{i=1}^n$, and without loss of generality, suppose that $w_i = 1$ for each community i. The optimal experiment solving (3) is as follows:*

1. *When n is even: The optimal experiment uniformly at random chooses $\frac{n}{2}$ communities and assigns them to treatment, while assigning the rest to control. The correlation coefficient of any two community assignments is $\sigma = -\frac{1}{n-1}$. In the worst-case potential outcome, $\frac{n}{2}$ communities take $y_i = 1$ and the other $\frac{n}{2}$ communities take $y_i = 0$. The worst-case variance of the estimator is $\frac{n^2}{n-1}$.*
2. *When n is odd: The optimal experiment uniformly at random chooses $\frac{n+1}{2}$ communities and assigns them to treatment with probability $\frac{1}{2}$, while assigning the remaining communities to control (or vice versa). The correlation coefficient of any two community assignments is $\sigma = -\frac{1}{n}$. In the worst-case potential outcome, $\frac{n+1}{2}$ communities take $y_i = 1$ and the other $\frac{n-1}{2}$ communities take $y_i = 0$. The worst-case variance of the estimator is $\frac{(n+1)^2}{n}$.*

We prove Proposition 3.1 in Appendix A.3. When the number of communities n is odd, we can add a dummy community with size zero to get a problem instance with an even number of communities. Then, the optimal experiment randomly assigns half of the communities to treatment, and the dummy community is selected with probability $\frac{1}{2}$. Importantly, this result not only characterizes the optimal experiment in this case, but also shows that there is an inherent gap between the variance of the policy that treats communities independently and the optimal experiment. Specifically, the worst-case variance of the optimal experiment asymptotically converges to n, while it is straightforward to see that the worst-case variance of the independent community-level assignment is exactly equal to $4n$. Hence, there is a multiplicative gap of 4 between the two, which characterizes the price of independence for community-level experiments.

3.2 Optimal Independent Blocks and Dynamic Programming

Proposition 3.1 shows that when communities have equal sizes, an optimal experiment treats communities in an identical way, and it minimizes the correlation between assignments to any two communities. This smallest (negative) correlation is achieved by uniformly at random assigning half of the communities to treatment. In general, when community sizes are different, although we would like the correlation to be small for any two communities, an optimal experiment will prioritize the negative correlation between some specific community pairs (e.g., when both communities are large) over some other pairs (when both communities are relatively small). It may even deliberately introduce positive correlation between some pairs of communities in order to attain larger negative correlation between other pairs. It is not clear what the optimal correlation among communities would look like or how to search for it in a computationally efficient way.

Overview of IBR Experiments Inspired by the case with equal community sizes, we consider a family of simple experiments which we refer to as *independent block randomization (IBR)* experiments. Specifically, in an IBR experiment, we first sort communities in decreasing order of size. We then partition them into blocks so that each block contains communities of similar sizes.[19] We then try to obtain assignments of any two communities in a block to treatment and control in a way that induces large negative correlation. To do this, we uniformly at random treat half of the communities in each block, and do so independently across blocks. Note that Proposition 3.1 implies that the aforementioned assignment attains the largest negative correlation among communities in a block, ignoring the difference in sizes. Thus, intuitively, this assignment ensures large negative correlation when the community sizes in a block are not too different. The independence of assignments across blocks, on the other hand, comes at the price of no correlation between communities of different blocks.

A careful design of the blocks mentioned above trades off between: (i) the benefit of a larger negative correlation within a block, and (ii) the cost of independence across blocks. Specifically, as a block contains more communities, more communities are negatively correlated with each other, but simultaneously, the correlation also becomes weaker (i.e., the magnitude of the correlation becomes smaller). Example 3.1 illustrates this point by focusing on two extreme cases. See also Figure 1 for an illustration of the structure of the correlation matrix of the community-level assignments generated by an IBR experiment. This structure shows how our IBR experiments control the induced negative correlation between different communities.

Example 3.1. If there is only one block that contains all communities, then the assignments of any two communities are negatively correlated, but the correlation is only $\Theta\left(-\frac{1}{n}\right)$. This turns to be optimal when the community sizes are equal by Proposition 3.1. If each block instead contains only two communities, then in the associated IBR experiment, the correlation between the assignments of these two communities is -1, which is the largest (in absolute value) possible. However, in this case, every community is only (negatively) correlated with the community in the same block, and is independent from the remaining $n-2$ communities.

[19]In fact, we sort *before* partitioning precisely because we would like to ensure similar sizes in the same block.

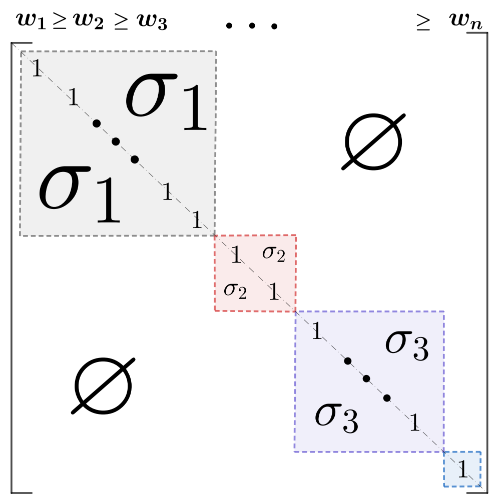


Figure 1: Example of the correlation matrix of an IBR experiment (with four blocks, where the last block is a single community). Note that for each block i, $\sigma_i < 0$ and $|\sigma_i|$ decreases as the size of the block increases.

We next show that the optimal partition of communities can be obtained through the solution of a simple dynamic program. Our approach builds on Lemma 3.2, which characterizes the worst-case potential outcome of a block in an IBR experiment.

Lemma 3.2 (*Worst-Case Potential Outcome of a Block*). *Consider a block with k communities sorted in decreasing order of sizes, i.e., $w_1 \geq w_2 \geq \cdots \geq w_k$. Let $\Sigma \in \Re$ denote the correlation matrix of assignments with off-diagonal entries equal to σ, and let p be the largest index such that $w_p \geq -2\sigma \sum_{i \leq p-1} w_i$. Let $y_i = w_i$ for any $i \leq p$, and $y_i = 0$ for any $i > p$. Then, $y = (y_i)_{i\in[k]}$ is a worst-case potential outcome of the block, i.e., it solves $\max_{y_i\in[0,w_i],\,\forall\, i\in[k]} 4y^{\mathrm{T}}\Sigma y$.*

We prove Lemma 3.2 in Appendix A.4. With an IBR experiment, the assignments to blocks are independent; thus, the worst-case variance is simply the sum of worst-case variances of different blocks. This fact and Lemma 3.2 together indicate that we can obtain an optimal community partition in polynomial time using dynamic programming.

Specifically, given the sorted list $w_1 \geq w_2 \geq \ldots \geq w_n$, we let $V_k(h)$ denote the continuation worst-case variance when there are $k \in [n]$ remaining communities $[n-k+1 : n] \triangleq \{n-k+1, \ldots, n\}$ to be partitioned and the next block contains $h \in [k]$ communities. Let $V_k = \min_{h\in[k]} V_k(h)$ be the worst-case variance with the optimal partition of these k communities. The Bellman equation is

$$V_k = \min_{h\in[k]} V_k(h) = \min_{h\in[k]} g_k(h) + V_{k-h}, \tag{4}$$

where $g_k(h)$ is the worst-case variance of the block that contains the first h of the remaining k

communities. Note that this quantity can be easily computed using Lemma 3.2. As a side note, let us point out that it is without loss of optimality (see Lemma B.2 in the Appendix) to focus on blocks that contain an even number of communities, except for the last block that contains the smallest communities (which has an odd number of communities when n is odd). This property halves the computational requirement for solving the DP. In the remainder of the paper, we let $V^{\text{DP}} \triangleq V_n$ denote the worst-case variance of the optimal IBR experiment.

We conclude this section with a remark that compares our independent block randomization experiments with stratified randomization experiments.

Remark 3.1 (*Comparison to Stratified Randomization*)**.** Stratified randomization experiments (Fisher 1935, Higgins et al. 2016) group (or stratify) the experimental units (usually based on their covariates) and then assign treatments to each group independently. We can interpret our independent block randomization experiment as stratified randomization in community sizes. However, the underlying reason for stratification is very different from conventional stratified randomization. Specifically, we stratify in community sizes not because we assume communities with similar sizes to have similar potential outcomes. Instead, we stratify in community sizes to minimize the worst-case variance of the estimator (see the analysis in Section 4). Actually, from Lemma 3.2, although communities in the same block have similar sizes, their potential outcomes take distinct values in the worst case: specifically, some potential outcomes take the extreme values w_i, whereas the rest potential outcomes take zero, which is the other extreme.

4 Performance Analysis of the IBR Experiments

In this section, we analyze the performance of our IBR experiments. When the community sizes are heterogeneous, the suboptimality of an IBR experiment comes from two sources:

(i) Assignments are independent across blocks; hence, we lose the opportunity to introduce negative correlation between communities of different blocks (which would yield lower total variance in the worst case outcome).

(ii) Within each block, the community sizes are not exactly the same, but we treat the communities in an identical way.

Our design of IBR experiments tries to mitigate both sources of suboptimality. In fact, our analysis relies on showing that these losses can indeed be substantially reduced through careful choices of the partitions. We consider both an approximation ratio analysis for any problem instance

(Section 4.1) and an asymptotic analysis when the number of communities is large (Section 4.2). In both cases, we show the worst-case variance only increases by a small amount compared to the worst-case variance V_c^{OPT} of an optimal community-level assignment.

Remark 4.1. Throughout this section, for notational convenience, we assume communities are sorted in decreasing order of sizes, i.e., $w_1 \geq w_2 \geq \cdots \geq w_n$.

In the analysis, instead of comparing the performance of IBR experiments to the optimal worst-case variance V_c^{OPT} directly, we compare it with a lower bound V^{LB} of (3). The lower bound is achieved by relaxing the outer problem to allow for any correlation matrix. More precisely, we have:

$$V^{\text{LB}} \triangleq \min_{\Sigma \in \Re} \max_{y \in \times_{i \in [n]}[0, w_i]} 4y^T \Sigma y \leq V_c^{\text{OPT}}. \tag{5}$$

The key difference of this problem from (3) is that in (5), Σ belongs to the set of all correlation matrices $\Re$, and it need not be attained by a joint binary assignment.[20] As a result of this relaxation, we have the following lemma.

Lemma 4.1 (*Relaxation of the Optimal Community-Level Experiment*)**.** *Suppose V_c^{OPT} and V^{LB} are the optimal objective values of the min-max optimizations* (3) *and* (5)*, respectively. Then, we have:*

$$V^{\text{LB}} \leq V_c^{\text{OPT}}.$$

4.1 Approximation Ratio Analysis

In this section, we show that a simple IBR experiment is a $\frac{7}{3}$-approximation for any problem instance, which in turn implies that the optimal IBR experiment from solving the DP in (4) is a $\frac{7}{3}$-approximation as well.

To start, we first consider a naive experiment that treats every community independently with probability $\frac{1}{2}$; this is equivalent to having one block for each community. The corresponding correlation matrix is $\Sigma = I$, and the worst-case variance is $4\sum_{i \in [n]} w_i^2$. We first provide a lower bound on V^{LB} and show this naive experiment is a 4-approximation.

Lemma 4.2 (*Approximation Ratio of the Independent Experiment*)**.** *The worst-case variance V_c^{OPT}*

[20] In fact, the set of all correlation matrices is equivalent to the set of symmetric positive semi-definite matrices with all diagonal entries being one—simply because any such matrix can be induced by a jointly Gaussian distribution. Note that this set is not a polyhedron. On the other hand, the set of correlation matrices $\Sigma(P)$ with randomized joint binary assignments $P \in \mathcal{P}_c$ is indeed a polyhedron.

of an optimal community-level experiment satisfies

$$\max\left\{4w_1^2, \sum_{i\in[n]} w_i^2\right\} \le V^{\text{LB}} \le V_c^{\text{OPT}} \le 4\sum_{i\in[n]} w_i^2.$$

The lower bound $4w_1^2 \le V^{\text{LB}}$ in Lemma 4.2 is trivial because $4y^{\text{T}}\Sigma y = 4w_1^2$ with the outcome vector y such that $y_1 = w_1$ and $y_i = 0$ for all $i \ge 2$, for any correlation matrix Σ. So, we only need to prove the inequality $\sum_{i\in[n]} w_i^2 \le V^{\text{LB}}$.

Proof of Lemma 4.2. Consider the following randomized potential outcomes with Y_i being either 0 or w_i with equal probability, and let these Y_i be independent. Then, $\mathbb{E}\left[Y_i^2\right] = w_i^2/2$ and $\mathbb{E}\left[Y_iY_k\right] = w_iw_k/4$ for any $i \neq k$. Let $Y = (Y_i)_{i\in[n]}$ be the concatenation of these randomized potential outcomes and $w = (w_i)_{i\in[n]}$ be the vector of community sizes. For any correlation matrix $\Sigma = (\sigma_{ik})_{i,k\in[n]} \in \Re$, we have

$$\max_{y\in\times_{i\in[n]}[0,w_i]} 4y^T\Sigma y \ge 4\mathbb{E}\Big[Y^{\text{T}}\Sigma Y\Big] = 2\sum_{i\in[n]} w_i^2 + \sum_{i\in[n]}\sum_{[n]\ni k\neq i} w_iw_k\sigma_{ik} = \sum_{i\in[n]} w_i^2 + w^{\text{T}}\Sigma w \ge \sum_{i\in[n]} w_i^2.$$

Thus, $V^{\text{LB}} \ge \sum_{i\in[n]} w_i^2$, which finishes the proof of the lemma. □

The lower bound $\sum_{i\in[n]} w_i^2 \le V^{\text{LB}}$ implies that the aforementioned naive policy is a 4-approximation. Note that the approximation ratio 4 is (asymptotically) tight. To see this, consider a problem instance with n communities where all community sizes are equal to one, and assume that n is even. Since $V_c^{\text{OPT}} = \frac{n^2}{n-1}$ from Proposition 3.1, we have $\frac{4\sum_{i\in[n]} w_i^2}{V_c^{\text{OPT}}} = \frac{4(n-1)}{n} \xrightarrow{n\to\infty} 4$.

We now use Lemma 4.2 as a building block to analyze a more advanced IBR experiment that groups every fixed k communities together, and then solves for the optimal partition for each group separately to find the final blocks. We call such an experiment a *k-partition IBR experiment*, and we let V^k denote the worst-case variance of this experiment.

More specifically, let $N = \left\lceil \frac{n}{k} \right\rceil$ be the number of groups. Every group $h \in [1 : N-1]$ contains exactly k communities $(h-1)k + i$ for $i = 1, 2, \ldots, k$, and the last group $h = N$ contains the remaining communities (hence, it can have fewer than k communities). Now we solve the DP for each group separately to obtain the optimal partition for the group, and we combine these partitions as the final partition in the k-partition IBR experiment. Here, we only consider k-partition IBR experiments for numbers $k \in \mathbb{N}$ that are even.[21] Lemma 4.3 bounds the approximation ratio of the

[21]We only consider even numbers for k, simply because by Lemma B.2 every block of an optimal partition contains an even number of communities.

k-partition IBR experiment for any even number k.

Lemma 4.3 (*Approximation Ratio of the k-Partition IBR Experiment*). *For an even number $k \in \mathbb{N}$, the worst-case variance V^k of a k-partition IBR experiment satisfies*

$$\frac{V^k}{V_c^{\mathrm{OPT}}} \leq \frac{V^k}{V^{\mathrm{LB}}} \leq \frac{k}{k-1} + k \cdot \frac{w_1^2}{V^{\mathrm{LB}}} \leq \frac{k}{k-1} + \frac{k}{4}.$$

Proof of Lemma 4.3. We start by finding an upper-bound on the the worst-case variance within a group. First, it is clear that the worst-case variance increases with the size w_i of any community i in the group because increasing w_i enlarges the potential outcomes' uncertainty set. Thus, for a group with at most an even number k of communities and the largest community size being equal to w, the worst-case variance is largest when there are exactly k communities and all community sizes are equal to w. From Proposition 3.1, the optimal partition is to have all the k communities in one block, and the corresponding worst-case variance is $\frac{k^2}{k-1} w^2$.

Next, let $N = \lceil \frac{n}{k} \rceil$ be the number of groups. From the above analysis, we have

$$V^k \leq \frac{k^2}{k-1} S_k, \tag{6}$$

where $S_k = \sum_{i=1}^N w_{(i-1)k+1}^2$ is the sum of squares of the largest community sizes in each group. We claim:

$$V^{\mathrm{LB}} \geq \sum_{i \in [n]} w_i^2 \geq k \cdot (S_k - w_1^2) + w_1^2. \tag{7}$$

The first inequality in (7) follows from Lemma 4.2. The second inequality in (7) holds, because

$$\begin{aligned}\sum_{i \in [n]} w_i^2 &\geq w_1^2 + \sum_{i=2}^{(N-1)k+1} w_i^2 = w_1^2 + \sum_{h=1}^{N-1} \sum_{i=(h-1)k+2}^{hk+1} w_i^2 \\ &\overset{(*)}{\geq} w_1^2 + k \cdot \sum_{h=1}^{N-1} w_{hk+1}^2 = w_1^2 + k \cdot (S_k - w_1^2),\end{aligned}$$

where the inequality $(*)$ above follows from the fact that communities $(h-1)k+2$ to hk have weakly larger sizes than the community $hk+1$. Now, by re-arranging the terms in (7), we have

$$S_k \leq \frac{V^{\mathrm{LB}}}{k} + \frac{k-1}{k} w_1^2.$$

Combining this with (6) yields:

$$\frac{V^k}{V^{\text{LB}}} \leq \frac{k^2}{k-1} \cdot \frac{S_k}{V^{\text{LB}}} \leq \frac{k}{k-1} + k \cdot \frac{w_1^2}{V^{\text{LB}}} \leq \frac{k}{k-1} + \frac{k}{4},$$

where the last inequality follows from $4w_1^2 \leq V^{\text{LB}}$ by Lemma 4.2. $\square$

Figure 2 illustrates how the upper bound in Lemma 4.3 changes as a function of k. It is easy to see that the minimum is achieved for $k = 4$, which yields $\frac{V^{k=4}}{V_c^{\text{OPT}}} \leq \frac{7}{3}$. As discussed earlier (and detailed in Lemma B.2 of the Appendix), in the optimal partition from solving the DP, all blocks will have an even number of communities (except for the last block that contains the smallest communities when n is odd). Thus, to obtain the optimal partition of a set of $k = 4$ communities (as required by the k-partition IBR experiment), we only need to compare the worst-case variances between two cases: (a) a block that contains all four communities, and (b) two blocks with one block containing the first two communities and the second one containing the last two communities. Thus, the 4-partition IBR experiment has a minimal computational requirement.

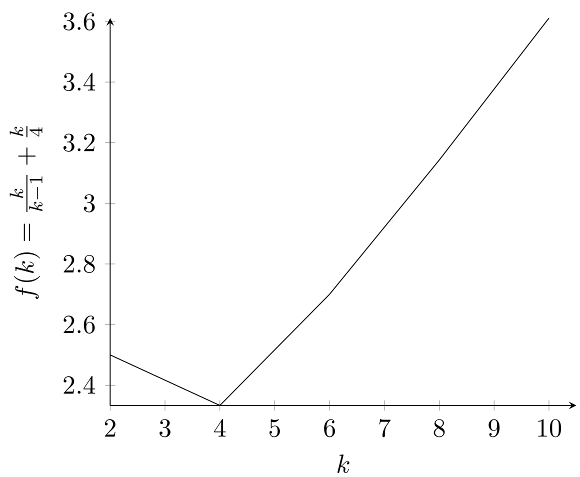


Figure 2: Approximation ratio of the k-partition IBR experiment; note that k only takes even integer values. The best ratio for such k is attained at $k = 4$ and it is equal to $f(4) = \frac{7}{3}$.

Lemma 4.3 directly implies that the optimal IBR experiment is $\frac{7}{3}$-approximation, as we formally state in Theorem 4.4.

Theorem 4.4 (*Approximation Ratio of the Optimal IBR Experiment*)**.** *The worst-case variance of the optimal IBR experiment from solving the DP satisfies*

$$\frac{V^{\text{DP}}}{V_c^{\text{OPT}}} \leq \frac{V^{\text{DP}}}{V^{\text{LB}}} \leq \frac{V^{k=4}}{V^{\text{LB}}} \leq \frac{7}{3}.$$

Remark 4.2 (*Theoretical vs. Numerical Performance Guarantees*)**.** In Section 7, we construct problem instances where the approximation ratio for the optimal IBR experiment cannot be better than approximately $1.5 < 7/3$. But we do not have an example where the approximation ratio is substantially closer to $7/3$. This gap can be partially attributed to the improvement of the optimal IBR experiment over the k-partition IBR experiment with $k = 4$. Furthermore, our numerical study in Section 7 demonstrates that the optimal IBR experiment often performs substantially better than the theoretical performance guarantee in Theorem 4.4. This is illustrated by focusing both on randomly generated examples and data-driven examples based on Facebook subnetworks.

Remark 4.3 (*Performance Guarantee for the Pair Matching Experiment*)**.** The pair matching experiment, or pair experiment for short, is one of the commonly used heuristics for community level randomization (Imai et al. 2009). In our setting, the pair experiment pairs communities of similar sizes together and randomly assigns one community from each pair to treatment; this is equivalent to an IBR experiment with each block containing two communities. The correlation matrix of the assignments is a block diagonal matrix with each block being 2×2 dimension, the diagonal entries being 1, and off-diagonal entries being -1. From Lemma 4.3 with $k = 2$, the approximation ratio of the pair experiment is 2.5, which is slightly larger than the $\frac{7}{3}$ approximation ratio guarantee for the optimal IBR experiment. On the other hand, for the special case of equal community sizes (and assuming n is even), the worst-case variance of the pair experiment is $2n$. Thus, Proposition 3.1 implies that the multiplicative gap is 2 relative of the optimal (IBR) experiment. The pair experiment is also not asymptotically optimal in this case as n grows large, whereas the optimal IBR experiment is asymptotically optimal under mild regularity conditions on community sizes, as discussed next.

4.2 Asymptotic Optimality

In our asymptotic analysis, we consider the regime where the number of communities n grows to infinity and there is no dominating community, in the sense that $w_1^2 = o\big(\sum_{i\in[n]} w_i^2\big)$. Since $V_c^{\text{OPT}} = \Theta\big(\sum_{i\in[n]} w_i^2\big)$ by Lemma 4.2, this intuitively means that no community is large enough to have a substantial effect on the variance of the estimator.

In this regime, Lemma 4.3 immediately implies that any k-partition IBR experiment with $k = \Theta\left(\sqrt{\frac{\sum_{i\in[n]} w_i^2}{w_1^2}}\right)$ is asymptotically optimal. To see this, note that from Lemma 4.2 and Lemma 4.3

we have

$$\frac{V^{\mathrm{DP}}-V^{\mathrm{LB}}}{V^{\mathrm{LB}}} \le \frac{V^{k}-V^{\mathrm{LB}}}{V^{\mathrm{LB}}} \le \frac{1}{k-1} + k\cdot \frac{w_1^2}{\sum_{i\in[n]} w_i^2} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0,$$

where the last equality is attained, e.g., if we let k be the even number that is closest to $\sqrt{\frac{\sum_{i\in[n]} w_i^2}{w_1^2}}$. Since the optimal IBR experiment achieves a smaller variance than the k-partition IBR experiment, it is asymptotically optimal as well, as we establish next.

Theorem 4.5 (*Asymptotic Performance of the Optimal IBR Experiment*). *The optimal IBR experiment is asymptotically optimal when the number of communities n grows large and $w_1^2 = o\big(\sum_{i\in[n]} w_i^2\big)$. Moreover, the convergence rate satisfies*

$$\frac{V^{\mathrm{DP}}-V^{\mathrm{LB}}}{V^{\mathrm{LB}}} = O\left(\sqrt{\frac{w_1^2}{\sum_{i\in[n]} w_i^2}}\right) \to 0.$$

We next illustrate by Example 4.1 that the aforementioned "no dominating community" condition is also necessary for any IBR experiment to be asymptotically optimal.

Example 4.1. Consider a problem instance where the community sizes form a geometric sequence, i.e., $w_i = \beta^{n-i}$ for community $i \in [n]$, and we let $\beta = \frac{5}{4}$. It can be shown that in the optimal partition, all blocks contain four communities. More precisely, supposing n is divisible by four, every block h contains communities $4h-3$ to $4h$ for $1 \le h \le n/4$. Note that each block essentially contains the same communities up to a scaling. Since the optimal IBR experiment randomly assigns half of the communities in a block to treatment, it increases the worst-case variance of each block by a constant fraction of 10.7% compared to an experiment that assigns treatment to communities in a block in an optimal way, and does so independently across blocks. Thus, the optimal IBR experiment cannot be not asymptotically optimal as n grows large. The no dominating community condition is violated because $\sum_{i=1}^{n} w_i^2 = \frac{\beta^{2n}-1}{\beta^2-1}$ and $w_1^2 = \beta^{2n-2}$ have the same order. We provide further details in Appendix A.5.

4.2.1 Asymptotic Optimality in the High-Multiplicity Model

The asymptotic optimality in Theorem 4.5 is for a general setting, and we can obtain a stronger result under additional assumptions. Specifically, for a high-multiplicity model where community sizes take values in a finite set and the number of communities of each size is a fixed proportion of the total number of communities n, a simple IBR experiment that has one block for each community size only increases the worst-case variance by an absolute constant that is independent of n, as we show in Theorem 4.6.

Theorem 4.6 (*Improved Performance Guarantee in the High-Multiplicity Model*). *Suppose community sizes only take K finite values $\{w_i\}_{i\in[K]}$, where $w_1 \geq w_2 \geq \cdots \geq w_K$. Let n_k be the number of communities of size w_k, and consider a simple IBR experiment that has one block for each community size. The worst-case variance, V, of this experiment satisfies*

$$V - V^{\text{LB}} \leq \sum_{k\in[K]} w_k^2 \cdot \left(5 + \min\left\{n_k, \frac{w_1^2 n}{w_k^2 n_k}\right\}\right),$$

where the right-hand side scales with the total number of communities n at most at a square-root rate (whereas, the lower bound satisfies $V^{\text{LB}} = \Theta(n)$).

If, in addition, each $n_k = \alpha_k n$ is a fixed proportion $\alpha_k \in (0,1)$ of n, then

$$V - V^{\text{LB}} \leq \sum_{k\in[K]} \left(5w_k^2 + \frac{w_1^2}{\alpha_k}\right),$$

which is a constant independent of n.

We prove Theorem 4.6 in Appendix A.6. When each block contains communities of equal sizes, there is no loss from ignoring the difference of community sizes in a block (i.e., the second source of the performance loss). Thus, Theorem 4.6 indicates that the loss from independent assignments across blocks (i.e., the first source of the performance loss) can be made small (and in fact asymptotically negligible) with an IBR experiment. Intuitively, if an experiment is close to the optimal experiment, the worst-case potential outcomes of the two experiments are close as well. Note that we have a complete characterization of the worst-case potential outcome with an IBR experiment (Proposition 3.1 and Lemma 3.2). In our proof, we bound the worst-case variance gap between the simple IBR experiment and the optimal experiment by considering both experiments against the worst-case potential outcome of the simple IBR experiment, and by using the fact that the correlation matrix of any experiment is positive semi-definite.

4.2.2 Simplicity vs. Complexity: the IBR Experiment with a Logarithmic Partition

Although the k-partition IBR experiment with $k = \Theta\left(\sqrt{\sum_{i\in[n]} w_i^2/w_1^2}\right)$ is asymptotically optimal under the no dominating community condition, to compute its partition, we need to solve a DP for every group that contains $k = \Omega(1)$ communities. We next show that a very simple logarithmic partition that requires no explicit optimization is sufficient for an IBR experiment to be asymptotically optimal. However, the convergence rate of the corresponding experiment could be slower and we require a slightly stronger version of the no dominating community condition.

Let us start by introducing the slightly stronger version of the no dominating community condition needed for our analysis of the aforementioned experiment: $\frac{\sum_{i=1}^{n} w_i^2}{w_1^2} = \Omega\big(n^c\big)$ for some constant $c > 0$. We consider a simple logarithmic partition in which the ratio of the largest to the smallest community sizes is upper-bounded across blocks. In particular, we define two parameters $\delta_1, \delta_2 \in (0, 1)$ to be specified later. We first include all communities with sizes $w_i \leq \bar{w} \triangleq \sqrt{\frac{\sum_{i=1}^{n} w_i^2}{n^{1+\delta_1}}}$ in one block, and label this as block zero. The communities in this block are small enough to have little effect on the variance of the estimator. We then iteratively go through the remaining communities in decreasing order of sizes to create blocks. Specifically, in each step we focus on the communities that are not yet assigned to a block, pick the largest one, and include all of the communities whose sizes are at least $\frac{1}{\alpha}$ times the size of this community in a block. Here, α is a parameter given by $\alpha = 1 + n^{-\delta_2}$. We label these blocks from one to K, in decreasing order of the block's largest community size. Theorem 4.7 shows that such logarithmic partition induces an asymptotically optimal IBR experiment when δ_1 and δ_2 are chosen properly.

Theorem 4.7 (*Asymptotic Performance of the IBR Experiment with a Logarithmic Partition*)**.** *Suppose* $\sum_{i=1}^{n} w_i^2 / w_1^2 = \Omega(n^c)$ *with some constant* $c > 0$*. Consider the above logarithmic partition with parameters* $\delta_1 = \delta_2 = c/4$*, and let* V *denote the worst-case variance of the logarithmic partitioning IBR experiment. This logarithmic partitioning IBR experiment is asymptotically optimal, and the convergence rate satisfies*

$$\frac{V - V^{\text{LB}}}{V^{\text{LB}}} = O\left(n^{-\frac{c}{4}} \ln n\right).$$

We prove Theorem 4.7 in Appendix A.7. Because of the specific partition we chose, communities in the same block have similar sizes (the ratio of sizes of the largest to the smallest communities in a block is at most α, which goes to one as n grows large). This makes the second source of performance loss small. In our proof, we consider a perturbed problem where we decrease all community sizes in a block to the minimum community size of the block. This is only a small perturbation and does not change the worst-case variance of an experiment much, as communities in a block are similar in size. After the perturbation, communities in a block have equal sizes, and we can adopt the analysis of Theorem 4.6.

5 Optimality of Community-Level Assignments

In the earlier sections, we conducted our analysis under two assumptions. First, we assumed that the users of a community are connected only to other users in the same community and there are no connections between communities. Second, we restricted attention to community-level experiments, which always assign all users in the same community to the same variant. Under the first assumption, such experiments ensure that each user either exhibits the treatment or the control outcome (i.e., she and her neighbors are all either assigned to treatment or to control), which in turn ensures that each user always provides a useful outcome for the Horvitz-Thompson estimator. We developed a family of IBR experiments whose variance approximates that of the optimal community-level experiment.

In this section, we relax the two assumptions of the earlier analysis highlighted above. Specifically, in Section 5.1, we still focus on settings where the network has the aforementioned community structure and allow the decision maker to use general experimental designs (that do not require community-level assignments). We show that under a mild regularity condition (which holds, e.g., when the communities are densely connected), our results on the optimal IBR experiment carry over.

Then, in Section 5.2, we relax the other assumption as well and allow for connections among different communities. Intuitively, the results should extend if there is a small number of connections between communities. In Section 5.2, we make this intuition precise. In particular, we assume that the network of users can be partitioned into multiple communities by dropping only a small number of connections, and each community satisfies the earlier regularity condition. In this case, we show that community-level experiments are close to optimal among all possible experiments.

5.1 More general designs

In this section, we assume that the network is partitioned into n disjoint communities with sets of users $\{S_i\}_{i\in[n]}$. For a given experiment and community $i \in [n]$, we let

$$\hat{\tau}_i = \sum_{j\in S_i} y_{j1}\frac{z_{j1}}{\mathbb{P}[z_{j1}=1]} - \sum_{j\in S_i} y_{j0}\frac{z_{j0}}{\mathbb{P}[z_{j0}=1]}$$

denote the Horvitz-Thompson estimator *restricted to community* i. We conduct our analysis under the following regularity condition:

Assumption 5.1 (*Optimality under Identical Outcomes*). For each community $i \in [n]$, when all users have identical outcomes (e.g., $y_{j1} = y_{j0} = 1$ for all $j \in S_i$), assigning all users to the same variant and to each variant with probability $\frac{1}{2}$, minimizes the variance $\mathbb{V}\mathrm{ar}\big[\hat{\tau}_i\big]$.

Note that this assumption exclusively focuses on cases where the users' outcomes are identical. In this case, it requires the Horvitz-Thompson estimator restricted to each community i to have the lowest variance when the decision maker relies on community-level randomization. It can be readily seen that the corresponding variance is $\mathbb{V}\mathrm{ar}\big[\hat{\tau}_i\big] = 4w_i^2$. It is worth noting that this assumption is silent on which design ensures low variance when users have different outcomes (which could be the case, e.g., for the worst-case outcomes associated with a given design).

We conduct our analysis under this abstract assumption, as it is sufficient for our subsequent results. That said, interestingly, this assumption holds for important general settings. One such setting is the case of densely connected communities.

Definition 5.1 (*Densely Connected Community*)**.** A community i is *densely connected* if any two nodes in the community are either connected or have a common neighbor.

Lemma 5.1. *Assumption 5.1 holds if all communities are densely connected.*

We prove Lemma 5.1 in Appendix A.8. Suppose that Assumption 5.1 holds. We next establish that V^{LB} from (5) is also a lower bound on the worst-case variance V^{OPT} of the optimal experiment (that is not restricted to community-level assignments – see (Min-Max Optimization)).

Lemma 5.2. *If Assumption 5.1 holds, then $V^{\text{LB}} \leq V^{\text{OPT}}$.*

We prove Lemma 5.2 in Appendix A.9. The critical property which we exploit for proving our results in Section 4 is the fact that V^{LB} is a lower bound on the variance V_c^{OPT} of the optimal community-level assignment. Lemma 5.2 establishes that V^{LB} is also a lower bound on the variance V^{OPT} of the optimal experiment (when communities are densely connected), thereby implying that our results in Section 4 readily extend (after replacing V^{OPT} with V_c^{OPT} in our results and analyses, and following an identical approach). Specifically, Theorems 4.4–4.7 continue to hold (after replacing V^{OPT} with V_c^{OPT}), indicating that the optimal IBR experiment is a 7/3-approximation of the optimal experiment obtained from solving (Min-Max Optimization), and it enjoys various asymptotic optimality properties (again relative to the optimal experiment from (Min-Max Optimization)).

5.2 Incorporating a Few Cross-Community Connections

We next allow for a small number of connections between communities. Specifically, suppose we can partition the network of users into communities $\{S_i\}_{i\in[n]}$, such that only a small number $\delta \ll m$ of users have connections to some other users in a different community. We call these users the contaminated users. Let $\bar{S}_i \subseteq S_i$ denote the set of uncontaminated users in community i, and $U = \cup_{i\in[n]}\bar{S}_i$ the set of uncontaminated users in the network. Let

$$\hat{\tau}' = \sum_{j\in U} y_{j1}\frac{z_{j1}}{\mathbb{P}[z_{j1}=1]} - \sum_{j\in U} y_{j0}\frac{z_{j0}}{\mathbb{P}[z_{j0}=1]}$$

be the Horvitz-Thompson estimator of the total market effect restricted to the uncontaminated users. In what follows, we focus on $\hat{\tau}'$ as an approximate estimator of the true total market effect τ. Note that this estimator is not unbiased. The bias is due to the contaminated users and satisfies $|\mathbb{E}[\hat{\tau}'] - \tau| \leq \delta$. In most interesting cases, the total market effect would scale with the size of the market, and hence $\tau = \Theta(m)$.[22] Since $\delta \ll m$, it follows that in such cases the bias is small relative to the total market effect.[23]

We next focus on the (robust) experimental design problem to minimize the variance $\mathbb{V}\mathrm{ar}\,[\hat{\tau}']$, and argue that the IBR experiment is near-optimal under a variant of Assumption 5.1. To that end, we consider a new problem instance where we drop all of the contaminated users. Then, the original network of users decomposes into multiple communities, and the size of each "residual" community i (of uncontaminated users) is $w_i = |\bar{S}_i|$, i.e., only the uncontaminated users in S_i remain. We require Assumption 5.1 to hold for residual communities (i.e., the assumption holds after replacing S_i with $\bar{S}_i$).

Fix a community-level experiment in the original problem. Since the Horvitz-Thompson estimator $\hat{\tau}'$ is restricted to the uncontaminated users, the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ of this community-level experiment in the original problem is the same as the the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ of the same experiment in the new problem, for any potential outcomes.[24]

Recall that Assumption 5.1 holds for residual communities. Thus, following an identical approach to the proof of Lemma 5.2, it follows that V^{LB} from solving (5) for the new problem remains

[22] For instance, this is the case when the difference between the treatment and control outcome has the same sign for all users, and is bounded away from zero.

[23] On the other hand, if $\tau = O(\delta)$, the difference of the treatment and control variants on total user outcomes is essentially negligible.

[24] To apply the same experiment to the new problem, we simply keep the previous randomized assignments for the uncontaminated users, and ignore the assignments of the contaminated users.

to be a lower bound on the worst-case variance $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ for any experiment in the original problem. This observation implies that both Lemma 5.2 and the subsequent discussion readily extend to this setting. In particular, the optimal IBR experiment that uses the number of uncontaminated users in each community as component sizes is approximately and asymptotically optimal in terms of the worst-case variance $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ in the original problem. These observations readily imply the following result:[25]

Proposition 5.3. *Suppose that Assumption 5.1 holds for residual communities. The Horvitz-Thompson estimator $\hat{\tau}'$ that is restricted to the uncontaminated users has a bias of at most δ. Consider the community-level experiment that minimizes the $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ in the worst case, and the (DP or logarithmic partition-based) IBR experiments obtained by focusing on the residual communities. These experiments satisfy Theorems 4.4–4.7. In particular, the IBR experiment is both approximately optimal (with an approximation factor of $7/3$) and asymptotically optimal.*

This result, in particular, implies that if the network of users consists of multiple densely-connected communities plus only a small number of connections between them, we may use the Horvitz-Thompson estimator $\hat{\tau}'$ restricted to uncontaminated users as an approximate estimator to the true total market effect τ. Doing so only incurs a small bias (bounded by δ). Moreover, the optimal IBR experiment for the uncontaminated users is still approximately and asymptotically optimal (compared to the optimal experiment that minimizes the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}']$ against the worst-case potential outcomes).

6 Extensions to Bipartite Networks

So far we have focused on a network model, where a node represents an experimental unit, edges capture the interference patterns, and the decision maker assigns treatment and control to each node to measure the total impact of the treatment on all nodes. In many applications, especially applications related to two-sided marketplaces, the experimental units (i.e., units to which the decision maker assigns treatment and control) and the analysis units (i.e., units whose outcomes the decision maker cares about) are different. In such cases, it may be more appropriate to consider a bipartite network model that has the experimental units on one side and the analysis units on the other. For example, a short-term lodging platform (e.g., Airbnb) may examine the effect of different displays of host listings (the experimental units) on the number of consumers who rent properties

[25] Note that this proposition readily implies the discussion at the end of the previous subsection (with $\delta = 0$). We did not state the earlier result formally to prevent repetition.

(the analysis units); or an online advertising platform (e.g., Google Ads) may care about how providing different amount of information about publishers' inventories (the experimental units) affects the bids of the advertisers (the analysis units). In this section, we show that our earlier results and analyses readily extend to such a bipartite network setting.

The Model We focus on a bipartite network $G = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, where $\mathcal{U}$ denotes the set of experimental units, $\mathcal{V}$ the set of analysis units, and $\mathcal{E}$ the connections between experimental and analysis units. For ease of exposition, in the remainder of this section, we refer to the experimental units as resources and the analysis units as users. Suppose the network has $m \in \mathbb{N}$ users, and without loss of generality, let $\mathcal{V} = [m]$. There is an (undirected) edge between a resource i and a user j if resource i's assignment to treatment/control possibly impacts user j's outcome.[26]

The decision maker aims to design a randomized binary experiment over the set of the resources. Given an assignment of the resources to treatment/control, the outcome of each user j will be determined by the assignments of her neighboring resources $\mathcal{N}_j \triangleq \{i \in \mathcal{U} : \{i, j\} \in \mathcal{E}\}$. In particular, for a user $j \in [m]$, we let $y_{j1} \in \mathbb{R}$ ($y_{j0} \in \mathbb{R}$) be the potential outcome when all of her neighboring resources receive the treatment (control) variant. If for some user j, the resources in $\mathcal{N}_j$ are not all assigned to the same variant, then no meaningful outcome is associated with that user (as in our model in Section 2.1). Finally, we assume that the potential outcomes are deterministic and non-negative as in Assumption 2.1, and we restrict attention to experiments that are symmetric between treatment and control as in Assumption 2.2.

Remark 6.1. Note that the network model studied earlier is a special case of the bipartite network model presented here. This is because, given the network model G of Section 2.1, we can formulate an equivalent bipartite network model as follows: both the set of resources and set of users of the bipartite network model are identical copies of the set of users in network G, and there is an edge between a resource i and a user j in the bipartite network model if and only if $i = j$ or users i and j are neighbors in G.

Experimental Design Problem The decision maker would like to estimate the total market effect τ given by:

$$\tau \triangleq \sum_{j \in [m]} y_{j1} - \sum_{j \in [m]} y_{j0},$$

[26]For example, in the short-term lodging platform example, an edge between a host i (a resource) and a user j might mean that host i fits user j's needs (e.g., in terms of the location of her property or number of bedrooms). In this case, host i is in user j's consideration set, and hence her treatment may change the outcome of the user.

which is the difference between the sum of the outcomes when all resources receive the treatment and when all resources receive the control. We assume that he uses the unbiased Horvitz-Thompson estimator $\hat{\tau}$ to estimate the total market effect τ:

$$\hat{\tau} \triangleq \sum_{j\in[m]} y_{j1} \frac{z_{j1}}{\mathbb{P}\big[z_{j1} = 1\big]} - \sum_{j\in[m]} y_{j0} \frac{z_{j0}}{\mathbb{P}\big[z_{j0} = 1\big]},$$

where, as before, z_{j1} is a binary random variable such that $z_{j1} = 1$ if user $j \in [m]$ has the treatment outcome and $z_{j1} = 0$ otherwise. Analogously, z_{j0} is a binary random variable which is one if user j has the control outcome.

The decision maker's objective is to design an experiment to minimize the variance of the Horvitz-Thompson unbiased estimator, against the worst-case value of the unknown potential outcomes. The optimization problem is almost identical to (Min-Max Optimization), with one minor difference: the set $\mathcal{P}$ of decision variables now represents the set of distributions of joint *assignments of resources* to treatment/control that satisfy Assumption 2.2.

Optimal Community-Level Assignments We first consider the case where the bipartite network consists of several disjoint subgraphs (each of which are bipartite). For consistency with earlier sections, we refer to these subgraphs as communities. The size w_i of community i is now the number of *users* in the community. For a graph that consists of several disjoint communities, we can again consider assignments at the community-level, i.e., assigning all resources in a community to the same variant with equal probability. For such experiments, Lemma 2.1 still characterizes the structure of the worst-case potential outcomes (with the definition of w_i given above). Hence, the optimization problem to solve the optimal community-level assignment is still given by (3).

Given the identical formulation, all our analyses and results in Section 4 readily extend to the bipartite network setting – the only difference is the definition of community sizes $\{w_i\}$. Specifically, the optimal IBR experiment is a $\frac{7}{3}$-approximation and enjoys various asymptotic optimality properties compared to the optimal community-level assignment.

Incorporating a Few Cross-Community Connections Suppose that the bipartite network can be partitioned into disjoint communities after ignoring only a small number of edges.[27] Suppose

[27] In the context of the applications highlighted earlier, this structural property would hold if the market is a collection of submarkets that share a few resources/users.

further that Assumption 5.1 holds after adapting it to the bipartite setting.[28] Analogous to Lemma 5.1 (and Definition 5.1), a sufficient condition for this assumption to hold is for all communities to be densely connected, i.e., any two users in the community share a common resource.

In this case, it can be readily checked that our arguments in Section 5 go through (again with the new definition of $\{w_i\}$) and all of our results in Section 5 continue to hold. Specifically, let δ be the number of contaminated users (i.e., users that connect to a resource from a different community) under the partition. A Horvitz-Thompson estimator restricted to the uncontaminated users incurs a bias of at most δ, which is small relative to the total network effect when the number of cross-community connections is small. Focus on minimizing the variance of this restricted Horvitz-Thompson estimator against the worst-case potential outcomes. The optimal IBR experiment is again approximately and asymptotically optimal compared to the optimal experiment from solving (Min-Max Optimization) (under the aforementioned variant of Assumption 5.1).

7 Numerical Examples

In this section, we examine the performance of our IBR experiments on three numerical examples: a synthetic example with randomly generated instances (Section 7.1) and two data-driven examples based on real Facebook subnetworks (Section 7.2). Moreover, we compare its performance with heuristic experiments such as (i) independent community-level randomization, (ii) the experiment that uniformly at random picks half of the communities and assigns them to treatment, and (iii) the pair matching experiment (Imai et al. 2009, Holtz and Aral 2020, Ugander and Yin 2020). We illustrate that the worst-case variance of our IBR experiment is small compared to these heuristic experiments and is close to the worst-case variance of the optimal community-level experiment.

In Appendix D, as opposed to focusing on the worst case, we assume that potential outcomes are random draws from a given distribution, and numerically compare the "average case" performances of these experiments. Our results indicate that the IBR experiment still reduces the variance substantially relative to independent community-level randomization and improves upon other heuristic experiments. We also compare the standard deviation of the estimator against the true total market effect in this setting, and show the statistical power of the Horvitz-Thompson unbiased estimator for total market effect, when it is coupled with our proposed IBR experiment.

[28] That is, in the setting of the assumption assigning all *resources* to the same variant with probability $1/2$ minimizes the variance.

7.1 Synthetic Examples with a Small Number of Communities

We first consider randomly generated instances of the community-level assignment problem with number of communities $n \in \{6, 8, 10, 12\}$. For each fixed n, we generate $i.i.d.$ community sizes, each uniformly at random between 1 and 100, i.e., $w_i \sim \text{Unif}\{1, \cdots, 100\}$. We create 10^4 samples for Monte-Carlo simulation. For each sample, we calculate: (a) the worst-case variance V^{DP} of the optimal IBR experiment (by solving the DP of Section 3), and (b) the worst-case variance V_c^{OPT} of the optimal community-level experiment (by solving (3)). As discussed earlier, the latter problem can be reformulated as a linear program with exponentially many variables/constraints as in (9) in the Appendix. The aforementioned values of n are small enough that we can solve this linear program and obtain the optimal worst-case variance V_c^{OPT} (and the corresponding experiment).

In Table 1, we report the average and the max value of the ratio $\frac{V^{\text{DP}}}{V_c^{\text{OPT}}}$ over the 10^4 samples for each fixed n, and we draw the box-plot of these ratios in Figure 3. Overall, the IBR experiment performs quite well and in most instances it only increases the worst-case variance by at most 50%. Notably, this is substantially better than the $\frac{7}{3}$-approximation guarantee of Theorem 4.4.

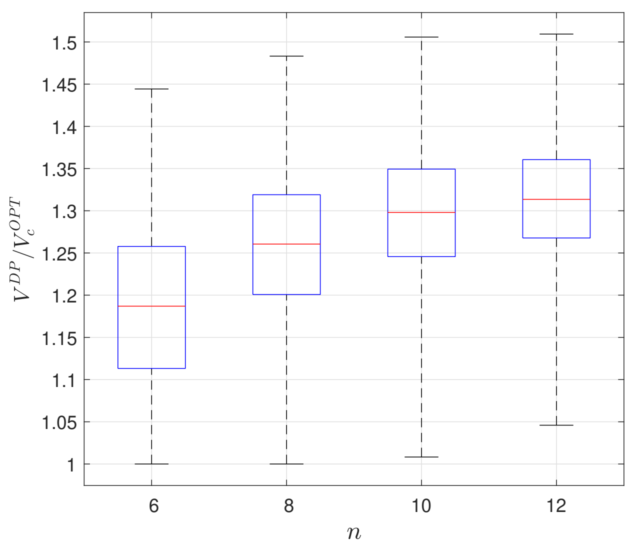


Figure 3: Box-plot of the ratios $V^{\text{DP}}/V_c^{\text{OPT}}$ using 10^4 samples for each fixed n. For each box, the central red edge indicates the median, the bottom and top blue edges of the box indicate the 25-th and 75-th percentiles, respectively, and the bottom and top black edges outside the box indicate the minimum and maximum extreme values, respectively.

7.2 Facebook Examples

Next, we focus on two data-driven examples based on Facebook data. While our first example in Section 7.2.1 is relatively smaller, the second example in Section 7.2.2 involves a larger Facebook

n	6	8	10	12
average of $V^{\mathrm{DP}}/V_c^{\mathrm{OPT}}$	1.186	1.258	1.297	1.314
max of $V^{\mathrm{DP}}/V_c^{\mathrm{OPT}}$	1.444	1.483	1.506	1.509

Table 1: Average and max values of the ratio $\frac{V^{\mathrm{DP}}}{V_c^{\mathrm{OPT}}}$ for randomly generated examples.

subnetwork of 100 US universities.

7.2.1 A First Example

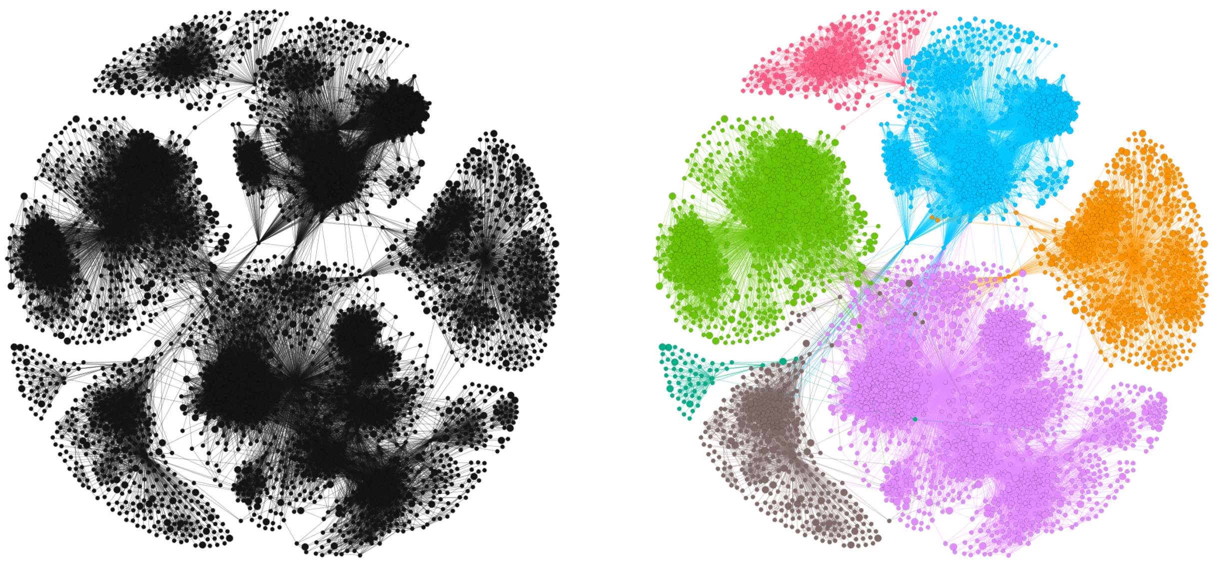
In our first example, we focus on a Facebook subnetwork from McAuley and Leskovec (2012). The dataset contains 4039 users. There is an edge between any two users if they are friends in this subnetwork, which indicates the possibility of interference between these two users.

We first partition the above network into communities by applying the well-known Louvain algorithm (Blondel et al. 2008).[29] We then combine two communities together if one community contaminates more than 10% of the nodes of the other one, to ensure that communities in the final partition are largely uncontaminated. This results in a partition of the network into $n = 7$ communities of different sizes. See Figure 4 for a visualisation and Table 2 for further details about our partition. As can be seen from this table, there is only a small number of cross community connections. By ignoring the contaminated users, we estimate the total market effect only on the uncontaminated users. As discussed in Section 5, this only introduces a small bias. The number of uncontaminated users in each community is provided in the second-to-last column in Table 2.

	1	2	3	4	5	6	7	# uncontaminated	# contaminated
1	1339	58	74	33	12	0	4	1190	149
2	4	752	1	0	2	0	0	747	5
3	38	2	780	0	4	1	0	741	39
4	7	0	0	548	1	5	0	537	11
5	12	23	11	1	354	0	0	315	39
6	0	0	1	2	0	206	0	203	3
7	1	0	0	0	0	0	60	59	1

Table 2: The partition of the network of Facebook users into 7 communities. Columns 1 to 7: entry (i, i) (highlighted in gray) is the number of users in community i; entry (i, k) with $i \neq k$ is the number of users in community i that are contaminated by community k (i.e., number of users in community i that have a connection with a user in community k). Columns 8 to 9: row i provides the number of users in community i that are contaminated by other communities, and the number of uncontaminated users.

[29] This is a classic approach to extract communities in large networks. It tries to construct a partition of nodes by maximizing modularity, i.e., the fraction of edges that remain within community/partition relative to a random distribution of edges. Modularity maximization is itself a computationally challenging problem, thus the algorithm relies on a greedy heuristic for this purpose. See Blondel et al. (2008) for details.



(a) (b)

Figure 4: (a) The Facebook subnetwork of 4039 users. (b) The partition of users into $n = 7$ communities, with each color representing a different community.

The worst-case variance of an optimal community-level experiment is $V_c^{\mathrm{OPT}} = 5.66 \times 10^6$. Although being optimal, this experiment results in a complex randomized assignment (see Table 4 in Appendix C). Specifically, the experiment randomizes over 36 different possible assignment vectors (where the number of treated communities varies between 2 and 5), and chooses different probabilities for these vectors without following any clear patterns. It even deliberately introduces some amount of positive correlation between small communities to attain a larger negative correlation between some pairs of large and small communities. In particular, the correlation matrix of the assignments under the optimal community-level experiment is

$$\Sigma_c^{\mathrm{OPT}} = \begin{pmatrix} 1 & -0.314 & -0.311 & -0.226 & -0.132 & -0.085 & -0.087 \\ -0.314 & 1 & -0.268 & 0 & 0 & 0 & -0.242 \\ -0.311 & -0.268 & 1 & -0.402 & -0.019 & 0 & 0.100 \\ -0.226 & 0 & -0.402 & 1 & 0 & 0 & 0 \\ -0.132 & 0 & -0.019 & 0 & 1 & 0 & 0.237 \\ -0.085 & 0 & 0 & 0 & 0 & 1 & 0 \\ -0.087 & -0.242 & 0.100 & 0 & 0.237 & 0 & 1 \end{pmatrix}.$$

As can be seen from the above matrix and the details in Appendix C, the optimal experiment has a fairly complicated correlation structure that makes it hard to interpret.

On the contrary, the optimal IBR experiment—which is obtained by solving the DP presented earlier—has only two blocks; it simply places the first four communities in one block and the last three communities in another block. Then, it randomly assigns half of the communities (chosen uniformly at random) in each block to treatment. Thus, unlike the optimal community-level experiment, the optimal IBR experiment is easy to interpret: it pools the large communities together in one block and the small communities together in another, and then in each block pretends that the community sizes are the same and runs the optimal community-level experiment accordingly inside the block. The worst-case variance of this experiment is $V^{\mathrm{DP}} = 6.06 \times 10^6$ and relative to the optimal community-level experiment[30], it increases the worst-case variance by $\frac{V^{\mathrm{DP}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 7\%$.

We also compare our design with three natural heuristics:

1. `HALF`: The simple experiment that randomly assigns half of the communities to treatment (i.e., an IBR experiment with only one block);
2. `PAIR`: the pair matching experiment that sorts the communities based on their sizes, pairs each community with an odd index in the sorted list to the next community with an even index, and finally randomly assigns one community from each pair to treatment (i.e., an IBR experiment with each block containing two communities);
3. `IND`: the naive experiment with an independent community-level assignment (which assigns each community to treatment or control independently with probability 1/2).

The `HALF` experiment has a worst-case variance $V^{\mathrm{half}} = 7.44 \times 10^6$, and this increases the worst-case variance by $\frac{V^{\mathrm{half}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 31.3\%$.[31] The `PAIR` experiment has a worst-case variance $V^{\mathrm{pair}} = 8.27 \times 10^6$, and this increases the worst-case variance of the optimal community-level assignment by $\frac{V^{\mathrm{pair}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 46.0\%$. Finally, the `IND` experiment has a worst-case variance $V^{\mathrm{ind}} = 1.18 \times 10^7$, and this increases the worst-case variance of the optimal community-level assignment by $\frac{V^{\mathrm{ind}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 108.5\%$.

This example illustrates that not only the IBR experiment is considerably easier to compute and implement than the optimal community-level assignment, it also admits a worst-case variance that almost matches that of the optimal community-level assignment. At the same time, other

[30] Note that we assume users' potential outcomes are in $[0, 1]$ by Assumption 2.1, so only the relative value of an experiment's variance is meaningful.

[31] Since the last community is substantially small compared to the other communities, we can also manually combine the last two communities together and randomly assign half of the final six communities to treatment. This has a worst-case variance $V^{\mathrm{half}'} = 6.47 \times 10^6$ and increases the worst-case variance of the optimal community-level assignment by $\frac{V^{\mathrm{half}'} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 14.3\%$.

simple heuristic designs have substantially worse performances and higher worst-case variance. For further details on this example see Appendix C.

7.2.2 Facebook Subnetworks of US Universities

Next, we focus on Facebook subnetworks of 100 US universities. Specifically, we leverage the data described in Section 2 of Traud et al. (2012), which can be accessed from Rossi and Ahmed (2015). We consider community-level experiments over these subnetworks, with the users from each university constituting one community. We assume users from different universities are only loosely connected (in contrast with the dense connection structure within each subnetwork). Hence, as discussed earlier, the interference among these subnetworks is negligible. See Traud et al. (2012) for the sizes of these communities.

With $n = 100$ communities, it is computationally prohibitive to obtain the optimal community-level experiment. The optimal IBR experiment, on the other hand, is fairly easy to compute by solving the DP in Section 3.2. Specifically, the optimal IBR experiment partitions the communities into 11 blocks, with the number of communities in each block being[32]

$$8, 10, 10, 12, 10, 12, 12, 10, 8, 4, 4.$$

The worst-case variance of the optimal IBR experiment is $V^{\text{DP}} = 2.93 \times 10^{10}$.

We again consider the three natural experiments in Section 7.2.1 for comparison. Specifically, the `HALF` experiment has a worst-case variance $V^{\text{half}} = 5.22 \times 10^{10}$, and this increases the worst-case variance by $\frac{V^{\text{half}} - V^{\text{DP}}}{V^{\text{DP}}} = 78.3\%$ relative to our IBR experiment. The `PAIR` experiment has a worst-case variance $V^{\text{pair}} = 4.77 \times 10^{10}$, and increases the worst-case variance by $\frac{V^{\text{pair}} - V^{\text{DP}}}{V^{\text{DP}}} = 62.9\%$. The `IND` experiment has a worst-case variance $V^{\text{ind}} = 9.08 \times 10^{10}$, and increases the worst-case variance by $\frac{V^{\text{ind}} - V^{\text{DP}}}{V^{\text{DP}}} = 210.0\%$. Thus, the IBR experiment again reduces the worst-case variance considerably when compared to other commonly used heuristic experiments.

8 Summary and Further Discussions

We have considered the problem of designing a randomized experiment for a network of users, with the objective of minimizing the variance of an unbiased Horvitz-Thompson estimator that estimates the total market effect. We formulate the problem as robust optimization against the

[32] Blocks are sorted in decreasing order of the largest community size.

adversarial selection of potential outcomes. We derive our results by first focusing on a network composed of multiple disjoint communities that satisfy a regularity condition (which holds, e.g., if each community is densely-connected). Then, we extend them to settings where communities involve a small number of connections among them. We establish that it is near-optimal to assign all users in a community to the same variant (treatment or control). An optimal community-level assignment is computationally expensive to solve, and can be difficult to implement due to the required complicated correlation structure. Motivated by this, we develop a family of simple independent block experiments that are easy to compute and interpret. These experiments are optimal when all communities have identical sizes. More generally, we show that IBR experiments constitute a $\frac{7}{3}$-approximation for any problem instance, and are asymptotically optimal (in the number of communities) under a mild no-dominating community condition. In general settings, the suboptimality originates from the loss both due to independence across blocks and due to ignoring the differences in sizes of communities within a block. Our results indicate that this suboptimality can be made small with a careful partitioning of communities to blocks.

8.1 Discussion for Practical Implementation

To complement our proposal of using the IBR experiments in the cluster-based experimental design pipeline, we next discuss a few practical considerations. We highlight some challenges and present interesting future directions related to implementing our experiments.

Constructing Communities Constructing communities to control interference can be a challenging task in cluster-based experimental design. In our numerical examples, we used off-the-shelf community detection algorithms. In practice, more sophisticated methods can be used (e.g., Rolnick et al. 2019 and Section 4.1.1 of Holtz et al. 2020). The number of communities generated by any such method essentially controls the trade-off between the bias and variance of the estimation. Notably, the interference model we consider in Section 2.1 is quite conservative in capturing interference, as it requires the user and all her neighbors to receive the same assignments for a meaningful non-interfered outcome. Nevertheless, as we pointed out in Remark 2.1, the solution to the optimal community-level assignment problem—and hence the results of Section 4—does not depend on the specific interference model once the communities are extracted (and the decision maker opts to ignore the interference between communities). Thus, if the interference is weaker in practice, this provides room to extract more communities while still controlling bias which in turn increases the statistical power of the experiment.

Alternative Estimators We focused on the Horvitz-Thompson estimator in this paper, mainly because it is unbiased. However, in practical settings, one might find it reasonable to use an estimator with lower variance at the price of adding small bias.[33] Importantly, with a more sophisticated estimator, the variance as a function of the randomized experiment and the potential outcomes can have a more complex form. Extending our approach to these estimators is an interesting avenue for future research.

Incorporating Covariate Information We considered a prior-free model of potential outcomes with minimal assumptions and no side information. However, in some practical scenarios, it is possible to have access to covariate information for the users or the communities, that is correlated with their potential outcomes. In this case, one promising approach is to use this covariate information to obtain the right "balance" between the treatment and control groups that delivers small variance (see, e.g, Bertsimas et al. 2015 and Bhat et al. 2020). However, in practice this task can be challenging due to several reasons. For example, the number of covariates can be quite large, or it can be difficult to incorporate all the relevant covariates (and there may even be latent covariates). In addition, the outcomes might be generated by complicated market interactions that are not well-summarized by simple covariate models (e.g., a linear model in treatments and covariates). In all of these cases, it is worth considering the robust design approach that takes into account the worst-case potential outcomes. Hence, extending our model to incorporate covariate information while considering the optimal robust experimental design remains to be a promising research direction.

Combining Different Design Approaches There are other experimental designs that are commonly used in practice. Some prominent examples are the *switchback experiments* (Bojinov et al. 2020) that alter the treatment and control assignments over time and the *stratified experiments* (Fisher 1935, Higgins et al. 2016) that group (or stratify) the experimental units based on their covariates and then assign treatments to each group independently.[34] It might be interesting to investigate ways of combining these experiments in practice, perhaps based on the insights from each individual design.

Other Practical Constraints In many operational settings, experimentation is not free of cost. Also, interference often stems from non-trivial market effects (e.g., as in Johari et al. 2020 and Wager

[33]For example, the Hajek estimator (Hájek 1971) or other parametric data-driven estimators (e.g., Khan and Ugander 2021).

[34]As we point out in Remark 3.1, our independent block randomization can be interpreted as stratified randomization in community sizes.

and Xu 2021). The design of experiments that carefully model (e.g., through a bipartite network) and exploit the interference structure, while taking into account the aforementioned market effects and experimentation costs, remain to be directions for future work on this subject.

Acknowledgements

We thank Gabriel Weintraub for his valuable comments on an earlier version of this work, which greatly improved the current revision. We also thank the attendees of the Revenue Management and Pricing conference and the Marketplace Innovation Workshop.

References

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

Asmussen, S. and Glynn, P. W. (2007), *Stochastic simulation: algorithms and analysis*, Vol. 57, Springer Science & Business Media.

Barajas, J., Akella, R., Holtan, M. and Flores, A. (2016), 'Experimental designs and estimation for online display advertising attribution in marketplaces', *Marketing Science* **35**(3), 465–483.

Berger, J. O. (2013), *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media.

Bertsimas, D., Johnson, M. and Kallus, N. (2015), 'The power of optimization over randomization in designing experiments involving small samples', *Operations Research* **63**(4), 868–876.

Bertsimas, D., Korolko, N. and Weinstein, A. M. (2019), 'Covariate-adaptive optimization in online clinical trials', *Operations Research* **67**(4), 1150–1161.

Bhat, N., Farias, V. F., Moallemi, C. C. and Sinha, D. (2020), 'Near-optimal ab testing', *Management Science* **66**(10), 4477–4495.

Bibaut, A., Chambaz, A., Dimakopoulou, M., Kallus, N. and van der Laan, M. (2021), 'Post-contextual-bandit inference', *arXiv preprint arXiv:2106.00418* .

Blake, T. and Coey, D. (2014), Why marketplace experimentation is harder than it seems: The role of test-control interference, *in* 'Proceedings of the fifteenth ACM conference on Economics and computation', pp. 567–582.

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

Bojinov, I., Simchi-Levi, D. and Zhao, J. (2020), 'Design and analysis of switchback experiments', *Available at SSRN 3684168* .

Chamandy, N. (2016), 'Lyft engineering: Experimentation in a ridesharing marketplace'. Available at https://eng.lyft.com/experimentation-in-a-ridesharing-marketplace-b39db027a66e#.djox1933t.

Cui, R., Li, J. and Zhang, D. (2017), 'Discrimination with incomplete information in the sharing economy: Evidence from field experiments on airbnb', *Harvard Business School* pp. 1–35.

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

Easley, D., Kleinberg, J. et al. (2010), *Networks, crowds, and markets*, Vol. 8, Cambridge university press Cambridge.

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

Fisher, R. A. (1935), *The design of experiments*, Oliver & Boyd.

Glynn, P., Johari, R. and Rasouli, M. (2020), 'Adaptive experimental design with temporal interference: A maximum likelihood approach', *arXiv preprint arXiv:2006.05591* .

Gui, H., Xu, Y., Bhasin, A. and Han, J. (2015), Network a/b testing: From sampling to estimation, *in* 'Proceedings of the 24th International Conference on World Wide Web', pp. 399–409.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S. and Athey, S. (2019), 'Confidence intervals for policy evaluation in adaptive experiments', *arXiv preprint arXiv:1911.02768* .

Hájek, J. (1971), 'Comment on "an essay on the logical foundations of survey sampling, part one"', *The foundations of survey sampling* **236**.

Harshaw, C., Eisenstat, D., Mirrokni, V. and Pouget-Abadie, J. (2021), 'Design and analysis of bipartite experiments under a linear exposure-response model', *arXiv preprint arXiv:2103.06392* .

Harshaw, C., Sävje, F., Spielman, D. and Zhang, P. (2019), 'Balancing covariates in randomized experiments using the gram-schmidt walk', *arXiv preprint arXiv:1911.03071* .

Higgins, M. J., Sävje, F. and Sekhon, J. S. (2016), 'Improving massive experiments with threshold blocking', *Proceedings of the National Academy of Sciences* **113**(27), 7369–7376.

Holtz, D. and Aral, S. (2020), 'Limiting bias from test-control interference in online marketplace experiments', *arXiv preprint arXiv:2004.12162* .

Holtz, D., Lobel, R., Liskovich, I. and Aral, S. (2020), 'Reducing interference bias in online marketplace pricing experiments', *arXiv preprint arXiv:2004.12489* .

Horvitz, D. G. and Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American statistical Association* **47**(260), 663–685.

Imai, K., King, G. and Nall, C. (2009), 'The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation', *Statistical Science* **24**(1), 29–53.

Imbens, G. W. and Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

Jackson, M. O. (2010), *Social and economic networks*, Princeton university press.

Johari, R., Li, H., Liskovich, I. and Weintraub, G. (2020), 'Experimental design in two-sided platforms: An analysis of bias', *arXiv preprint arXiv:2002.05670* .

Kallus, N. (2018), 'Optimal a priori balance in the design of controlled experiments series b statistical methodology', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(1), 85–112.

Karrer, B., Shi, L., Bhole, M., Goldman, M., Palmer, T., Gelman, C., Konutgan, M. and Sun, F. (2020), 'Network experimentation at scale', *arXiv preprint arXiv:2012.08591* .

Khan, S. and Ugander, J. (2021), 'Adaptive normalization for ipw estimation', *arXiv preprint arXiv:2106.07695* .

Kohavi, R., Tang, D. and Xu, Y. (2020), *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press.

Koutra, V. (2017), Designing experiments on networks, PhD thesis, University of Southampton.

Lattimore, T. and Szepesvári, C. (2020), *Bandit algorithms*, Cambridge University Press.

McAuley, J. and Leskovec, J. (2012), 'Snap datasets: Stanford large network data set collection'. Available at http://snap.stanford.edu/data/ego-Facebook.html.

Newman, M. E. (2006), 'Modularity and community structure in networks', *Proceedings of the national academy of sciences* **103**(23), 8577–8582.

Ostrovsky, M. and Schwarz, M. (2011), Reserve prices in internet advertising auctions: A field experiment, *in* 'Proceedings of the 12th ACM conference on Electronic commerce', pp. 59–60.

Owen, A. B. (2020), 'A first course in experimental design notes from stat 263/363', *Lecture notes* . Available at https://statweb.stanford.edu/~owen/courses/363/doenotes.pdf.

Pouget-Abadie, J., Aydin, K., Schudy, W., Brodersen, K. and Mirrokni, V. (2019), 'Variance reduction in bipartite experiments through correlation clustering', *Advances in Neural Information Processing Systems* **32**, 13309–13319.

Rolnick, D., Aydin, K., Pouget-Abadie, J., Kamali, S., Mirrokni, V. and Najmi, A. (2019), Randomized experimental design via geographic clustering, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', pp. 2745–2753.

Rossi, R. A. and Ahmed, N. K. (2015), 'The network data repository'. Available at http://networkrepository.com/socfb.php.

Slivkins, A. (2019), 'Introduction to multi-armed bandits', *Foundations and Trends® in Machine Learning* **12**(1-2), 1–286.

Traud, A. L., Mucha, P. J. and Porter, M. A. (2012), 'Social structure of facebook networks', *Physica A: Statistical Mechanics and its Applications* **391**(16), 4165–4180.

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

Wager, S. and Xu, K. (2021), 'Experimenting in equilibrium', *Management Science* .

Williamson, D. P. and Shmoys, D. B. (2011), *The design of approximation algorithms*, Cambridge university press.

Xiong, R., Athey, S., Bayati, M. and Imbens, G. W. (2019), 'Optimal experimental design for staggered rollouts', *Available at SSRN 3483934* .

Yu, Q., Zhang, Y. and Zhou, Y.-P. (2020), 'Delay information in virtual queues: A large-scale field experiment on a ride-sharing platform', *Available at SSRN 3687302* .

Zhang, D. J., Dai, H., Dong, L., Qi, F., Zhang, N., Liu, X., Liu, Z. and Yang, J. (2020), 'The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on alibaba', *Management Science* **66**(6), 2589–2609.

Zigler, C. M. and Papadogeorgou, G. (2021), 'Bipartite causal inference with interference', *Statistical science: a review journal of the Institute of Mathematical Statistics* **36**(1), 109.

A Omitted Proofs

A.1 Proof of Lemma 2.1

By Assumption 2.2, we have $\mathbb{P}[z_{j1} = 1, z_{\ell 1} = 1] = \mathbb{P}[z_{j0} = 1, z_{\ell 0} = 1] = \mathbb{P}[z_i = 1] = \frac{1}{2}$ and $\mathbb{P}[z_{j1} = 1, z_{\ell 0} = 1] = \mathbb{P}[z_{j0} = 1, z_{\ell 1} = 1] = 0$ for any two users $j, \ell \in S_i$ and community i. Analogously, for any two users $j \in S_i$ and $\ell \in S_k$ from two different communities $i, k \in [n], i \neq k$, $\mathbb{P}[z_{j1} = 1, z_{\ell 1} = 1] = \mathbb{P}[z_{j0} = 1, z_{\ell 0} = 1] = \mathbb{P}[z_i = 1, z_k = 1] = \mathbb{P}[z_i = 0, z_k = 0] \triangleq \pi_{ik}$ and $\mathbb{P}[z_{j1} = 1, z_{\ell 0} = 1] = \mathbb{P}[z_{j0} = 1, z_{\ell 1} = 1] = \mathbb{P}[z_i = 1, z_k = 0] = \mathbb{P}[z_i = 0, z_k = 1] \triangleq \bar{\pi}_{ik}$ for some $\pi_{ik}, \bar{\pi}_{ik} \in [0, 1]$. Thus, from (2) we have

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\hat{\tau}] &= y_1^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_1\tilde{z}_1^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big)y_1 + y_0^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_0\tilde{z}_0^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big)y_0 - 2 \cdot y_1^{\mathrm{T}}\Big(\mathbb{E}\big[\tilde{z}_1\tilde{z}_0^{\mathrm{T}}\big] - \mathbf{1}\mathbf{1}^{\mathrm{T}}\Big)y_0 \\
&= \sum_{i \in [n]} (y_{i1} + y_{i0})^2 + \sum_{i \in [n]} \sum_{[n] \ni k \neq i} \Big[y_{i1}y_{k1} \cdot (4\pi_{ik} - 1) + y_{i0}y_{k0} \cdot (4\pi_{ik} - 1) \\
&\quad - y_{i1}y_{k0} \cdot (4\bar{\pi}_{i\ell} - 1) - y_{i0}y_{k1} \cdot (4\bar{\pi}_{i\ell} - 1)\Big],
\end{aligned}$$

where $y_{i1} = \sum_{j \in S_i} y_{j1}$ and $y_{i0} = \sum_{j \in S_i} y_{j0}$ are the sum of the treatment and control potential outcomes, respectively, over the users in community i. Note that since each z_i is a symmetric Bernoulli random variable with $\mathbb{P}[z_i = 1] = \mathbb{P}[z_i = 0] = \frac{1}{2}$, we have $4\pi_{ik} - 1 = 1 - 4\bar{\pi}_{ik} = \sigma_{ik}$, where σ_{ik} is the correlation coefficient of z_i and z_k. Thus,

$$\mathbb{V}\mathrm{ar}[\hat{\tau}] = \sum_{i \in [n]} (y_{i1} + y_{i0})^2 + \sum_{i \in [n]} \sum_{[n] \ni k \neq i} \sigma_{ik}\Big(y_{i1}y_{k1} + y_{i0}y_{k0} + y_{i1}y_{k0} + y_{i0}y_{k1}\Big). \tag{8}$$

Note that $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ is a quadratic convex function of y_{i1} and $y_{i1} \in [0, w_i]$, for all i. As a result, in the worst case, each y_{i1} takes a value of either 0 or w_i (which corresponds to all of the nodes in community i having the same potential outcome, i.e., $y_{j1} = 0$ or $y_{j1} = 1$ for all $j \in S_i$). Analogously, in the worst case, each y_{i0} takes a value that is either 0 or w_i (and this corresponds to all of the nodes $j \in S_i$ either having outcome $y_{j0} = 0$ or $y_{j0} = 1$ at the same time).

We claim that in the worst case, $y_{i1} = y_{i0} = y_i \in \{0, w_i\}$ for any community i. Suppose not, and without loss of generality assume that $y_{i1} = w_i$ and $y_{i0} = 0$ for some community i. Note that if we set $y_{i1} = y_{i0} = w_i$, the variance changes by

$$(\Delta_{w_i}) \triangleq 3w_i^2 + 2\sum_{k \neq i} \sigma_{ik}w_i \cdot \Big(y_{k1} + y_{k0}\Big).$$

Similarly, if we set $y_{i1} = y_{i0} = 0$, the variance changes by

$$(\Delta_0) \triangleq -w_i^2 - 2\sum_{k \neq i} \sigma_{ik}w_i \cdot \Big(y_{k1} + y_{k0}\Big).$$

Since $(\Delta_{w_i}) + (\Delta_0) = 2w_i^2 > 0$, at least one of the two above deviations strictly increases the variance. Thus, in the worst case, it has to be the case that $y_{i1} = y_{i0} = y_i \in \{0, w_i\}$ for any community i. The corresponding variance of the Horvitz-Thompson estimator is

$$\mathbb{V}\mathrm{ar}[\hat{\tau}] = 4\sum_{i \in [n]} y_i^2 + 4\sum_{i \in [n]} \sum_{[n] \ni k \neq i} \sigma_{ik}y_iy_k = 4y^T\Sigma y,$$

as claimed. $\square$

A.2 Details of Remark 2.2

If the uncertainty set of each user j's potential outcomes is $y_{j1}, y_{j0} \in [-1, 1]$, then the constraint in (3) requires $y_i \in [-w_i, w_i]$ for any community i. In this case, we claim that the optimal experiment is to simply assign treatment to each community independently. The correlation matrix with such independent assignment is the identity matrix, i.e., $\Sigma^* = I$. The worst-case potential outcome is $y_i = w_i$ for any community i, and the worst-case variance is $4y^T\Sigma^* y = 4\sum_{i\in[n]} w_i^2$.

We now show no experiment can achieve a worst-case variance strictly smaller than $4\sum_{i\in[n]} w_i^2$. To see this, consider any feasible experiment and let Σ denote the corresponding correlation matrix. Consider the following randomized potential outcomes with Y_i being either w_i or $-w_i$ both with probability $\frac{1}{2}$, and let these Y_i be independent. Then, $\mathbb{E}\big[Y_i^2\big] = w_i^2$ and $\mathbb{E}\big[Y_iY_k\big] = 0$ for any $i \neq k$. Thus, letting $Y = (Y_i)_{i\in[n]} \in \mathbb{R}^n$ be the concatenation, we have

$$\max_{y \in \times_{i\in[n]}[0,w_i]} 4y^T\Sigma y \geq 4\mathbb{E}\Big[Y^T\Sigma Y\Big] = 4\sum_{i\in[n]} w_i^2,$$

which clearly shows the optimality of independently assigning treatments to each community.

A.3 Proof of Proposition 3.1

Since the set of community-level joint assignment distributions $\mathcal{P}_c$ is a polyhedron and $\Sigma\big(P\big)$ is a linear map of the distribution $P \in \mathcal{P}_c$, we can write (3) into a linear program with exponentially many decision variables and constraints as in (9),

$$\begin{aligned} \underset{P\in\mathcal{P}_c,\, z\in\mathbb{R}}{\text{minimize}} \quad & z \\ \text{subject to} \quad & z \geq y^{\mathrm{T}}\Sigma\big(P\big)y,\ \forall\, y \in \bigtimes_{i\in[n]} \{0, w_i\}, \end{aligned} \tag{9}$$

where we have polyhedral constraints of $\mathcal{P}_c$ and we have one constraint for each extreme point of the potential outcomes' uncertainty set. This formulation also implies that a convex combination of optimal experiments is an optimal experiment as well.

Since communities have equal sizes, this observation implies that given any optimal experiment, we can always construct another optimal experiment that treats communities in an identical way. As a result, it is without loss of generality to assume $\sigma_{ik} = \sigma$ for any two communities. Thus, the variance of the estimator with a given set of potential outcomes becomes

$$4y^{\mathrm{T}}\Sigma y = 4\sum_{i\in[n]} y_i^2 + 4\sigma \sum_{i\in[n]} \sum_{[n]\ni k\neq i} y_i y_k.$$

Since all of the potential outcomes are non-negative, the optimal experiment tries to make σ as small as possible. Let $S = \sum_{i\in[n]} z_i$ be the total number of communities that receive the treatment. We have

$$\mathbb{V}\mathrm{ar}[S] = \frac{n}{4} + n(n-1)\cdot\frac{\sigma}{4},$$

simply because $\mathbb{C}\mathrm{ov}[z_i, z_j] = \frac{\sigma}{4}$ for any two communities $i \neq j$. When n is even, since $\mathbb{V}\mathrm{ar}[S] \geq 0$, we have $\sigma \geq -\frac{1}{n-1}$. The equality is attained when S is constant with probability one, and this can be achieved by randomly assigning $\frac{n}{2}$ communities to treatment. When n is odd, since S is an integer whereas $\mathbb{E}[S] = \sum_{i\in[n]} \mathbb{E}[z_i] = \frac{n}{2}$ is fractional, we have $|S - \mathbb{E}[S]| \geq \frac{1}{2}$ and thus $\mathbb{V}\mathrm{ar}[S] \geq \frac{1}{4}$. As a result, $\sigma \geq -\frac{1}{n}$. The equality is attained when $|S - \mathbb{E}[S]| = \frac{1}{2}$ with probability one, and this is achieved by randomly partitioning communities into two groups with cardinality $\frac{n+1}{2}$ and $\frac{n-1}{2}$, and randomly assigning one group to treatment.

We next study the worst-case potential outcome and the worst-case variance. Let $h \in \mathbb{N}$ be the number of communities that take an outcome $y_i = 1$, and suppose that the other communities take the outcome $y_i = 0$. Then

$$\mathbb{V}\mathrm{ar}\big[\hat{\tau}\big] = 4h + 4h(h-1)\sigma.$$

Denote by $h^* \in \mathbb{N}$ the integer that maximizes this quantity. This quantity satisfies $1 - \frac{1}{2\sigma} \geq h^* \geq -\frac{1}{2\sigma}$. When n is even, since $\sigma = -\frac{1}{n-1}$, we have $\frac{n+1}{2} \geq h \geq \frac{n-1}{2}$. Hence, $h^* = \frac{n}{2}$ and the corresponding variance is $\frac{n^2}{n-1}$. When n is odd, since $\sigma = -\frac{1}{n}$, we have $\frac{n}{2} + 1 \geq h^* \geq \frac{n}{2}$. Hence, $h^* = \frac{n+1}{2}$ and the corresponding variance is $\frac{(n+1)^2}{n}$. $\square$

A.4 Proof of Lemma 3.2

Since the objective of $\max_{y_i\in[0,w_i]} 4y^{\mathrm{T}}\Sigma y$ is jointly convex in y, in the worst case, $y_i \in \{0, w_i\}$ for each community $i \in [k]$. We first prove that there exists a worst-case potential outcome y such that $y_i = w_i$ for $i \leq p$ for some integer $p \in [n]$, and $y_i = 0$ for $i > p$. If not, then for any worst-case potential outcome y there exists indices $i < j$ such that $y_i = 0$ and $y_j = w_j$, whereas $w_i \geq w_j$. Since all the off-diagonals of the correlation matrix Σ have the same value of σ, the objective does not change if we instead swap y_i and y_j, and set $y_i = w_j \in [0, w_i]$ and $y_j = 0$. We can further (weakly) increase the objective by setting y_i to its extreme values, i.e., $y_i = w_i$ or $y_i = 0$. By iterating this process, we end up with a worst-case outcome that satisfies our desired property.

It remains to determine the value of p. Let y denote the worst-case potential outcome vector such that $y_i = w_i$ for $i \leq p$ and $y_i = 0$ for $i > p$. Note that

$$y^{\mathrm{T}}\Sigma y = \sum_{i\in[k]} y_i^2 + \sigma \cdot \sum_{i\in[k]} \sum_{[k]\ni j\neq i} y_i y_j.$$

Observe that if we update $y_i = 0$ for some community $i \leq p$, the variance changes by

$$-w_i^2 - 2\sigma \sum_{j\leq p \text{ and } j\neq i} w_i w_j \leq 0.$$

Similarly, if we set $y_i = w_i$ for some community $i > p$, the variance changes by

$$w_i^2 + 2\sigma \sum_{j=1}^{p} w_i w_j \leq 0.$$

Together these imply that

$$\forall\, i \leq p: \qquad w_i \geq -2\sigma \sum_{j\leq p \text{ and } j\neq i} w_j, \tag{10}$$

$$\forall\, i > p: \qquad w_i \leq -2\sigma \sum_{j=1}^{p} w_j. \tag{11}$$

Let $p^\star$ be the largest index such that $w_{p^\star} \geq -2\sigma \sum_{i\leq p^\star-1} w_i$. Since the community sizes are decreasing, $p^\star$ satisfies (10) and (11). Moreover, if $w_{p^\star} > -2\sigma \sum_{i\leq p^\star-1} w_i$, $p = p^\star$ is the only integer that satisfies both (10) and (11), and hence corresponds to a worst-case potential outcome. If $w_{p^\star} = -2\sigma \sum_{i\leq p^\star-1} w_i$, then both $p = p^\star$ and $p = p^\star - 1$ satisfy (10) and (11), and they both constitute a worst-case potential outcome. $\square$

A.5 Details and Proof of Correctness of Example 4.1

Throughout this section, we focus on the setting of Example 4.1. For notational convenience, we index the blocks of an IBR experiment in the decreasing order of the size of the largest community in each block. Also, for simplicity, we assume an even number n of communities. We first show in Lemma A.1 that when a block contains four or more communities, at least $p \geq 2$ communities take positive values (i.e., non-zero outcomes) in the worst-case potential outcome $y = (y_i)_{i\in[n]}$.

Lemma A.1. *Suppose a block contains an even number $k \geq 4$ communities, with community sizes $w_i = \beta^{k-i} w_k$ for any $i \in [k]$. Then, at least $p \geq 2$ communities take positive values in the worst-case potential outcome $y = (y_i)_{i\in[n]}$.*

Proof. Since the number of communities k is even, the correlation between any two communities is $\sigma = -\frac{1}{k-1}$. From Lemma 3.2, p is the largest integer that satisfies

$$\beta^{k-p} w_k = w_p \geq \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} w_i = \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} \beta^{k-i} w_k = \frac{2}{k-1} \frac{\beta^p - \beta}{\beta - 1} \cdot \beta^{k-p} w_k.$$

This implies that $p = \left\lfloor \frac{\ln\left(\frac{k-1}{2}(\beta-1)+\beta\right)}{\ln \beta} \right\rfloor \geq 2$ when $k \geq 4$. $\square$

We next show in Lemma A.2 that with the optimal partition, all blocks contain either two or four communities.

Lemma A.2. *Suppose the number of communities n is even. All blocks in an optimal partition contain either 2 or 4 communities.*

Proof. From Lemma B.2, each block contains an even number of communities. Suppose a block instead contains k communities where k is an even number satisfying $k \geq 6$. Without loss of generality, we assume that $w_i = \beta^{k-i}$ for each community $i \in [k]$ (since we can always normalize community sizes by the size of the smallest community). We claim that we can further partition this block into two smaller blocks to reduce the worst-case variance. Specifically, the first block contains the first $k-2$ communities of the original block, and the second block contains the last two blocks.

First, consider the worst-case variance with the two new blocks, denoted by V_a. Let p be the number of positive values in the worst-case potential outcome of the first block. By Lemma A.1, $p \geq 2$ because block one contains $k-2 \geq 4$ communities. We have

$$V_a = 4\left(\sum_{i\in[p]} w_i^2 - \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \frac{1}{k-3} w_i w_j\right) + 4\beta^2.$$

Here $4\beta^2$ is simply the worst-case variance of the second block, because in the worst case, community $k-1$ takes a positive outcome and community k takes outcome zero.

Let V_b denote the worst-case variance of the original block. It satisfies

$$V_b \geq 4\left(\sum_{i\in[p]} w_i^2 - \sum_{i\in[p]}\sum_{[p]\ni j\neq i} \frac{1}{k-1} w_i w_j\right),$$

because the correlation between any two communities in the original block is $-\frac{1}{k-1}$, and p is not necessarily the number of positive values in the worst-case potential outcome of the original block. Thus,

$$\begin{aligned}\frac{V_b - V_a}{4} &\geq \sum_{i\in[p]}\sum_{[p]\ni j\neq i}\left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_i w_j - \beta^2 \\ &\geq 2\left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_1 w_2 - \beta^2 \\ &= \frac{4\beta^{2k-3}}{(k-1)(k-3)} - \beta^2,\end{aligned}$$

which is non-negative when $k \geq 6$. Hence, splitting the large block into two smaller blocks reduces the worst-case variance. □

Finally, we show in Lemma A.3 that all blocks contain four communities in the optimal partition.

Lemma A.3. *Suppose the number of communities n is even. Then the optimal partition of an IBR experiment satisfies the following:*

- *If n is divisible by 4, all blocks contain exactly 4 communities;*
- *Otherwise, the last block contains 2 communities, and all the other blocks contain exactly 4 communities.*

Proof. By Lemma A.2, each block contains either two or four communities. For a two-community block, the worst-case outcome is simply the large community taking a positive outcome and the small community taking the zero outcome. For a four-community block, by the proof of Lemma A.1, the worst-case potential outcome is only the two largest communities taking positive outcomes.

Let K be the number of blocks. It suffices to show that all of the first $K-1$ blocks contain four communities. Suppose towards a contradiction that block $h \leq K-1$ contains two communities, k and $k+1$. If block $h+1$ contains two communities $(k+2,\, k+3)$ as well, then the observations above imply that merging blocks h and $h+1$ decreases the worst-case variance by

$$4w_{k+3}^2 \cdot \left\{(\beta^6 + \beta^2) - \left(\beta^6 + \beta^4 - \frac{2}{3}\cdot\beta^5\right)\right\} > 0.$$

Now suppose block $h+1$ contains four communities. Suppose we reconstruct blocks h and $h+1$ by assigning communities k to $k+3$ to block h and communities $k+4$ and $k+5$ to block $h+1$. Then, the worst-case variance decreases by

$$4w_{k+5}^2 \cdot \left\{\left(\beta^{10} + \beta^6 + \beta^4 - \frac{2}{3}\cdot\beta^5\right) - \left(\beta^{10} + \beta^8 - \frac{2}{3}\cdot\beta^9 + \beta^2\right)\right\} > 0.$$

Thus, it follows that it is not optimal for any block $h \leq K-1$ to contain two communities. □

We now show that the optimal IBR experiment is asymptotically suboptimal. Consider a block with four communities with sizes $w_i = \beta^{4-i}$ for $i \in [4]$. The worst-case variance with randomly assigning half of the communities to treatment is $v^{\text{half}} = 4 \cdot \left(\beta^6 + \beta^4 - \frac{2}{3} \cdot \beta^5\right) = 16.886$. The worst-case variance with the optimal randomized joint assignment (given in Table 3) is $v^{\text{OPT}} = 15.259$. Now, let us revisit Example 4.1. For simplicity, we assume the number of communities n is divisible by four. Since every block in the optimal partition contains 4 communities by Lemma A.3 and the blocks are identical up to a scaling, the worst-case variance with the optimal IBR experiment is

$$V^{\text{DP}} = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{half}}.$$

For an experiment that assigns communities in a block to treatment in an optimal way, and does so independently across blocks, the worst-case variance, denoted by V, is

$$V = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{OPT}}.$$

Thus,

$$\frac{V^{\text{DP}} - V_c^{\text{OPT}}}{V_c^{\text{OPT}}} \geq \frac{V^{\text{DP}} - V}{V} = \frac{v^{\text{half}} - v^{\text{OPT}}}{v^{\text{OPT}}} = 10.7\%,$$

which implies that the optimal IBR experiment is strictly suboptimal asymptotically.

	1	2	3	4	probability
1	×	√	×	√	0.1411
2	√	×	√	×	0.1411
3	×	√	√	×	0.1356
4	√	×	×	√	0.1356
5	×	×	√	√	0.1211
6	√	√	×	×	0.1211
7	×	√	√	√	0.0734
8	√	×	×	×	0.0734
9	×	×	×	√	0.0289
10	√	√	√	×	0.0289

Table 3: The optimal randomized joint assignment of treatment to four communities, with community sizes being $w_i = \beta^{4-i}$ for $i \leq 4$ and $\beta = \frac{5}{4}$. Each of the 10 rows corresponds to an assignment, where $\surd$ denotes treatment and $\times$ denotes control.

A.6 Proof of Theorem 4.6

Throughout the proof, we refer to a community with size w_k as a community of type k. First, using a similar argument to the one in the proof of Proposition 3.1, any convex combination of optimal correlation matrices in (5) (which is the optimization problem that defines V^{LB}) is an optimal correlation matrix as well. Thus, there exists an optimal correlation matrix attaining V^{LB}

that takes the following form:

$$\Sigma = \begin{pmatrix} \sigma_{11}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{11})I & \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{22}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{22})I & & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{KK}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{KK})I \end{pmatrix}, \tag{12}$$

where σ_{kk} is the correlation coefficient of the treatment assignments for any two different communities of the same type k, and $\sigma_{k\ell}$ is the correlation coefficient of the treatment assignments for any two communities of types k and ℓ with $k \neq \ell$, respectively. Since the correlation matrix Σ needs to be positive semi-definite, for any scalars $x_k \in \mathbb{R}$ for $k \in [K]$, we have

$$\begin{aligned} 0 &\leq \begin{pmatrix} x_1\mathbf{1}^{\mathrm{T}} & x_2\mathbf{1}^{\mathrm{T}} & \cdots & x_K\mathbf{1}^{\mathrm{T}} \end{pmatrix} \Sigma \begin{pmatrix} x_1\mathbf{1} \\ x_2\mathbf{1} \\ \vdots \\ x_K\mathbf{1} \end{pmatrix} \\ &= \sum_{k=1}^{K} \Big[\sigma_{kk}x_k^2 n_k^2 + (1-\sigma_{kk})x_k^2 n_k\Big] + \sum_{k=1}^{K}\sum_{\ell\neq k} \sigma_{k\ell}x_k x_\ell n_k n_\ell \\ &= \begin{pmatrix} x_1 n_1 & x_2 n_2 & \cdots & x_K n_K \end{pmatrix} \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \begin{pmatrix} x_1 n_1 \\ x_2 n_2 \\ \vdots \\ x_K n_K \end{pmatrix}. \end{aligned}$$

Thus, the matrix

$$\tilde{\Sigma} = \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \succeq 0 \tag{13}$$

needs to be positive semi-definite.

For the IBR experiment with one block for each set of communities of the same type, the corresponding correlation matrix takes the form of (12) as well, with $\sigma_{kk} = -\frac{1}{n_k-1}$ if n_k is even and $\sigma_{kk} = -\frac{1}{n_k}$ if n_k is odd, and $\sigma_{k\ell} = 0$ for all $k \neq \ell$. From Proposition 3.1, in the worst-case potential outcome for such an IBR experiment, $\bar{h}_k$ communities of type k take values w_k and the other type k communities take values 0, with $\bar{h}_k = \frac{n_k}{2}$ if n_k is even and $\bar{h}_k = \frac{n_k+1}{2}$ if n_k is odd. Now, let Σ^* denote an optimal correlation matrix attaining V^{LB} that takes the form of (12), and let $\{\sigma^*_{k\ell}\}$ denote the corresponding correlation coefficients between community types. We have

$$\begin{aligned} V^{\mathrm{LB}} &= 4 \cdot \max_{h_k \in [n_k]} \left\{ \sum_{k\in[K]} w_k^2 \Big[h_k + h_k(h_k-1)\sigma^*_{kk}\Big] + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell h_k h_\ell \sigma^*_{k\ell} \right\} \\ &\geq 4 \cdot \left\{ \sum_{k\in[K]} w_k^2 \Big[\bar{h}_k + \bar{h}_k(\bar{h}_k-1)\sigma^*_{kk}\Big] + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell} \right\}, \end{aligned}$$

where we plug the worst-case potential outcome of the IBR experiment to obtain the inequality. As a result,

$$
\begin{aligned}
V - V^{\mathrm{LB}} \le & -4 \cdot \left\{ \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell} + \sum_{k\in[K]:n_k \text{ even}} w_k^2 \cdot \bar{h}_k \big(\bar{h}_k - 1\big) \left(\sigma^*_{kk} + \frac{1}{n_k - 1}\right) \right. \\
& \left. + \sum_{k\in[K]:n_k \text{ odd}} w_k^2 \cdot \bar{h}_k \big(\bar{h}_k - 1\big) \left(\sigma^*_{kk} + \frac{1}{n_k}\right) \right\} \\
= & -4 \cdot \left\{ \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell} + \sum_{k\in[K]} \frac{1 + (n_k - 1)\sigma^*_{kk}}{n_k} w_k^2 \bar{h}_k^2 \right\} \\
& + \sum_{k\in[K]:n_k \text{ even}} w_k^2 \left(\frac{n_k}{n_k - 1} + n_k \sigma^*_{kk}\right) + \sum_{k\in[K]:n_k \text{ odd}} w_k^2 \left(\frac{2(n_k+1)}{n_k} + \frac{n_k^2 - 1}{n_k}\sigma^*_{kk}\right) \\
\le & \sum_{k\in[K]:n_k \text{ even}} w_k^2 \left(\frac{n_k}{n_k - 1} + n_k \sigma^*_{kk}\right) + \sum_{k\in[K]:n_k \text{ odd}} w_k^2 \left(\frac{2(n_k+1)}{n_k} + \frac{n_k^2 - 1}{n_k}\sigma^*_{kk}\right) \\
\le & \sum_{k\in[K]} w_k^2 \cdot \Big(5 + n_k \sigma^*_{kk}\Big),
\end{aligned}
\tag{14}
$$

where the second inequality follows from the fact that with $u = \big(w_k \bar{h}_k\big)_{k\in[K]} \in \mathbb{R}^K$,

$$
\sum_{k\in[K]} \frac{1 + (n_k - 1)\sigma^*_{kk}}{n_k} w_k^2 \bar{h}_k^2 + \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell} = u^{\mathrm{T}} \tilde{\Sigma} u \ge 0,
$$

because $\tilde{\Sigma}$ is positive semi-definite by (13). To bound the value of $n_k \sigma^*_{kk}$ from above, note that

$$
V^{\mathrm{LB}} \le V = \sum_{k\in[K]:n_k \text{ even}} w_k^2 \cdot \frac{n_k^2}{n_k - 1} + \sum_{k\in[K]:n_k \text{ odd}} w_k^2 \cdot \frac{(n_k+1)^2}{n_k} \le 4 w_1^2 \cdot n, \tag{15}
$$

where the equality follows from Proposition 3.1 and the fact that the worst-case variance is additive across blocks. Also, note that

$$
\begin{aligned}
V^{\mathrm{LB}} &= 4 \max_{h_k \in [n_k]} \left\{ \sum_{k\in[K]} w_k^2 \Big[h_k + h_k(h_k - 1)\sigma^*_{kk}\Big] + \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell h_k h_\ell \sigma^*_{k\ell} \right\} \\
&\ge 4 w_k^2 \Big[n_k + n_k(n_k - 1)\sigma^*_{kk}\Big] \\
&\ge 4 w_k^2 n_k^2 \sigma^*_{kk},
\end{aligned}
\tag{16}
$$

where we take $h_k = n_k$ and $h_\ell = 0$ for all $\ell \neq k$ to obtain the first inequality. Combining the inequalities (15) and (16) gives

$$
n_k \sigma^*_{kk} \le \min\left\{ n_k, \frac{w_1^2 n}{w_k^2 n_k} \right\}.
$$

Plugging this last inequality into the right-hand-side of the chain of inequalities in (14) yields the desired bound, and hence finishes the proof. □

A.7 Proof of Theorem 4.7

First, note that we have

$$K \leq \left\lceil \frac{\ln w_1/\bar{w}}{\ln \alpha} \right\rceil \leq \frac{\ln w_1/\bar{w}}{\ln \alpha} + 1 \leq \frac{(1+\delta_1)\ln n}{2\ln(1+n^{-\delta_2})} + 1 = O\Big(n^{\delta_2} \cdot \ln n\Big).$$

Let $S_k \subseteq [n]$ denote the set of communities in block k and $n_k = |S_k|$ denote this block's cardinality. Consider a new problem instance in which we first drop block zero, and then for each block $k \in [K]$ we decrease all of the community sizes to the minimum community size of the block. We denote the minimum community size of the block k by $\underline{w}_k \triangleq \min_{i \in S_k} w_i$. For this new problem instance, let $\tilde{V}^{\text{LB}}$ denote the lower bound on the worst-case variance of the optimal community-level experiment (i.e., the optimal value of (5)) and $\tilde{V}$ denote the worst-case variance of the IBR experiment. We proceed by bounding the differences of the original and the new problem instances in terms of the worst-case variances of their IBR experiments and the corresponding lower bounds.

Lemma A.4. *$\tilde{V}^{\text{LB}} \leq V^{\text{LB}}$ and $0 \leq V - \tilde{V} \leq 4n^{-\delta_1} \sum_{i=1}^n w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2$.*

Proof. Since community sizes are weakly smaller in the new problem instance, the set of potential outcomes is more restricted, and hence $\tilde{V}^{\text{LB}} \leq V^{\text{LB}}$ and $\tilde{V} \leq V$. For an IBR experiment, the worst-case variance is the sum of the worst-case variances of each block. Let V_k and $\tilde{V}_k$ denote the worst-case variances of block k in the original and the new problem instances, respectively. We have

$$\begin{aligned} V - \tilde{V} &= V_0 + \sum_{k \in [K]} \Big(V_k - \tilde{V}_k\Big) \\ &\leq 4 \sum_{i \in S_0} w_i^2 + \sum_{k \in [K]} \Big(\alpha^2 \tilde{V}_k - \tilde{V}_k\Big) \\ &\leq 4n^{-\delta_1} \sum_{i \in [n]} w_i^2 + (\alpha^2 - 1)\tilde{V} \\ &\leq 4n^{-\delta_1} \sum_{i \in [n]} w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2. \end{aligned} \tag{17}$$

Here, the first inequality uses the fact that for the specific logarithmic partition, we have $w_i \leq \alpha \underline{w}_k$ for any community $i \in S_k$. This implies that the worst-case variance V_k is no larger than the worst-case variance when all community sizes in block k are equal to $\alpha \underline{w}_k$. The latter quantity is $\alpha^2 \tilde{V}_k$, and hence $V_k \leq \alpha^2 \tilde{V}_k$. The first and third inequalities also make use of the inequalities $V_0 \leq 4 \sum_{i \in S_0} w_i^2$ and $\tilde{V} \leq V \leq 4 \sum_{i=1}^n w_i^2$, respectively. These hold because an IBR experiment always has a smaller worst-case variance than the naive experiment that assigns treatment to each community independently. Finally, the second inequality follows from the fact that $w_i \leq \bar{w}$ for all communities $i \in S_0$ and that $n_0 = |S_0| \leq n$. $\square$

We next analyze the new problem instance, and bound the gap $\tilde{V} - \tilde{V}^{\text{LB}}$. In the new problem instance, communities in a block have equal sizes. Therefore, we can adopt the analysis for Theorem 4.6. Specifically, from (14) we have

$$\tilde{V} - \tilde{V}^{\text{LB}} \leq \sum_{k \in [K]} \underline{w}_k^2 \cdot \Big(5 + n_k \sigma_{kk}^*\Big),$$

where σ^*_{kk} is the optimal correlation coefficient of any two communities in block k obtained from the solution of (5) (that has the form (12)) for the new problem instance. By (16),

$$\underline{w}_k^2 n_k \sigma^*_{kk} \le \min\left\{\underline{w}_k^2 n_k, \frac{\tilde{V}^{\text{LB}}}{4n_k}\right\} \le \frac{\underline{w}_k\sqrt{\tilde{V}^{\text{LB}}}}{2} \le \frac{\underline{w}_k\sqrt{V^{\text{LB}}}}{2}.$$

Thus, we have

$$\tilde{V} - \tilde{V}^{\text{LB}} \le \sum_{k\in[K]} 5\cdot \underline{w}_k^2 + \frac{1}{2}\cdot \underline{w}_k\sqrt{V^{\text{LB}}} \le K\cdot\left(5w_1^2 + \frac{1}{2}w_1\sqrt{V^{\text{LB}}}\right) \le 3Kw_1\sqrt{V^{\text{LB}}}, \tag{18}$$

where the last inequality follows since $4w_1^2 \le V^{\text{LB}}$ by Lemma 4.2.

Combining (17) and (18), we have

$$\begin{aligned}
\frac{V - V^{\text{LB}}}{V^{\text{LB}}} &\le \frac{V - \tilde{V} + \tilde{V} - \tilde{V}^{\text{LB}}}{V^{\text{LB}}} \\
&\le \frac{4n^{-\delta_1}\sum_{i=1}^n w_i^2 + 4(\alpha^2-1)\sum_{i=1}^n w_i^2 + 3Kw_1\sqrt{V^{\text{LB}}}}{V^{\text{LB}}} \\
&\le \frac{4n^{-\delta_1}\sum_{i=1}^n w_i^2 + 4(\alpha^2-1)\sum_{i=1}^n w_i^2 + 6Kw_1\sqrt{\sum_{i=1}^n w_i^2}}{\sum_{i=1}^n w_i^2} \\
&= 4n^{-\delta_1} + 4(\alpha^2-1) + 6K\cdot\sqrt{\frac{w_1^2}{\sum_{i=1}^n w_i^2}} \\
&= O\left(n^{-\delta_1}\right) + O\left(n^{-\delta_2}\right) + O\left(n^{-\frac{c}{2}+\delta_2}\cdot \ln n\right),
\end{aligned}$$

where the third inequality follows from the lower and upper bounds of V^{LB} in Lemma 4.2. Taking $\delta_1 = \delta_2 = \frac{c}{4}$ results in

$$\frac{V - V^{\text{LB}}}{V^{\text{LB}}} = O\left(n^{-\frac{c}{4}}\ln n\right),$$

which finishes the proof of the theorem. □

A.8 Proof of Lemma 5.1

For a community i, we let $\tilde{z}_1^i = \sum_{j\in S_i}\frac{z_{j1}}{\mathbb{P}[z_{j1}=1]}$ and $\tilde{z}_0^i = \sum_{j\in S_i}\frac{z_{j0}}{\mathbb{P}[z_{j0}=1]}$ be the sum of normalized indicators for users in community i, and let $\tilde{z}_i = \tilde{z}_1^i - \tilde{z}_0^i$. Note that when any two users in community i are either connected or have a common neighbor, no user in this community can have the control (treatment) outcome if one user has the treatment (control) outcome. Thus, when $y_{j1} = y_{j0} = 1$ for all nodes $j \in S_i$,

$$\mathbb{V}\text{ar}\left[\hat{\tau}_i\right] = \mathbb{V}\text{ar}\left[\tilde{z}_1^i - \tilde{z}_0^i\right] = \mathbb{V}\text{ar}\left[\tilde{z}_1^i\right] + \mathbb{V}\text{ar}\left[\tilde{z}_0^i\right] - 2\cdot\mathbb{C}\text{ov}\left[\tilde{z}_1^i, \tilde{z}_0^i\right] = \mathbb{V}\text{ar}\left[\tilde{z}_1^i\right] + \mathbb{V}\text{ar}\left[\tilde{z}_0^i\right] + 2w_i^2,$$

where the last equality holds because $\mathbb{E}\left[\tilde{z}_1^i\right] = \mathbb{E}\left[\tilde{z}_0^i\right] = w_i$ and $\mathbb{E}\left[\tilde{z}_1^i\tilde{z}_0^i\right] = 0$. We next show that an experiment that assigns all users in a community to the same variant with equal probability minimizes $\mathbb{V}\text{ar}\left[\tilde{z}_1^i\right]$ and $\mathbb{V}\text{ar}\left[\tilde{z}_0^i\right]$ simultaneously, and hence, it minimizes the variance of the estimator $\hat{\tau}_i$.

To see this, let $A_i \in \{0, 1, \varnothing\}$ be a random variable for community i such that $A_i = 1$ if any user in the community has the treatment outcome, $A_i = 0$ if any user has the control outcome, and $A_i = \varnothing$ otherwise. Note that we have $\{A_i = 1\} \cap \{A_i = 0\} = \varnothing$. By Assumption 2.2, $\mathbb{P}[A_i = 1] = \mathbb{P}[A_i = 0]$. We denote this probability by $p_i \leq \frac{1}{2}$. Since $\mathbb{E}\big[\tilde{z}_1^i\big] = w_i$ and $\mathbb{E}\big[\tilde{z}_1^i \big| A_i = 0\big] = \mathbb{E}\big[\tilde{z}_1^i \big| A_i = \varnothing\big] = 0$, $\mathbb{E}\big[\tilde{z}_1^i \big| A_i = 1\big] = \frac{w_i}{p_i}$. By the law of total variance,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\big[\tilde{z}_1^i\big] &= \mathbb{E}\Big[\mathbb{V}\mathrm{ar}\big[\tilde{z}_1^i \big| A_i\big]\Big] + \mathbb{V}\mathrm{ar}\Big[\mathbb{E}\big[\tilde{z}_1^i \big| A_i\big]\Big] \\
&\geq \mathbb{V}\mathrm{ar}\Big[\mathbb{E}\big[\tilde{z}_1^i \big| A_i\big]\Big] \\
&= (1 - p_i) \cdot w_i^2 + p_i \cdot \left(\frac{w_i}{p_i} - w_i\right)^2 = \frac{w_i^2}{p_i} - w_i^2 \geq w_i^2.
\end{aligned}$$

The first inequality holds with equality if $\mathbb{V}\mathrm{ar}\big[\tilde{z}_1^i \big| A_i = 1\big] = \mathbb{V}\mathrm{ar}\big[\tilde{z}_1^i \big| A_k = 0\big] = 0$. Similarly, the second inequality holds with equality if $p_i = \frac{1}{2}$. Both of these conditions are satisfied when one assigns all of the users simultaneously to either the treatment or the control variant, with probability $\frac{1}{2}$. Thus, the aforementioned experiment minimizes $\mathbb{V}\mathrm{ar}\big[\tilde{z}_1^i\big]$. Analogously, this experiment minimizes $\mathbb{V}\mathrm{ar}\big[\tilde{z}_0^i\big]$ as well, and in turn the variance of the estimator $\hat{\tau}_i$. The corresponding variance is $\mathbb{V}\mathrm{ar}\big[\hat{\tau}_i\big] = 4w_i^2$.

A.9 Proof of Lemma 5.2

Following the notation in Appendix A.8, for a community i, we let $\tilde{z}_1^i = \sum_{j \in S_i} \frac{z_{j1}}{\mathbb{P}[z_{j1}=1]}$ and $\tilde{z}_0^i = \sum_{j \in S_i} \frac{z_{j0}}{\mathbb{P}[z_{j0}=1]}$ be the sum of normalized indicators for the users in community i, and we let $\tilde{z}_i = \tilde{z}_1^i - \tilde{z}_0^i$. Now, suppose users in the same community have the same potential outcomes, i.e., $y_{j1} = y_{j0} = \tilde{y}_i \in [0, 1]$ for all $j \in S_i$. Then, the Horvitz-Thompson estimator is

$$\hat{\tau} = \sum_{i \in [n]} \sum_{i \in S_i} y_{j1} \frac{z_{j1}}{\mathbb{P}\big[z_{j1} = 1\big]} - \sum_{i \in [n]} \sum_{i \in S_i} y_{j0} \frac{z_{j0}}{\mathbb{P}\big[z_{j0} = 1\big]} = \sum_{i \in [n]} \tilde{y}_i \tilde{z}_1^i - \tilde{y}_i \tilde{z}_0^i = \sum_{i \in [n]} \tilde{y}_i \tilde{z}_i,$$

and its variance is

$$\mathbb{V}\mathrm{ar}\big[\hat{\tau}\big] = \mathbb{V}\mathrm{ar}\left[\sum_{i \in [n]} \tilde{y}_i \tilde{z}_i\right] = \tilde{y}^{\mathrm{T}} \mathbb{C}\mathrm{ov}[\tilde{z}] \tilde{y} = \tilde{y}^{\mathrm{T}} \mathrm{diag}\left(\left(\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}\right)_{i \in [n]}\right) \mathbb{C}\mathrm{orr}[\tilde{z}] \mathrm{diag}\left(\left(\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}\right)_{i \in [n]}\right) \tilde{y},$$

where $\tilde{y} = (\tilde{y}_i)_{i \in [n]}$ is the vector of potential outcomes, $\mathbb{C}\mathrm{ov}[\tilde{z}]$ and $\mathbb{C}\mathrm{orr}[\tilde{z}]$ are the covariance and correlation matrices of the random vector $\tilde{z} = (\tilde{z}_i)_{i \in [n]}$, and $\mathrm{diag}\left(\left(\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}\right)_{i \in [n]}\right)$ is a diagonal matrix with the i-th diagonal entry equal to $\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}$ for all $i \in [n]$.

By Assumption 5.1 and the following discussion, we have $\mathbb{V}\mathrm{ar}[\tilde{z}_i] \geq 4w_i^2$. Let $y_i = w_i \tilde{y}_i \in [0, w_i]$ be the aggregate outcome for a community i, and $y = (y_i)_{i \in [n]}$. We have

$$\begin{aligned}
\max_{y_{j0}, y_{j1} \in [0,1], j \in [m]} \mathbb{V}\mathrm{ar}\big[\hat{\tau}\big] &\geq \max_{\tilde{y} \in [0,1]^n} \tilde{y}^{\mathrm{T}} \mathrm{diag}\left(\left(\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}\right)_{i \in [n]}\right) \mathbb{C}\mathrm{orr}[\tilde{z}] \mathrm{diag}\left(\left(\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]}\right)_{i \in [n]}\right) \tilde{y} \\
&\geq \max_{y \in \times_{i \in [n]} [0, w_i]} 4 y^{\mathrm{T}} \mathbb{C}\mathrm{orr}[\tilde{z}] y,
\end{aligned}$$

where the first inequality holds because we restrict potential outcomes of all users in the same community to take the same value, and the second inequality follows because $\sqrt{\mathbb{V}\mathrm{ar}[\tilde{z}_i]} \geq 2w_i$. As a result,

$$V^{\text{OPT}} = \min_{P\in\mathcal{P}} \max_{y_{j0},y_{j1}\in[0,1],j\in[m]} \mathbb{V}\text{ar}\big[\hat{\tau}\big] \geq \min_{P\in\mathcal{P}} \max_{y\in\times_{i\in[n]}[0,w_i]} 4y^{\mathrm{T}}\mathbb{C}\text{orr}[\tilde{z}]y \geq \min_{\Sigma\in\Re} \max_{y\in\times_{i\in[n]}[0,w_i]} 4y^T\Sigma y = V^{\text{LB}},$$

which finishes the proof of the lemma. □

B Structural Properties of IBR Experiments

In this section we shed light on two structural properties of the IBR experiments. Our first property establishes that under an IBR experiment, at most half of the communities in a block can take a positive value in the worst-case potential outcome.

Lemma B.1. *Consider a block with k communities, and let p be the number of communities that take a positive value in the worst-case potential outcome. Then, $p \leq \frac{k}{2}$ if k is even, and $p \leq \frac{k+1}{2}$ if k is odd.*

Proof. To see this, note that when the correlation σ is negative, Lemma 3.2 implies that

$$w_p \geq -2\sigma \sum_{i\leq p-1} w_i \geq -2(p-1)\sigma \cdot w_p,$$

which in turn implies $p \leq -\frac{1}{2\sigma} + 1$. When the block contains an even number k of communities, $\sigma = -\frac{1}{k-1}$ by Proposition 3.1. Thus, p is an integer no larger than $\frac{k}{2}$. Analogously, when the number of communities k is odd, $\sigma = -\frac{1}{k}$, and hence, p is an integer no larger than $\frac{k+1}{2}$. □

For the next result, we index the blocks of an IBR experiment in the decreasing order of the size of the largest community of the block. We show in Lemma B.2 that in the optimal partition obtained from solving the DP, all but the last block contain an even number of communities.

Lemma B.2. *There exists an optimal community partition obtained by solving the DP such that (i) if the number of communities n is even, then all blocks contain an even number of communities, and (ii) if n is odd, then all but the last block contain an even number of communities.*

Proof. We first focus on a single block with k communities with community sizes $w_1 \geq w_2 \geq \cdots \geq w_k$. We highlight three observations. First, clearly, if the vector of community sizes entry-wise decreases, the worst-case variance of this block will weakly decrease. Second, if we drop any of the communities from the block, the worst-case variance of the block will weakly decrease as well because assignments to the remaining communities become more negatively correlated. Third, if the number of communities k is odd, adding a community of size $w' \leq w_k$ does not change the worst-case variance of this block. To see this, note that after the addition, the correlation between the assignments to any two communities in the block does not change (and remains to be $-\frac{1}{k}$ by Proposition 3.1). Moreover, the worst-case potential outcome does not change as well, due to Lemma 3.2 and Lemma B.1.

Now, we turn to the optimal partition by solving the DP. Let K denote the number of blocks, and $S_k = \{w_1^k \geq w_2^k \geq \cdots \geq w_{n_k}^k\}$ denote the set of communities (sorted in decreasing order of the sizes) in block $k \in [K]$. If a block $h \leq K-1$ has an odd number of communities, consider a new partition $\{S'_k\}_{k\in[K]}$ as follows: $S'_k = S_k$ for any $k \leq h-1$, $S'_h = S_h \cup \{w_1^{h+1}\}$, $S'_k = S_k \setminus \{w_1^k\} \cup \{w_1^{k+1}\}$ for any $h+1 \leq k \leq K-1$, and $S'_K = S_K \setminus \{w_1^K\}$. By the former discussion, the worst-case variance of each block weakly decreases. Thus, $\{S'_k\}_{k\in[K]}$ is a weakly better partition than $\{S_k\}_{k\in[K]}$. Iterating this last step finishes the proof. □

C More on the First Facebook Example in Section 7.2.1

We elaborate more on the small Facebook example of Section 7.2.1 in this appendix. As a reminder, check Figure 4 and Table 2 to see the community-level structure, community sizes, and the number of users in each community contaminated by other users of other communities. The optimal community-level experiment for this example is provided in Table 4. The induced correlation structure was reported in Section 7.2.1. Note that the optimal experiment deliberately introduces some positive correlations between small communities to attain a larger negative correlation between a small and a large community. The worst-case potential outcome is $y^{\text{OPT}} = [1190, 747, 0, 537, 0, 203, 0]^{\text{T}}$. The worst-case variance is $V_c^{\text{OPT}} = 5.66 \times 10^6$. The optimal community-level experiment or the corresponding worst-case potential outcome may not to be unique.

The correlation matrix of the optimal IBR experiment (computed by solving a DP) is

$$\Sigma^{\text{DP}} = \begin{pmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 \end{pmatrix}.$$

The worst-case community-level potential outcome for this experiment is $y^{\text{DP}} = [1190, 0, 0, 0, 315, 0, 0]^{\text{T}}$. The worst-case variance is $V^{\text{DP}} = 6.06 \times 10^6$, which increases the worst-case variance of the actual optimal community-level experiment by $\frac{V^{\text{DP}} - V_c^{\text{OPT}}}{V_c^{\text{OPT}}} = 7\%$.

We now compare our IBR experiment with the three heuristic experiments `HALF`, `PAIR`, and `IND`—introduced in Section 7.2.1. First, the correlation matrix of the `HALF` experiment is

$$\Sigma^{\text{half}} = \frac{8}{7} I - \frac{1}{7} \mathbf{1}\mathbf{1}^{\text{T}} = \begin{pmatrix} 1 & -1/7 & -1/7 & -1/7 & -1/7 & -1/7 & -1/7 \\ -1/7 & 1 & -1/7 & -1/7 & -1/7 & -1/7 & -1/7 \\ -1/7 & -1/7 & 1 & -1/7 & -1/7 & -1/7 & -1/7 \\ -1/7 & -1/7 & -1/7 & 1 & -1/7 & -1/7 & -1/7 \\ -1/7 & -1/7 & -1/7 & -1/7 & 1 & -1/7 & -1/7 \\ -1/7 & -1/7 & -1/7 & -1/7 & -1/7 & 1 & -1/7 \\ -1/7 & -1/7 & -1/7 & -1/7 & -1/7 & -1/7 & 1 \end{pmatrix}.$$

The worst-case community-level potential outcome is $y^{\text{half}} = [1190, 747, 741, 0, 0, 0, 0]^{\text{T}}$. The worst-case variance is $V^{\text{half}} = 7.44 \times 10^6$, and this increases the worst-case variance of the optimal community-level experiment by $\frac{V^{\text{half}} - V_c^{\text{OPT}}}{V_c^{\text{OPT}}} = 31.3\%$. If we instead combine the last two communities together and randomly assign half of the final six communities to treatment, the correlation

	1	2	3	4	5	6	7	probability
1	×	×	√	√	√	√	√	0.0539
2	√	√	×	×	×	×	×	0.0539
3	×	√	×	√	√	√	√	0.0532
4	√	×	√	×	×	×	×	0.0532
5	×	√	√	×	√	×	√	0.0502
6	√	×	×	√	×	√	×	0.0502
7	×	√	×	√	√	×	×	0.0462
8	√	×	√	×	×	√	√	0.0462
9	×	√	√	×	×	√	×	0.0366
10	√	×	×	√	√	×	√	0.0366
11	×	√	×	√	×	√	×	0.0359
12	√	×	√	×	√	×	√	0.0359
13	×	√	√	√	×	×	×	0.0336
14	√	×	×	×	√	√	√	0.0336
15	×	√	√	×	√	√	×	0.0313
16	√	×	×	√	×	×	√	0.0313
17	×	√	×	√	×	×	√	0.0236
18	√	×	√	×	√	√	×	0.0236
19	×	×	√	√	×	×	√	0.0226
20	√	√	×	×	√	√	×	0.0226
21	×	×	√	√	×	√	×	0.0209
22	√	√	×	×	√	×	√	0.0209
23	×	×	√	×	√	×	×	0.0180
24	√	√	×	√	×	√	√	0.0180
25	×	×	√	×	×	×	√	0.0180
26	√	√	×	√	√	√	×	0.0180
27	×	√	√	×	×	√	√	0.0179
28	√	×	×	√	√	×	×	0.0179
29	×	×	√	×	√	√	√	0.0139
30	√	√	×	√	×	×	×	0.0139
31	×	×	√	√	√	×	√	0.0109
32	√	√	×	×	×	√	×	0.0109
33	×	×	×	×	×	√	√	0.0077
34	√	√	√	√	√	×	×	0.0077
35	×	×	×	√	√	×	×	0.0056
36	√	√	√	×	×	√	√	0.0056

Table 4: The randomized joint assignment of the optimal community-level experiment. Each of the 36 rows corresponds to a possible assignment, where $\surd$ denotes treatment and $\times$ denotes control.

matrix of the experiment becomes

$$\Sigma^{\mathrm{half}'} = \begin{pmatrix} 1 & -1/5 & -1/5 & -1/5 & -1/5 & -1/5 & -1/5 \\ -1/5 & 1 & -1/5 & -1/5 & -1/5 & -1/5 & -1/5 \\ -1/5 & -1/5 & 1 & -1/5 & -1/5 & -1/5 & -1/5 \\ -1/5 & -1/5 & -1/5 & 1 & -1/5 & -1/5 & -1/5 \\ -1/5 & -1/5 & -1/5 & -1/5 & 1 & -1/5 & -1/5 \\ -1/5 & -1/5 & -1/5 & -1/5 & -1/5 & 1 & 1 \\ -1/5 & -1/5 & -1/5 & -1/5 & -1/5 & 1 & 1 \end{pmatrix}.$$

The worst-case community-level potential outcome is $y^{\mathrm{half}'} = [1190, 747, 0, 0, 0, 0, 0]^{\mathrm{T}}$. The worst-case variance is $V^{\mathrm{half}'} = 6.47 \times 10^6$, and this increases the worst-case variance of the optimal community-level experiment by $\frac{V^{\mathrm{half}'} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 14.3\%$.

Second, the correlation matrix of the `PAIR` experiment is

$$\Sigma^{\mathrm{pair}} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The worst-case community-level potential outcome is $y^{\mathrm{pair}} = [1190, 0, 741, 0, 315, 0, 59]^{\mathrm{T}}$. The worst-case variance is $V^{\mathrm{pair}} = 8.27 \times 10^6$, and this increases the worst-case variance of the optimal community-level experiment by $\frac{V^{\mathrm{pair}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 46.0\%$.

Finally, the correlation matrix of the `IND` experiment is the identity matrix $\Sigma^{\mathrm{ind}} = I$. The worst-case community-level potential outcome is $y^{\mathrm{ind}} = [1190, 747, 741, 537, 315, 203, 59]^{\mathrm{T}}$, the worst-case variance is $V^{\mathrm{ind}} = 1.18 \times 10^7$, and this increases the worst-case variance of the optimal community-level experiment by $\frac{V^{\mathrm{ind}} - V_c^{\mathrm{OPT}}}{V_c^{\mathrm{OPT}}} = 108.5\%$.

D Average-Case Analysis

In this section, we conduct an average-case analysis by comparing different experiments' variances on the same potential outcomes (in contrast with their respective worst-case outcomes as in Section 7), where the potential outcomes are drawn randomly from a given distribution.

From the expression of the variance of the Horvitz-Thompson estimator for a community-level experiment (see (8) in Appendix A.1), we can see that the variance only depends on the users' potential outcomes $\{y_{j1}\}$ and $\{y_{j0}\}$ through their summations in each community, i.e., $\sum_{j \in S_i} y_{j1}$ and $\sum_{j \in S_i} y_{j0}$, where S_i is the set of users in community i. Hence, it suffices to consider random draws for the total outcomes of each community in our average-case analysis instead of sampling each individual user's potential outcomes. In the remaining of this section, with some notational abuse, we denote the total treatment and control outcomes by y_{i1} and y_{i0} respectively, for each community i.

D.1 The Case with Equal-Size Communities

We first consider the case when communities have equal sizes, and we compare the `HALF`, `PAIR` and `IND` experiments as described in Section 7.2.1 on the same potential outcomes, randomly drawn from a given distribution. From Proposition 3.1, the `HALF` experiment is optimal with respect to minimizing the variance against the worst-case potential outcomes, and our optimal IBR experiment reduces to it. The `PAIR` experiment has an approximation ratio no larger than 2.5 by Lemma 4.3. With some abuse of notation, we let V^{half}, V^{pair} and V^{ind} denote the variances of the Horvitz-Thompson estimator under the `HALF`, `PAIR` and `IND` experiments, respectively; these are random variables that depend on the value of the potential outcomes. Since the potential outcomes are nonnegative, V^{half} and V^{pair} are always smaller than V^{ind} (due to the negative correlation induced by these experiments between the treatment assignments of different communities). We are interested in quantitatively measuring the extent of this variance reduction for different realizations of the randomized potential outcomes.

For simplicity and exposition of our main point, we consider the following three cases for the underlying distribution of the total potential outcomes $\{y_{i1}, y_{i0}\}$:

Case 1: Sample $y_{i1}, y_{i0} \sim \text{Unif}[0, 1]$ for each community i, all *i.i.d.*;

Case 2: Sample $y_{i0} \sim \text{Unif}[0, 1]$, *i.i.d.*, and let $y_{i1} = y_{i0} + 0.1$ for each community i; and

Case 3: Sample $y_{i1}, y_{i0} \sim \text{Unif}[0.5, 1.5]$ for each community i, all *i.i.d.*.

In all three cases, we consider potential outcomes that are independent across different communities, simply to avoid assuming a specific stylized correlation structure with no roots in practice. We further let the sampling distributions to be identical across communities for simplicity.

In the simulation, we vary the number of communities n from 20 to 100 with a step size of 20. For each case and each value of n, we sample the potential outcomes 10^4 times and plot the box-plots for the quantities V^{ind}/n, V^{pair}/n and V^{half}/n and the ratio $V^{\text{ind}}/V^{\text{half}}$ in Figure 5.

As can be seen from Figure 5, the `HALF` and `PAIR` experiments reduce the variance drastically compared to the `IND` experiment in all the three cases and for all of the realization of potential outcomes—especially when the potential outcomes are away from zero (i.e., Case 3). The `HALF` and `PAIR` experiments have comparable variances under the three sampling distributions (see the right column of Figure 5). Notably, the variance of the `HALF` experiment is more concentrated around the median. Moreover, as we have shown in our theoretical investigation, this experiment yields a smaller worst-case variance compared to the `PAIR` experiment. Hence, it is more robust to the choice of values for the unknown potential outcomes.

D.2 Facebook Subnetworks of US Universities

In this section, we revisit the Facebook subnetworks of US universities in Section 7.2.2 and conduct an average-case analysis. Specifically, we compare the optimal IBR experiment and the `HALF`, `PAIR` and `IND` experiments on the same potential outcomes, again randomly drawn from a given distribution. Let V^{DP}, V^{half}, V^{pair} and V^{ind} denote the variances of the Horvitz-Thompson estimator for the optimal IBR experiment, the `HALF` experiment, the `PAIR` experiment, and the `IND` experiment, respectively. These are random variables that depend on the value of the potential outcomes.

We consider the following three cases for the underlying distribution of the total potential outcomes $\{y_{i1}, y_{i0}\}$:

Case 1: Sample $y_{i1}, y_{i0} \sim \text{Unif}[0, w_i]$ for each community i, all *i.i.d.*;

Case 2: Sample $y_{i0} \sim \text{Unif}\left[0, \frac{w_i}{2}\right]$, *i.i.d.*, and let $y_{i1} = y_{i0} + 0.2w_i$ for each community i; and

Case 3: Sample $y_{i0} \sim \text{Unif}\left[0, \frac{w_i}{2}\right]$, *i.i.d.*, and let $y_{i1} = y_{i0} + 0.4w_i$ for each community i.

Now, let $\tau_a \triangleq \frac{\tau}{m}$ denote the average treatment effect, where $m = \sum_{i\in[n]} w_i$ is the number of users and τ is the total market effect. Note that τ_a is approximately zero in Case 1, $\tau_a = 0.2$ in Case 2 and $\tau_a = 0.4$ in Case 3. We further let $\sigma_{\text{DP}} \triangleq \frac{\sqrt{V^{\text{DP}}}}{m}$ denote the standard deviation of the Horvitz-Thompson estimator for τ_a under the optimal IBR experiment. Analogously, we let $\sigma_{\text{half}} \triangleq \frac{\sqrt{V^{\text{half}}}}{m}$, $\sigma_{\text{pair}} \triangleq \frac{\sqrt{V^{\text{pair}}}}{m}$, and $\sigma_{\text{ind}} \triangleq \frac{\sqrt{V^{\text{ind}}}}{m}$ denote the standard deviations of the Horvitz-Thompson estimator for τ_a under the `HALF`, `PAIR` and `IND` experiments, respectively. In the simulation, we randomly draw the potential outcomes 10^4 times for each case and we plot the box-plots of the values σ_{DP}, σ_{half}, σ_{pair} and σ_{ind} and the ratios $\sigma_{\text{half}}/\sigma_{\text{DP}}$ and $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ in Figure 6.

As can be seen from Figure 6, the optimal IBR experiment reduces the variance substantially compared to the `HALF` and `IND` experiments in all of the cases and under all realizations of the potential outcomes. The `PAIR` experiment has a marginally smaller variance on average under the three sampling distributions[35] (see the right column of Figure 6). On the other hand, the optimal IBR experiment attains a smaller worst-case variance, its variance is more concentrated around the median, and hence it is more robust to the unknown potential outcomes. We also highlight that the observation that the `PAIR` experiment has a smaller variance (on average) than our IBR experiment is substantially an artifact of the choice of the sampling distribution. In particular, the assumption that the potential outcomes are independent across communities is indeed the source of this observation. When, for example, the potential outcomes are negatively correlated between communities of similar sizes, the variance of the `PAIR` experiment is in general larger than the variance of the optimal IBR experiment.[36] From all three cases in this example, it can be seen that although the optimal IBR experiment is designed with the goal of minimizing the worst-case variance, it maintains the same performance or even reduces the variance on average in comparison to other heuristic experiments.

Finally, we plot the box-plot of the ratio $\sigma_{\text{DP}}/\tau_a$, which is the standard deviation of the Horvitz-Thompson estimator of the average treatment effect over the true value, for Cases 2 and 3 in Figure 7. From Figure 7, the standard deviation is relatively small compared to the true average treatment effect, and this demonstrates the statistical power of the estimator. We elaborate more on this point in Appendix E.

E Statistical Inference from IBR Experiments

In this section, we discuss a way to construct the confidence interval for estimating the total market effect with an IBR experiment. The Horvitz-Thompson estimator $\hat{\tau}$ is unbiased.[37] Now fix the potential outcomes for treatment and control. When the number of blocks is large, by the central limit theorem and the fact that assignments are independent across blocks, the distribution of the estimator $\hat{\tau}$ is approximately normal. When the number of blocks is small, we assume $\hat{\tau}$ is approximately normal as well. Hence, a level α-confidence interval for the total market effect can

[35] Specifically, the median of the ratio $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ is 0.98 in Cases 1 and 2 and 0.97 in cases 3.

[36] This can happen when the potential outcomes are correlated through (possibly unobserved) covariates and these covariates are quite different across communities of similar sizes.

[37] Or almost unbiased when restricting to the uncontaminated users, as discusses in Section 5.2.

be given by $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\mathbb{V}\mathrm{ar}[\hat{\tau}]}, \hat{\tau} + z_{\alpha/2}\sqrt{\mathbb{V}\mathrm{ar}[\hat{\tau}]}\right]$, where $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, with $\Phi(\cdot)$ being the CDF of a standard normal distribution.

The problem, however, is that we are not able to compute the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$, as it depends on *all* the values of potential outcomes, which cannot be observed simultaneously (the expression of $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ is given in (8)). We may use the worst-case variance of the IBR experiment as a surrogate to $\mathbb{V}\mathrm{ar}[\hat{\tau}]$, but this can be quite loose especially when some of the potential outcomes are observed after the experiment.

Analogous to Section 4.3 of Imai et al. (2009) and Section 4.2 of Bojinov et al. (2020), we consider a conservative estimator for the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$. Specifically, for each community $i \in [n]$, we let $y_i^{\mathrm{obs}} \triangleq \mathbb{1}[z_i = 1] \cdot y_{i1} + \mathbb{1}[z_i = 0] \cdot y_{i0}$ denote the observed outcome for community i. We use the following estimator $\hat{\sigma}^2$ for the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$, with

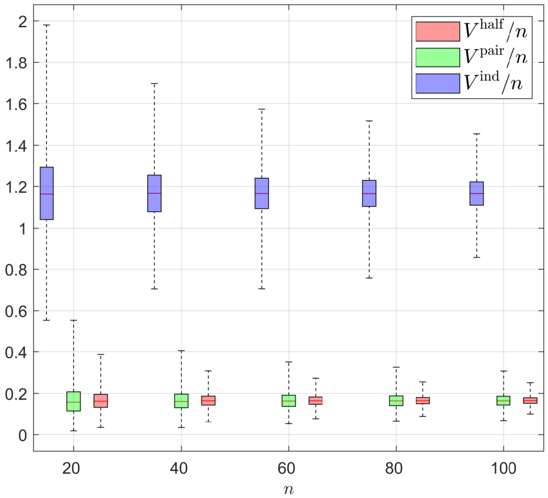
$$\hat{\sigma}^2 \triangleq 4 \sum_{i\in[n]} \left(y_i^{\mathrm{obs}}\right)^2 + 2 \sum_{i\in[n]} \sum_{[n]\ni k\neq i} \sigma_{ik} \left(\left(y_i^{\mathrm{obs}}\right)^2 + \left(y_k^{\mathrm{obs}}\right)^2\right),$$

where σ_{ik} is the correlation between communities i and k as in (8). The mean of $\hat{\sigma}^2$ provides an upper bound on the variance $\mathbb{V}\mathrm{ar}[\hat{\tau}]$ because

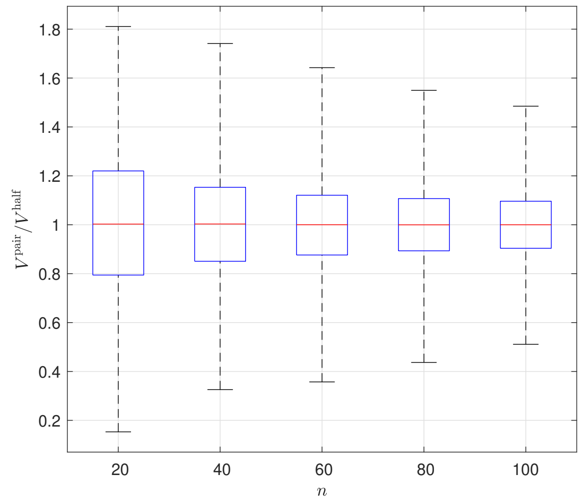
$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2\right] &= 2 \sum_{i\in[n]} \left(y_{i1}^2 + y_{i0}^2\right) + \sum_{i\in[n]} \sum_{[n]\ni k\neq i} \sigma_{ik}\Big(y_{i1}^2 + y_{i0}^2 + y_{k1}^2 + y_{k0}^2\Big) \\
&\geq \sum_{i\in[n]} \left(y_{i1} + y_{i0}\right)^2 + \sum_{i\in[n]} \sum_{[n]\ni k\neq i} \sigma_{ik}\Big(y_{i1}y_{k1} + y_{i0}y_{k0} + y_{i1}y_{k0} + y_{i0}y_{k1}\Big) \\
&= \mathbb{V}\mathrm{ar}[\hat{\tau}],
\end{aligned}$$

where the inequality follows from the basic inequality $2xy \leq x^2 + y^2$ and the second equation follows from (8).

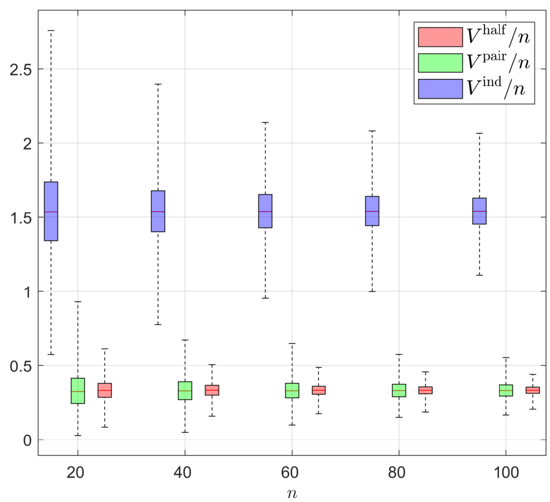
Following Imai et al. (2009) and Bojinov et al. (2020), we suggest using $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\hat{\sigma}^2}, \hat{\tau} + z_{\alpha/2}\sqrt{\hat{\sigma}^2}\right]$ for the level α-confidence interval of the total market effect τ. This is a common heuristic, and we leave its formal analysis for future work.



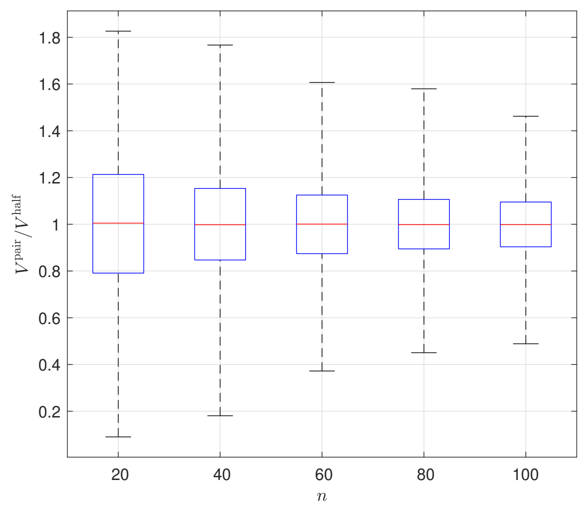
(a) Case 1: V^{ind}/n, V^{pair}/n and V^{half}/n



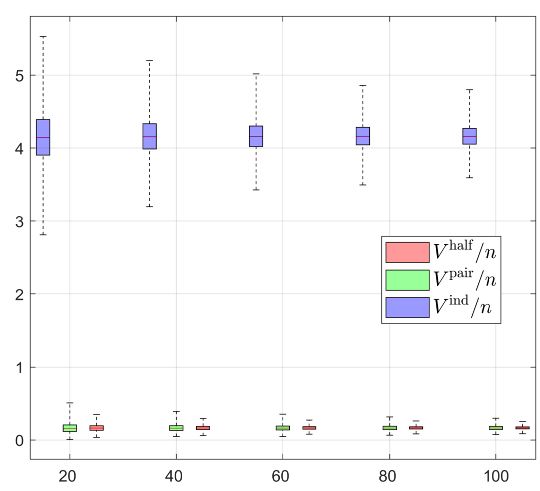
(b) Case 1: $V^{\mathrm{pair}}/V^{\mathrm{half}}$



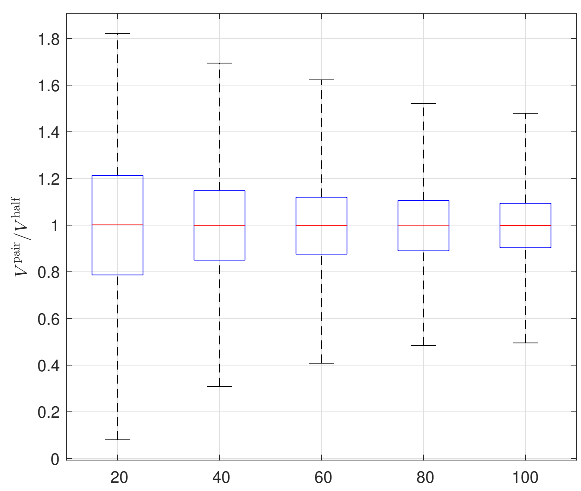
(c) Case 2: V^{ind}/n, V^{pair}/n and V^{half}/n



(d) Case 2: $V^{\mathrm{pair}}/V^{\mathrm{half}}$

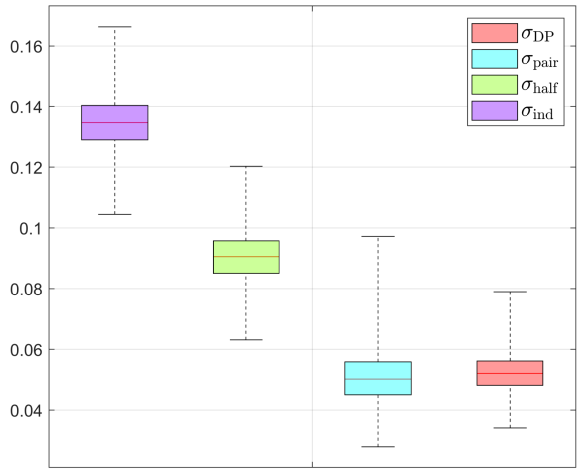


(e) Case 3: V^{ind}/n, V^{pair}/n and V^{half}/n

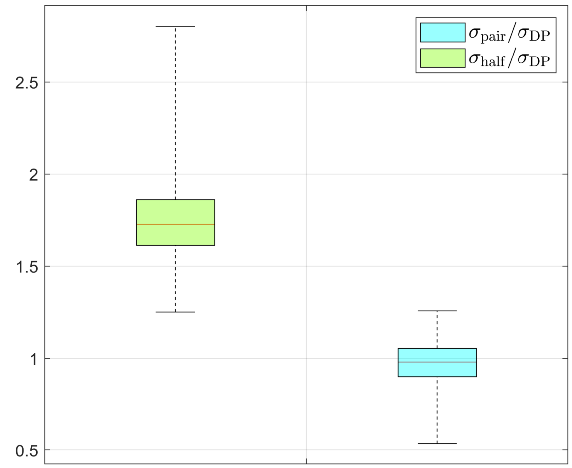


(f) Case 3: $V^{\mathrm{pair}}/V^{\mathrm{half}}$

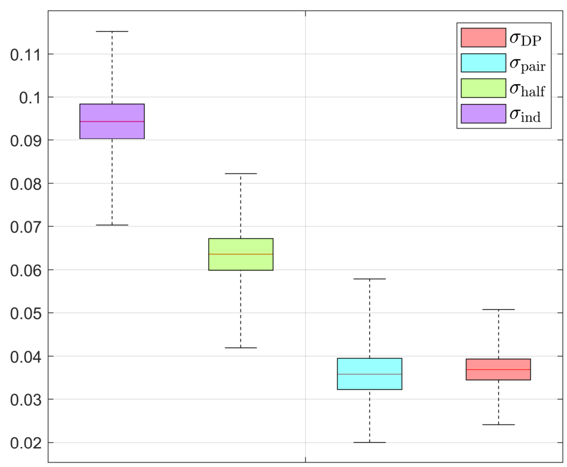
Figure 5: Box-plots of the values V^{ind}/n, V^{pair}/n and V^{half}/n (left column) and the ratio $V^{\mathrm{ind}}/V^{\mathrm{half}}$ (right column) over 10^4 samples for each case. The interpretation of the box-plots is the same as in Figure 3.



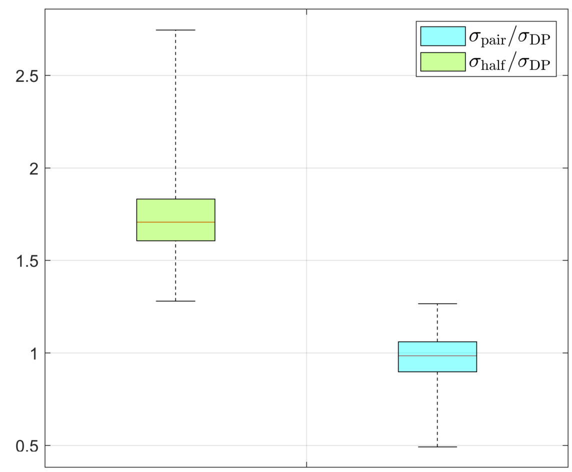
(a) Case 1: σ_{DP}, σ_{half}, σ_{pair} and σ_{ind}



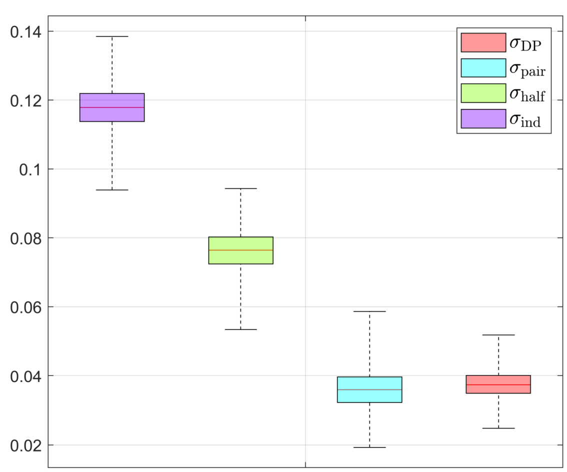
(b) Case 1: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



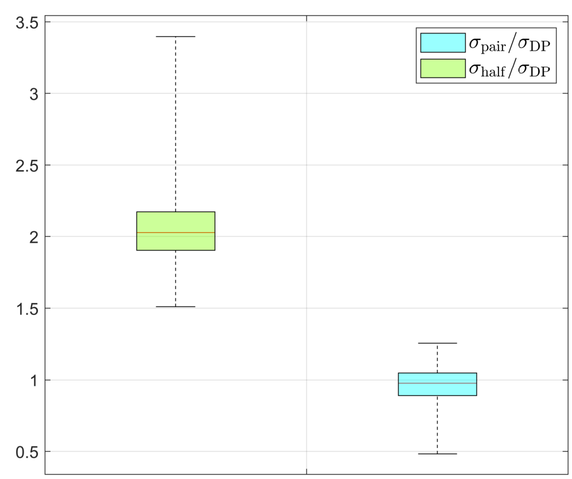
(c) Case 2: σ_{DP}, σ_{half}, σ_{pair} and σ_{ind}



(d) Case 2: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



(e) Case 3: σ_{DP}, σ_{half}, σ_{pair} and σ_{ind}



(f) Case 3: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$

Figure 6: Box-plots of the values σ_{DP}, σ_{half}, σ_{pair} and σ_{ind} (left column) and the ratios $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$ (right column) over 10^4 samples for each case. The interpretation of the box-plots is the same as in Figure 3.

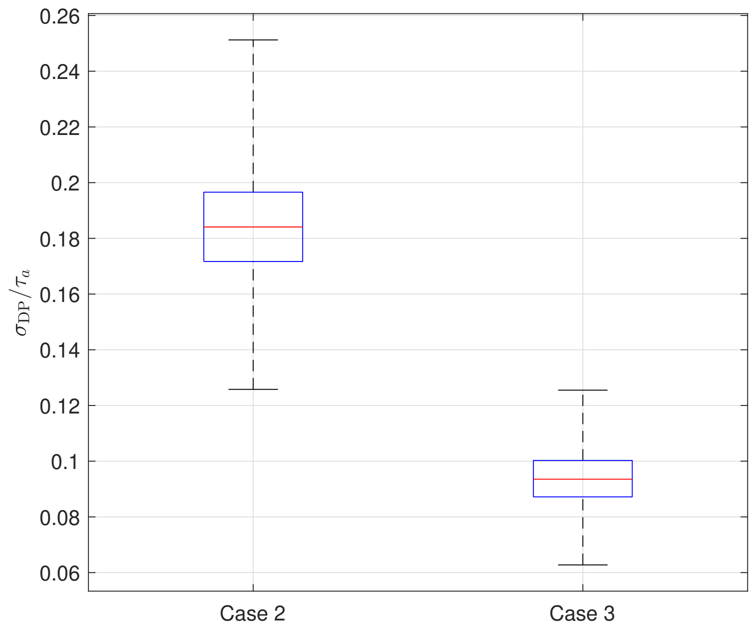


Figure 7: Box-plot of the ratio $\sigma_{\mathrm{DP}}/\tau_a$ over 10^4 samples for Cases 2 and 3. The interpretation of the box-plot is the same as in Figure 3.