CIFellows 2020-2021

Computing Innovation Fellows

Restricted Randomizations to Reduce the Variance of Causal Inference Estimator in Network-based Experiments[1]

Amir Ghasemian,

Human Nature Lab, Yale University

I. Introduction

- randomized trials are methodologically justified in order to achieve an unbiased estimate in causal inference, however, an estimator even unbiased can be inefficient if it has a large variance,
- standard difference-in-means estimator using Bernoulli or complete randomization outputs large variance estimates and the variance is even larger on networks, where the systematic relation among units can increase the variance of inferences through mechanisms such as Homophily and interference,
- to reduce the variance of estimators, we restrict the randomizations using a model-assisted design.

II. Model-Assisted Design[2]

model:

$$X_i \mid g_i \sim N(\mu_{g_i}, \sigma^2)$$

$$Y_i(0) \mid g_i, A \sim N\left(\mathbf{hom}((X_j)_{j \in \tilde{\mathcal{N}}_i}) + \beta_0 \mathbf{intrf}((Z_j)_{j \in \mathcal{N}_i}), \gamma^2\right)$$

$$Y_i(1) = Y_i(0) + \tau$$

g_i : community that node i belongs to

A : adjacency matrix

$\mathcal{N}_i / \tilde{\mathcal{N}}_i$: neighborhood of node i in-/ex- cluding i itself

τ : additive causal effect

estimator (direct effect):

$$\hat{\tau}(Y|Z) = \sum_i \frac{Y_i Z_i}{n_t} - \frac{Y_i(1 - Z_i)}{n_c}$$

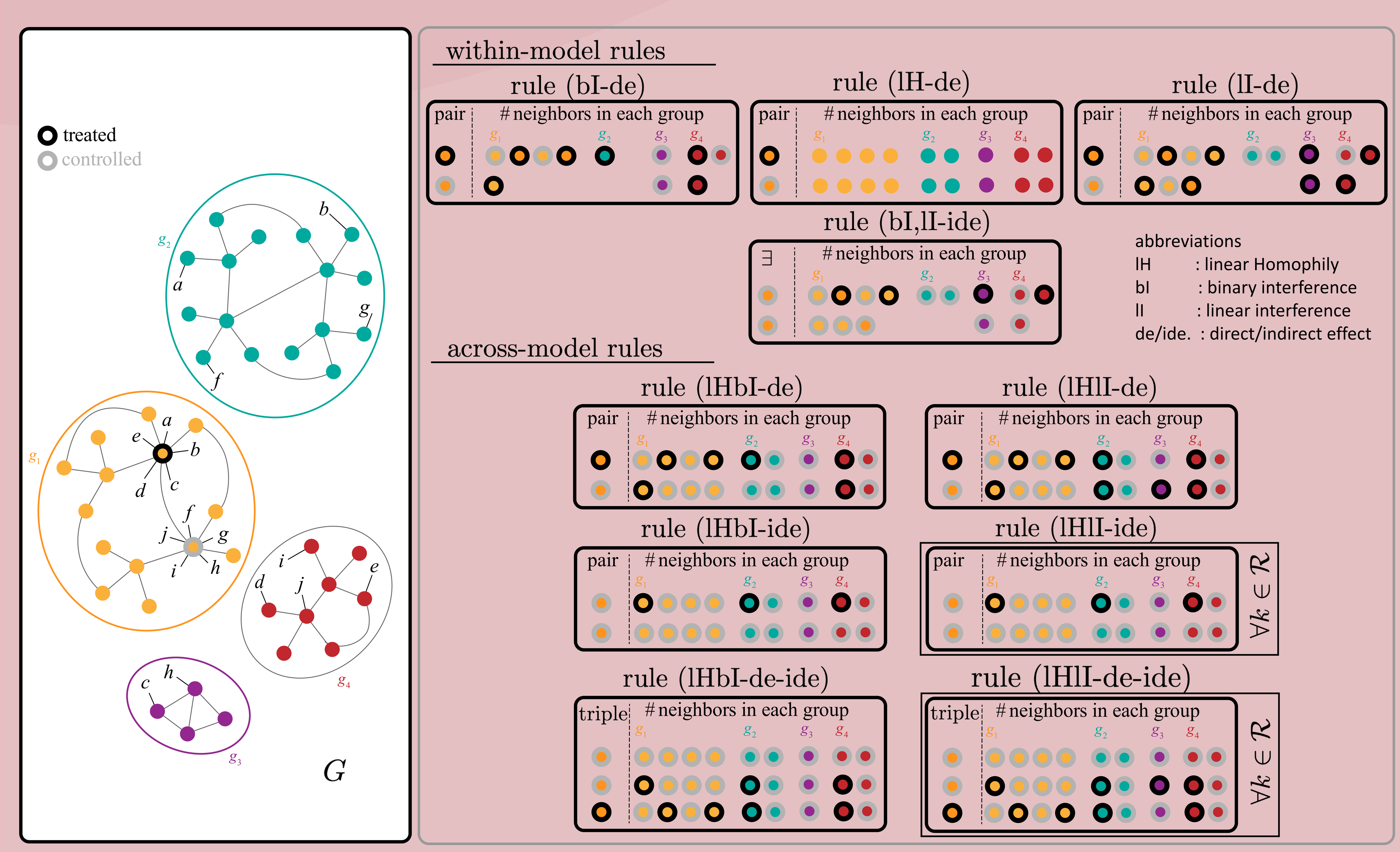
(indirect effect is removed.)

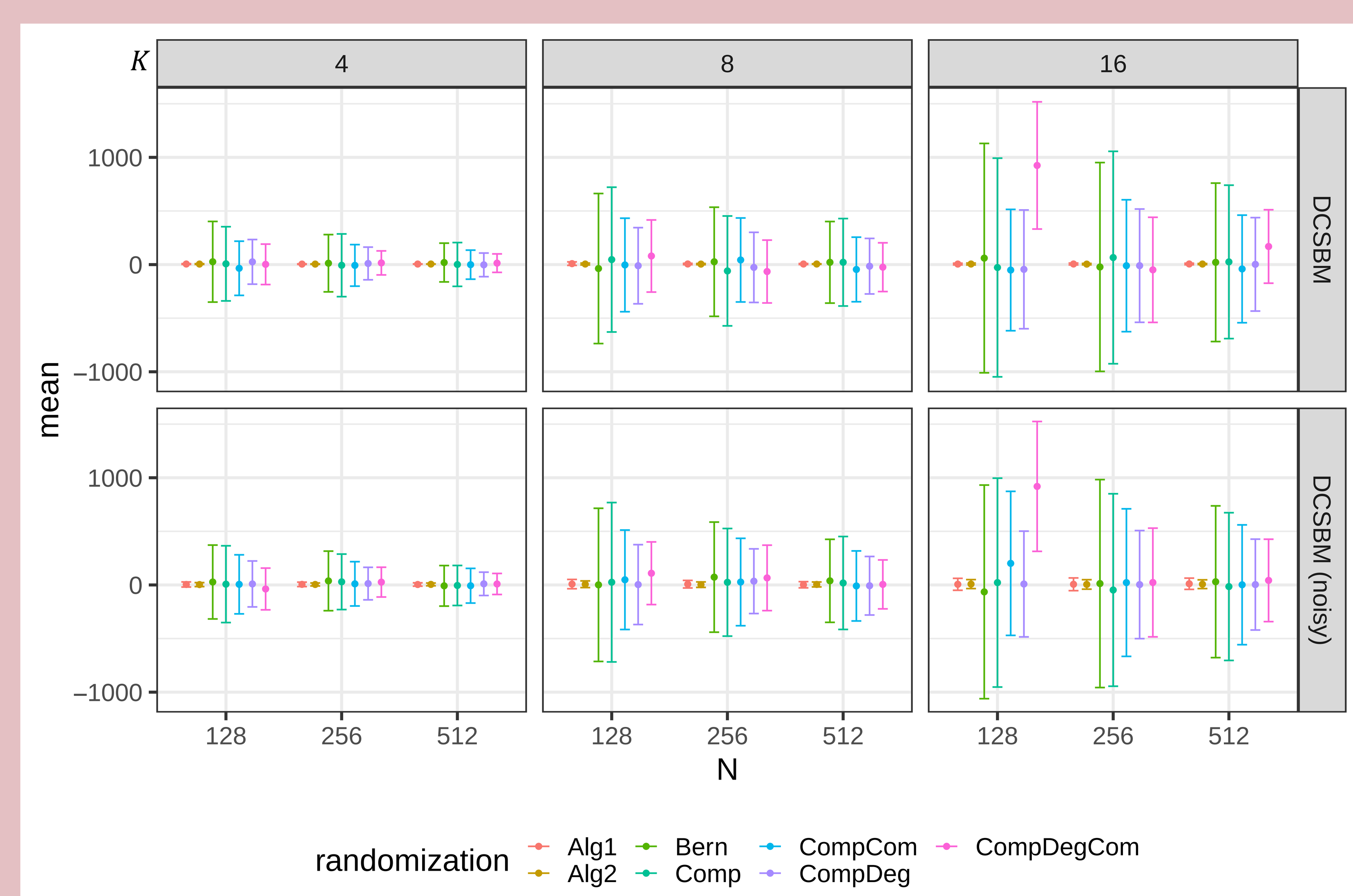
objective:

$$\begin{aligned} \mathrm{MMSE}(\hat{\tau}) &= E[(\tau - \hat{\tau})^2] = E[\mathrm{MSE}(\hat{\tau} \mid Z)] \\ &= E[\mathrm{bias}(\hat{\tau} \mid Z)^2 + \mathrm{var}(\hat{\tau} \mid Z)] \qquad (1) \end{aligned}$$

design: minimizing the mean squared error using simple rule-based designs or using MCMC approach.

III. Rule-based Designs and Results





The mean and standard error of average treatment effect (ATE) using a difference-in-means estimator through different randomizations (results are presented only for lH model). The Alg. 1 is the algorithm corresponding to the rule (lH-de) and the Alg. 2 is the Metropolis–Hastings algorithm to minimize Eq. (1). In Alg. 1, we select the assignment corresponding to minimum conditional bias in Eq. (1) among 20 randomizations (a rerandomization technique). Other randomizations are Bern: Bernoulli randomization, Comp: complete randomization, CompCom: complete randomization in cluster level, CompDeg: complete randomization in degree quantile strata, and CompDegCom: complete randomization in the degree quantile strata for each cluster level.

V. Conclusion

- due to the limitations of two traditional paradigms of design-based and model-based, the new mixed approach is taking advantages of both to introduce novel restricted randomizations with desired properties such as unbiasedness and minimum variance for difference-in-means estimator,
- the proposed method is useful for developing model-assisted design strategies for estimating other causal effects in more complicated settings,
- a crucial feature of the proposed model is that the computation of marginal MSE is analytically tractable, and we can use the results for sample size calculations.

References:

1. Amir Ghasemian, Minzhengxiong Zhang, and Edoardo M Airoldi. Restricted randomizations and approximate marginal MSE in the presence of homophily and interference. (under preparation)
2. Guillaume W Basse and Edoardo M Airoldi. Model-assisted design of experiments in the presence of network-correlated outcomes. Biometrika, 105(4):849–858, 2018.

CRA Computing Research Association

CCC Computing Community Consortium Catalyst

NSF