Low-Degree Outcomes and Clustered Designs

Combining Design and Analysis for Causal Inference under Interference

Samir Khan [1] Matthew Eichhorn [2] Johan Ugander [1] Christina Lee Yu [2]

[1]Stanford University [2]Cornell University

Introduction

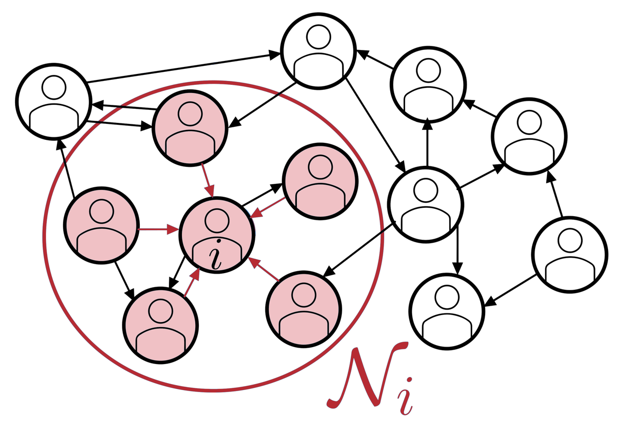
Two lines of work on causal inference under interference:

1. **Estimator Design:** Posit a structured potential outcome model that informs an estimation strategy
2. **Experimental Design:** Posit structural assumptions on the interference that inform an experimental design

Research Question
What happens when we combine sophisticated estimation strategies with complex experimental designs?

Set-up and Notation

- n units, treatment assignments $\mathbf{z} = (z_1, \cdots, z_n) \in \{0,1\}^n$
- Potential outcomes $Y_i(\mathbf{z})$ depend on treatments of unit i's in-neighborhood $\mathcal{N}_i$ in interference graph G



- Degree-β outcomes as in Cortez et al. [2022]:

$$Y_i(\mathbf{z}) = \sum_{\mathcal{S}\in\mathcal{S}_i^\beta} c_{i,\mathcal{S}} \prod_{j\in\mathcal{S}} z_j =: \langle \mathbf{c}_i, \tilde{\mathbf{z}}_i\rangle,$$

where $\mathcal{S}_i^\beta = \{\mathcal{S}\subseteq\mathcal{N}_i \colon |\mathcal{S}|\leq\beta\}$, $\tilde{\mathbf{z}}_i$ is a vector indicating treatments of subsets, and $\mathbf{c}_i$ are model coefficients

- We would like to estimate

$$\mathrm{TTE} := \frac{1}{n}\sum_{i=1}^n Y_i(\mathbf{1}) - Y_i(\mathbf{0}) = \frac{1}{n}\sum_{i=1}^n \sum_{\mathcal{S}\in\mathcal{S}_i^\beta\setminus\{\varnothing\}} c_{i,\mathcal{S}}.$$

The Pseudoinverse Estimator

- Using the pseudoinverse estimator of Swaminathan et al. [2017], we can estimate each $\mathbf{c}_i$ by $\hat{\mathbf{c}}_i = Y_i(\mathbf{z})\mathbb{E}_{\mathbf{z}}\left[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\intercal\right]^\dagger\tilde{\mathbf{z}}_i$.
- This leads to the estimator

$$\widehat{\mathrm{TTE}} := \frac{1}{n}\sum_{i=1}^n \langle\hat{\mathbf{c}}_i, \boldsymbol{\theta}_i\rangle,$$

where $\boldsymbol{\theta}_i = (0, 1, \cdots, 1)$.

Graph Cluster Randomization (GCR)

- Partition units into m clusters $\mathcal{C} = \{C_1, \ldots, C_m\}$
- $\mathcal{C}(i) :=$ cluster containing node i, $\quad \mathcal{C}(\mathcal{S}) := \{\mathcal{C}(i) \colon i\in\mathcal{S}\}$.
- Sample independent Ber(p) random variables $W_1, \cdots, W_m$, and set each $z_i = W_{\mathcal{C}(i)}$ [Ugander et al., 2013]

Variance Bounds for Arbitrary Designs

Theorem
Suppose that $\|\mathbf{c}_i\|_2 \leq M$ for all i and that $\boldsymbol{\theta}_i^\intercal\mathbf{c}_i$ has the same sign for all units i. Then

$$\mathrm{var}(\widehat{\mathrm{TTE}}) \leq \frac{M^2}{n^2}\sum_{i,j=1}^n \gamma_i\gamma_j \mathbf{1}\{\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j\},$$

where $\gamma_i = \sqrt{|\mathcal{S}_i^\beta| \cdot \boldsymbol{\theta}_i^\intercal \mathbb{E}[\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^\intercal]^\dagger\boldsymbol{\theta}_i}$.

- γ_i captures the contribution of node i to the variance
- $\mathbf{1}\{\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j\}$ captures the graph structure's effect on variance

Implications for Clustered Designs

We specialize our results for the case of clustered designs by computing

$$\gamma_i^2 \leq \begin{cases} 2d_i^\beta \cdot p^{-|\mathcal{C}(\mathcal{N}_i)|} & |\mathcal{C}(\mathcal{N}_i)| < \beta, \\ 2d_i^\beta \cdot |\mathcal{C}(\mathcal{N}_i)|^\beta p^{-\beta} & |\mathcal{C}(\mathcal{N}_i)| \geq \beta. \end{cases}$$

Substituting in particular clusterings and choices of β gives:

Setting	Prior work	Our work
Bernoulli	$O(M^2d^{2\beta}p^{-\beta}/n)$	$O(M^2d^{2\beta}p^{-\beta}/n)$
3-net GCR, $\beta=1$	$O(d\kappa^5p^{-\kappa^6})$	$O(d^2\kappa^6BM^2p^{-1}/n)$
arbitrary GCR, $\beta=2$	—	$O\left(\frac{M^2B}{n}\cdot\frac{n_1}{n}d^4p^{-3/2}\right) + O\left(\frac{M^2B}{n}\cdot\frac{n_{>1}}{n}d^6p^{-2}\right)$

Key quantities:

- d, maximum degree
- κ, restricted growth coefficient
- B, maximum cluster size
- n_1, number of nodes whose neighborhood is contained in a single cluster

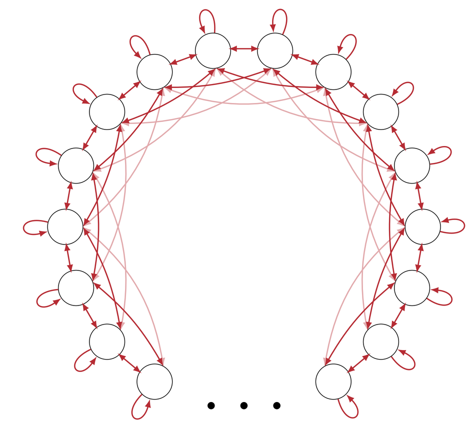
Our results:

- recover the bounds of Cortez et al. [2022] for Bernoulli designs and arbitrary β
- improve on the bounds of Ugander et al. [2013] for structured clusterings by leveraging the $\beta = 1$ assumption
- give new bounds for arbitrary clusterings that depend on the quality of the clusterings when $\beta = 2$

Experiments

Causal Network:

- Cycle network with $n = 120$ vertices, degree $d = 7$
- Vertex connected to themselves and 6 closest neighbors

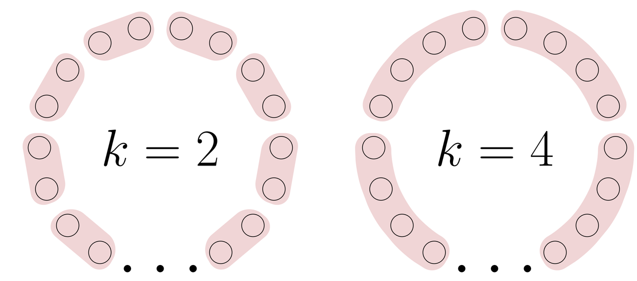


Potential Outcomes:

- Low-degree models with $\beta\in\{1,2,3,4\}$
- Coefficients $c_{i,\mathcal{S}} \sim \mathrm{Unif}\left(0, \frac{10}{2^{|\mathcal{S}|}}\right)$

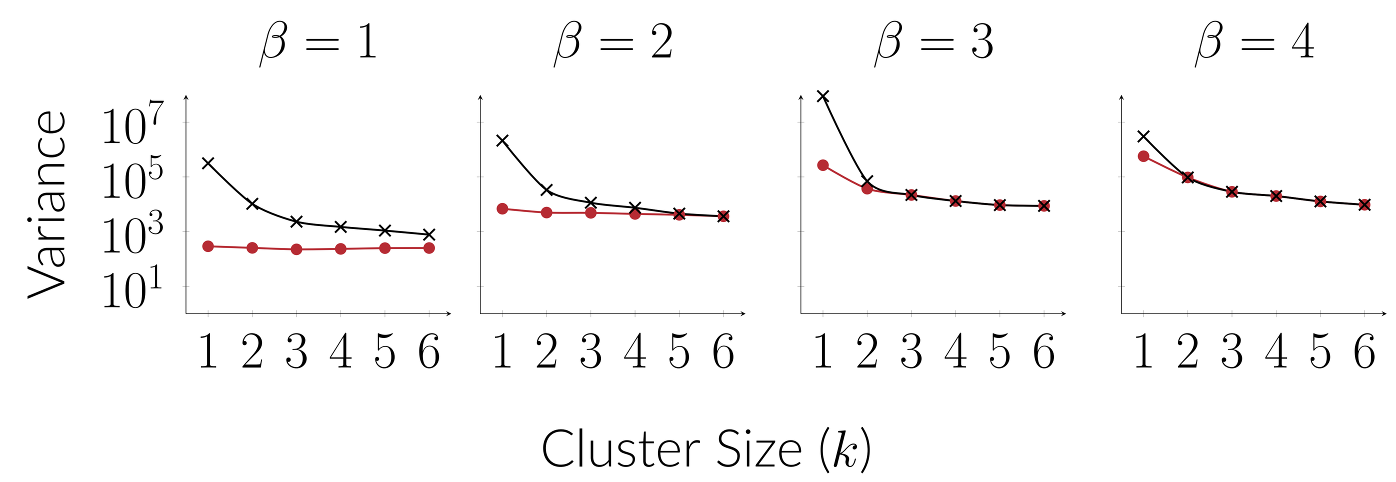
Treatment:

- GCR, cluster sizes $k\in\{1,\ldots,6\}$
- Treatment probability $p = 0.25$



Results:

Horvitz-Thompson (HT) vs. Pseudoinverse (PI)



Zooming in on a subset of the results:

	Method		
β	HT, k=6	PI, k=6	PI, k=1
1	7.77×10^2	2.52×10^2	2.19×10^2
4	9.59×10^3	9.59×10^3	5.71×10^5

Main Takeaway
Combining a low-degree outcome model with a clustered design gives ***best of both worlds*** behavior, with small variance when either β is small or the clustering is good.

References.

M. Cortez, M. Eichhorn, and C. L. Yu. Exploiting neighborhood interference with low order interactions under unit randomized design. *arXiv preprint arXiv:2208.05553*, 2022.

A. Swaminathan, A. Krishnamurthy, A. Agarwal, M. Dudik, J. Langford, D. Jose, and I. Zitouni. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems*, 30, 2017.

J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337, 2013.

Samir's Website: stanford.edu/~samirk/

Matt's Website: maeichho.github.io