Low-Degree Outcomes and Clustered Designs:

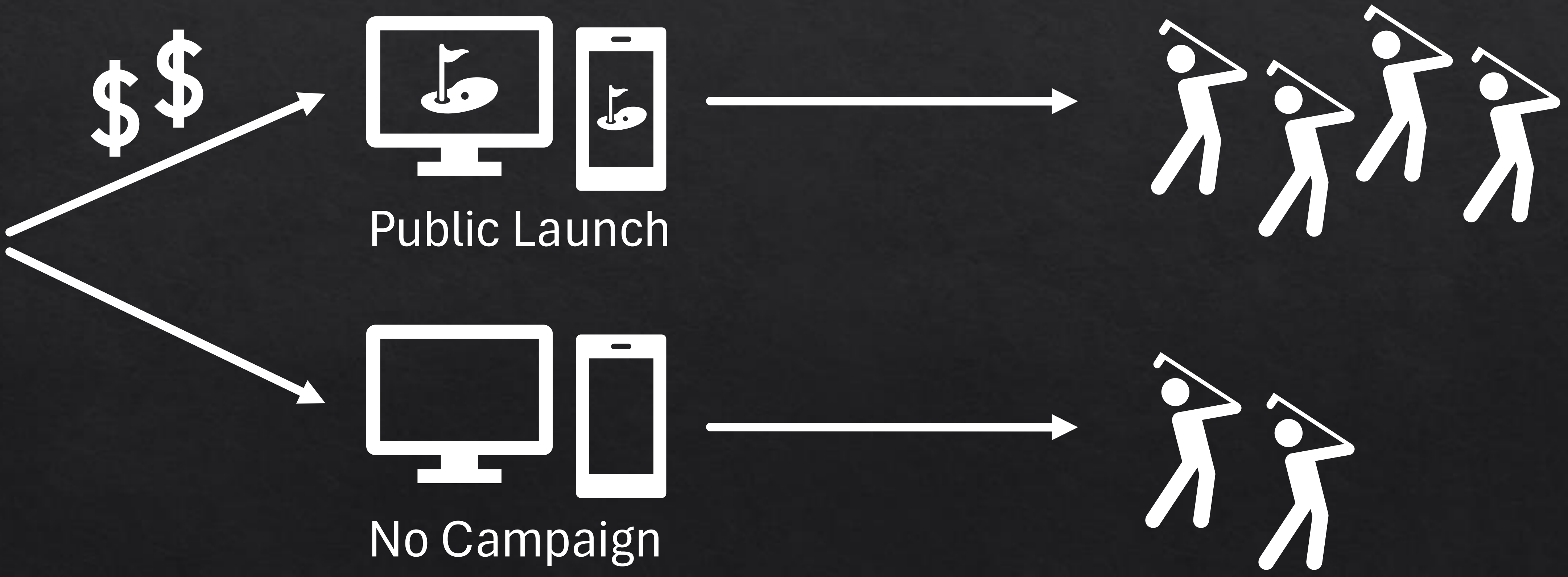
A Combined Approach for Causal Inference Under Interference




Matt Eichhorn
Samir Khan
Johan Ugander
Christina Lee Yu

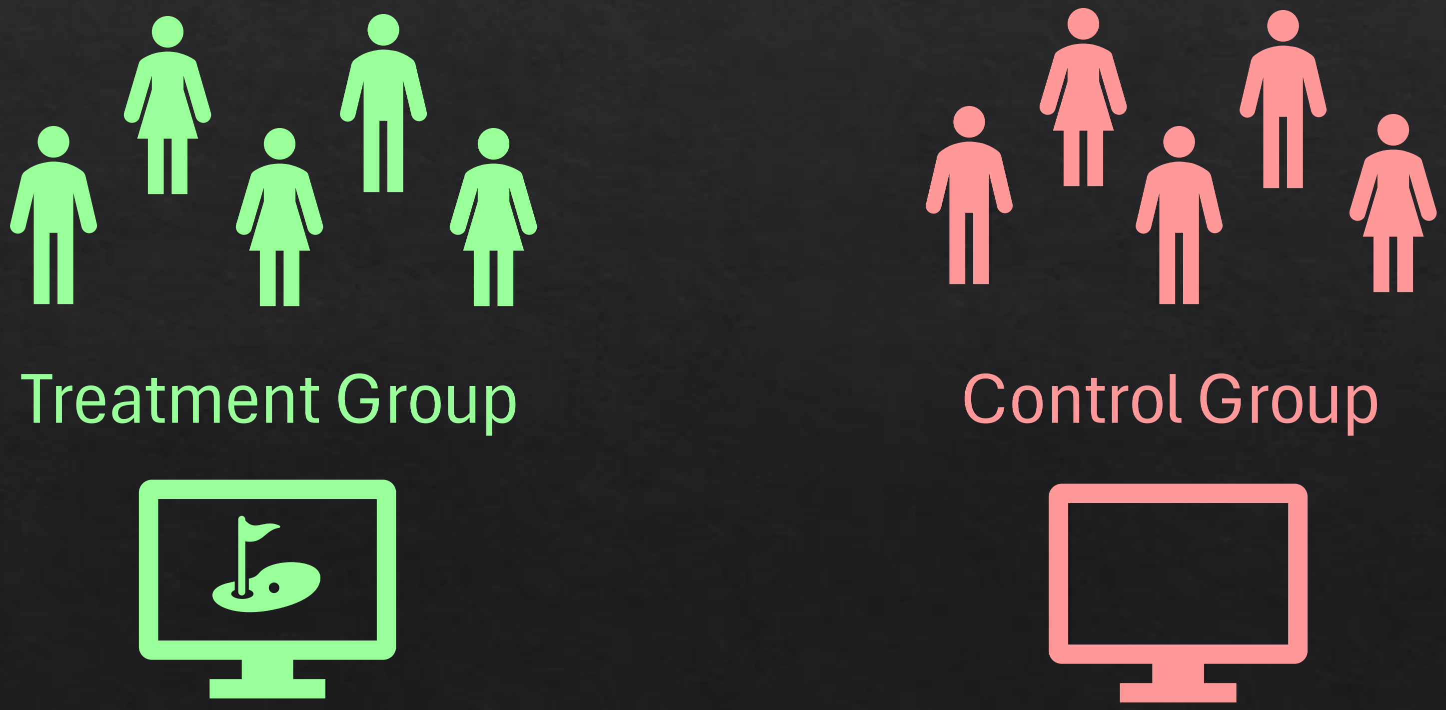
Motivation: Advertising

A golf course is deciding whether to run an advertising campaign



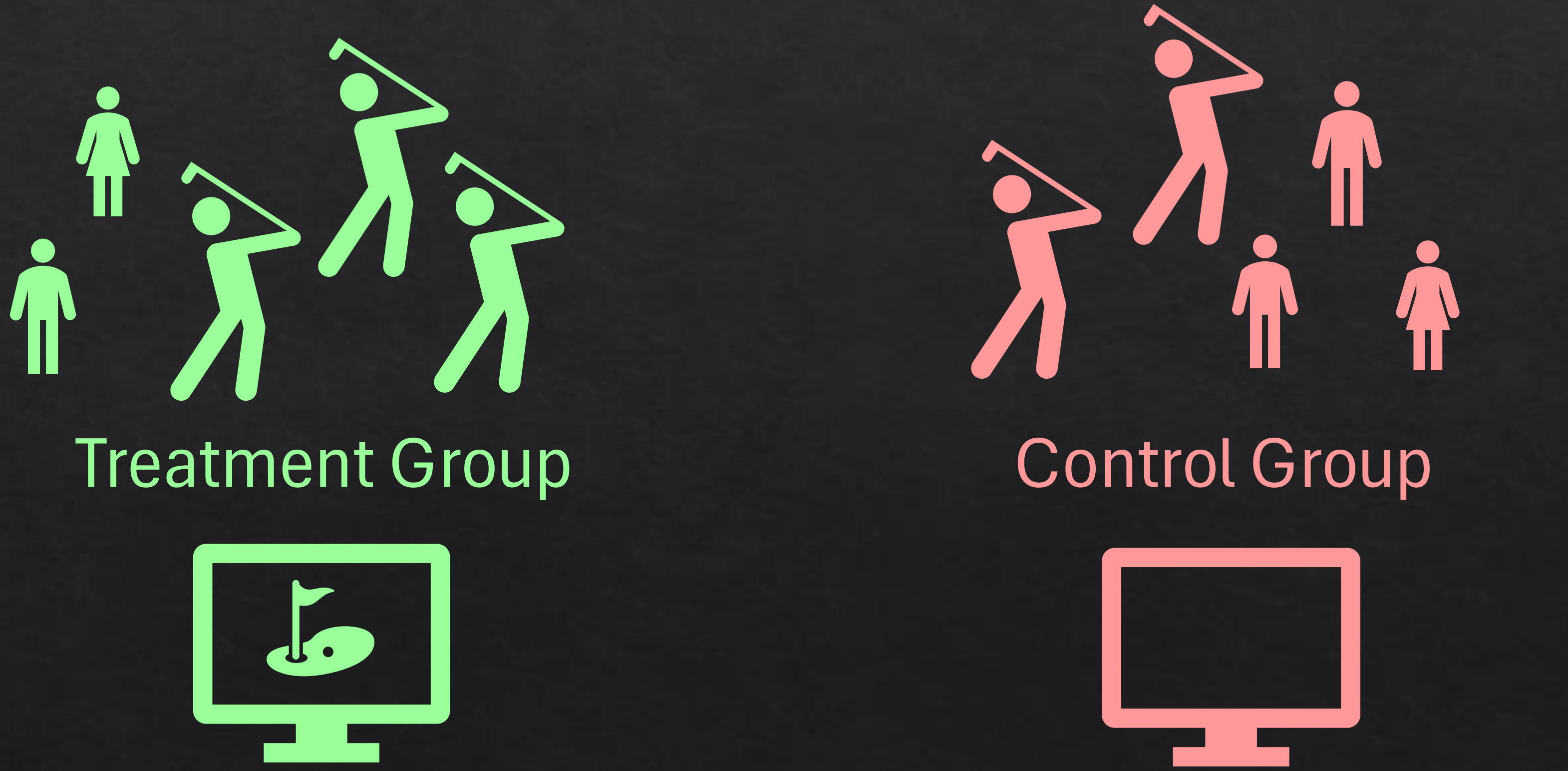
Motivation: Advertising

Use a (small) randomized experiment to inform their decision



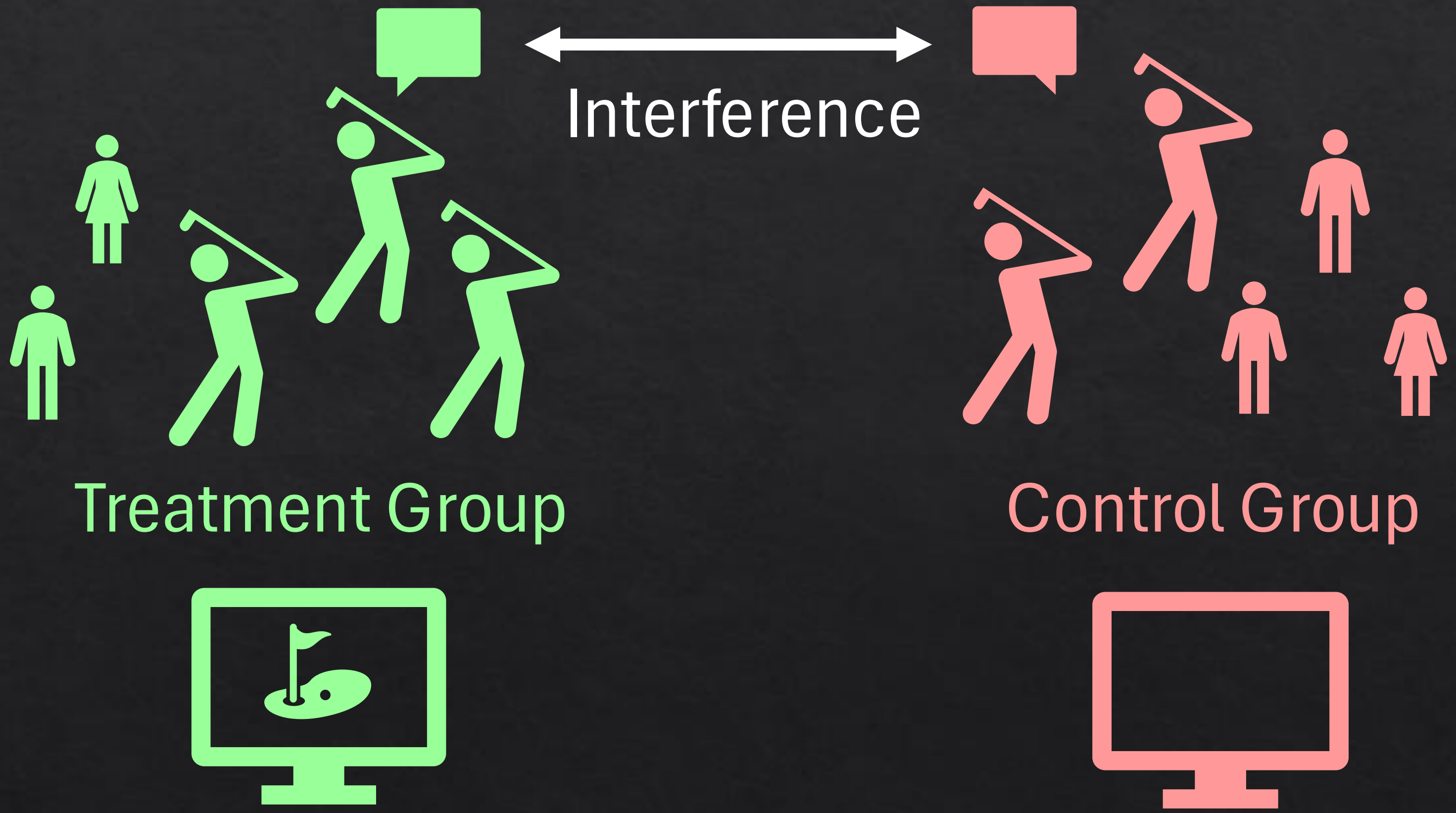
Motivation: Advertising

Use a (small) randomized experiment to inform their decision



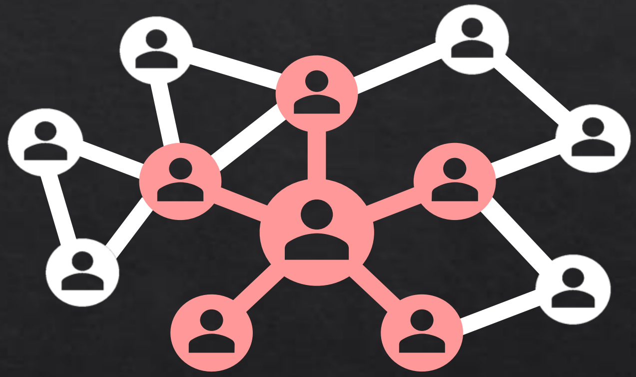
Motivation: Advertising

Use a (small) randomized experiment to inform their decision



Formalizing the Problem

- n individuals
- Treatment Assignments $z_i \in \{0,1\}$ (0 = control, 1 = treatment)
- Potential Outcomes $Y_i(\bar{\mathbf{z}}) : \{0,1\}^n \rightarrow \mathbb{R}$
- Neighborhood Interference
- Goal: Estimate Total Treatment Effect



$$TTE = \frac{1}{n} \sum_{i \in [n]} \left(Y_i(\bar{\mathbf{1}}) - Y_i(\bar{\mathbf{0}}) \right)$$

A Simple Unbiased Estimator

- Independent Treatment Assignments $z_j \sim \text{Bernoulli}(p)$
- Horvitz-Thompson Estimator

$$\widehat{TTE}_{\text{HT}} = \frac{1}{n} \sum_{i \in [n]} Y_i(\bar{\mathbf{z}}) \left(\prod_{j \in N_i} \frac{z_j}{p} - \prod_{j \in N_i} \frac{1 - z_j}{1 - p} \right)$$

0 unless entire neighborhood treated

0 unless entire neighborhood untreated

Prohibitively High Variance $O\left(\frac{1}{p^d}\right)$

Two Approaches for Variance Reduction

1. Better Experimental Design

- Under Bernoulli treatment, most neighborhoods are partially treated
- Smarter designs (e.g., with clustering) increases prevalence of fully treated neighborhoods
- Variance reduction relies on structural assumptions of **causal network**

Two Approaches for Variance Reduction

1. Better Experimental Design

- Under Bernoulli treatment, most neighborhoods are partially treated
- Smarter designs (e.g., with clustering) increases prevalence of fully treated neighborhoods
- Variance reduction relies on structural assumptions of **causal network**

2. Better Estimator Design

- Horvitz-Thompson Estimator ignores a lot of useful observations
- More clever estimators can utilize measurements from partially treated neighborhoods
- Variance reduction relies on structural assumptions of **potential outcomes**

Two Approaches for Variance Reduction

1. Better Experimental Design

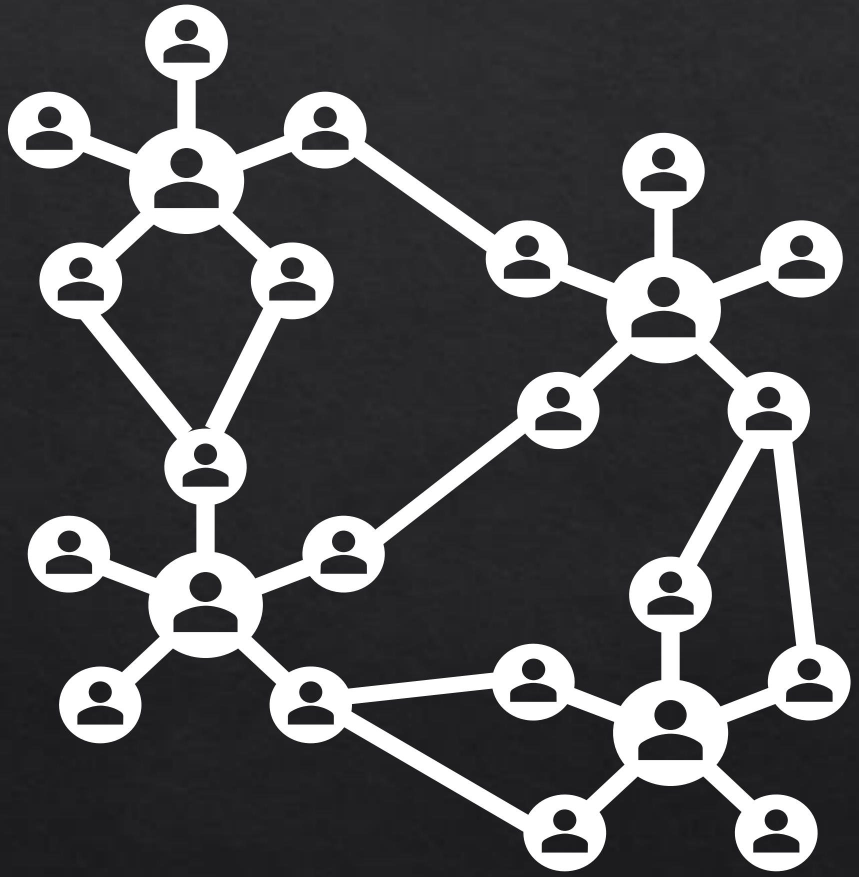
- Under Bernoulli treatment, most neighborhoods are partially treated
- Smarter designs (e.g., with clustering) increases prevalence of fully treated neighborhoods
- Variance reduction relies on structural assumptions of **causal network**

2. Better Estimator Design

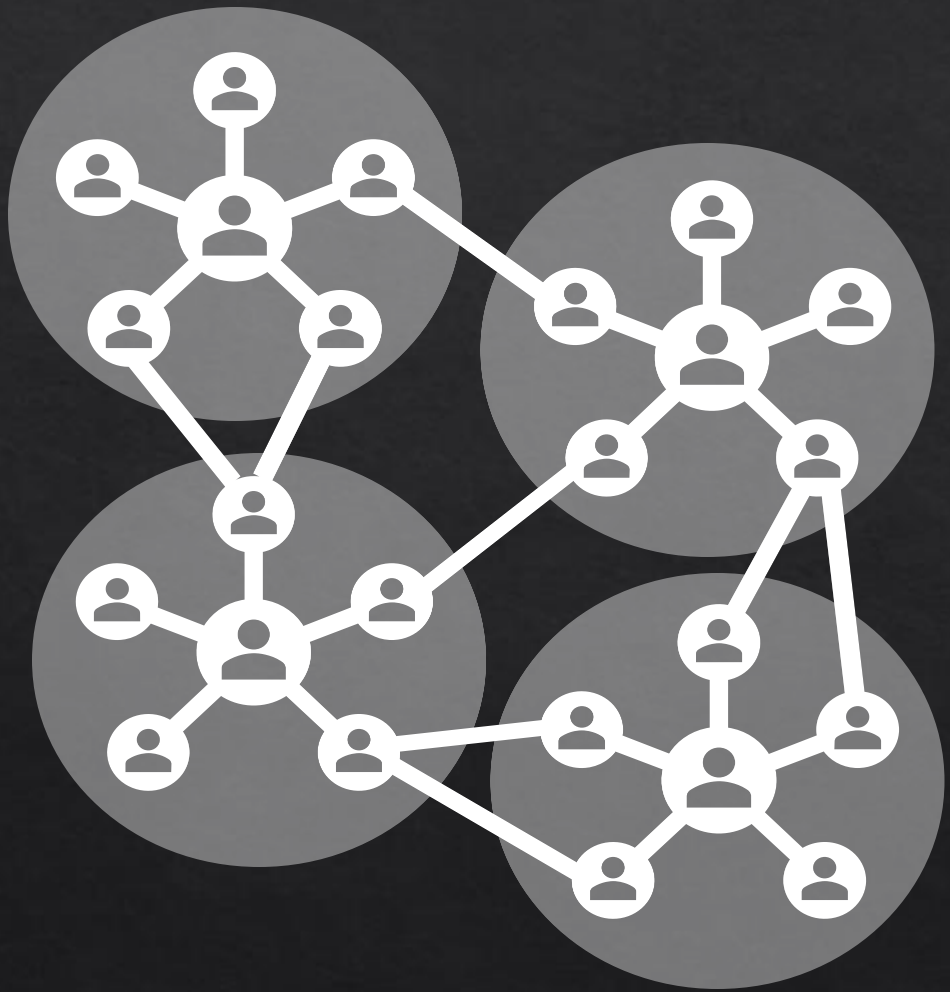
- Horvitz-Thompson Estimator ignores a lot of useful observations
- More clever estimators can utilize measurements from partially treated neighborhoods
- Variance reduction relies on structural assumptions of **potential outcomes**

Why not both?

Graph Cluster Randomized Design

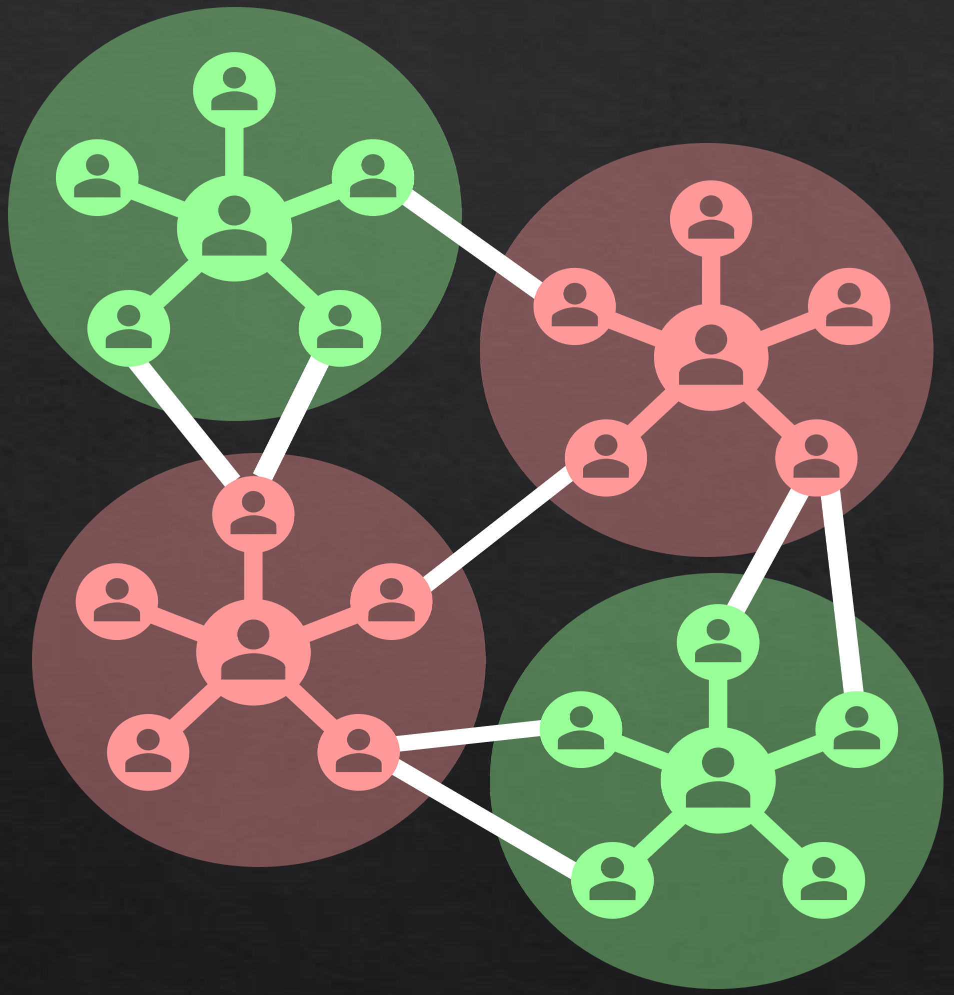


Graph Cluster Randomized Design



- Partition a graph into k clusters
  - $\mathcal{C} = \{ \mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_\mathrm{k} \}$
  - $\mathcal{C}(i)$ = cluster of individual i
  - $\mathcal{C}(N_i) = \bigcup_{j \in N_i} \mathcal{C}(j)$

Graph Cluster Randomized Design



- Partition a graph into k clusters
  - $\mathcal{C} = \{\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_\mathrm{k}\}$
  - $\mathcal{C}(i) =$ cluster of individual i
  - $\mathcal{C}(N_i) = \bigcup_{j \in N_i} \mathcal{C}(j)$
- Sample cluster treatments
  - $w_{C_\ell} = \mathbb{I}(C_\ell \text{ treated}) \sim \text{Bernoulli}(p)$
- Assign individual treatments
  - $x_i = w_{C(i)}$

Variance Bounds for GCR

- For Bernoulli (unit randomized) Design:

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{HT}}\right) = O\left(n^{-1} p^{-d}\right)$$

Network Degree

- For Graph Cluster Randomization:

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{HT}}\right) = O\left(n^{-1} p^{-\max_i |\mathcal{C}(N_i)|}\right)$$

“Cluster Degree”

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

Variance Bounds for GCR

- For Bernoulli (unit randomized) Design:

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{HT}}\right) = O\left(n^{-1}p^{-d}\right)$$

Network Degree

- For Graph Cluster Randomization:

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{HT}}\right) = O\left(n^{-1}p^{-\max_i |\mathcal{C}(N_i)|}\right)$$

"Cluster Degree"

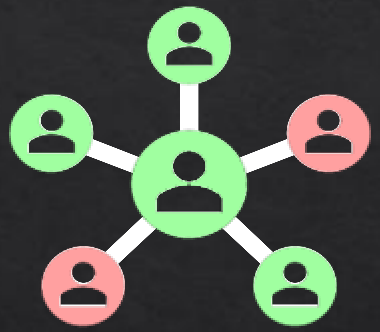
- 3-Net Clustering on κ-restricted growth graphs

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{HT}}\right) = O\left(n^{-1}d\,\kappa^5 p^{-\kappa^6}\right)$$

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

Alternate Estimator Designs

- Incorporate information from partially-treated neighborhoods
- Need assumptions on potential outcomes



Alternate Estimator Designs

- Incorporate information from partially-treated neighborhoods
- Need assumptions on potential outcomes
- Low-order Interactions Model

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

$\mathcal{S}_i^\beta$ = subsets of N_i with size $\leq \beta$

Cortez-Rodriguez, M., Eichhorn, M., and Yu C.L., “Exploiting neighborhood interference with low order interactions under unit randomized design”, Journal of Causal Inference (JCI), 2023.

The Pseudoinverse Estimator

- Under Low-order Interactions, we estimate $\hat{\mathbf{c}}_i = Y_i(\bar{\mathbf{z}})\ \mathbb{E}\left[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^{\mathrm{T}}\right]^{\dagger} \tilde{\mathbf{z}}_i$

Cortez-Rodriguez, M., Eichhorn, M., and Yu C.L., "Exploiting neighborhood interference with low order interactions under unit randomized design", Journal of Causal Inference (JCI), 2023.

The Pseudoinverse Estimator

- Under Low-order Interactions, we estimate $\hat{\mathbf{c}}_i = Y_i(\bar{\mathbf{z}})\ \mathbb{E}\left[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^{\mathrm{T}}\right]^{\dagger} \tilde{\mathbf{z}}_i$
- Extend by linearity to estimate $\widehat{TTE}_{\mathrm{PI}} = \frac{1}{n} \sum_{i \in [n]} \langle \hat{\mathbf{c}}_i, \theta_i \rangle \qquad \theta_i = (0,1,\ldots,1)$

Cortez-Rodriguez, M., Eichhorn, M., and Yu C.L., "Exploiting neighborhood interference with low order interactions under unit randomized design", Journal of Causal Inference (JCI), 2023.

The Pseudoinverse Estimator

- Under Low-order Interactions, we estimate $\widehat{\mathbf{c}}_i = Y_i(\bar{\mathbf{z}})\ \mathbb{E}\left[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^{\mathrm{T}}\right]^{\dagger} \tilde{\mathbf{z}}_i$
- Extend by linearity to estimate $\widehat{TTE}_{\mathrm{PI}} = \frac{1}{n} \sum_{i \in [n]} \langle \widehat{\mathbf{c}}_i, \theta_i \rangle \qquad \theta_i = (0,1, \dots, 1)$
- For Bernoulli (unit randomized) Design:

$$\widehat{TTE}_{\mathrm{PI}} = \frac{1}{n} \sum_{i \in [n]} Y_i(\bar{\mathbf{z}}) \sum_{S \in S_i^{\beta}} \left(\frac{1}{p^{|S|}} - \frac{1}{(1-p)^{|S|}} \right) \prod_{j \in S} (z_j - p)$$

- Unbiased with $\mathrm{Var}\left(\widehat{TTE}_{\mathrm{PI}}\right) = O\left(\frac{d^2}{n} \left(\frac{ed}{\beta p^2} \right)^{\beta} \right)$

Cortez-Rodriguez, M., Eichhorn, M., and Yu C.L., “Exploiting neighborhood interference with low order interactions under unit randomized design”, Journal of Causal Inference (JCI), 2023.

Combining these Approaches

- PI estimator works for any experimental design:

Main Theorem:

$\widehat{TTE}_{\mathrm{PI}}$ is unbiased when each θ_i lies in the column space of $\mathbb{E}\left[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^{\mathrm{T}}\right]^{\dagger}$ with:

$$\mathrm{Var}\left(\widehat{TTE}_{\mathrm{PI}}\right) = O\left(\frac{1}{n^2} \sum_{i,j \in [n]} \gamma_i \gamma_j \ \mathbb{I}\left(\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j\right)\right)$$

where $\gamma_i = \sqrt{\left|S_i^{\beta}\right| \theta_i^{\mathrm{T}} \mathbb{E}\left[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^{\mathrm{T}}\right]^{\dagger} \theta_i}$ is i's contribution to the variance

and $\mathbb{I}\left(\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j\right)$ captures the graph structure's effect on the variance

GCR + PI Estimator

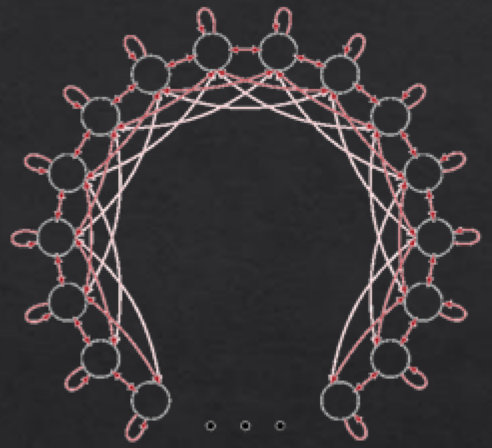
$$\gamma_i^2 \leq \begin{cases} 2 d_i^{\beta} \cdot p^{-|\mathcal{C}(N_i)|} & |\mathcal{C}(N_i)| < \beta \\ 2\, d_i^{\beta} \cdot |\mathcal{C}(N_i)|^{\beta} \cdot p^{-\beta} & |\mathcal{C}(N_i)| \geq \beta \end{cases}$$

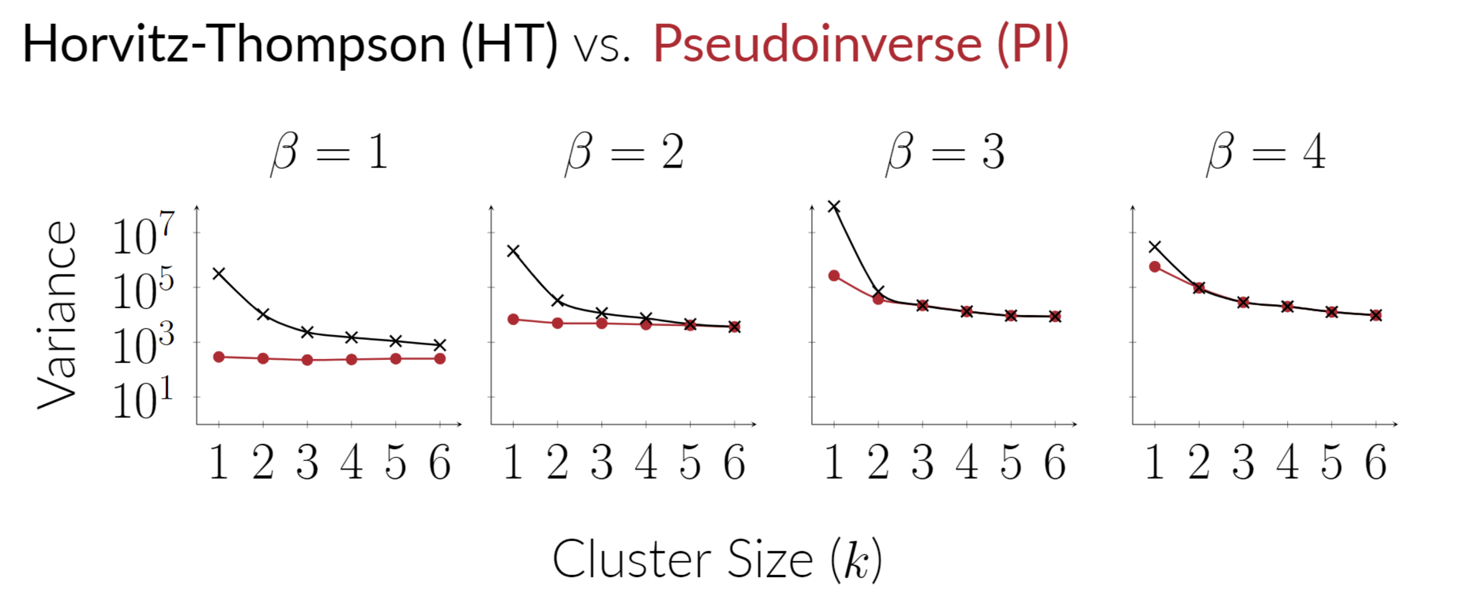
- When clustering quality "better than" potential outcomes complexity, variance depends exponentially on clustering quality
- When potential outcomes complexity is "better than" clustering quality, variance depends exponentially on interaction size parameter β

"Best of Both Worlds" Trade-off!

Experimental Results

- Cycle network with $n = 120$ and degree $d = 7$
- Effect coefficients sampled independently with $c_{i,S} \sim \text{Unif}\left(0, 10 \cdot 2^{-|S|}\right)$
- Treatment probability $p = 0.25$
- Cluster vertices into adjacent groups
- Vary cluster size k and interaction size parameter β





Thanks for Listening!