Causal Inference under Low-Order Interference

SINM Symposium
July 17, 2024

Mayleen
Cortez-Rodriguez

Matt
Eichhorn

Samir
Khan

Johan
Ugander

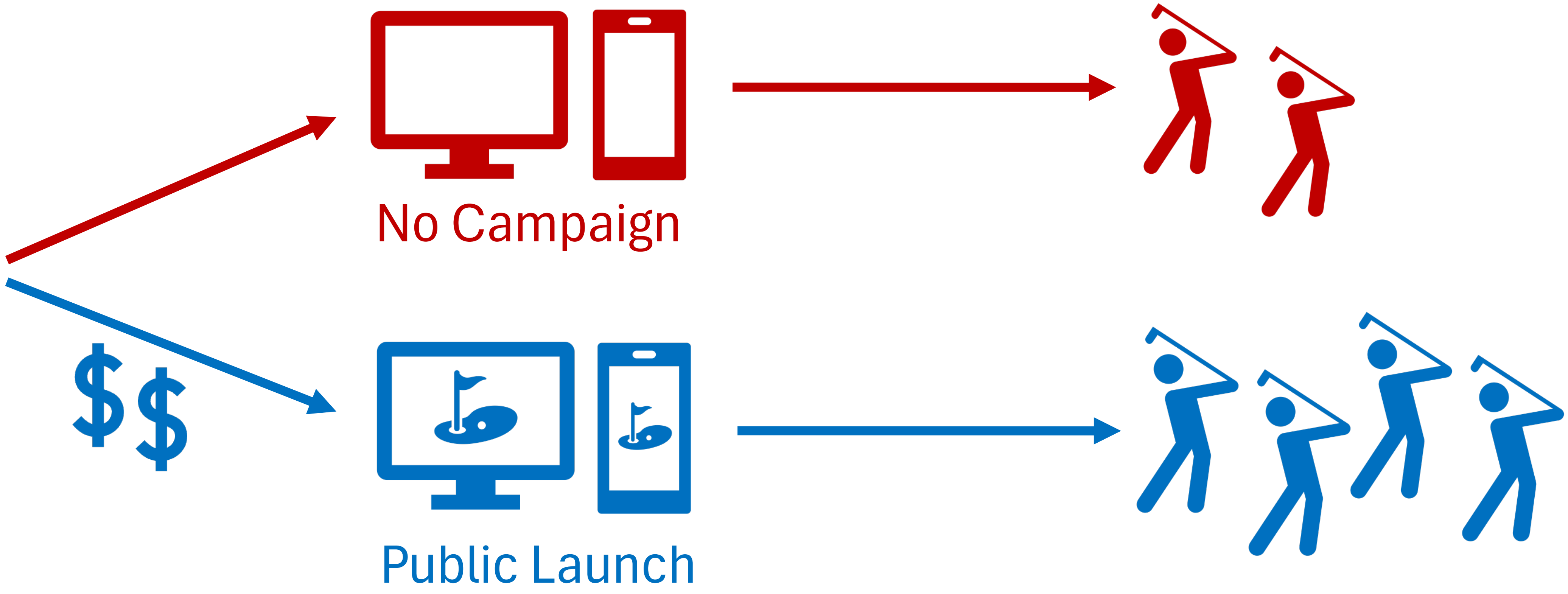
Christina
Lee Yu

Talk Structure

1. Motivation: causal inference under interference
2. β-Order interactions
3. Estimation with Bernoulli experimental designs
4. Estimation with arbitrary experimental designs
5. Selecting an experimental design

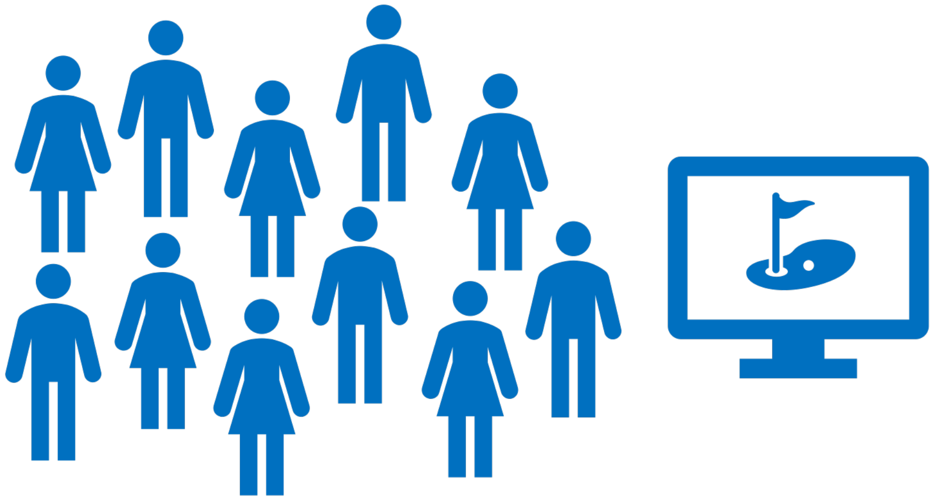
Motivating Example: Advertising

A golf course is deciding whether to run an advertising campaign



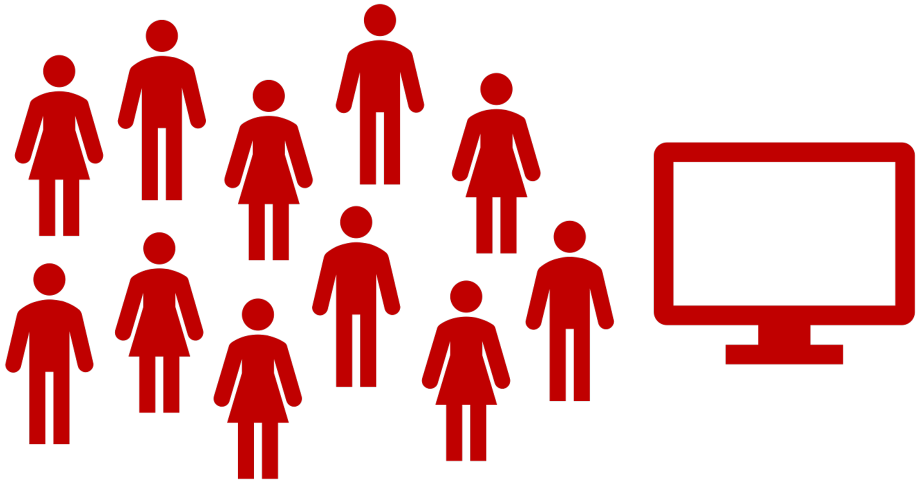
Total Treatment Effect

Difference in *average outcome* (e.g., monthly spending at the course) under two possible *global actions*:



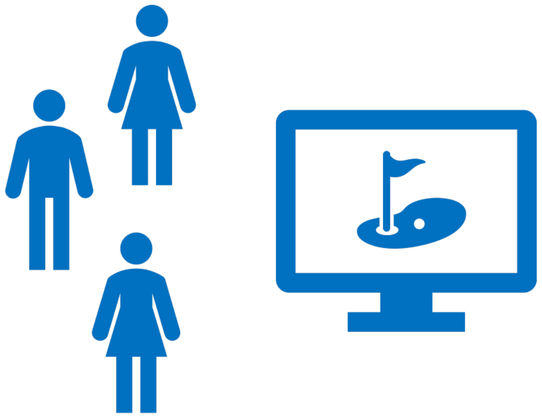
Everybody Treated

vs.

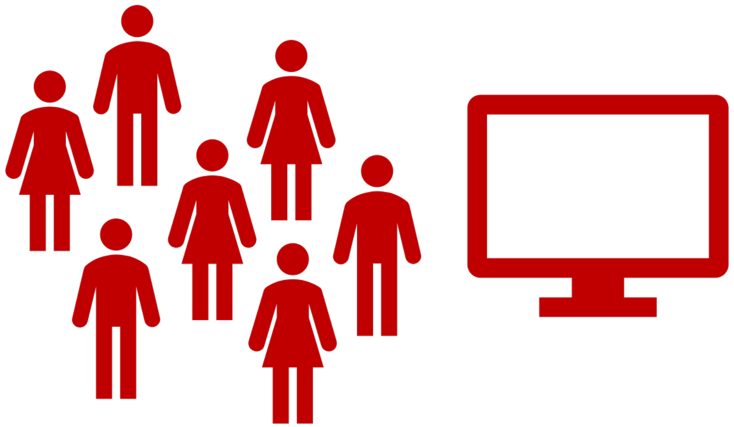


Nobody Treated

Randomized Experiment



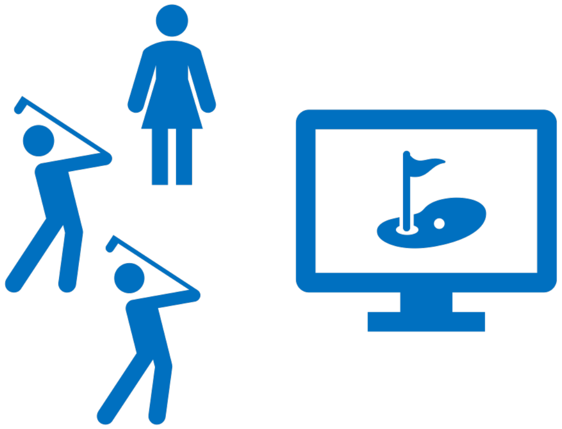
Treatment Group



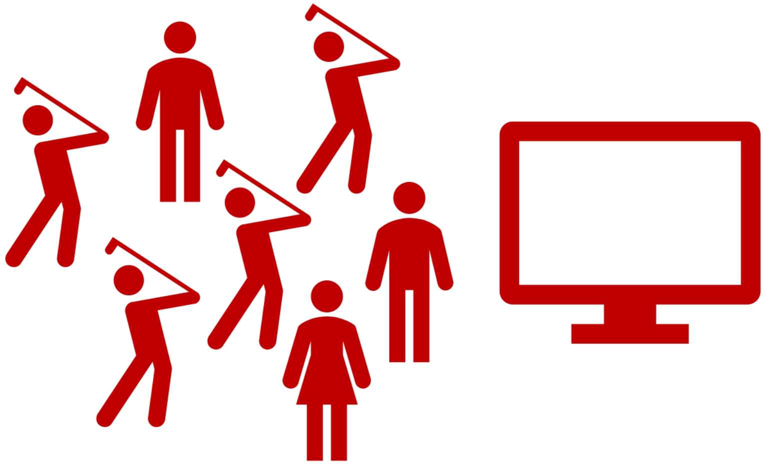
Control Group

*** Assume the marginal probability p of being in the treatment group is small.

Randomized Experiment



Treatment Group



Control Group

*** Assume the marginal probability p of being in the treatment group is small.

Randomized Experiment

Treatment Group

Control Group

*** Assume the marginal probability p of being in the treatment group is small.

Difference in Means Estimator:

$$\widehat{\text{TTE}}_{\text{DM}} = \begin{array}{l}\text{Average Outcome}\\ \text{in Treatment Group}\end{array} - \begin{array}{l}\text{Average Outcome}\\ \text{in Control Group}\end{array}$$

Interference

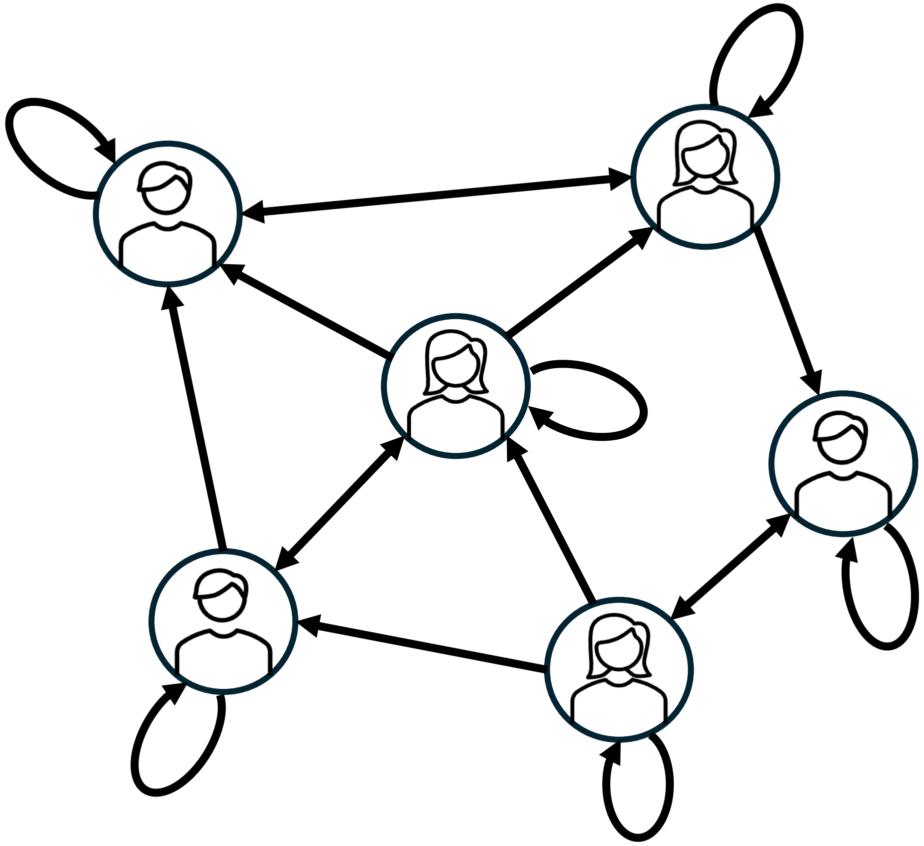
Individuals’ outcomes may change even if they are not treated

Interference

Individuals' outcomes may change even if they are not treated

Introduces Bias into DM Estimator

Modeling Interference



Directed Interference Graph $G = (V, A)$

V = n individuals

$(j, i) \in A \Rightarrow j'$s treatment affects i's outcome

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Aronow, Peter M., and Cyrus Samii. "Estimating average causal effects under general interference, with application to a social network experiment." (2017): 1912-1947.

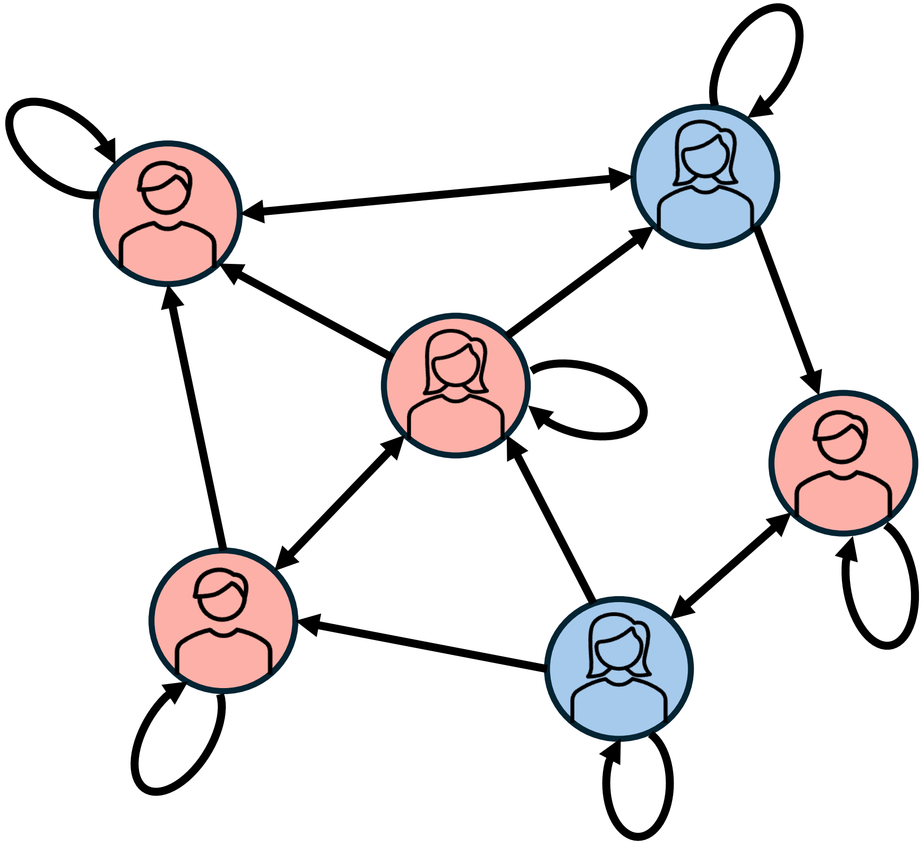
Modeling Interference

Directed Interference Graph $G = (V, A)$

V = n individuals

$(j, i) \in A \Rightarrow j'$s treatment affects i's outcome

Treatment Assignments $\boldsymbol{z} \in \{0,1\}^n$



Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Aronow, Peter M., and Cyrus Samii. "Estimating average causal effects under general interference, with application to a social network experiment." (2017): 1912-1947.

Modeling Interference

Directed Interference Graph $G = (V, A)$

V = n individuals

$(j, i) \in A \Rightarrow j'$s treatment affects i's outcome

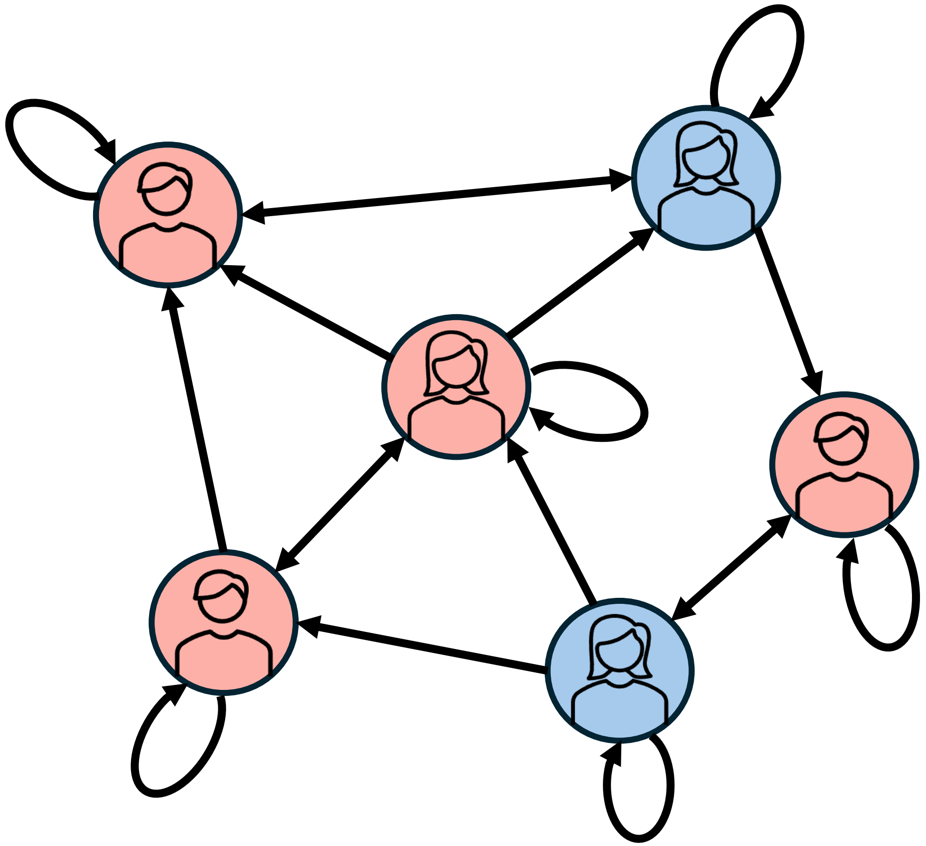
Treatment Assignments $\boldsymbol{z} \in \{0,1\}^n$

Potential Outcomes $Y_i(\boldsymbol{z}) : \{0,1\}^n \rightarrow \mathbb{R}$

* We'll assume these functions are *bounded*

Neighborhood Interference Assumption:

$z_j = z_j'$ for all $j \in N_i \quad \Rightarrow \quad Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}')$



Sussman, Daniel L., and Edoardo M. Airoldi. "Elements of estimation theory for causal effects in the presence of network interference." *arXiv preprint arXiv:1702.03578* (2017).

Modeling Interference

Directed Interference Graph $G = (V, A)$

V = n individuals

$(j, i) \in A \Rightarrow j'$s treatment affects i's outcome

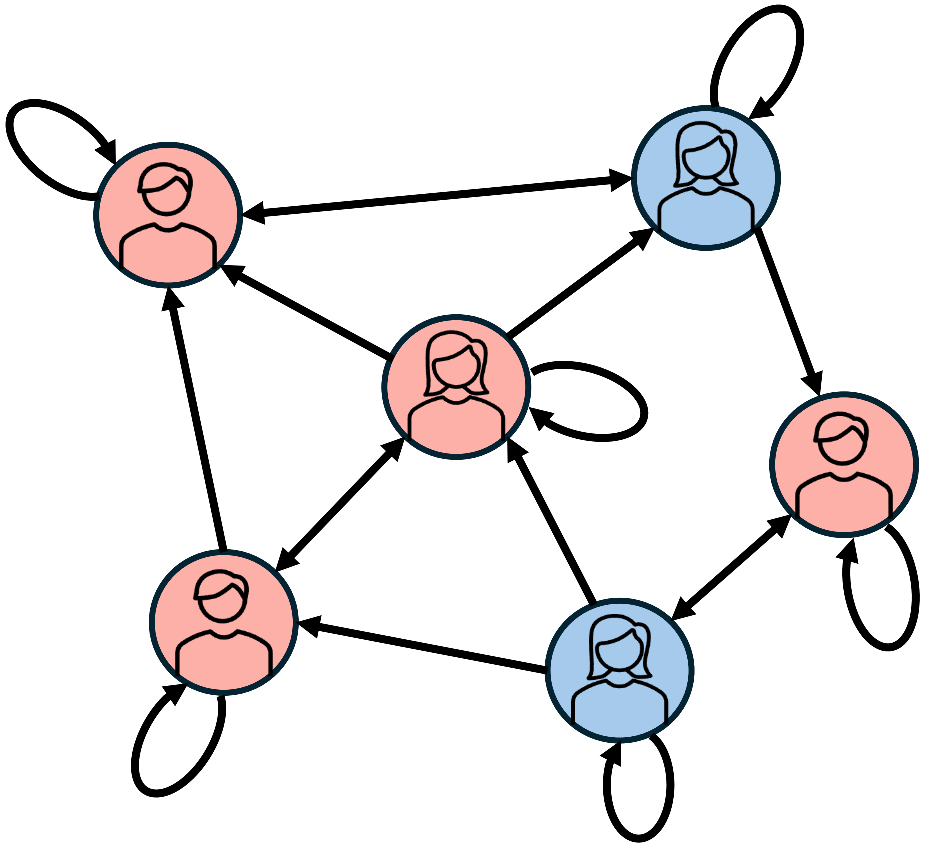
Treatment Assignments $\boldsymbol{z} \in \{0,1\}^n$

Potential Outcomes $Y_i(\boldsymbol{z}) : \{0,1\}^n \rightarrow \mathbb{R}$

* We'll assume these functions are *bounded*.

Neighborhood Interference Assumption:

$z_j = z_j'$ for all $j \in N_i \Rightarrow Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}')$



Total Treatment Effect:

$$\mathrm{TTE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i(\mathbf{1}) - Y_i(\mathbf{0}))$$

Horvitz-Thompson Estimator

$$\widehat{\mathrm{TTE}}_{\mathrm{HT}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\left(\frac{\mathbb{I}(N_i \text{ fully treated})}{\Pr(N_i \text{ fully treated})} - \frac{\mathbb{I}(N_i \text{ fully untreated})}{\Pr(N_i \text{ fully untreated})}\right)$$

Horvitz, Daniel G., and Donovan J. Thompson. "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association* 47.260 (1952): 663-685.

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Horvitz-Thompson Estimator

$$\widehat{\text{TTE}}_{\text{HT}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\left(\frac{\mathbb{I}(N_i \text{ fully treated})}{\Pr(N_i \text{ fully treated})} - \frac{\mathbb{I}(N_i \text{ fully untreated})}{\Pr(N_i \text{ fully untreated})}\right)$$

Under Independent Treatment Assignments $z_j \sim \text{Bernoulli}(p)$:

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\left(\underbrace{\prod_{j\in N_i}\frac{z_j}{p}}_{\text{0 unless entire neighborhood treated}} - \underbrace{\prod_{j\in N_i}\frac{1-z_j}{1-p}}_{\text{0 unless entire neighborhood untreated}}\right)$$

Horvitz, Daniel G., and Donovan J. Thompson. "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association* 47.260 (1952): 663-685.

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Horvitz-Thompson Estimator

$$\widehat{\mathrm{TTE}}_{\mathrm{HT}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z}) \left(\frac{\mathbb{I}(N_i \text{ fully treated})}{\Pr(N_i \text{ fully treated})} - \frac{\mathbb{I}(N_i \text{ fully untreated})}{\Pr(N_i \text{ fully untreated})} \right)$$

Under Independent Treatment Assignments $z_j \sim \mathrm{Bernoulli}(p)$:

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z}) \left(\underbrace{\prod_{j \in N_i} \frac{z_j}{p}}_{\text{0 unless entire neighborhood treated}} - \underbrace{\prod_{j \in N_i} \frac{1-z_j}{1-p}}_{\text{0 unless entire neighborhood untreated}} \right)$$

- Unbiased estimator
- Prohibitive $O(p^{-d})$ variance

Horvitz, Daniel G., and Donovan J. Thompson. "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association* 47.260 (1952): 663-685.

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Variance Reduction Strategies

1. Smarter Estimator Design

- Horvitz-Thompson Estimator ignores a lot of useful observations
- Smarter estimators incorporate measurements from partially treated neighborhoods
- Relies on structural assumptions on **potential outcomes**

2. Smarter Experimental Design

- Under Bernoulli treatment, most neighborhoods are partially treated
- Smarter designs increase prevalence of fully treated neighborhoods
- Relies on structural assumptions on **interference network**

Related Work

Network Structure

Outcomes Structure

	C Disconnected Subcommunities	**κ-Restricted Growth**	**Fully General**
Linear	Directions for Future Work		OLS Estimation Bernoulli Design
Generalized Linear			Feature Regression Bernoulli Design
β-Order Interactions	Pseudoinverse Estimators for General Designs		SNIPE Estimator Bernoulli Designs
General Neighborhood Interference	HT Estimator Cluster Designs	HT Estimator (Randomized) Cluster Designs	HT Estimator Bernoulli Design

Many citations arranged in final bibliography

β-Order Interactions

Cortez, Mayleen, **Matthew Eichhorn**, and Christina Yu. "Staggered rollout designs enable causal inference under interference without network knowledge." *Advances in Neural Information Processing Systems* 35 (2022): 7437-7449.

Cortez-Rodriguez, Mayleen, **Matthew Eichhorn**, and Christina Lee Yu. "Exploiting neighborhood interference with low-order interactions under unit randomized design." *Journal of Causal Inference* 11.1 (2023): 20220051.

Potential Outcomes under Network Interference

Since treatments are binary, $z_j \in \{0,1\}$, we can write:

$$Y_i(\boldsymbol{z}) = \sum_{T \subseteq N_i} a_{i,T} \underbrace{\prod_{j \in T} z_j}_{T \text{ fully } \textbf{treated}} \underbrace{\prod_{j' \in N_i \setminus T} (1 - z_{j'})}_{N_i \setminus T \text{ fully } \textbf{control}}$$

Potential Outcomes under Network Interference

Since treatments are binary, $z_i \in \{0,1\}$, we can write:

$$Y_i(\boldsymbol{z}) = \sum_{T \subseteq N_i} a_{i,T} \underbrace{\prod_{j \in T} z_j}_{T \text{ fully } \textbf{treated}} \underbrace{\prod_{j' \in N_i \setminus T} (1 - z_{j'})}_{N_i \setminus T \text{ fully } \textbf{control}}$$

Re-parameterize in the *monomial basis*:

$$Y_i(\boldsymbol{z}) = \sum_{S \subseteq N_i} c_{i,S} \prod_{j \in S} z_j$$

$c_{i,S} =$ additive effect on i when S is treated
(regardless of other treatments)

β-Order Interactions

Intuition: Influence comes from small subsets of neighbors

Imposes a sparsity assumption on the $c_{i,S}$ coefficients

$$c_{i,S} = 0 \text{ whenever } |S| > \beta$$

$$Y_i(\boldsymbol{z}) = \sum_{\substack{S \subseteq N_i \\ |S| \leq \beta}} c_{i,S} \prod_{j \in S} z_j = \sum_{S \in S_i^{\beta}} c_{i,S} \, (\tilde{\boldsymbol{z}}_i)_S = \langle \boldsymbol{c}_i, \tilde{\boldsymbol{z}}_i \rangle$$

β-Order Interactions

Intuition: Influence comes from small subsets of neighbors

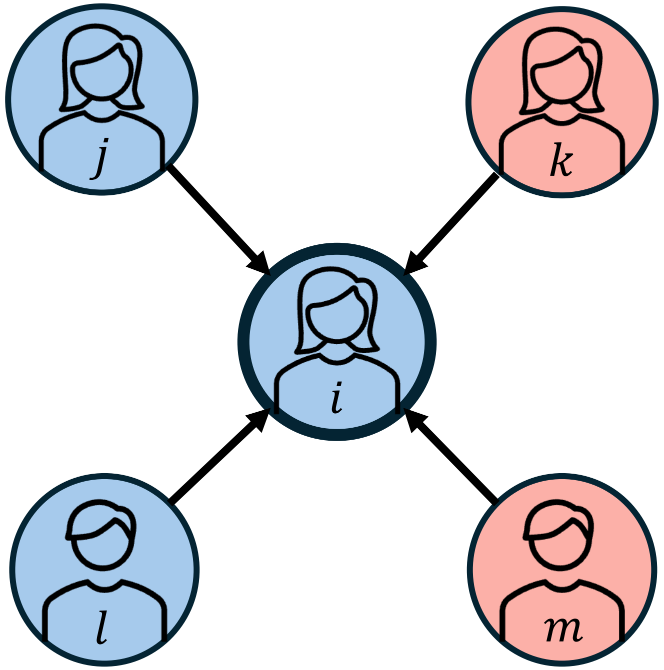
Imposes a sparsity assumption on the $c_{i,S}$ coefficients

$$c_{i,S} = 0 \text{ whenever } |S| > \beta$$

$$Y_i(\boldsymbol{z}) = \sum_{\substack{S \subseteq N_i \\ |S| \leq \beta}} c_{i,S} \prod_{j \in S} z_j = \sum_{S \in S_i^\beta} c_{i,S} \, (\tilde{\mathbf{z}}_i)_S = \langle \boldsymbol{c}_i, \tilde{\mathbf{z}}_i \rangle$$

$(\tilde{\mathbf{z}}_i)_S = \prod_{j \in S} z_j$ indicates if *everyone* in S is treated

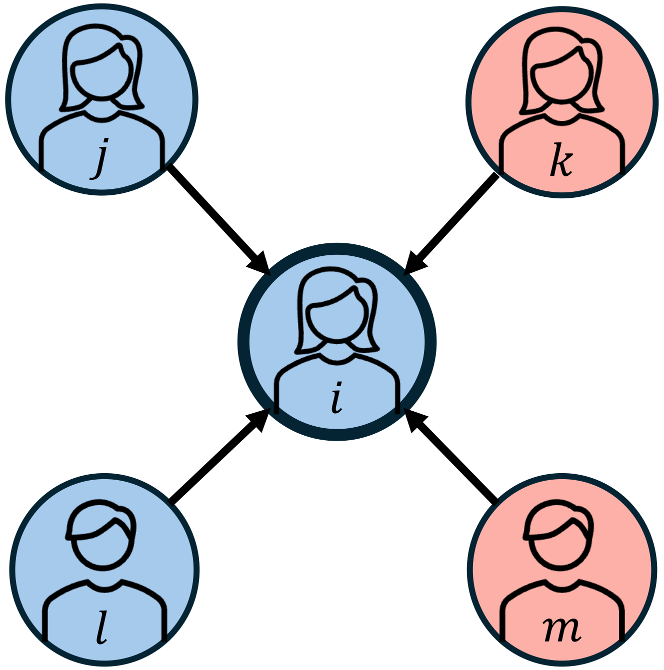
Example



$\boldsymbol{\beta} = \mathbf{1}$:

$$Y_i(z) = \underbrace{c_{i,\emptyset}}_{\text{Baseline}} + \underbrace{c_{i,\{i\}} + c_{i,\{j\}} + c_{i,\{l\}}}_{\text{Treatment Effect}}$$

Example



$\boldsymbol{\beta} = \mathbf{1}$:

$$Y_i(z) = \underbrace{c_{i,\emptyset}}_{\text{Baseline}} + \underbrace{c_{i,\{i\}} + c_{i,\{j\}} + c_{i,\{l\}}}_{\text{Treatment Effect}}$$

$\boldsymbol{\beta} = \mathbf{2}$:

$$\begin{aligned} Y_i(z) = c_{i,\emptyset} + c_{i,\{i\}} + c_{i,\{j\}} + c_{i,\{l\}} \\ + c_{i,\{i,j\}} + c_{i,\{i,l\}} + c_{i,\{j,l\}} \end{aligned}$$

Interpreting β

Most Restrictive

Most General

$\boldsymbol{\beta = 1}$

Linear (Heterogeneous) Outcomes Models

Additive effects from treated neighbors

$\boldsymbol{\beta = 2}$

Dyadic Interactions

Outcomes depend on interactions between pairs of individuals

$\boldsymbol{\beta = 4}$

Causal Network Motifs

Outcomes depend on treatment patterns in small graph motifs (closed/open triangles, tetrads, etc.)

$\boldsymbol{\beta} = \max_{\boldsymbol{i}} \boldsymbol{d_i}$

Arbitrary Potential Outcomes under Neighborhood Interference

Yu, Christina Lee, et al. "Estimating the total treatment effect in randomized experiments with unknown network structure." *PNAS* 119.44 (2022):

Deng, Lu, et al. "Unbiased Estimation for Total Treatment Effect Under Interference Using Aggregated Dyadic Data." *arXiv preprint arXiv:2402.12653* (2024).

Yuan, Yuan, Kristen Altenburger, and Farshad Kooti. "Causal network motifs: Identifying heterogeneous spillover effects in a/b tests." *Proceedings of the Web Conference 2021*.

Total Treatment Effect

$$\mathrm{TTE} = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i(\mathbf{1}) - Y_i(\mathbf{0})\big) \quad = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{S\subseteq N_i \\ 1\leq|S|\leq\beta}} c_{i,S} \quad = \frac{1}{n}\sum_{i=1}^{n}\langle \mathbf{c}_i, \underbrace{\boldsymbol{\theta}_i}_{\substack{(\theta_i)_{\emptyset}= 0 \\ (\theta_i)_S= 1}} \rangle$$

We’ll develop an estimator for each $\mathbf{c}_i$ that can be extended by linearity to a TTE estimator.

Estimating TTE under Bernoulli randomization

Cortez-Rodriguez, Mayleen, **Matthew Eichhorn**, and Christina Lee Yu. "Exploiting neighborhood interference with low-order interactions under unit randomized design." *Journal of Causal Inference* 11.1 (2023): 20220051.

Designing an Estimator

Imagine we could replicate our randomized experiment R times

$$\underbrace{\begin{bmatrix} Y_i(\boldsymbol{z}^{(1)}) \\ Y_i(\boldsymbol{z}^{(2)}) \\ \vdots \\ Y_i(\boldsymbol{z}^{(R)}) \end{bmatrix}}_{\mathbf{Y}_{\boldsymbol{i}}} = \underbrace{\begin{bmatrix} \leftarrow \tilde{\mathbf{z}}_i{}^{(1)} \rightarrow \\ \leftarrow \tilde{\mathbf{z}}_i{}^{(2)} \rightarrow \\ \vdots \\ \leftarrow \tilde{\mathbf{z}}_i{}^{(R)} \rightarrow \end{bmatrix}}_{\mathbf{Z}_{\boldsymbol{i}}^{\boldsymbol{R}}} \begin{bmatrix} \uparrow \\ \boldsymbol{c}_i \\ \downarrow \end{bmatrix}$$

Designing an Estimator

Imagine we could replicate our randomized experiment R times

$$\underbrace{\begin{bmatrix} Y_i(\boldsymbol{z}^{(1)}) \\ Y_i(\boldsymbol{z}^{(2)}) \\ \vdots \\ Y_i(\boldsymbol{z}^{(R)}) \end{bmatrix}}_{\mathbf{Y}_{\boldsymbol{i}}} = \underbrace{\begin{bmatrix} \leftarrow \tilde{\boldsymbol{z}}_i{}^{(1)} \rightarrow \\ \leftarrow \tilde{\boldsymbol{z}}_i{}^{(2)} \rightarrow \\ \vdots \\ \leftarrow \tilde{\boldsymbol{z}}_i{}^{(R)} \rightarrow \end{bmatrix}}_{\mathbf{Z}_{\boldsymbol{i}}^{\boldsymbol{R}}} \begin{bmatrix} \uparrow \\ \boldsymbol{c}_i \\ \downarrow \end{bmatrix}$$

Left-Multiply by $\frac{1}{R}\left(\mathbf{Z}_i^R\right)^\intercal$

$$\begin{tikzcd}
\frac{1}{R}\sum_{r=1}^{R} Y_i\left(\boldsymbol{z}^{(r)}\right) \cdot \tilde{\boldsymbol{z}}_i^{(r)} \arrow[d, "\text{a.s.}"] & = & \frac{1}{R}\sum_{r=1}^{R} \tilde{\boldsymbol{z}}_i^{(r)} \cdot \left(\tilde{\boldsymbol{z}}_i^{(r)}\right)^\intercal \boldsymbol{c}_{\boldsymbol{i}} \arrow[d, "\text{a.s.}"] \\
\mathbb{E}[Y_i(\boldsymbol{z})\ \tilde{\boldsymbol{z}}_i] & = & \mathbb{E}\left[\ \tilde{\boldsymbol{z}}_i\ \tilde{\boldsymbol{z}}_i^\intercal\right]\ \boldsymbol{c}_i
\end{tikzcd}$$

Designing an Estimator

Imagine we could replicate our randomized experiment R times

$$\underbrace{\begin{bmatrix} Y_i(\boldsymbol{z}^{(1)}) \\ Y_i(\boldsymbol{z}^{(2)}) \\ \vdots \\ Y_i(\boldsymbol{z}^{(R)}) \end{bmatrix}}_{\mathbf{Y}_i} = \underbrace{\begin{bmatrix} \leftarrow \tilde{\mathbf{z}}_i{}^{(1)} \rightarrow \\ \leftarrow \tilde{\mathbf{z}}_i{}^{(2)} \rightarrow \\ \vdots \\ \leftarrow \tilde{\mathbf{z}}_i{}^{(R)} \rightarrow \end{bmatrix}}_{\mathbf{Z}_i^R} \begin{bmatrix} \uparrow \\ \boldsymbol{c}_i \\ \downarrow \end{bmatrix}$$

Left-Multiply by $\frac{1}{R}\left(\mathbf{Z}_i^R\right)^\intercal$

$$\begin{tikzcd}
\frac{1}{R}\sum_{r=1}^{R} Y_i\left(\boldsymbol{z}^{(r)}\right) \cdot \tilde{\mathbf{z}}_i^{(r)} \arrow[d, "\text{a.s.}"] & = & \frac{1}{R}\sum_{r=1}^{R} \tilde{\mathbf{z}}_i^{(r)} \cdot \left(\tilde{\mathbf{z}}_i^{(r)}\right)^\intercal \boldsymbol{c}_i \arrow[d, "\text{a.s."}] \\
\mathbb{E}[Y_i(\boldsymbol{z})\, \tilde{\mathbf{z}}_i] & = & \underbrace{\mathbb{E}\left[\, \tilde{\mathbf{z}}_i\, \tilde{\mathbf{z}}_i^\intercal \right]}_{\text{“Design Matrix”}} \boldsymbol{c}_i
\end{tikzcd}$$

The Design Matrix: $\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]$

Entries indexed by subsets of N_i:

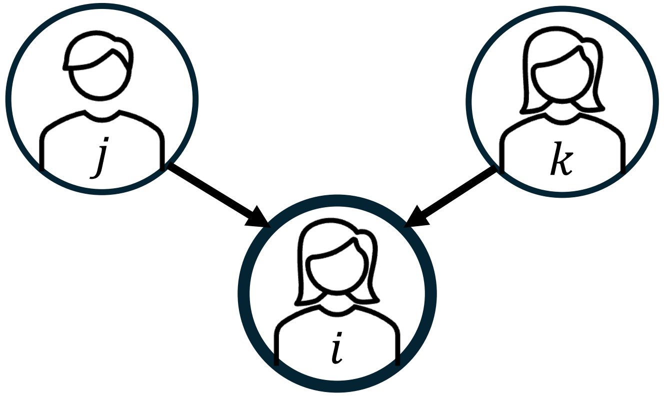
$$\left(\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]\right)_{S,T} = \Pr(S \cup T \text{ fully treated})$$

*** Depends only on the experimental design, not the observed outcomes

For (independent) Bern(p) treatment assignments

$$\left(\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]\right)_{S,T} = p^{|S \cup T|}$$

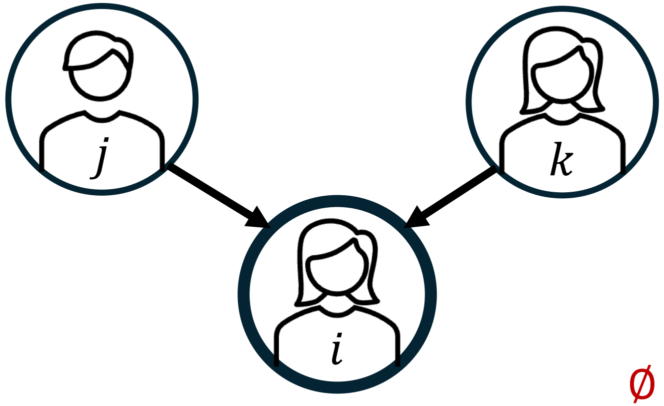
Example:



$\beta = 1$:

$$\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal] = \begin{array}{cc} \begin{matrix} \emptyset & \{i\} & \{j\} & \{k\} \end{matrix} & \\ \begin{bmatrix} 1 & p & p & p \\ p & p & p^2 & p^2 \\ p & p^2 & p & p^2 \\ p & p^2 & p^2 & p \end{bmatrix} & \begin{matrix} \emptyset \\ \{i\} \\ \{j\} \\ \{k\} \end{matrix} \end{array}$$

Example:



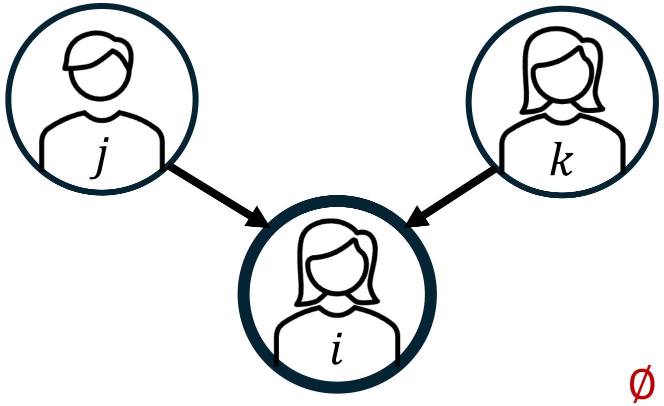
$\beta = 1$:

$$\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal\right] = \begin{array}{c@{}c} \begin{array}{cccc} \emptyset & \{i\} & \{j\} & \{k\} \end{array} & \\ \begin{bmatrix} 1 & p & p & p \\ p & p & p^2 & p^2 \\ p & p^2 & p & p^2 \\ p & p^2 & p^2 & p \end{bmatrix} & \begin{array}{c} \emptyset \\ \{i\} \\ \{j\} \\ \{k\} \end{array} \end{array}$$

$\beta = 2$:

$$\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal\right] = \begin{array}{c@{}c} \begin{array}{ccccccc} \emptyset & \{i\} & \{j\} & \{k\} & \{i,j\} & \{i,k\} & \{j,k\} \end{array} & \\ \begin{bmatrix} 1 & p & p & p & p^2 & p^2 & p^2 \\ p & p & p^2 & p^2 & p^2 & p^2 & p^3 \\ p & p^2 & p & p^2 & p^2 & p^3 & p^2 \\ p & p^2 & p^2 & p & p^3 & p^2 & p^2 \\ p^2 & p^2 & p^2 & p^3 & p^2 & p^3 & p^3 \\ p^2 & p^2 & p^3 & p^2 & p^3 & p^2 & p^3 \\ p^2 & p^3 & p^2 & p^2 & p^3 & p^3 & p^2 \end{bmatrix} & \begin{array}{c} \emptyset \\ \{i\} \\ \{j\} \\ \{k\} \\ \{i,j\} \\ \{i,k\} \\ \{j,k\} \end{array} \end{array}$$

Example:



$\beta = 1$:

$$\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal\right] = \begin{array}{c} \begin{array}{cccc} \emptyset & \{i\} & \{j\} & \{k\} \end{array} \\ \begin{bmatrix} 1 & p & p & p \\ p & p & p^2 & p^2 \\ p & p^2 & p & p^2 \\ p & p^2 & p^2 & p \end{bmatrix} \end{array} \begin{array}{c} \\ \emptyset \\ \{i\} \\ \{j\} \\ \{k\} \end{array}$$

$\beta = 2$:

$$\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal\right] = \begin{array}{c} \begin{array}{ccccccc} \emptyset & \{i\} & \{j\} & \{k\} & \{i,j\} & \{i,k\} & \{j,k\} \end{array} \\ \begin{bmatrix} 1 & p & p & p & p^2 & p^2 & p^2 \\ p & p & p^2 & p^2 & p^2 & p^2 & p^3 \\ p & p^2 & p & p^2 & p^2 & p^3 & p^2 \\ p & p^2 & p^2 & p & p^3 & p^2 & p^2 \\ p^2 & p^2 & p^2 & p^3 & p^2 & p^3 & p^3 \\ p^2 & p^2 & p^3 & p^2 & p^3 & p^2 & p^3 \\ p^2 & p^3 & p^2 & p^2 & p^3 & p^3 & p^2 \end{bmatrix} \end{array} \begin{array}{c} \\ \emptyset \\ \{i\} \\ \{j\} \\ \{k\} \\ \{i,j\} \\ \{i,k\} \\ \{j,k\} \end{array}$$

Theorem: Under independent Bernoulli design, each design matrix is invertible.

Designing an Estimator

$$\mathbb{E}[Y_i(\boldsymbol{z})\, \tilde{\boldsymbol{z}}_i] = \mathbb{E}\left[\, \tilde{\boldsymbol{z}}_i\, \tilde{\boldsymbol{z}}_i^\intercal\right]\boldsymbol{c}_i \qquad \Longrightarrow \qquad \boldsymbol{c}_i = \mathbb{E}\left[\, \tilde{\boldsymbol{z}}_i\, \tilde{\boldsymbol{z}}_i^\intercal\right]^{-1}\mathbb{E}[Y_i(\boldsymbol{z})\, \tilde{\boldsymbol{z}}_i]$$

Designing an Estimator

$$\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\boldsymbol{z}}_i] = \mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\top\big]\boldsymbol{c}_i \qquad \Longrightarrow \qquad \boldsymbol{c}_i = \mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\top\big]^{-1}\underbrace{\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\boldsymbol{z}}_i]}_{\text{Replace with single realization}}$$

$$\hat{\boldsymbol{c}}_i = Y_i(\boldsymbol{z})\,\mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\top\big]^{-1}\,\tilde{\boldsymbol{z}}_i$$

Designing an Estimator

$$\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\mathbf{z}}_i] = \mathbb{E}\big[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\top\big]\boldsymbol{c}_i \qquad \Longrightarrow \qquad \boldsymbol{c}_i = \mathbb{E}\big[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\top\big]^{-1}\underbrace{\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\mathbf{z}}_i]}_{\text{Replace with single realization}}$$

$$\hat{\boldsymbol{c}}_i = Y_i(\boldsymbol{z})\;\mathbb{E}\big[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\top\big]^{-1}\,\tilde{\mathbf{z}}_i$$

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i=1}^{n}\langle\hat{\mathbf{c}}_i, \boldsymbol{\theta}_i\rangle = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\,\Big\langle \mathbb{E}\big[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\top\big]^{-1}\boldsymbol{\theta}_i, \tilde{\mathbf{z}}_i\Big\rangle$$

Designing an Estimator

$$\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\boldsymbol{z}}_i] = \mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^{\intercal}\big]\boldsymbol{c}_i \qquad \Longrightarrow \qquad \boldsymbol{c}_i = \mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^{\intercal}\big]^{-1}\underbrace{\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\boldsymbol{z}}_i]}_{\text{Replace with single realization}}$$

$$\hat{\boldsymbol{c}}_i = Y_i(\boldsymbol{z})\,\mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^{\intercal}\big]^{-1}\,\tilde{\boldsymbol{z}}_i$$

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i=1}^{n}\langle\hat{\mathbf{c}}_i, \boldsymbol{\theta}_i\rangle = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\,\left\langle \mathbb{E}\big[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^{\intercal}\big]^{-1}\boldsymbol{\theta}_i, \tilde{\boldsymbol{z}}_i\right\rangle$$

$$\Big\downarrow\ \text{… Lots of Algebra …}$$

$$\boxed{\widehat{\mathrm{TTE}}_{\mathrm{SNIPE}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z}) \sum_{S\in\mathcal{S}_i^{\beta}}\left(\prod_{j\in S}\frac{z_j - p}{p} - \prod_{j\in S}\frac{z_j - p}{p-1}\right)}$$

Structured

Neighborhood **I**nterference

Polynomial **E**stimator

Properties of the SNIPE Estimator

- Unbiased
- Variance $= O\left(\frac{d^2}{n} \cdot \left(\frac{ed}{\beta p(1-p)}\right)^{\beta}\right)$

n = graph size
d = maximum vertex degree
β = model degree
p = treatment probability

Properties of the SNIPE Estimator

- Unbiased
- Variance $= O\left(\frac{d^2}{n} \cdot \left(\frac{ed}{\beta p(1-p)}\right)^{\beta}\right)$

n = graph size
d = maximum vertex degree
β = model degree
p = treatment probability

<u>Interpreting the Variance:</u>

- $\frac{1}{n}$ ensures consistency of estimator
- Polynomial scaling in d, exponential scaling in β
- Minimax $\Omega\left(\frac{1}{np^{\beta}}\right)$ lower bound on variance
- Compare with $\Theta\left(\frac{1}{np^{d}}\right)$ variance of HT estimator

Analyzing the SNIPE Estimator

Dataset

- Vertices: 19,828 DVDs sold on Amazon
- Arcs: Each DVD connected to 5 most frequent co-purchases

Potential Outcomes Model

- Variation on Ugander Yin model:

$$Y_i(\mathbf{0}) = (\alpha + \beta h_i + \sigma) \cdot \frac{d_i}{\bar{d}} \qquad Y_i(\mathbf{z}) = Y_i(\mathbf{0}) \left(\delta_i z_i + \sum_{S \in S_i^{\beta}} \gamma_{|S|} \cdot \binom{d_i}{|S|}^{-1} \tilde{\mathbf{z}}_S \right)$$

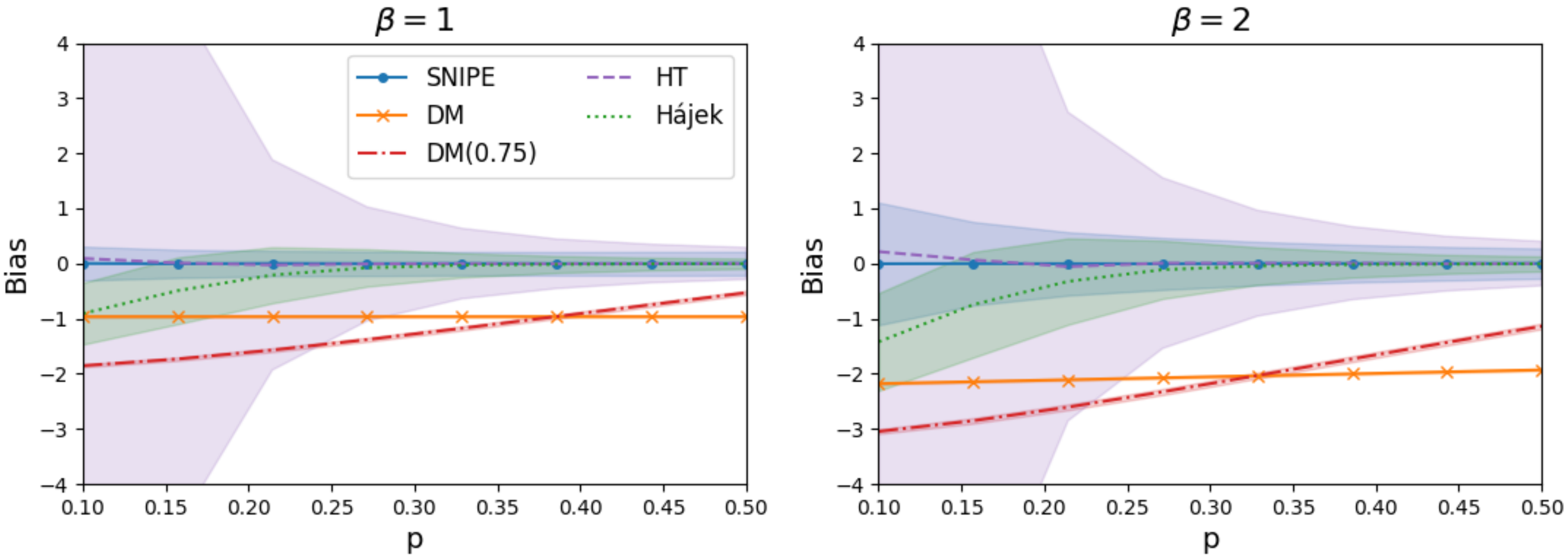
Incorporates:

- Homophily, heterogeneous treatments, degree-dependent effects

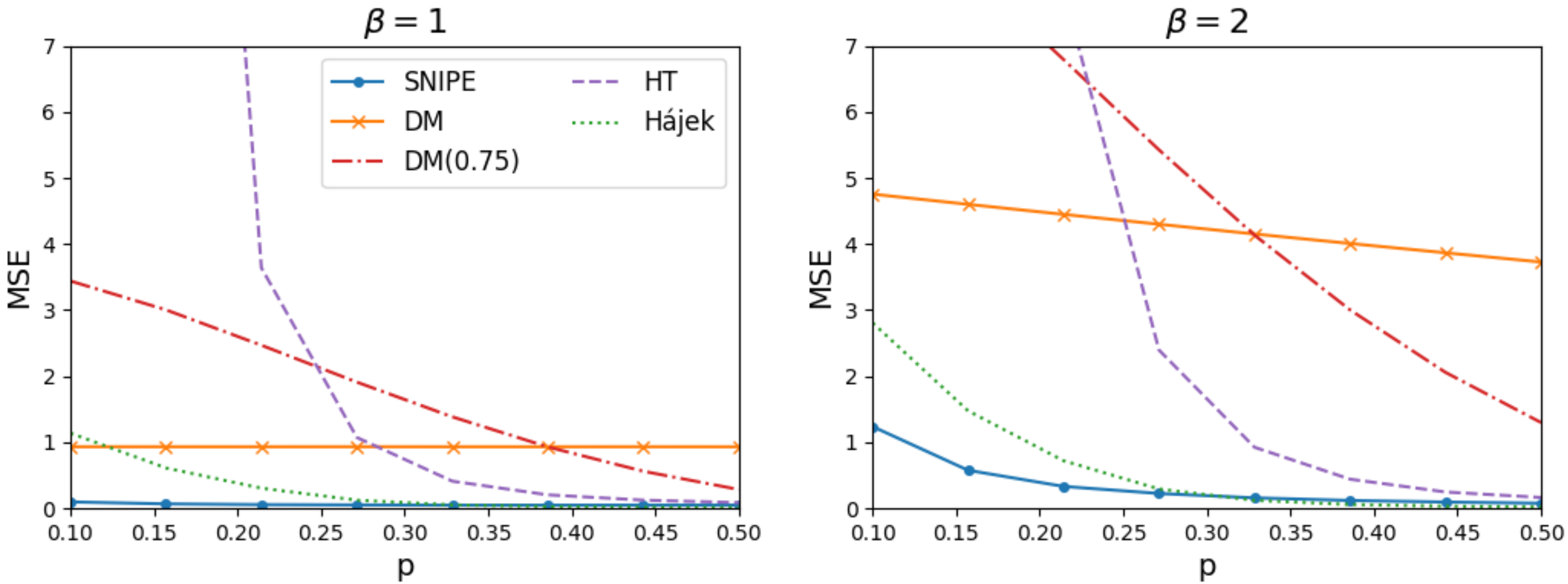
Leskovec, Jure , Andrej Krevl. "SNAP Datasets: Stanford Large Network Dataset Collection." . (2014).

Ugander, Johan, and Hao Yin. "Randomized graph cluster randomization." *Journal of Causal Inference* 11.1 (2023): 20220014.

Analyzing the SNIPE Estimator



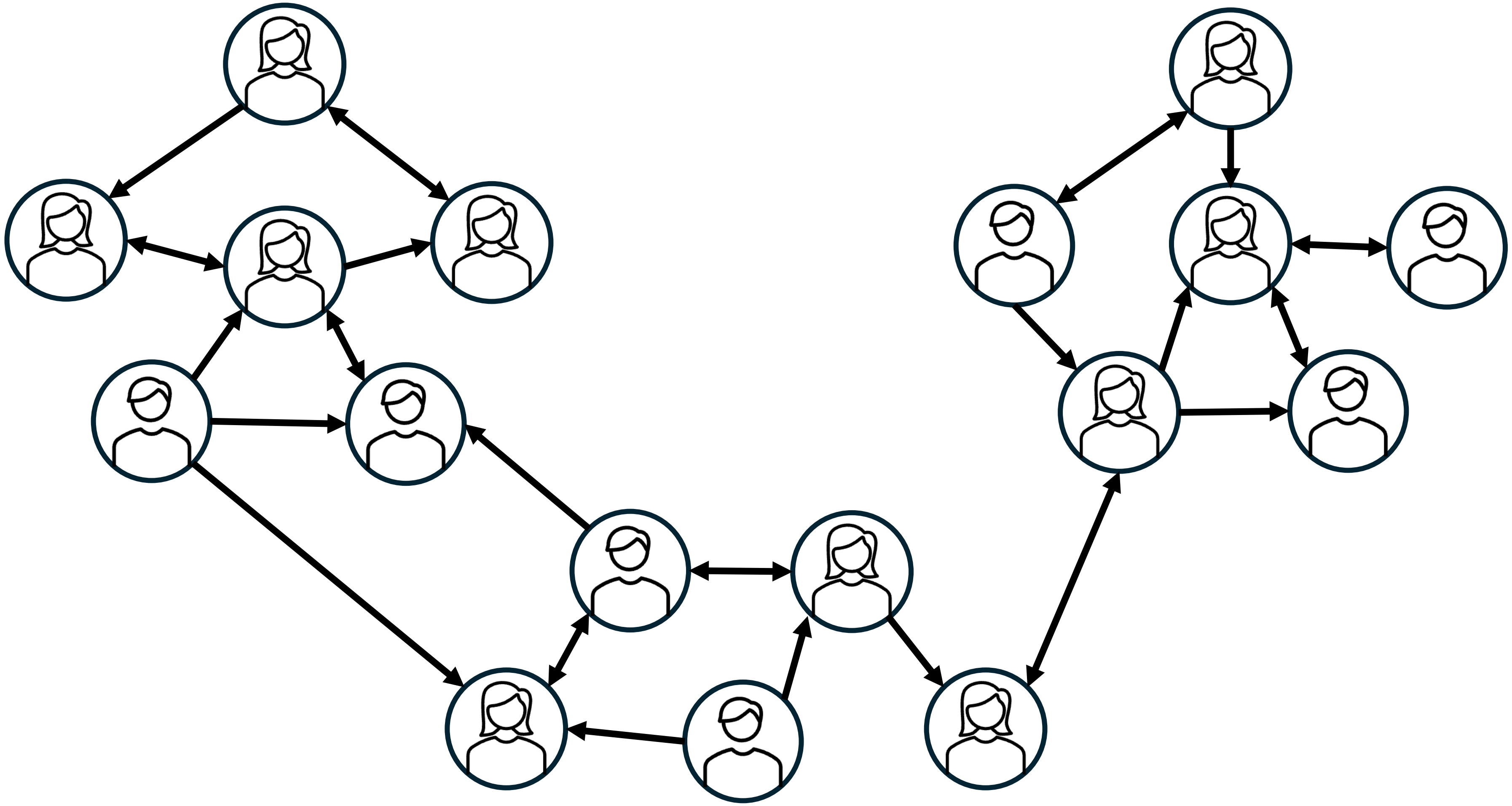
Analyzing the SNIPE Estimator



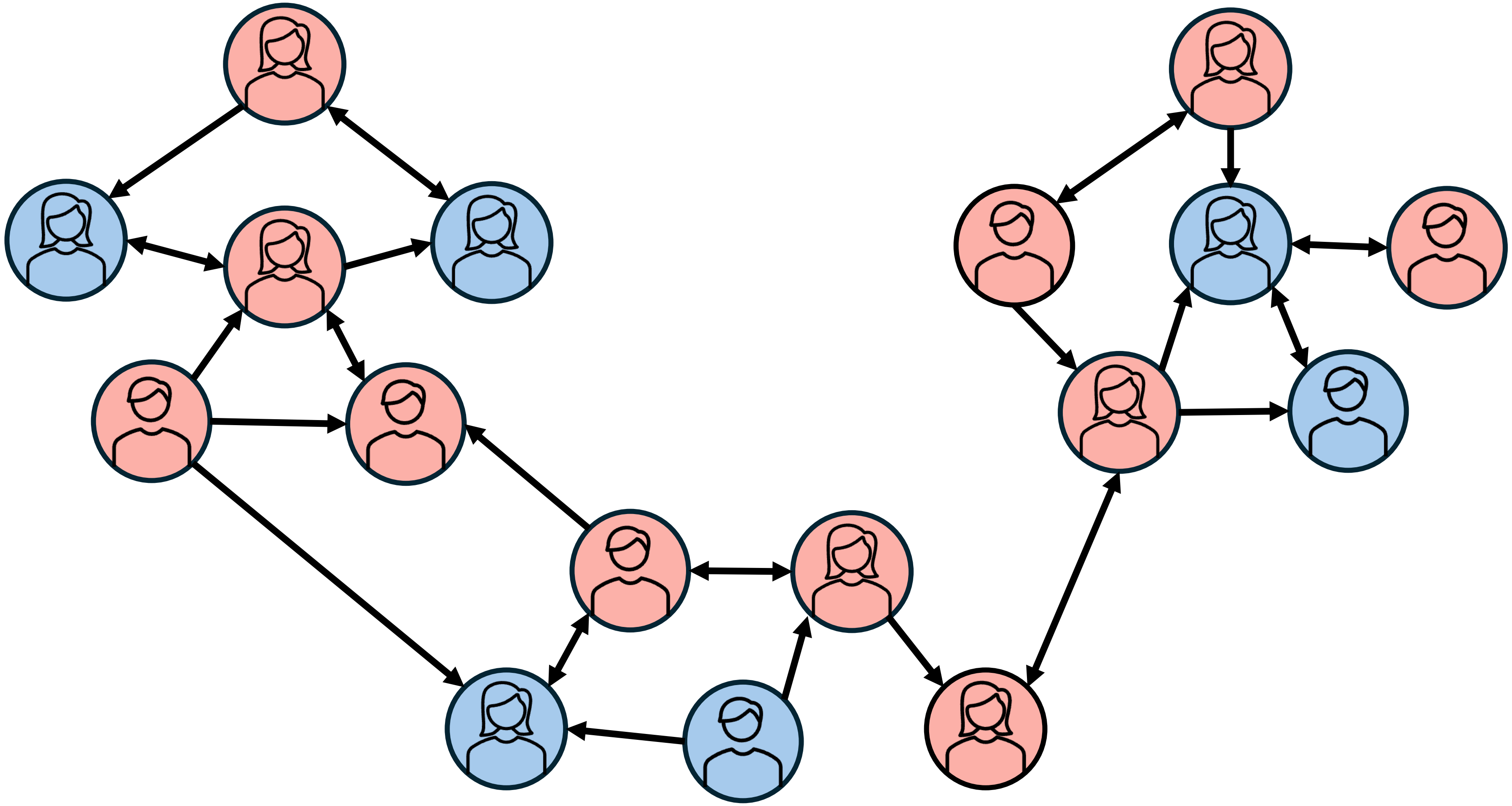
Estimating TTE under arbitrary experimental designs

Eichhorn, Matthew, Samir Khan, Johan Ugander, and Christina Lee Yu. "Low-order outcomes and clustered designs: combining design and analysis for causal inference under network interference." *arXiv preprint arXiv:2405.07979* (2024).

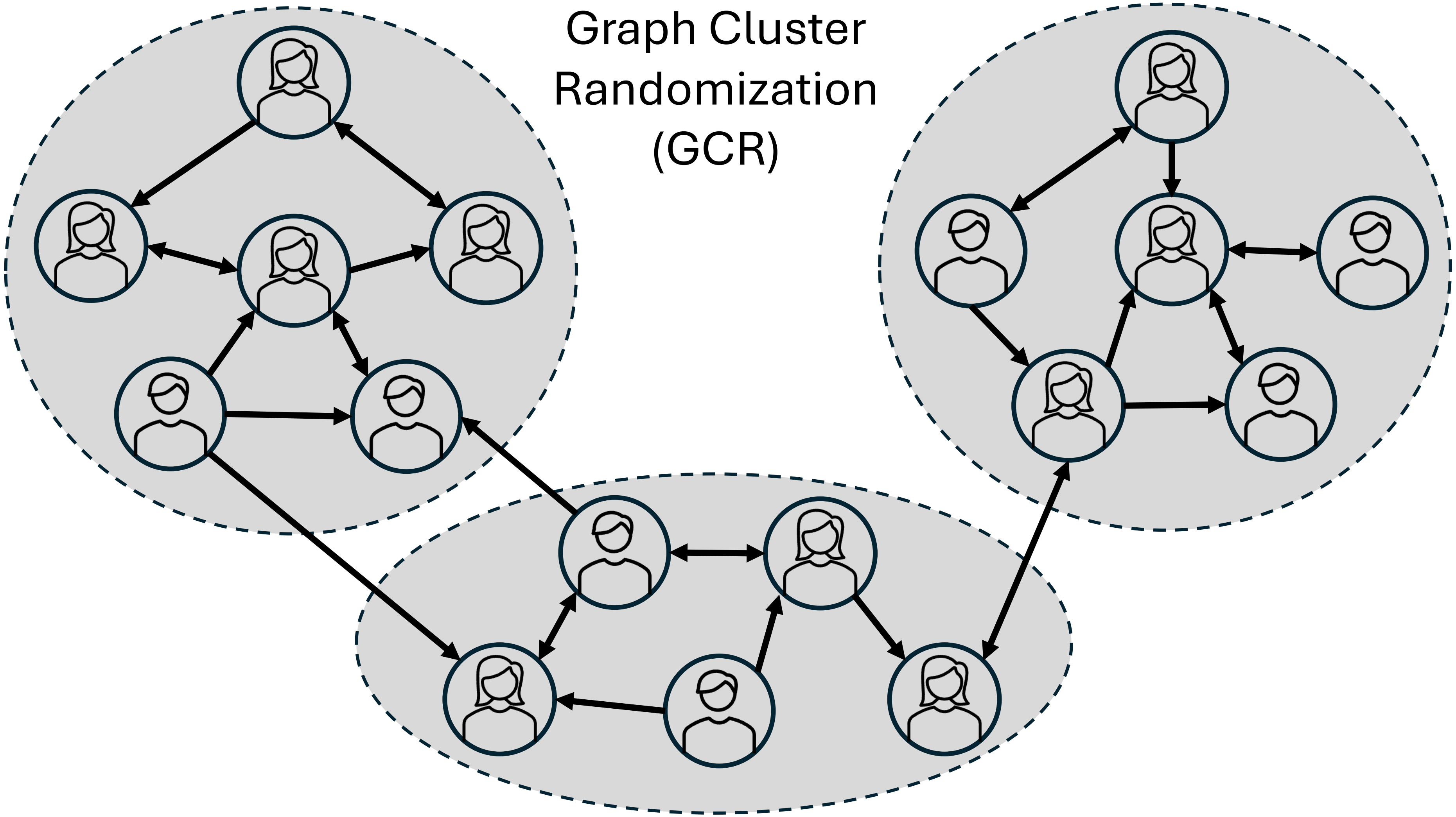
Other Experimental Designs



Other Experimental Designs

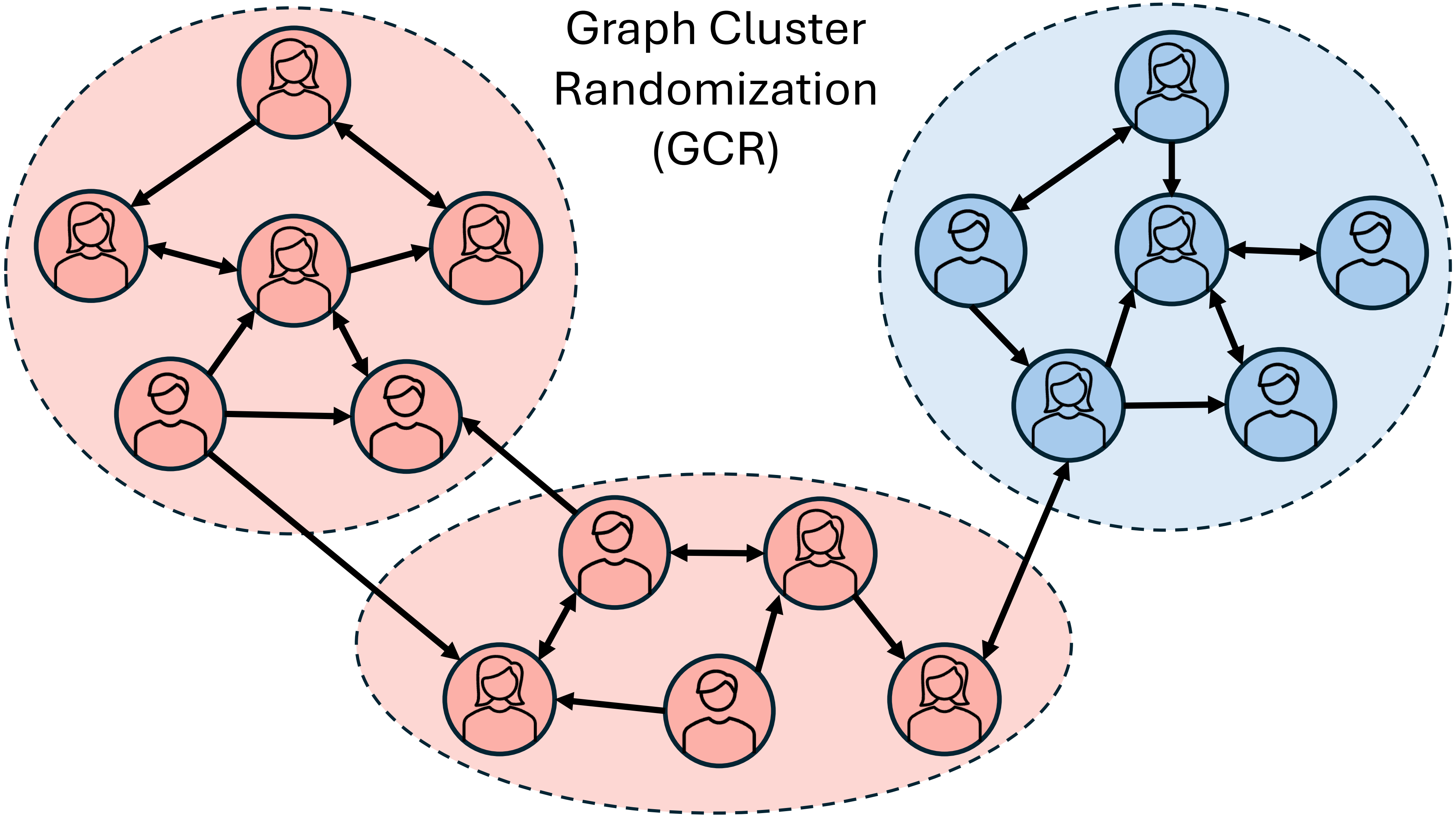


Other Experimental Designs



Graph Cluster
Randomization
(GCR)

Other Experimental Designs



Graph Cluster Randomization (GCR)

Generalizing the SNIPE Estimator

The identity $\mathbb{E}[Y_i(\boldsymbol{z})\,\tilde{\boldsymbol{z}}_i]$= $\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]\boldsymbol{c}_i$ holds for any experimental design

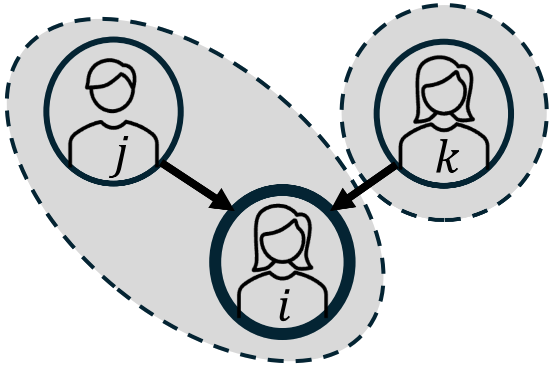
The design matrix $\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]$ may not be invertible

Use the (Moore-Penrose) pseudoinverse to estimate $\hat{\boldsymbol{c}}_i = Y_i(\boldsymbol{z})\,\mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]^\dagger\,\tilde{\boldsymbol{z}}_i$

Pseudoinverse Estimator:

$$\widehat{\mathrm{TTE}}_{\mathrm{PI}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\,\left\langle \mathbb{E}[\,\tilde{\boldsymbol{z}}_i\,\tilde{\boldsymbol{z}}_i^\intercal]^\dagger \boldsymbol{\theta}_i, \tilde{\boldsymbol{z}}_i \right\rangle$$

Example: Bernoulli GCR



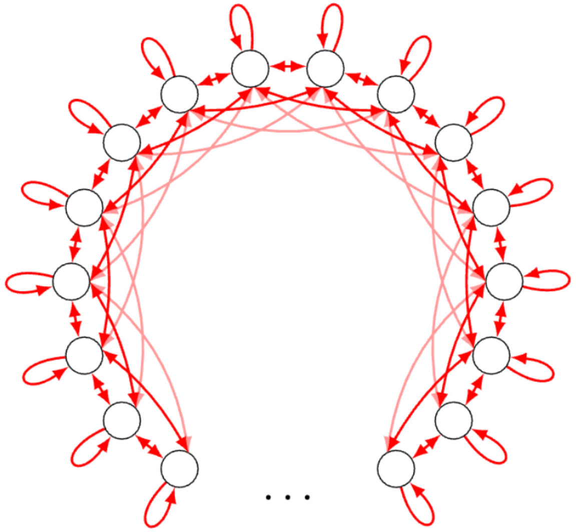
$\beta = 1$:

$$\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal] = \begin{array}{cccc|c} \emptyset & \{i\} & \{j\} & \{k\} & \\ \hline 1 & p & p & p & \emptyset \\ p & p & p & p^2 & \{i\} \\ p & p & p & p^2 & \{j\} \\ p & p^2 & p^2 & p & \{k\} \end{array}$$

$\beta = 2$:

$$\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal] = \begin{array}{ccccccc|c} \emptyset & \{i\} & \{j\} & \{i,j\} & \{k\} & \{i,k\} & \{j,k\} & \\ \hline 1 & p & p & p & p & p^2 & p^2 & \emptyset \\ p & p & p & p & p^2 & p^2 & p^2 & \{i\} \\ p & p & p & p & p^2 & p^2 & p^2 & \{j\} \\ p & p & p & p & p^2 & p^2 & p^2 & \{i,j\} \\ p & p^2 & p^2 & p^2 & p & p^2 & p^2 & \{k\} \\ p^2 & p^2 & p^2 & p^2 & p^2 & p^2 & p^2 & \{i,k\} \\ p^2 & p^2 & p^2 & p^2 & p^2 & p^2 & p^2 & \{j,k\} \end{array}$$

PI Estimator with Bernoulli GCR Designs

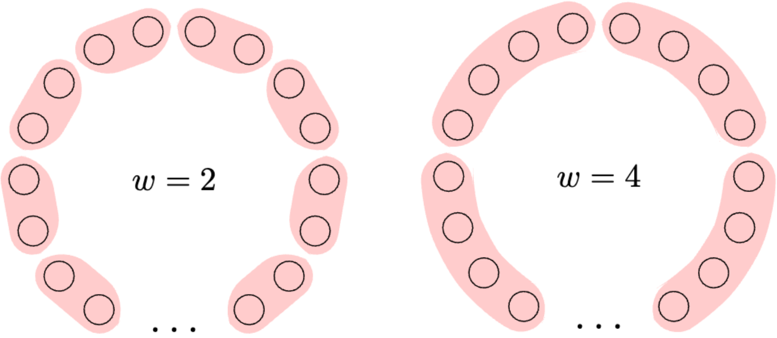


Toy experiment on powers of a cycle graph

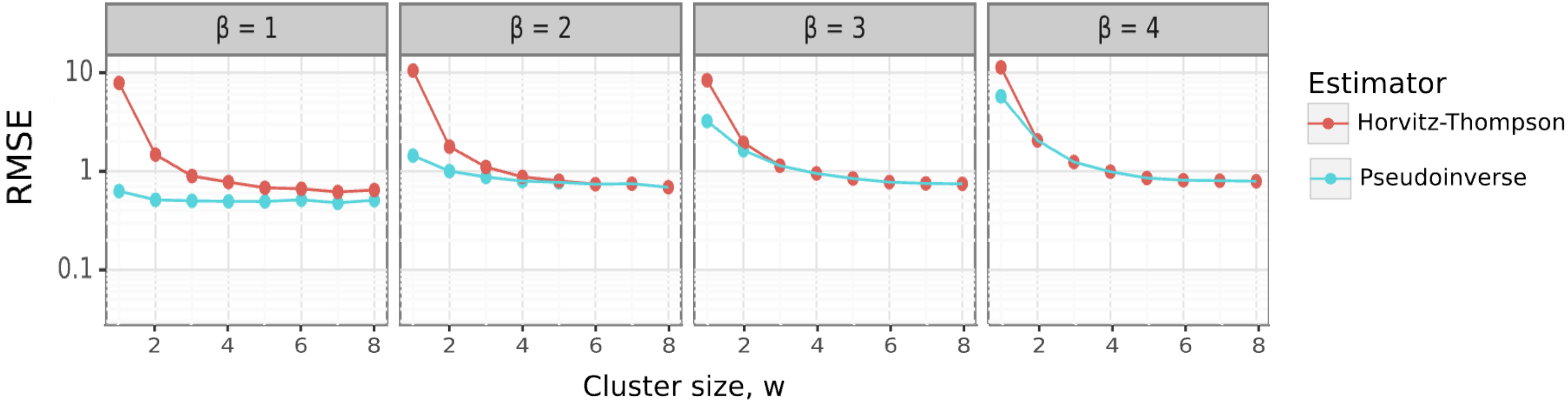
- Restricted growth graphs
- Clustering reduces error of HT Estimator

Compare against PI estimator for different values of β

Consider *contiguous clusterings* of various sizes w



PI Estimator with Bernoulli GCR Designs



Larger clusters = Fewer vertices with "cross cluster" neighbors (HT improves)

Larger β = Weaker Modeling Assumptions (PI gets worse, HT unaffected)

"Best of Both Worlds" Tradeoff

Theoretical Results

Bias:

$$\left|\mathbb{E}\left[\widehat{\mathrm{TTE}}_{\mathrm{PI}}\right] - \mathrm{TTE}\right| \leq O\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\left(\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}\right]^{\dagger}\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}\right] - I\right)\theta_i\right\|_2\right)$$

Unbiased when each θ_i lies in the column space of $\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}\right]$

Variance:

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

$\gamma_i \;=\; \sqrt{\left|S_i^{\beta}\right| \cdot \theta_i^{\intercal}\mathbb{E}\left[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}\right]^{\dagger}\theta_i}$ measures "sensitivity" of unit *i* to randomized design

Specialized to Bernoulli GCR

- $\widehat{\mathrm{TTE}}_{\mathrm{PI}}$ is unbiased
- $\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j$ when i and j have neighbors in the same cluster
- $\gamma_i = \begin{cases} O\left(d_i^{\beta} \cdot p^{-|C(N_i)|}\right) & |C(N_i)| < \beta \\ O\left(d_i^{\beta} \cdot |C(N_i)|^{\beta} \cdot p^{-\beta}\right) & |C(N_i)| \geq \beta \end{cases}$ Stronger assumption on graph / Stronger assumption on outcomes

Variance	Unit Randomization	Cluster Randomization
General Interference	$\exp(d)$	$\exp(\lvert C(N_i)\rvert)$
β-Order Interactions	$\exp(\beta)$	$\exp(\min(\beta, \lvert C(N_i)\rvert))$

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

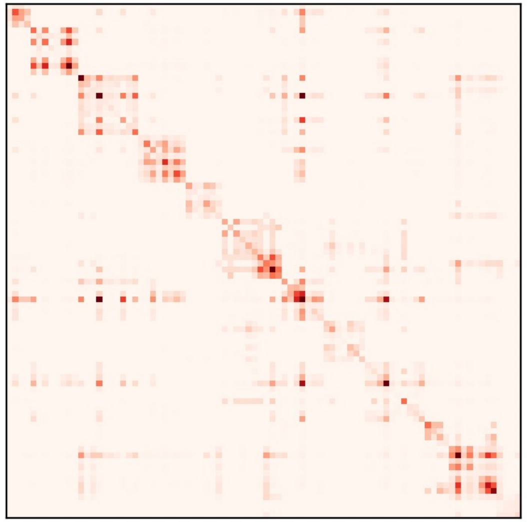
Selecting an Experimental Design

Eichhorn, Matthew, Samir Khan, Johan Ugander, and Christina Lee Yu. "Low-order outcomes and clustered designs: combining design and analysis for causal inference under network interference." *arXiv preprint arXiv:2405.07979* (2024).

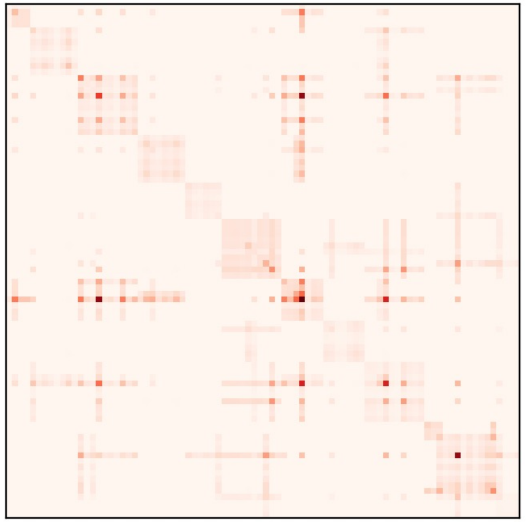
Visualizing the Variance

Variance contribution of each vertex pair in a small collaboration network

$$\mathrm{Var}\left(\widehat{\mathrm{TTE}}_{\mathrm{PI}}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(w_iY_i, w_jY_j)$$



Unit Bernoulli



Bernoulli GCR

Ryan A. Rossi, & Nesreen K. Ahmed (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.

Visualizing the Variance

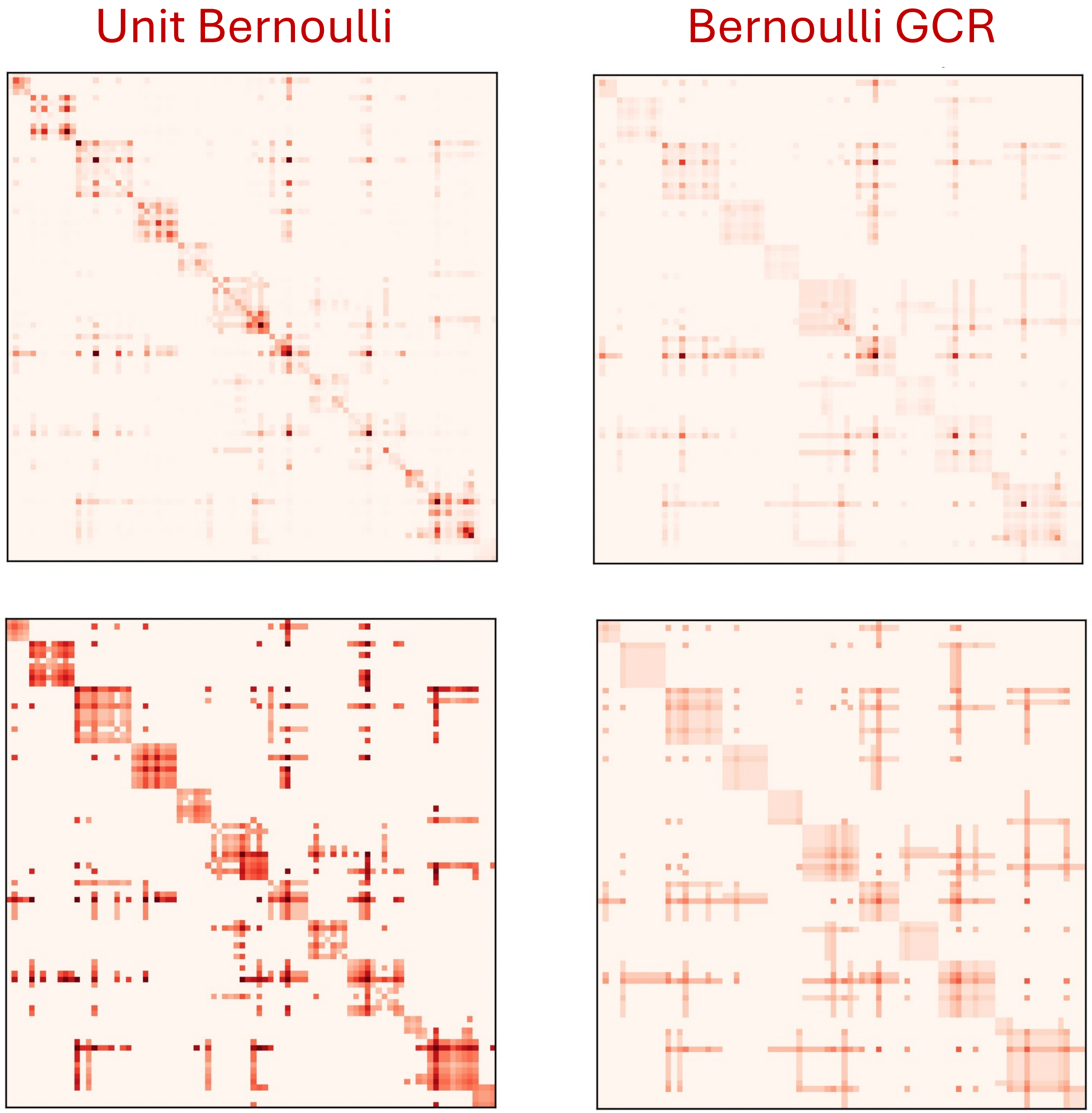
Variance contribution of each vertex pair in a small collaboration network

$$\mathrm{Var}\left(\widehat{\mathrm{TTE}}_{\mathrm{PI}}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(w_i Y_i, w_j Y_j)$$

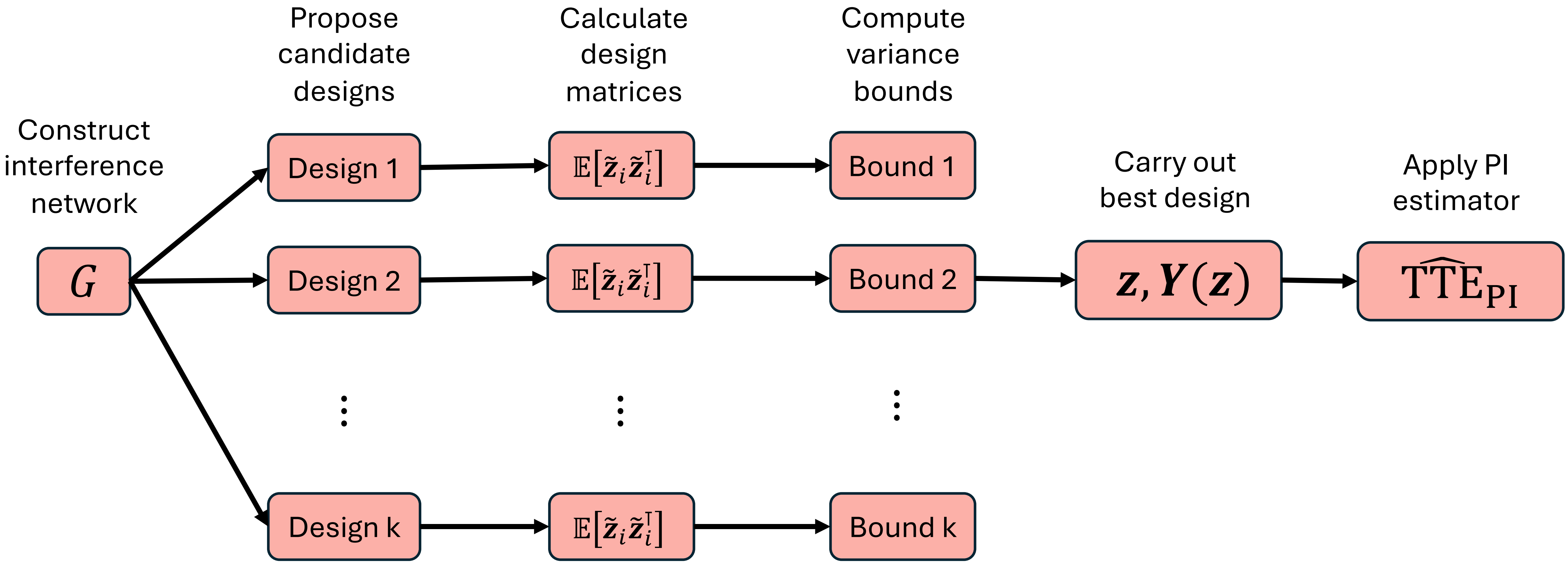
Contributions to the Variance Bound

$$\gamma_i \cdot \gamma_j \cdot \mathbb{I}(\tilde{\mathbf{z}}_i \perp \tilde{\mathbf{z}}_j)$$

Ryan A. Rossi, & Nesreen K. Ahmed (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.



Experimental Pipeline



Construct interference network

G

Propose candidate designs

Design 1

Design 2

⋮

Design k

Calculate design matrices

$\mathbb{E}[\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\intercal]$

$\mathbb{E}[\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\intercal]$

⋮

$\mathbb{E}[\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\intercal]$

Compute variance bounds

Bound 1

Bound 2

⋮

Bound k

Carry out best design

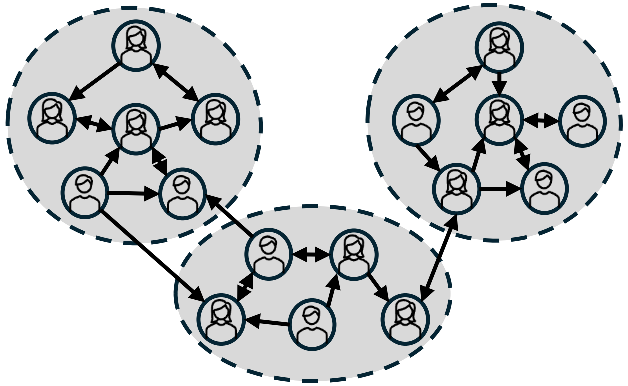
$\boldsymbol{z}, \boldsymbol{Y}(\boldsymbol{z})$

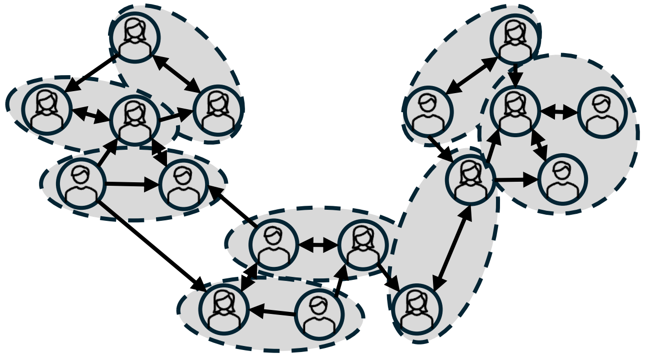
Apply PI estimator

$\widehat{\mathrm{TTE}}_{\mathrm{PI}}$

Example: Clustering Stochastic Block Models

At what *Louvain clustering resolution* does the $\widehat{\mathrm{TTE}}_{\mathrm{PI}}$ estimator with Bernoulli GCR have minimum variance?





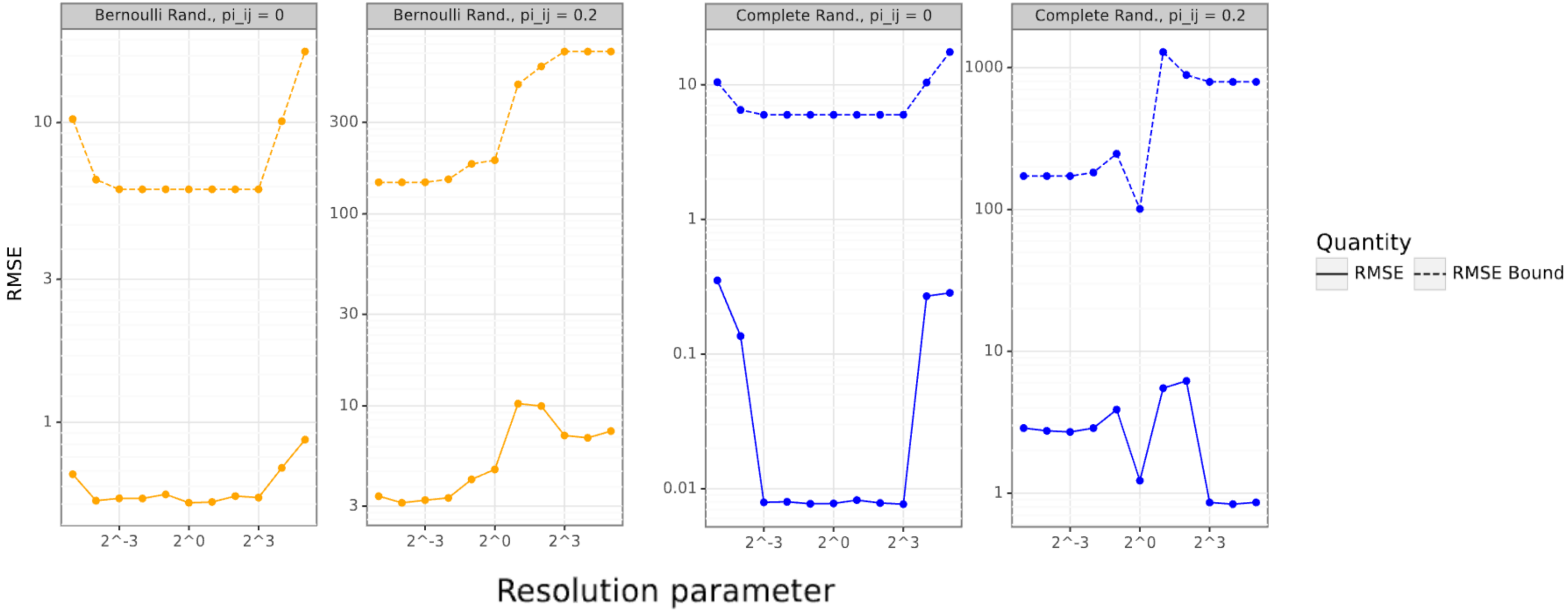
Resolution

Low Resolution ⟷	**High Resolution**
Few larger clusters	Many smaller clusters
More dependent vertex pairs	γ_i tend the be larger

Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.

Example: Clustering Stochastic Block Models

Though the theoretical bounds are loose, they capture the behavior of the estimator



Main Takeaways

- β-order interactions
  - Rich framework for modeling interference
  - Hierarchy of sparse bases for outcome parameterization

- Pseudoinverse estimators
  - Leverage outcome structure to give improvements over existing approaches
  - Can be adapted to arbitrary experimental designs

- Novel bias and variance results in terms of properties of the design
  - Provide a principled way to select an experimental design

Ongoing Question:

How can we best select a (design, estimator) pair?

References

Our Work:

Cortez, Mayleen, Matthew Eichhorn, and Christina Yu. "Staggered rollout designs enable causal inference under interference without network knowledge." *Advances in Neural Information Processing Systems* 35 (2022): 7437-7449.

Cortez-Rodriguez, Mayleen, Matthew Eichhorn, and Christina Lee Yu. "Exploiting neighborhood interference with low-order interactions under unit randomized design." *Journal of Causal Inference* 11.1 (2023): 20220051.

Eichhorn, Matthew, et al. "Low-order outcomes and clustered designs: combining design and analysis for causal inference under network interference." *arXiv preprint arXiv:2405.07979* (2024).

SNIPE Estimator

Pseudoinverse Estimators

References

Network interference:

Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Aronow, Peter M., and Cyrus Samii. "Estimating average causal effects under general interference, with application to a social network experiment." (2017): 1912-1947.

Sussman, Daniel L., and Edoardo M. Airoldi. "Elements of estimation theory for causal effects in the presence of network interference." *arXiv preprint arXiv:1702.03578* (2017).

Regression-based estimators (green in chart):

Chin, Alex. "Regression adjustments for estimating the global treatment effect in experiments with interference." *Journal of Causal Inference* 7.2 (2019): 20180026.

Toulis, Panos, and Edward Kao. "Estimation of causal peer influence effects." *International conference on machine learning*. PMLR, 2013.

Basse, Guillaume W., and Edoardo M. Airoldi. "Model-assisted design of experiments in the presence of network-correlated outcomes." *Biometrika* 105.4 (2018): 849-858.

Parker, Ben M., Steven G. Gilmour, and John Schormans. "Optimal design of experiments on connected units with application to social networks." *Journal of the Royal Statistical Society Series C: Applied Statistics* 66.3 (2017): 455-480.

References

Novel experimental designs under general interference (purple in chart):

Tchetgen, Eric J. Tchetgen, and Tyler J. VanderWeele. "On causal inference in the presence of interference." *Statistical methods in medical research* 21.1 (2012): 55-75.

Hudgens, Michael G., and M. Elizabeth Halloran. "Toward causal inference with interference." *Journal of the American Statistical Association* 103.482 (2008): 832-842.

Rosenbaum, Paul R. "Interference between units in randomized experiments." *Journal of the american statistical association* 102.477 (2007): 191-200.

Eckles, Dean, Brian Karrer, and Johan Ugander. "Design and analysis of experiments in networks: Reducing bias from interference." *Journal of Causal Inference* 5.1 (2017): 20150021.

Ugander, Johan, and Hao Yin. "Randomized graph cluster randomization." *Journal of Causal Inference* 11.1 (2023): 20220014.

Horvitz-Thompson estimation (gray in chart):

Horvitz, Daniel G., and Donovan J. Thompson. "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association* 47.260 (1952): 663-685.

References

Low-order outcomes in other work:

Yu, Christina Lee, et al. "Estimating the total treatment effect in randomized experiments with unknown network structure." *PNAS* 119.44 (2022):

Deng, Lu, et al. "Unbiased Estimation for Total Treatment Effect Under Interference Using Aggregated Dyadic Data." *arXiv preprint arXiv:2402.12653* (2024).

Yuan, Yuan, Kristen Altenburger, and Farshad Kooti. "Causal network motifs: Identifying heterogeneous spillover effects in a/b tests." *Proceedings of the Web Conference 2021*.

Experiments:

Leskovec, Jure , Andrej Krevl. "SNAP Datasets: Stanford Large Network Dataset Collection." . (2014).

Ryan A. Rossi, & Nesreen K. Ahmed (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.

Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.10 (2008): P10008.