Low-order Outcomes and Clustered Designs

Combining Design and Analysis for Causal Inference under Network Interference

Matthew Eichhorn

Samir Khan

Johan Ugander

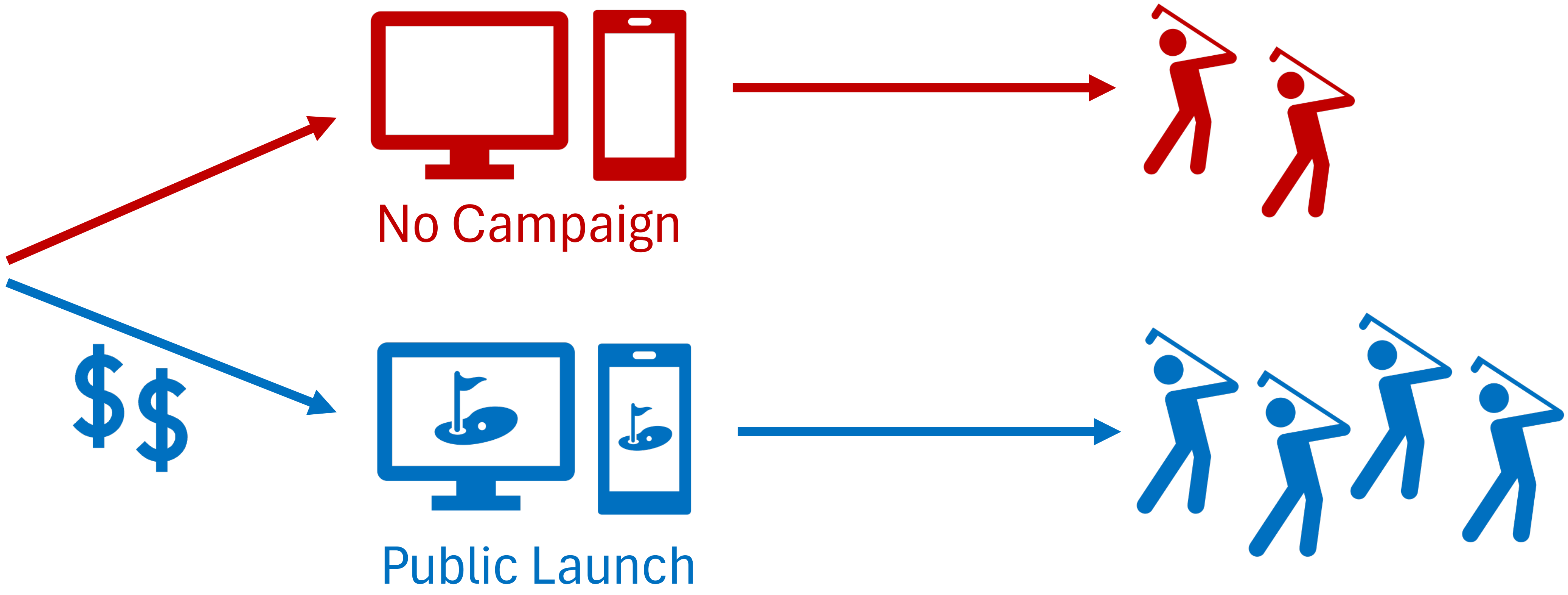
Christina Lee Yu

Statistical Society of Canada Annual Meeting

May 26, 2025

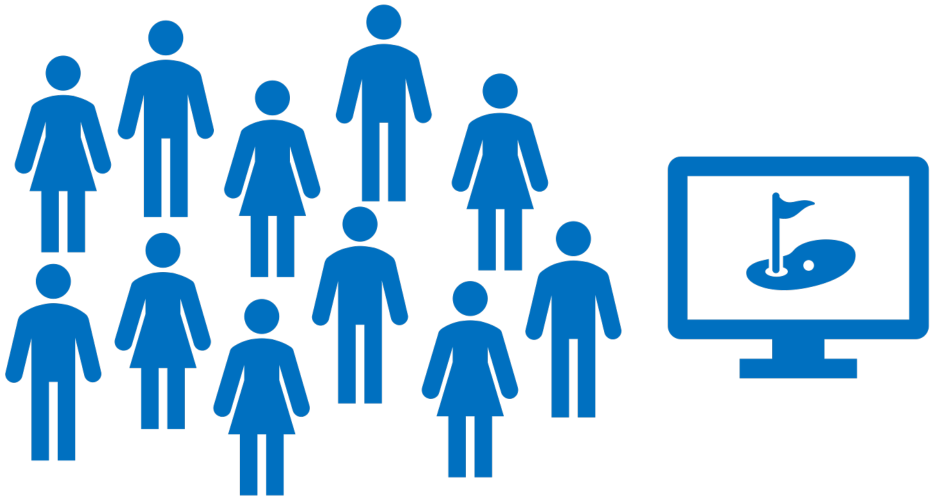
Motivating Example: Advertising

A golf course is deciding whether to run an advertising campaign



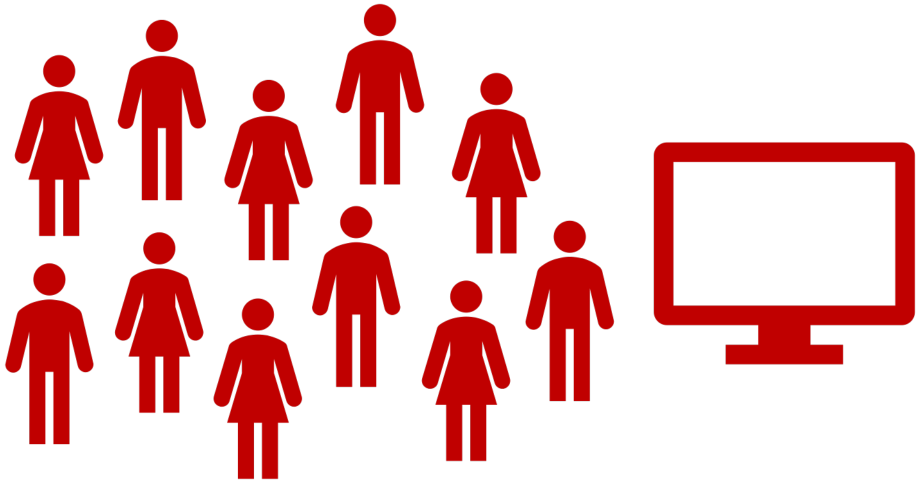
Total Treatment Effect

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



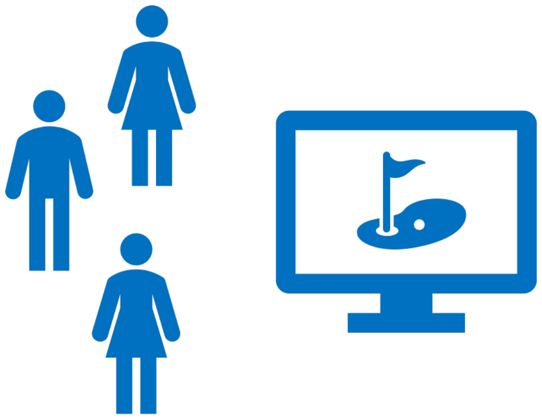
Everybody Treated

vs.

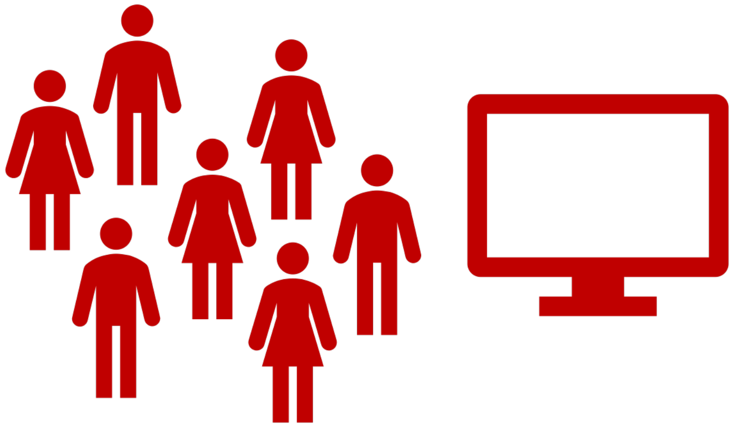


Nobody Treated

Randomized Experiment



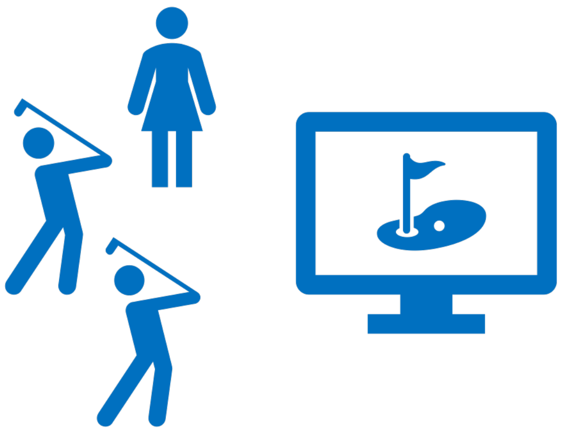
Treatment Group



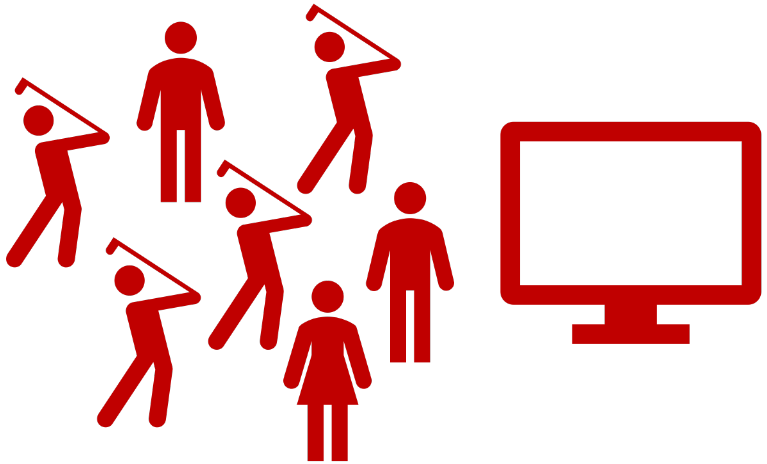
Control Group

*** 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.

Difference in Means Estimator:

$$\widehat{\mathrm{TTE}}_{\mathrm{DM}} = \text{Average Outcome in Treatment Group} - \text{Average Outcome in Control Group}$$

Interference

Individuals' outcomes may change even if they are not treated

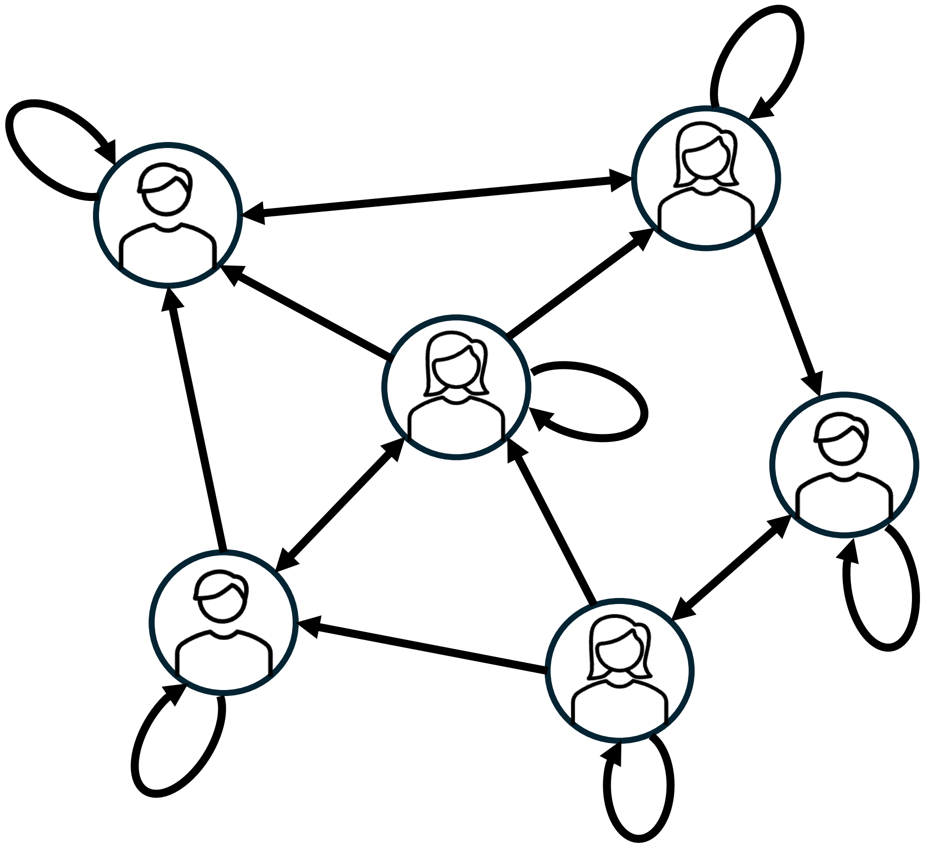
Introduces Bias into DM Estimator

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.

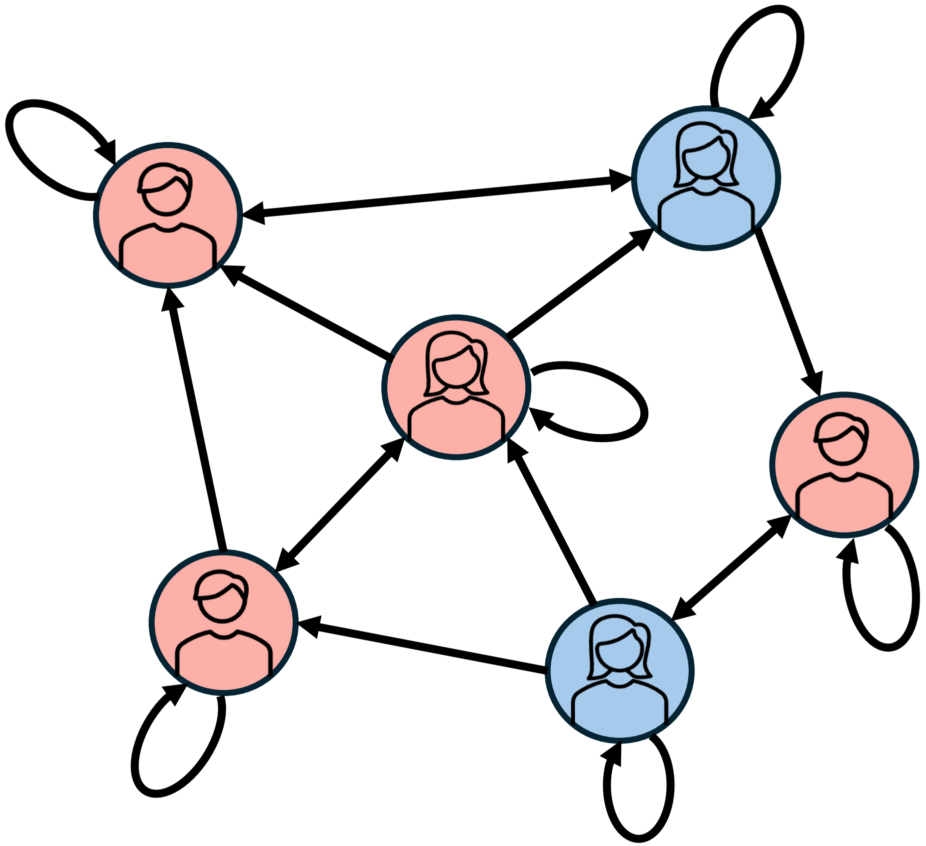
Modeling Interference

Directed <u>Interference Graph</u> $G = (V, A)$

V = n individuals

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

<u>Treatment Assignments</u> $\boldsymbol{z} \in \{0,1\}^n$



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.

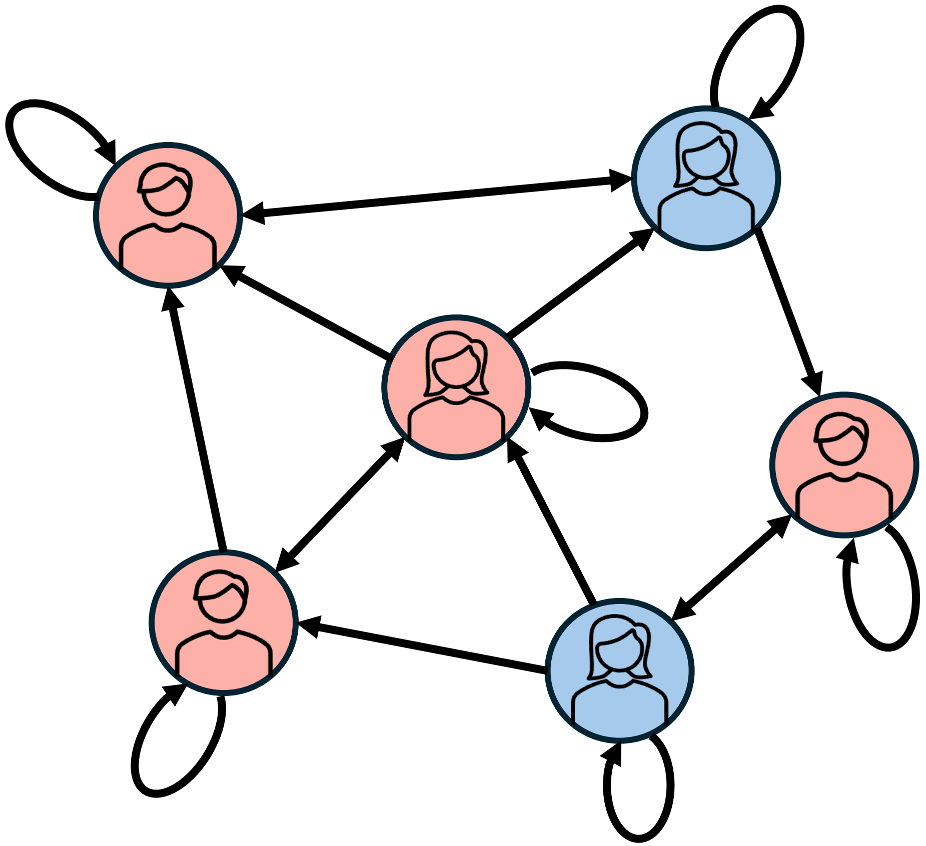
Aronow, Peter M., and Cyrus Samii. "Estimating average causal effects under general interference, with application to a social network experiment." (2017): 1912-1947.

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

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

Neighborhood Interference Assumption:

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

Total Treatment Effect:

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

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 $z_j \sim \text{Bernoulli}(p)$:

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z})\left(\prod_{j\in N_i}\frac{z_j}{p} - \prod_{j\in N_i}\frac{1-z_j}{1-p}\right)$$

0 unless entire neighborhood treated

0 unless entire neighborhood untreated

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 $z_j \sim \text{Bernoulli}(p)$:

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(\boldsymbol{z}) \left(\prod_{j \in N_i} \frac{z_j}{p} - \prod_{j \in N_i} \frac{1-z_j}{1-p} \right)$$

0 unless entire neighborhood treated

0 unless entire neighborhood untreated

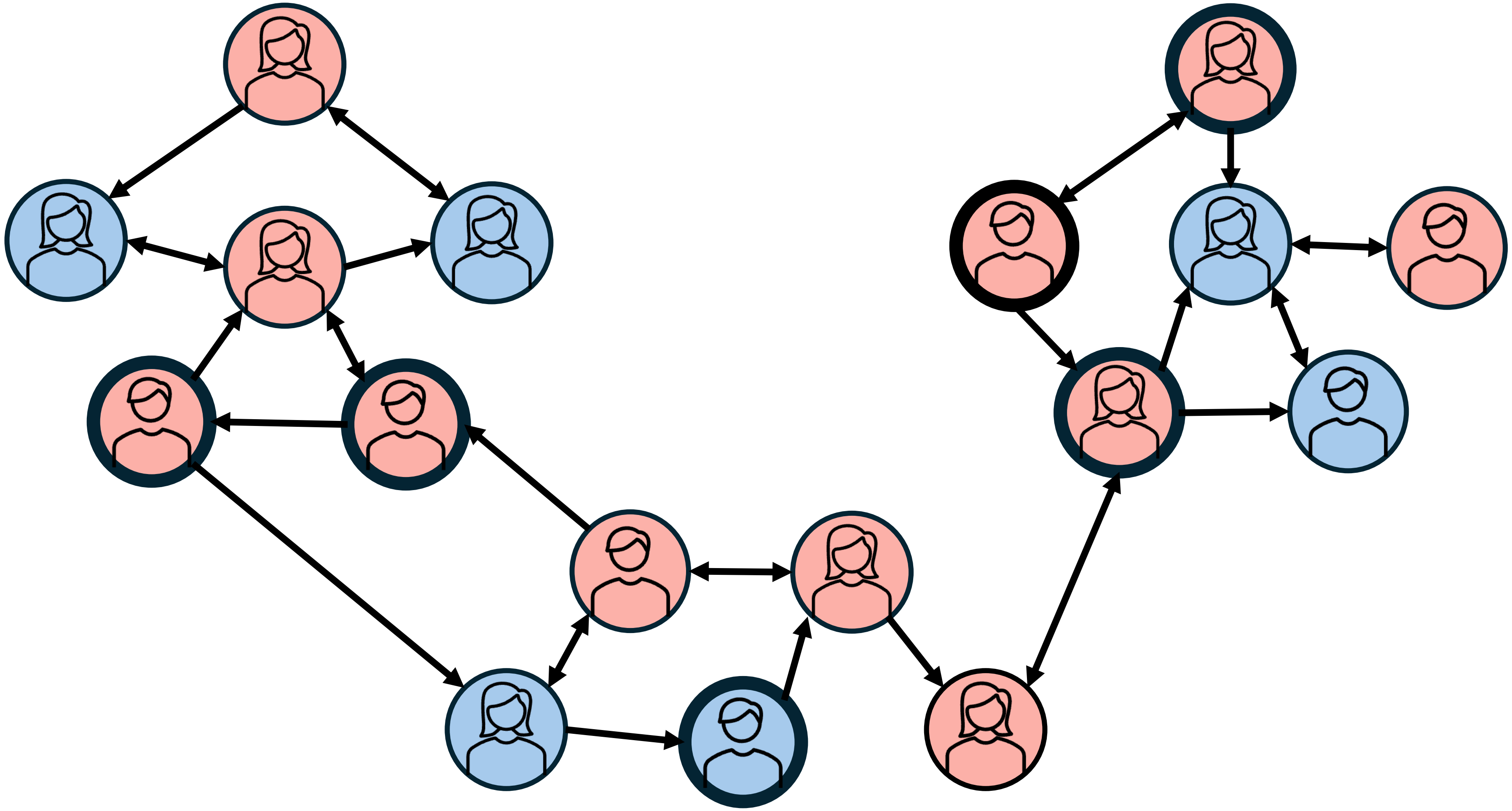
- Unbiased estimator
- Prohibitive $O(p^{-d})$ variance

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.

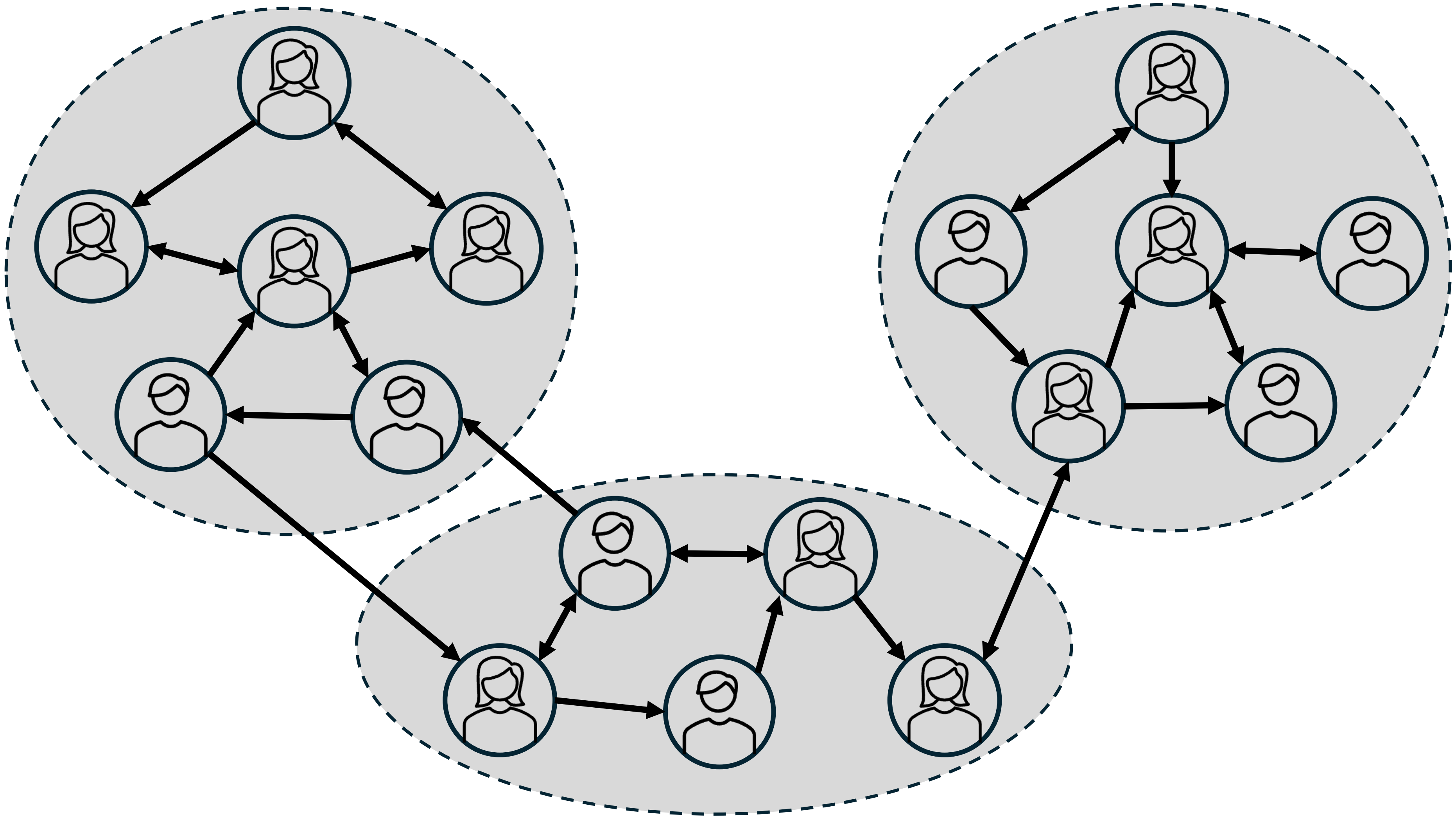
Ugander, Johan, et al. "Graph cluster randomization: Network exposure to multiple universes." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2013.

Variance Reduction 1:
Change the Experimental Design

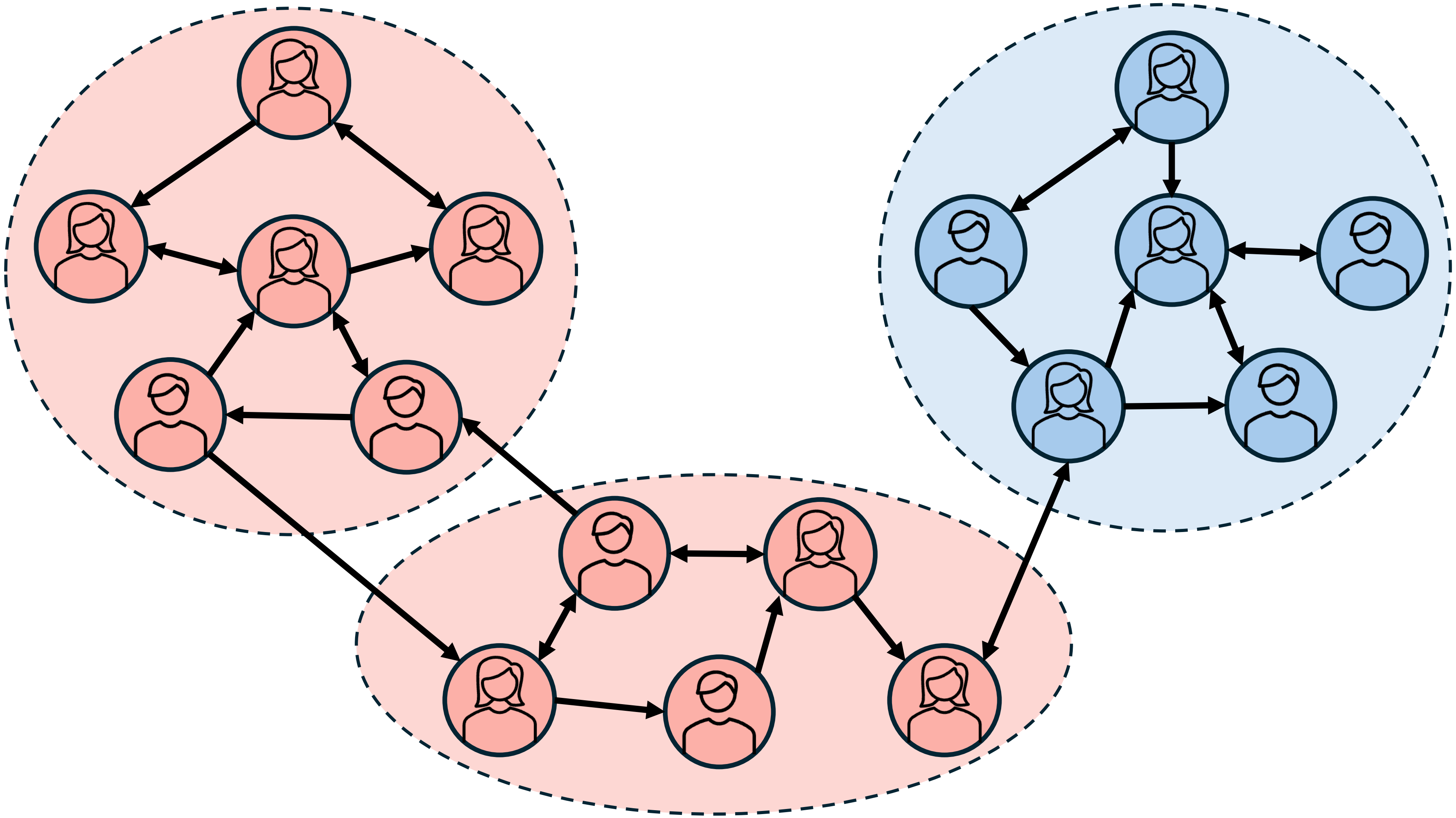
Unit Randomized Design: $p = \frac{1}{3}$



Graph Cluster Randomization (GCR)



Graph Cluster Randomization (GCR)



Variance Reduction 2: Change the Estimator

Structured Potential Outcomes

General Neighborhood Interference: *i's outcome* has $2^{|N_i|}$ parameters

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

In full generality, Horvitz-Thompson is the only unbiased estimator

Reduce parameter count by sparsifying basis in a nice way.

β-Order Interactions

Idea: Sparsify in the monomial basis

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

Intuition: Influence comes from *small* subsets of neighbors

$$Y_i(\boldsymbol{z}) = \langle \boldsymbol{c}_i, \tilde{\boldsymbol{z}}_i \rangle$$

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

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.)

$\boldsymbol{\beta = \max_i d_i}$

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

$$\mathrm{TTE} = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i(\mathbf{1}) - Y_i(\mathbf{0})\big) = \frac{1}{n}\sum_{i=1}^{n}\langle \mathbf{c}_i, \underbrace{\boldsymbol{\theta}_i}_{} \rangle$$

TTE coordinates: $(\theta_i)_\emptyset = 0$, $(\theta_i)_S = 1$

Pseudoinverse Estimator

$$\widehat{\mathrm{TTE}}_{\mathrm{PI}} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{Y_i(\boldsymbol{z})}_{\text{Outcome}} \left\langle \overbrace{\mathbb{E}\left[\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\top}\right]^{\dagger}}^{\text{Design}} \cdot \underbrace{\boldsymbol{\theta}_i}_{\text{Estimand}},\ \overbrace{\tilde{\boldsymbol{z}}_i}^{\text{Treatment}} \right\rangle$$

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

Entries indexed by subsets of N_i:

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

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

For GCR Design

$$\left(\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^\intercal]\right)_{S,T} = p^{\#\text{ Clusters containing } S \cup T}$$

Theoretical Results

Bias: $\widehat{\mathrm{TTE}}_{\mathrm{PI}}$ is unbiased when each θ_i lies in the column space of $\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}]$
Always true for GCR designs

Variance:

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

Continuous component:

$\gamma_i = \sqrt{\theta_i^{\intercal}\mathbb{E}[\,\tilde{\mathbf{z}}_i\,\tilde{\mathbf{z}}_i^{\intercal}]^{\dagger}\theta_i}$ measures "sensitivity" of unit *i*'s outcome to design

Discrete component:

$\mathbb{I}(\,\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j)$ models correlated treatment of neighborhoods

Specialized to Bernoulli GCR

- $\tilde{\mathbf{z}}_i \not\perp \tilde{\mathbf{z}}_j$ when i and j have neighbors in the same cluster
- $$\gamma_i = \begin{cases} O\left(p^{-|C(N_i)|}\right) & |C(N_i)| < \beta \\ O\left(|C(N_i)|^{\beta} \cdot p^{-\beta}\right) & |C(N_i)| \geq \beta \end{cases}$$

i internal to cluster, GCR gives good guarantee

i at cluster boundary, fall back on β-order

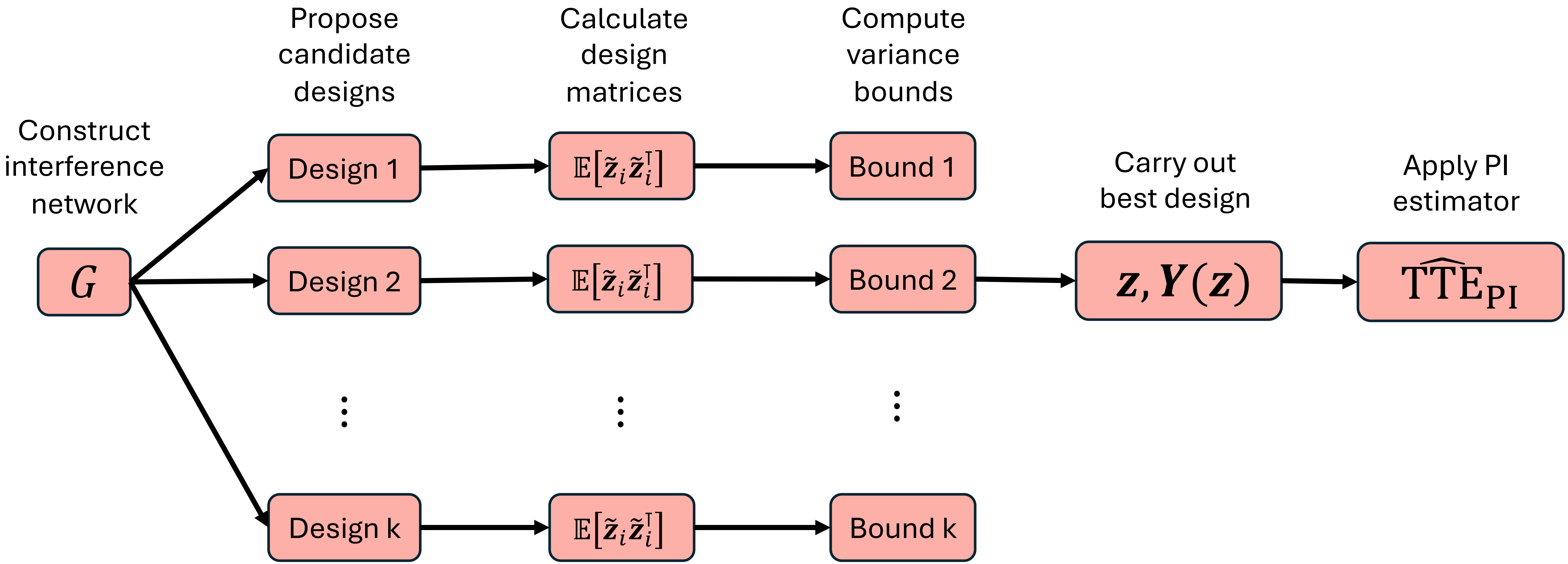
Variance	Unit Randomization	Cluster Randomization
General Interference	$\exp(d)$	$\exp(\|C(N_i)\|)$
β-Order Interactions	$\exp(\beta)$	$\exp(\min(\beta, \|C(N_i)\|))$

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).

Experimental Pipeline



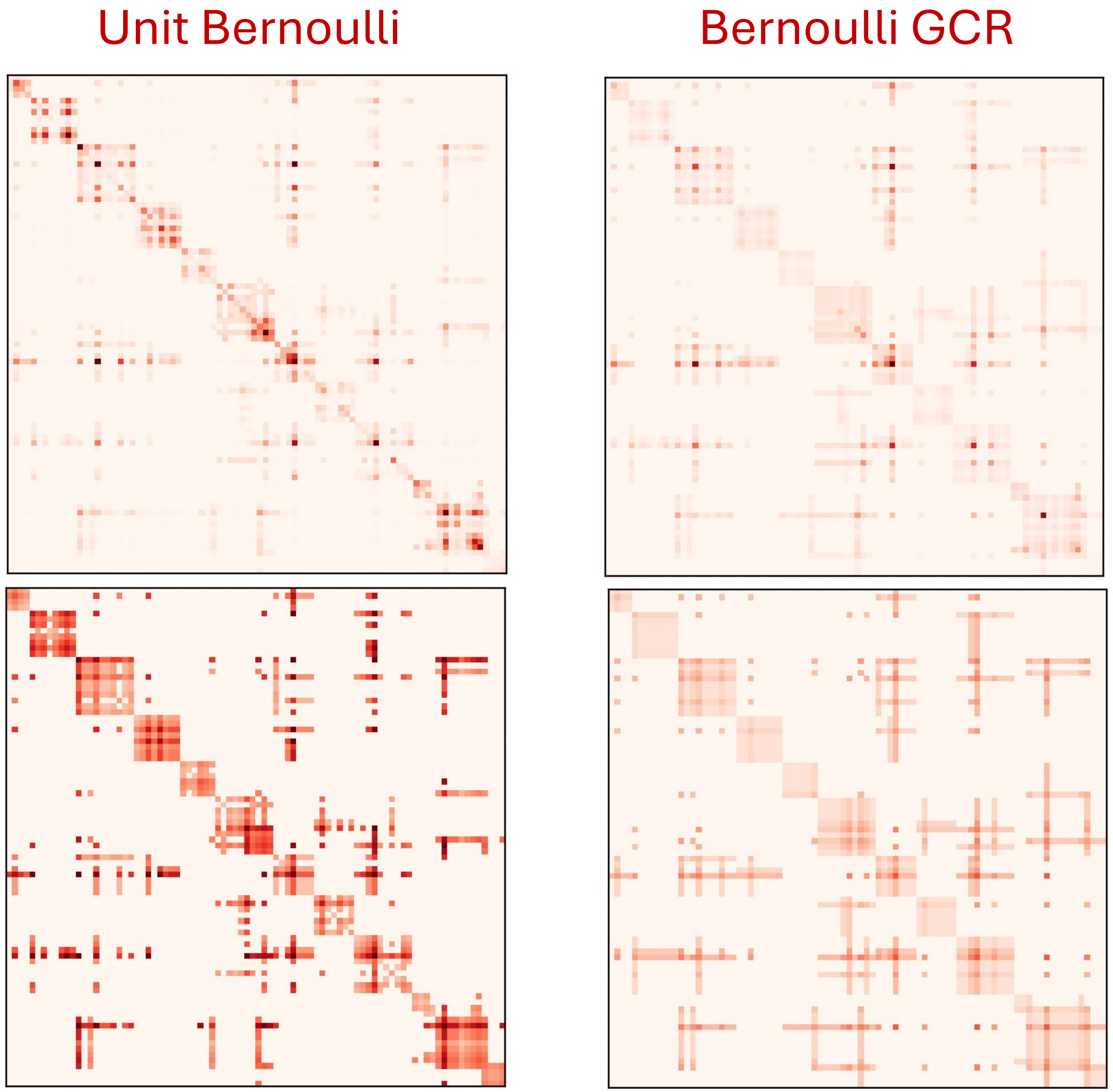
Visualizing the Variance Bound

Variance contribution of each vertex pair in a small collaboration network

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

Contributions to the Variance Bound

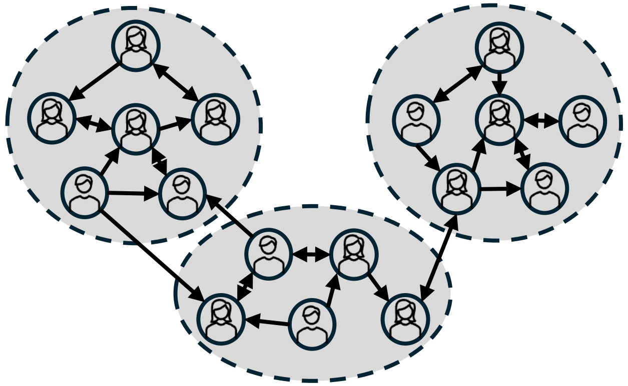
$$\gamma_i \cdot \gamma_j \cdot \mathbb{I}(\tilde{\mathbf{z}}_i \perp \tilde{\mathbf{z}}_j)$$

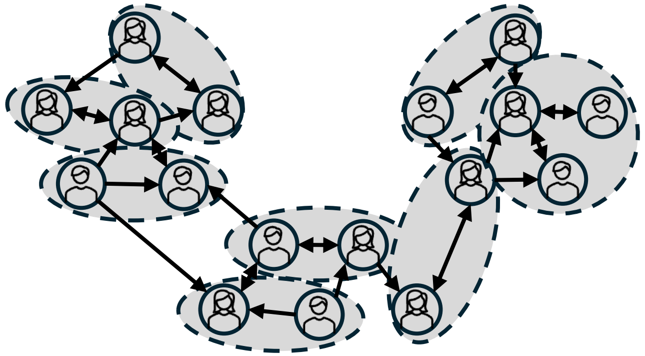


Ryan A. Rossi, & Nesreen K. Ahmed (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.

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

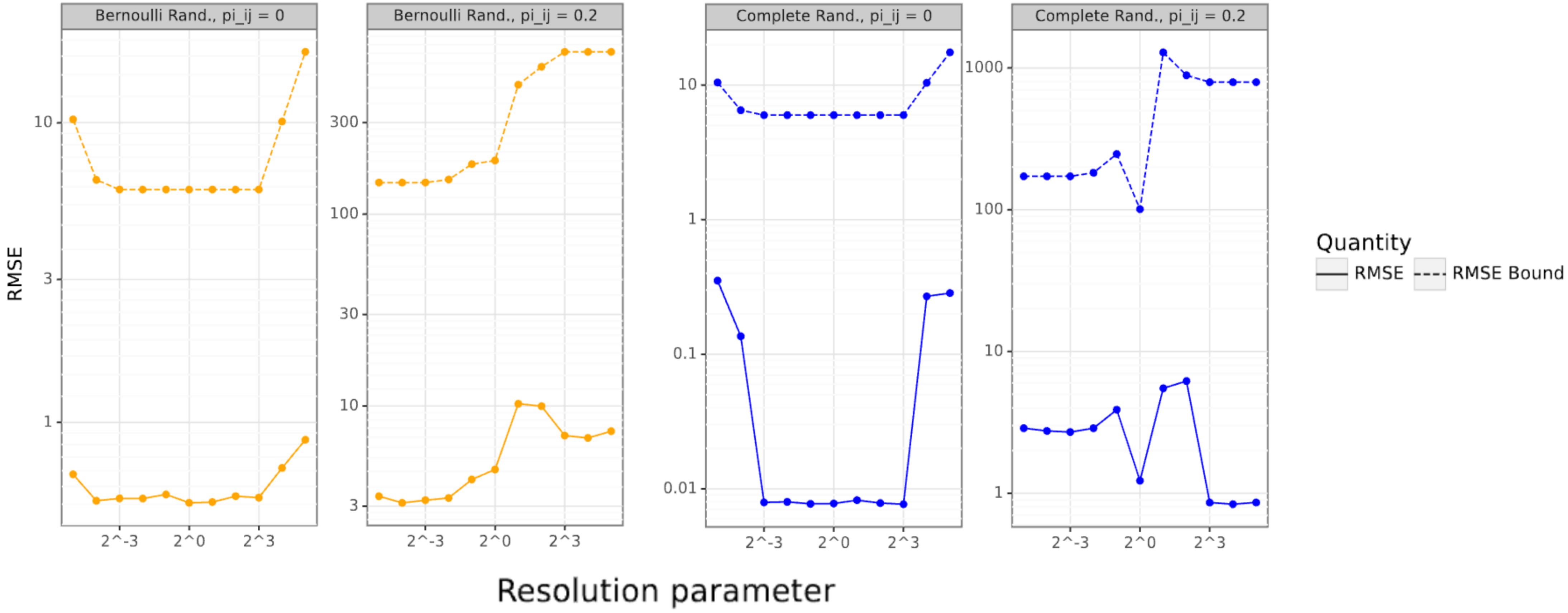
Low Resolution ⟷ **High Resolution**

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-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, 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).

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).

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.