Combining Rollout Designs and Clustering for Causal Inference under Low-order Interference

Matthew Eichhorn

Mayleen Cortez-Rodriguez

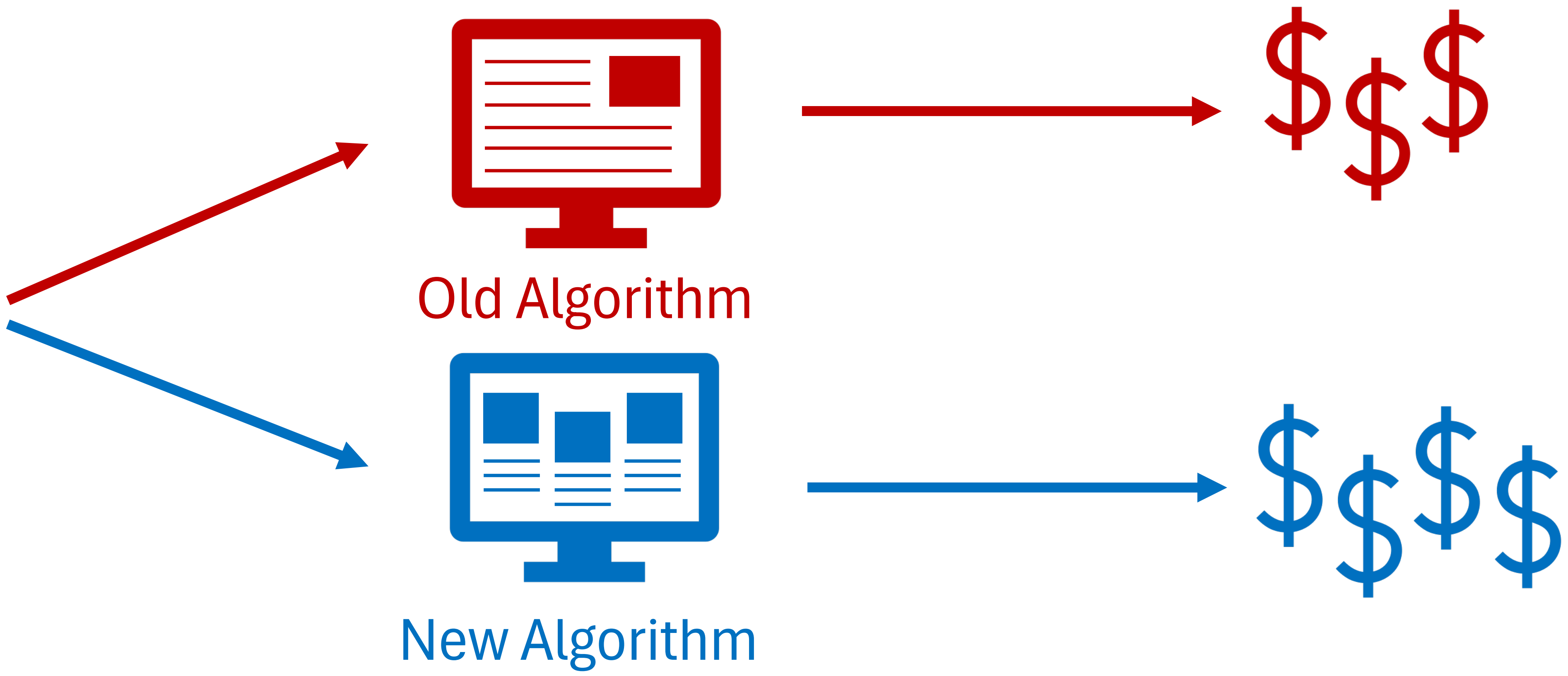
Christina Lee Yu

JSM

August 3, 2025

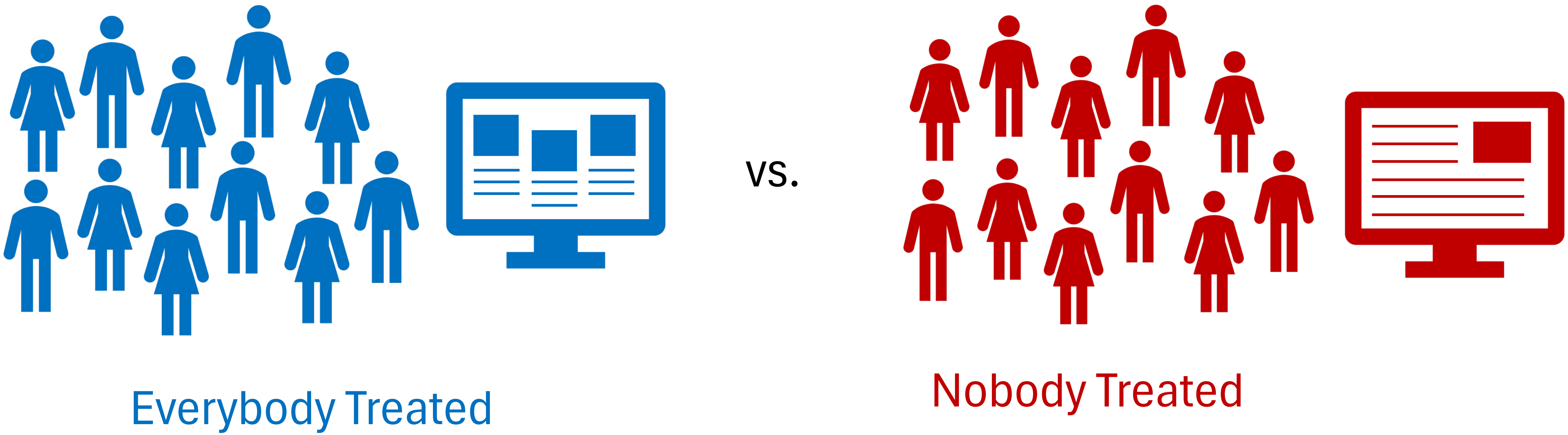
Motivating Example: Online Marketplace

An online marketplace wants to understand if a new product recommendation algorithm will increase sales

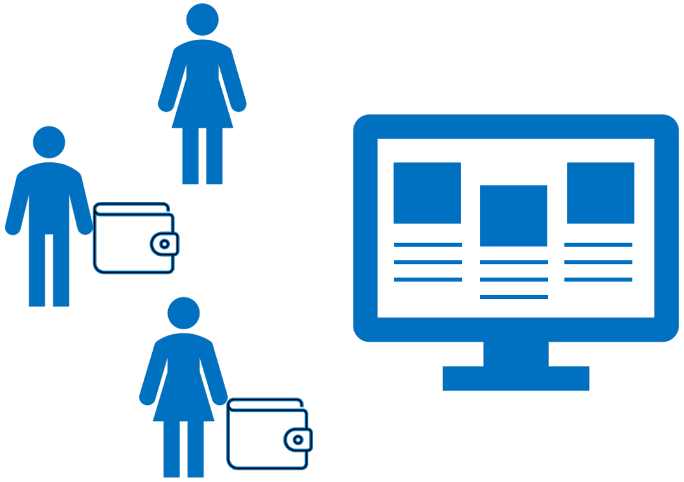


Total Treatment Effect

Difference in *average outcome* (e.g., monthly per-user spending) under two possible *global actions*:



Randomized Experiment



Treatment Group

Control Group

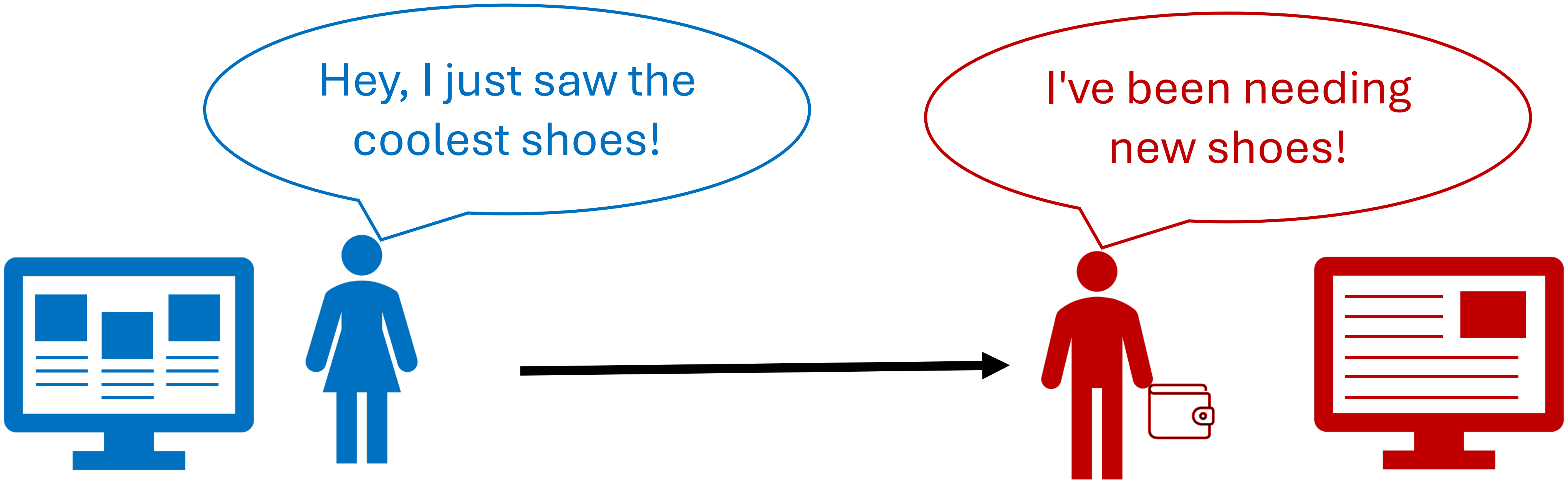
*** Marginal treatment probability p is small.

Difference in Means Estimator:

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

Interference

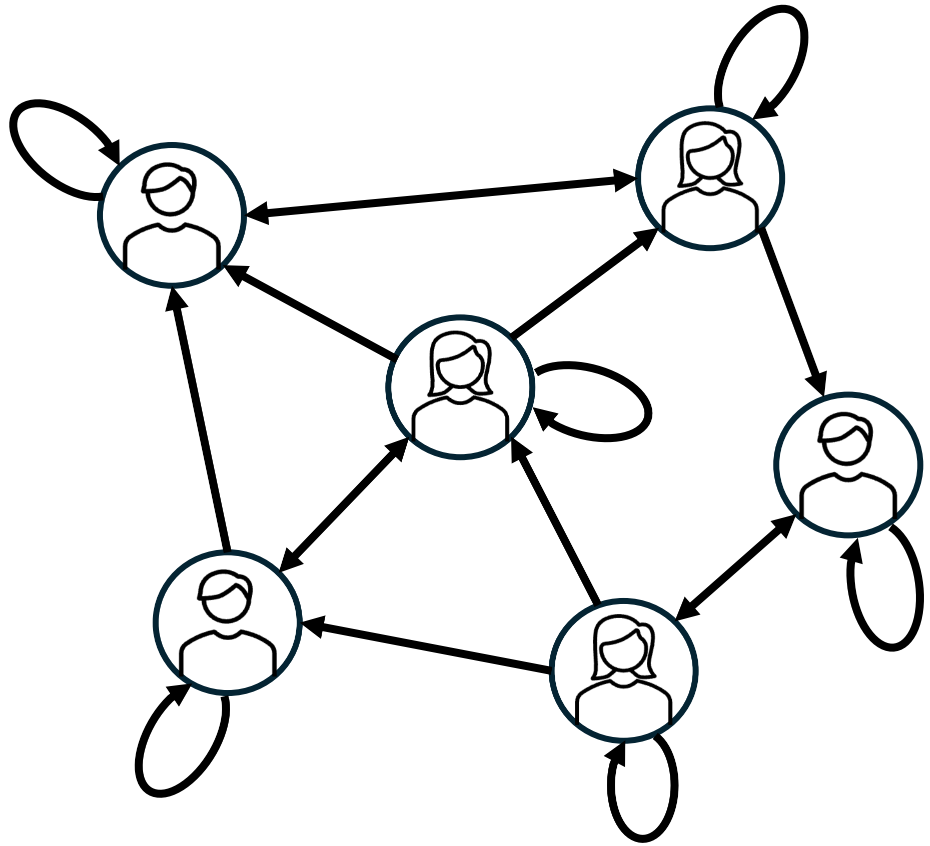
Individuals' outcomes may change even if they are not treated



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." *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).

Sussman, Daniel L., and Edoardo M. Airoldi. "Elements of estimation theory for causal effects in the presence of network interference."

Modeling Interference

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

$V = n$ individuals
$(j, i) \in A \Rightarrow j$'s treatment affects i's outcome

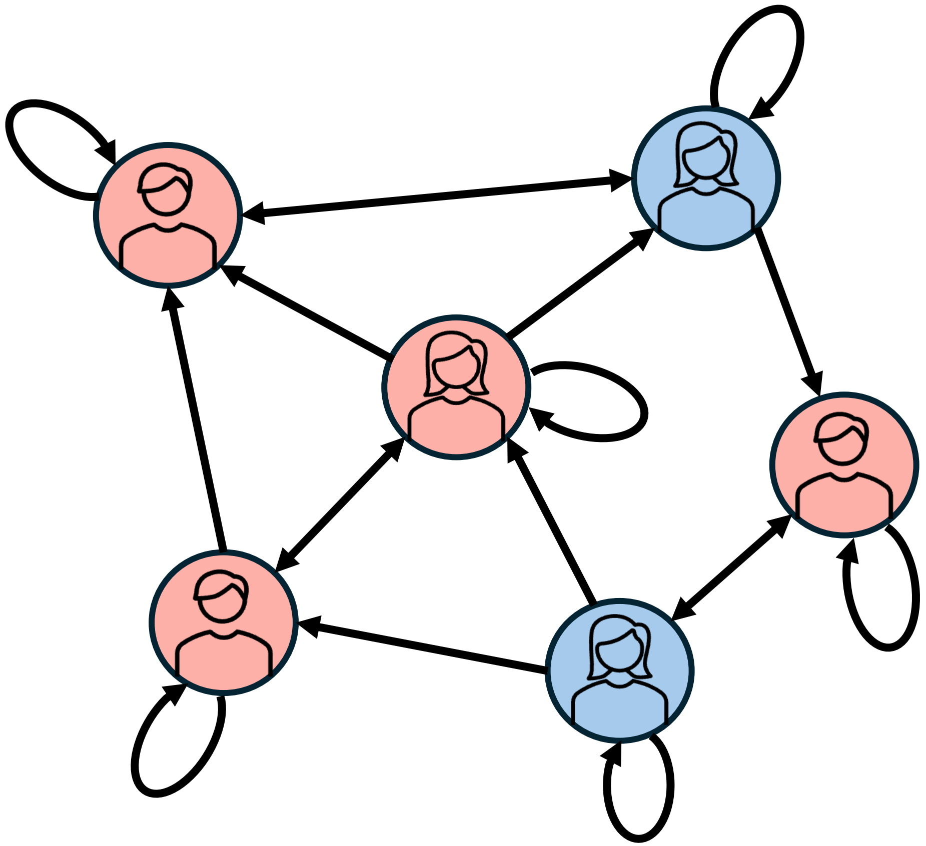
Treatment Assignments $\boldsymbol{z} \in \{0,1\}^n$

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

Neighborhood Interference:

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

Total Treatment Effect $\text{TTE} = \dfrac{1}{n} \sum_{i=1}^{n} (Y_i(\mathbf{1}) - Y_i(\mathbf{0}))$



Modeling Interference

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

$V = n$ individuals

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

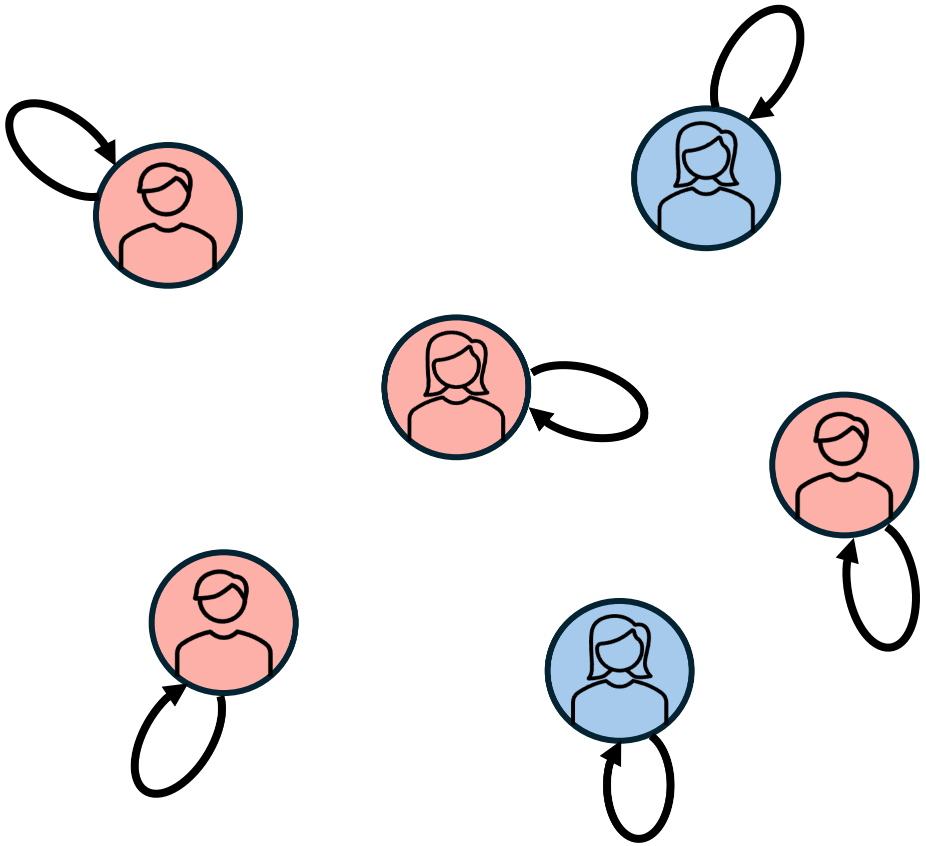
Treatment Assignments $\boldsymbol{z} \in \{0,1\}^n$

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

Neighborhood Interference:

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

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



Interference Graph may be Unknown

β-Order Interactions

An individual's outcome is a sum of **additive effects** that "turn on" when a **small subset** of their neighborhood is **fully treated**.

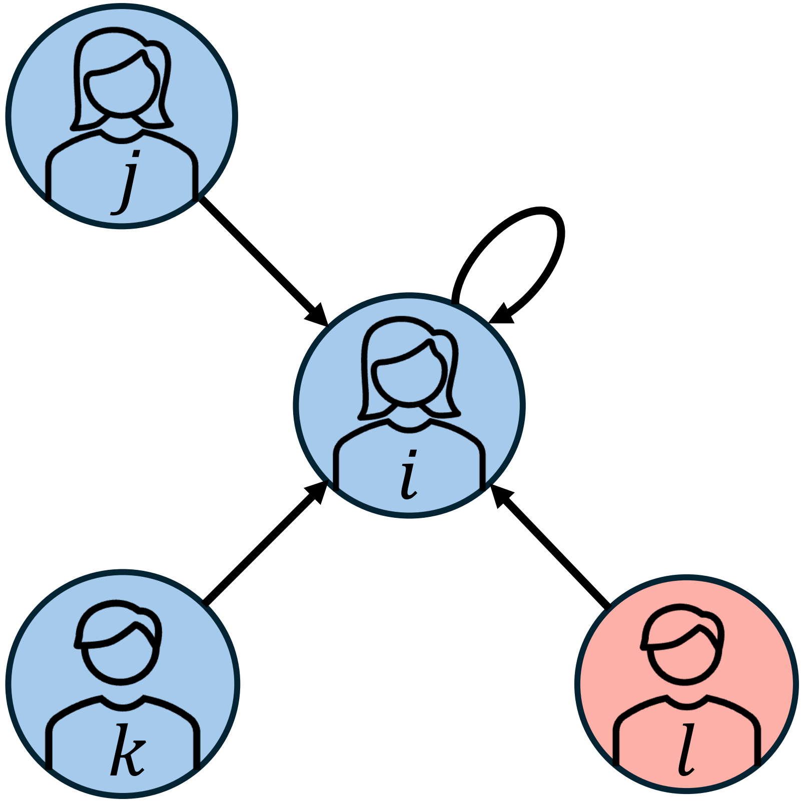
β = maximum set size that has an effect

$\beta = 1$:

$$Y_i(\mathbf{z}) = \underbrace{c_{i,\emptyset}}_{\text{Baseline}} + \underbrace{c_{i,\{i\}}}_{\text{Direct}} + \underbrace{c_{i,\{j\}} + c_{i,\{k\}}}_{\text{Individual Spillovers}}$$

$\beta = 2$:

$$Y_i(\mathbf{z}) = c_{i,\emptyset} + c_{i,\{i\}} + c_{i,\{j\}} + c_{i,\{k\}} + \underbrace{c_{i,\{i,j\}} + c_{i,\{i,k\}} + c_{i,\{j,k\}}}_{\text{Joint Effects}}$$



β-Order Interactions

An individual's outcome is a sum of **additive effects** that "turn on" when a **small subset** of their neighborhood is **fully treated**.

β = maximum set size that has an effect

$$Y_i(\mathbf{z}) = \sum_{\substack{S \subseteq N_i \\ |S| \leq \beta}} c_{i,S} \underbrace{\prod_{j \in S} z_j}_{S \text{ fully treated}}$$

$$\text{TTE} = \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{S \subseteq N_i \\ 1 \leq |S| \leq \beta}} c_{i,S}$$

Without knowledge of the interference graph, we can't measure these effects directly

Mayleen Cortez-Rodriguez, **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.

Aggregate Measurements

If we treat each individual with marginal probability x (e.g. under a Bernoulli or completely randomized design), the quantity

$$F(x) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \sum_{\substack{S \subseteq N_i \\ |S| \leq \beta}} c_{i,S} \prod_{j \in S} z_j\right]$$

is a polynomial in x with degree at most β.

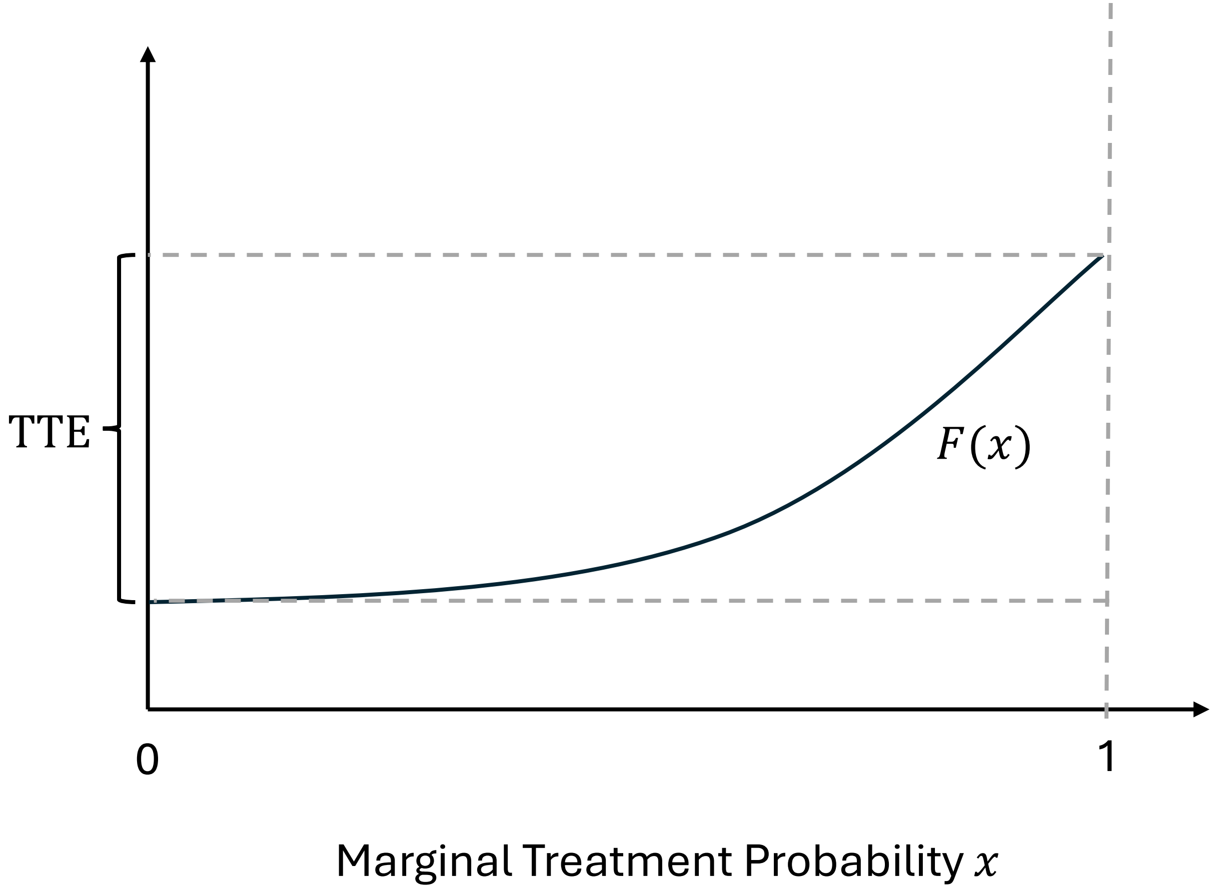
$$\text{TTE} = F(1) - F(0)$$

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

Polynomial Interpolation Estimator

<u>Staggered Rollout Design:</u>

- Select $\beta + 1$ treatment levels.
- Rollout treatment at these levels, measuring average outcome between each "ramp up".
- Interpolate a polynomial $\hat{F}(x)$ through these points.
- Estimate $\widehat{\mathrm{TTE}}_{\mathrm{PI}} = \hat{F}(1) - \hat{F}(0)$

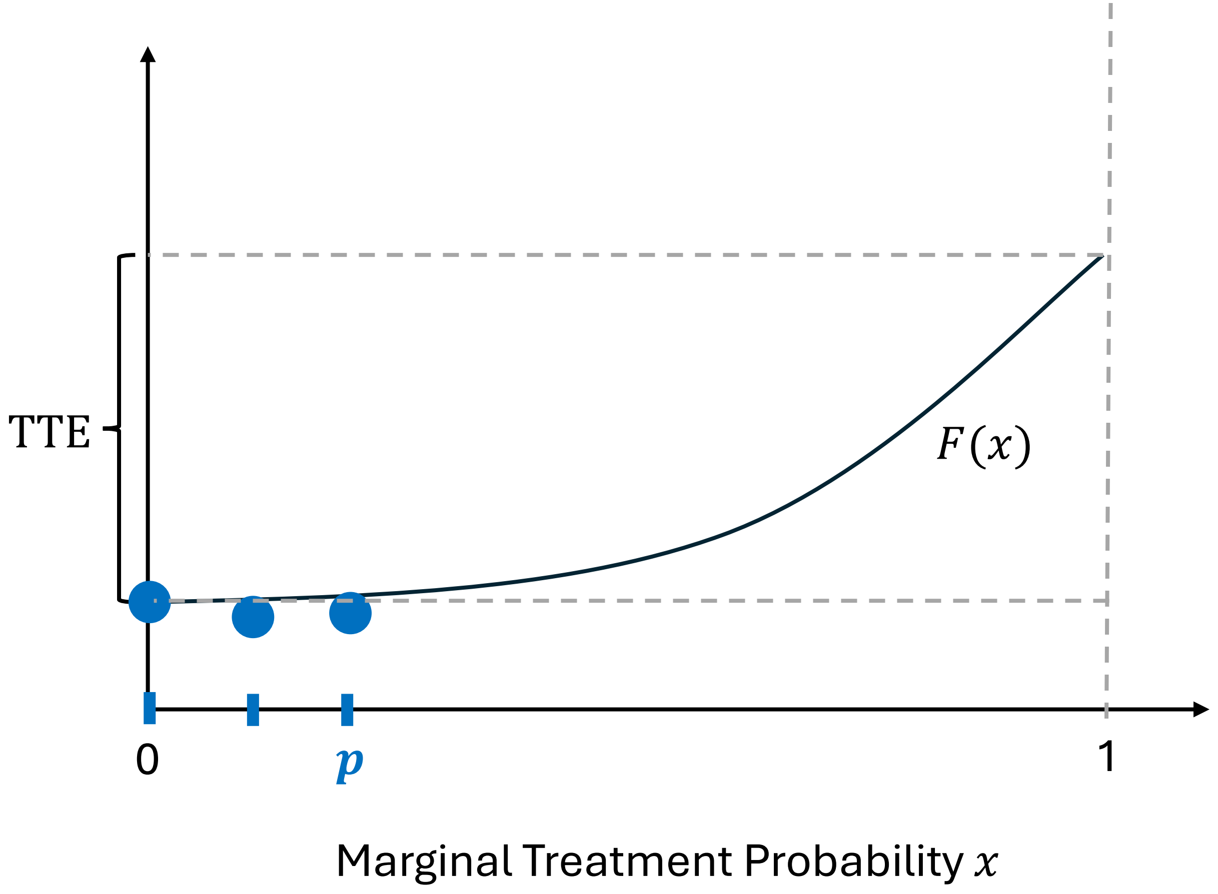


Mayleen Cortez-Rodriguez, **Matthew Eichhorn**, and Christina Lee Yu. "Staggered rollout designs enable causal inference under interference without network knowledge", Advances in Neural Information Processing Systems (NeurIPS), 2022.

Polynomial Interpolation Estimator

Staggered Rollout Design:

- Select $\beta + 1$ treatment levels.
- Rollout treatment at these levels, measuring average outcome between each "ramp up".
- Interpolate a polynomial $\hat{F}(x)$ through these points.
- Estimate $\widehat{\text{TTE}}_{\text{PI}} = \hat{F}(1) - \hat{F}(0)$

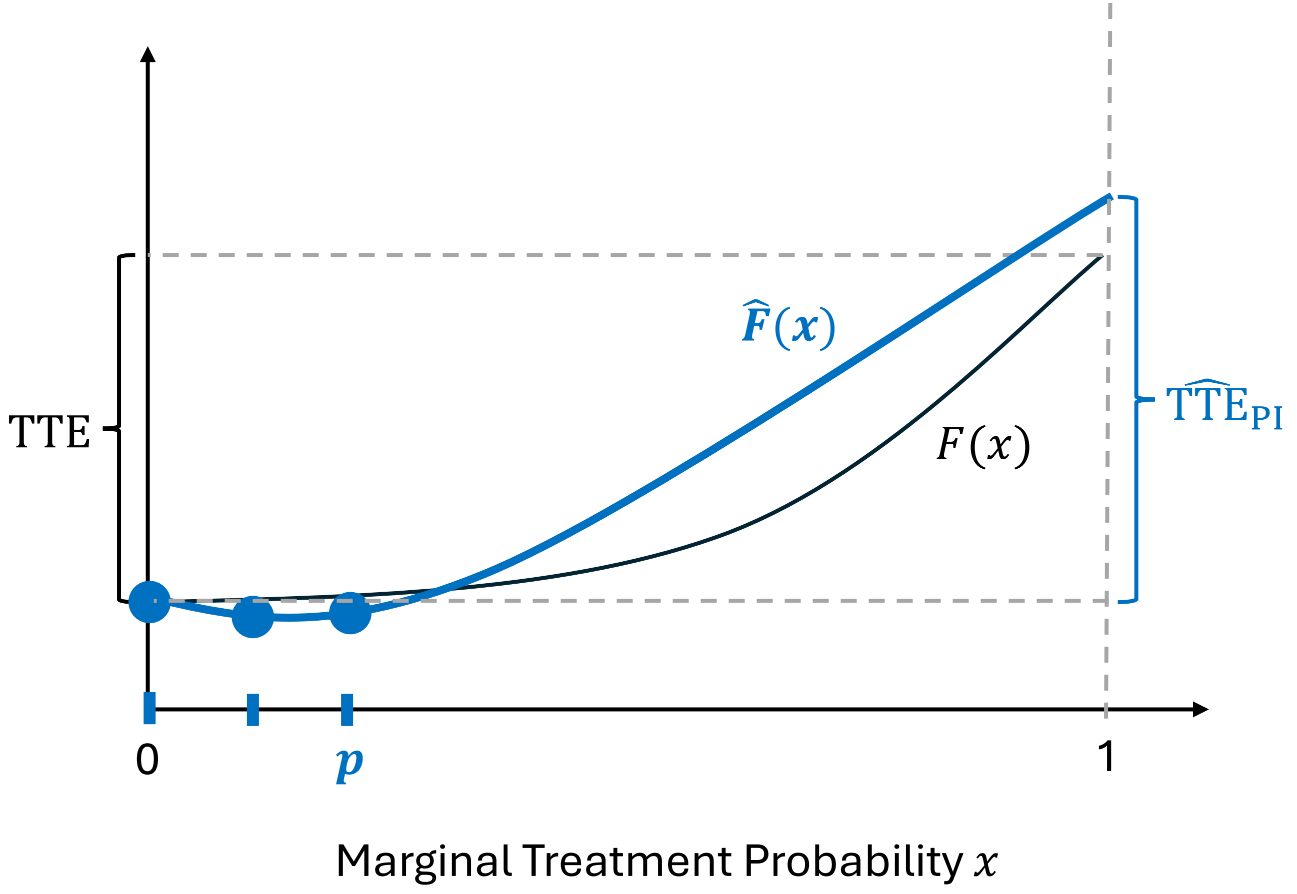


Mayleen Cortez-Rodriguez, **Matthew Eichhorn**, and Christina Lee Yu. "Staggered rollout designs enable causal inference under interference without network knowledge", Advances in Neural Information Processing Systems (NeurIPS), 2022.

Polynomial Interpolation Estimator

<u>Staggered Rollout Design:</u>

- Select $\beta + 1$ treatment levels.
- Rollout treatment at these levels, measuring average outcome between each "ramp up".
- Interpolate a polynomial $\hat{F}(x)$ through these points.
- Estimate $\widehat{\mathrm{TTE}}_{\mathrm{PI}} = \hat{F}(1) - \hat{F}(0)$



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

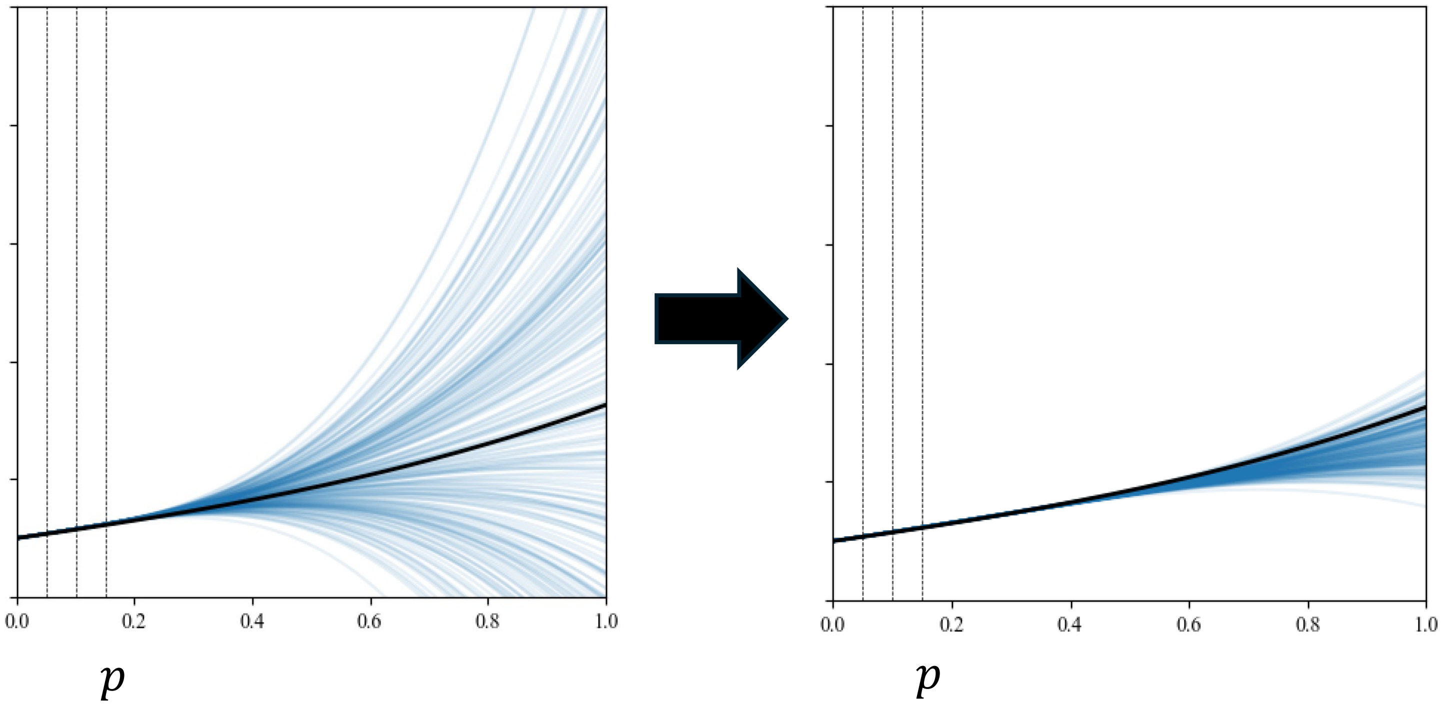
Polynomial Interpolation Estimator

Unbiased estimator

Extrapolation **far from the sampled points** $(p \ll 1)$ causes **prohibitive** $\mathbf{O}(\boldsymbol{p}^{-\boldsymbol{\beta}})$ **variance**.

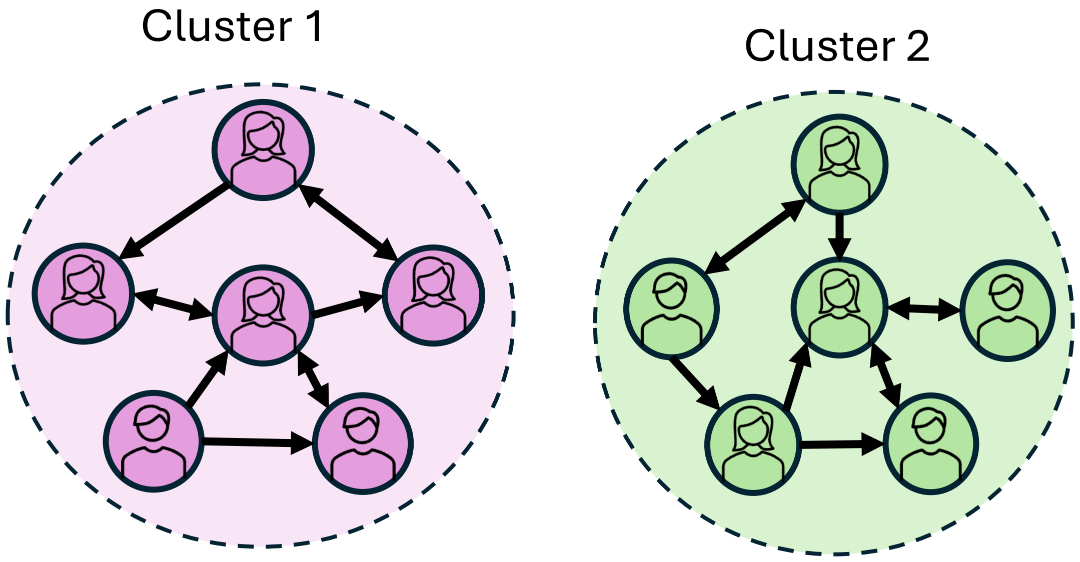
To respect **treatment budget** p, we must "artificially" push measurements to the right.

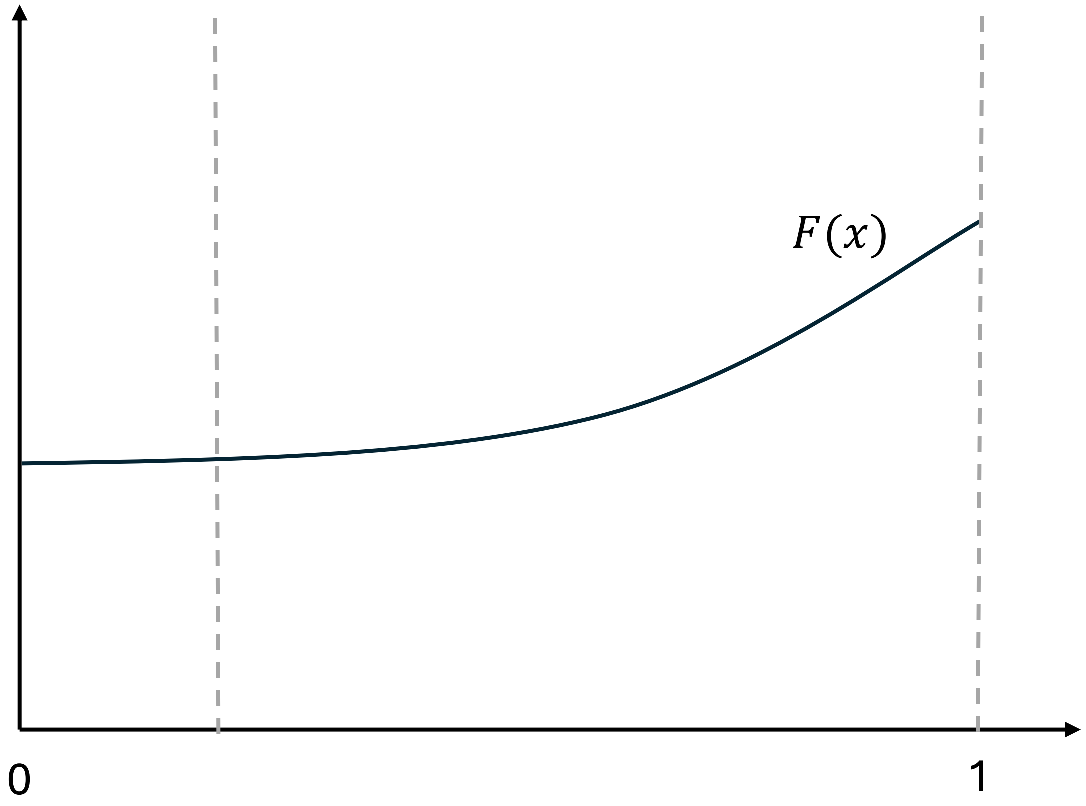
Subsampling lets us trade variance for some bias.



Subsampling (Motivation)

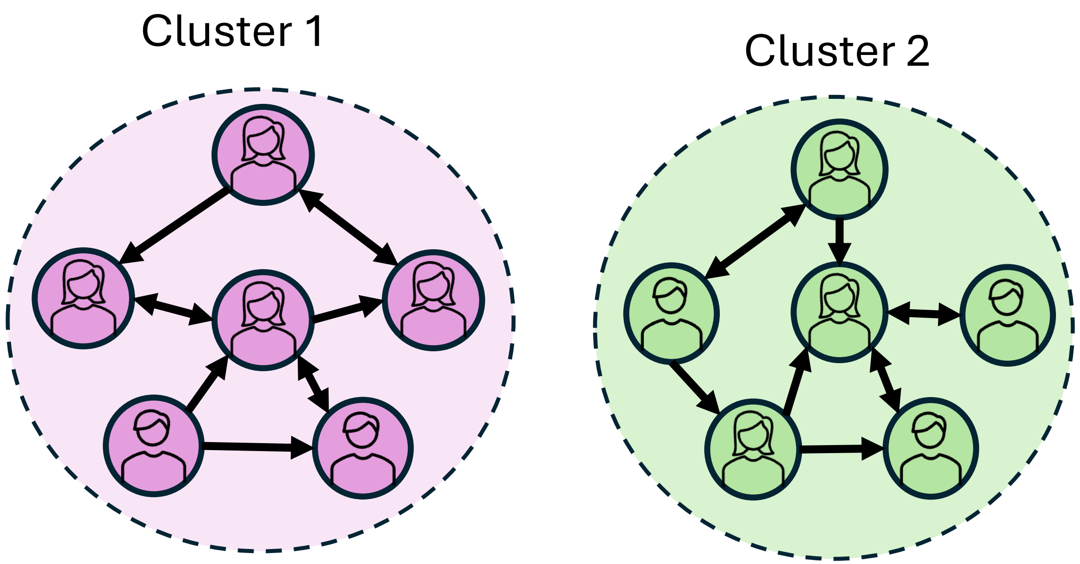
Imagine our interference graph splits into two, equal-sized disjoint clusters:



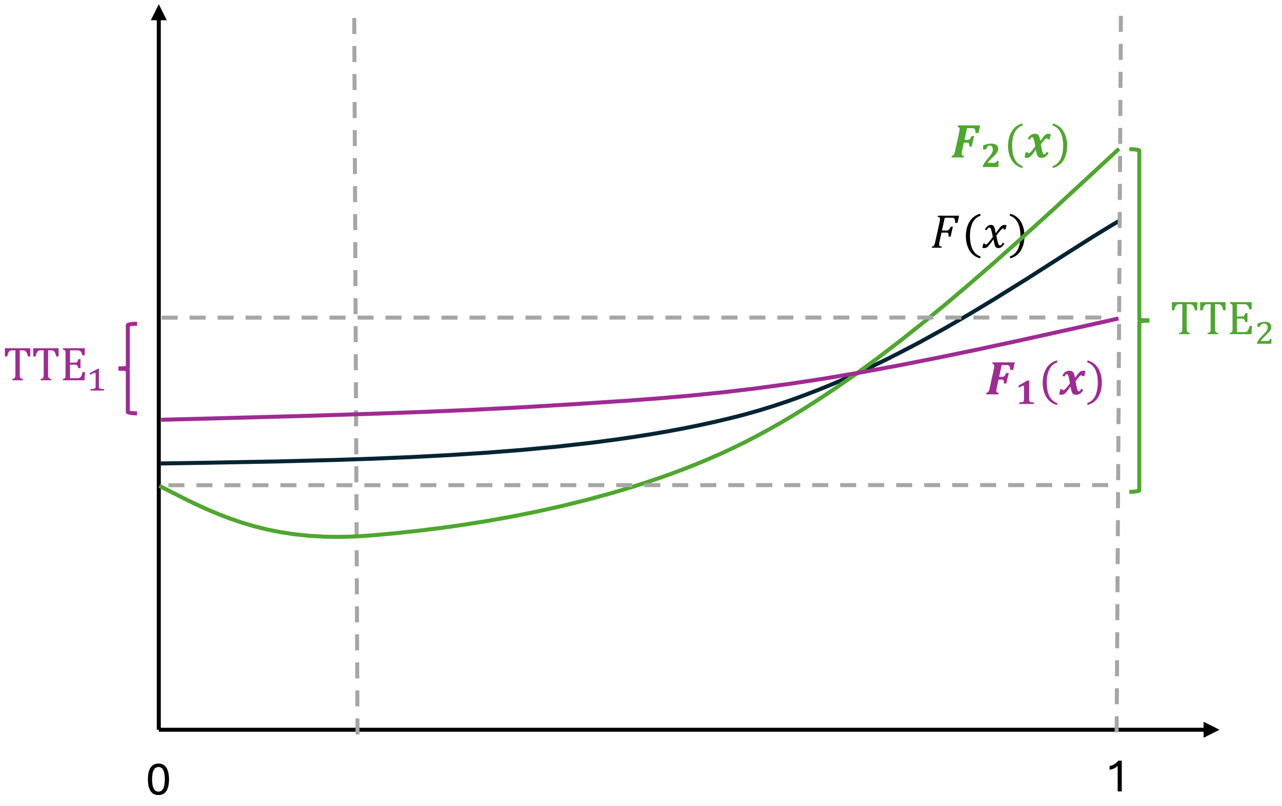


Subsampling (Motivation)

Imagine our interference graph splits into two, equal-sized disjoint clusters:

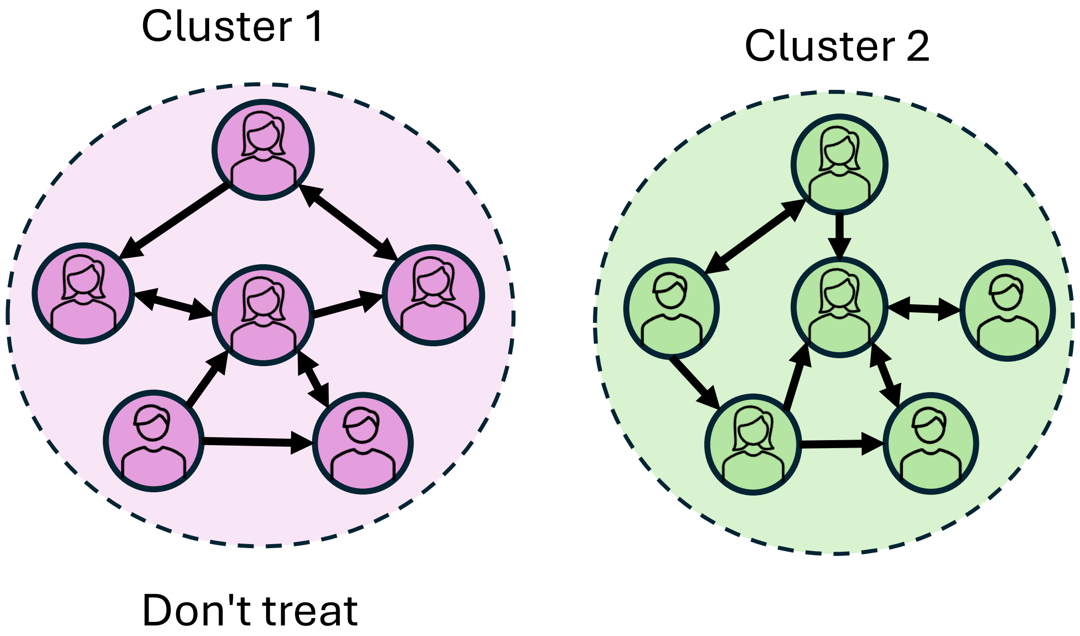


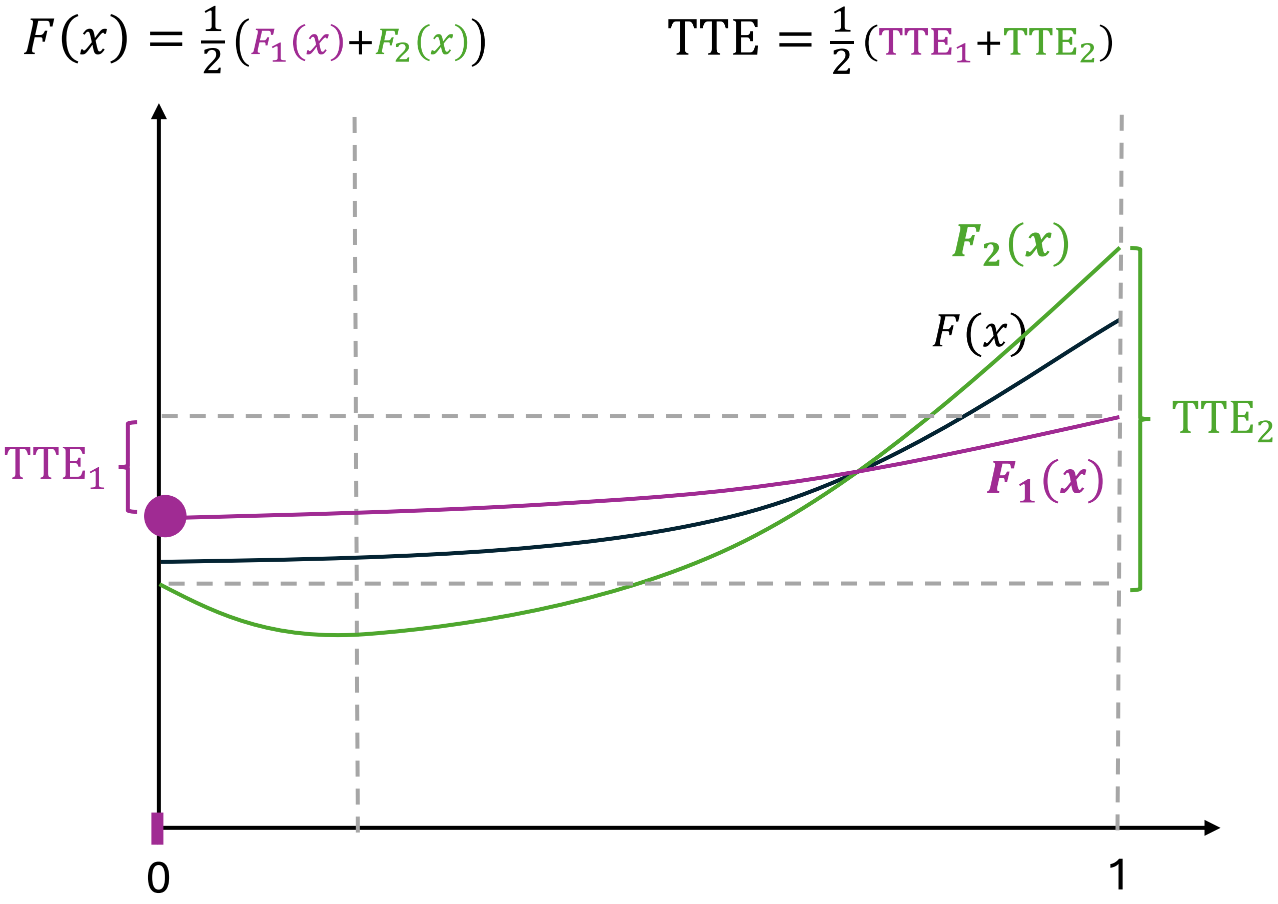
$F(x) = \frac{1}{2}(F_1(x) + F_2(x))$ $\quad$ $\text{TTE} = \frac{1}{2}(\text{TTE}_1 + \text{TTE}_2)$



Subsampling (Motivation)

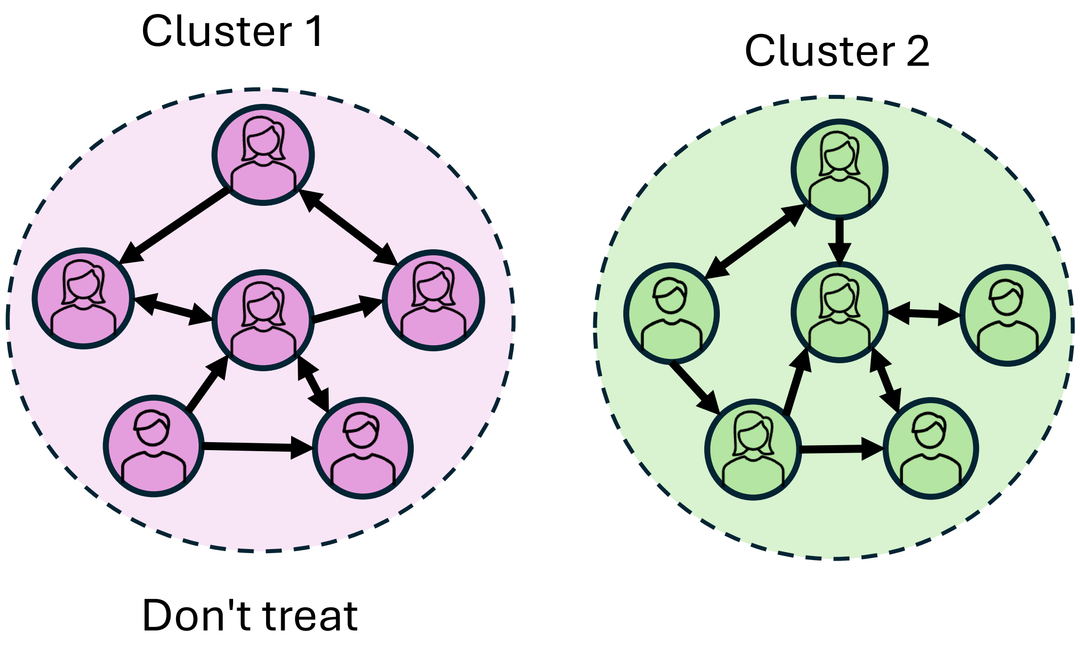
Imagine our interference graph splits into two, equal-sized disjoint clusters:



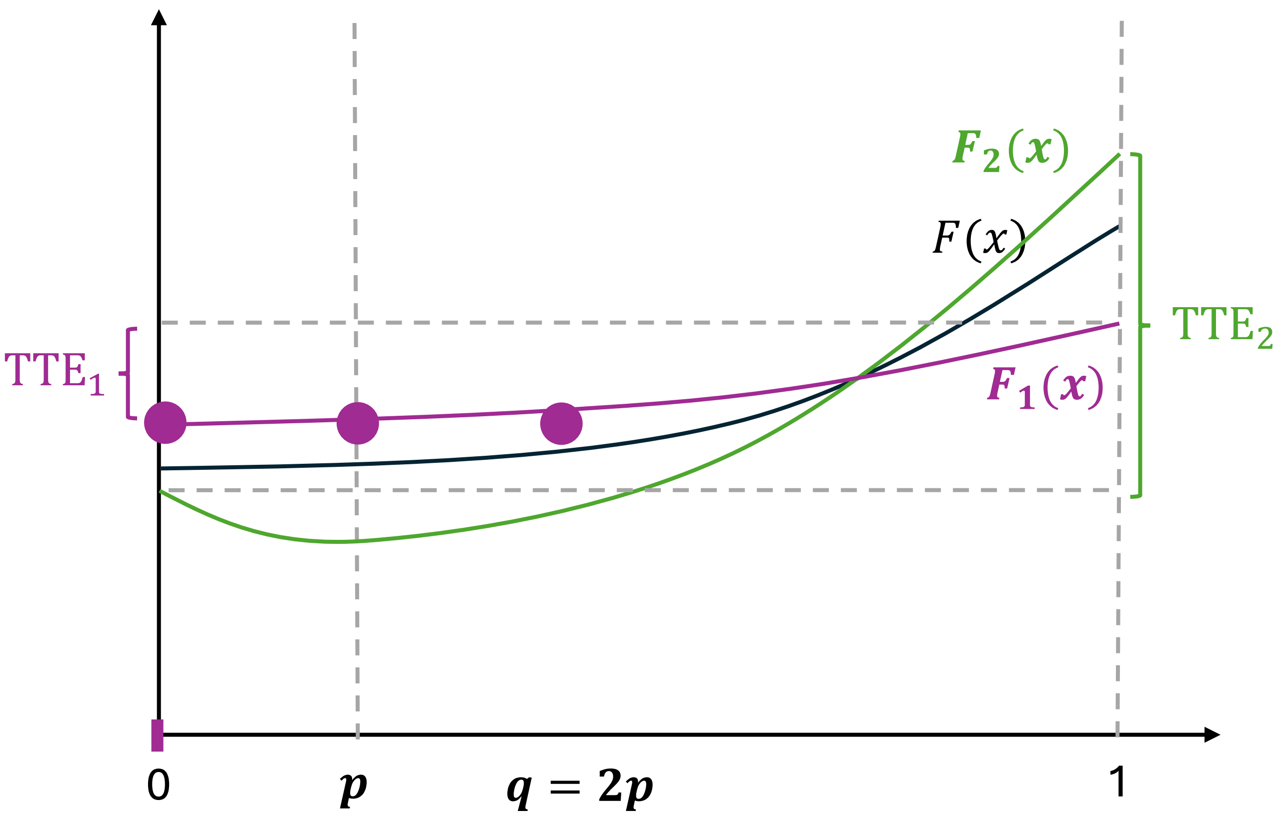


Subsampling (Motivation)

Imagine our interference graph splits into two, equal-sized disjoint clusters:

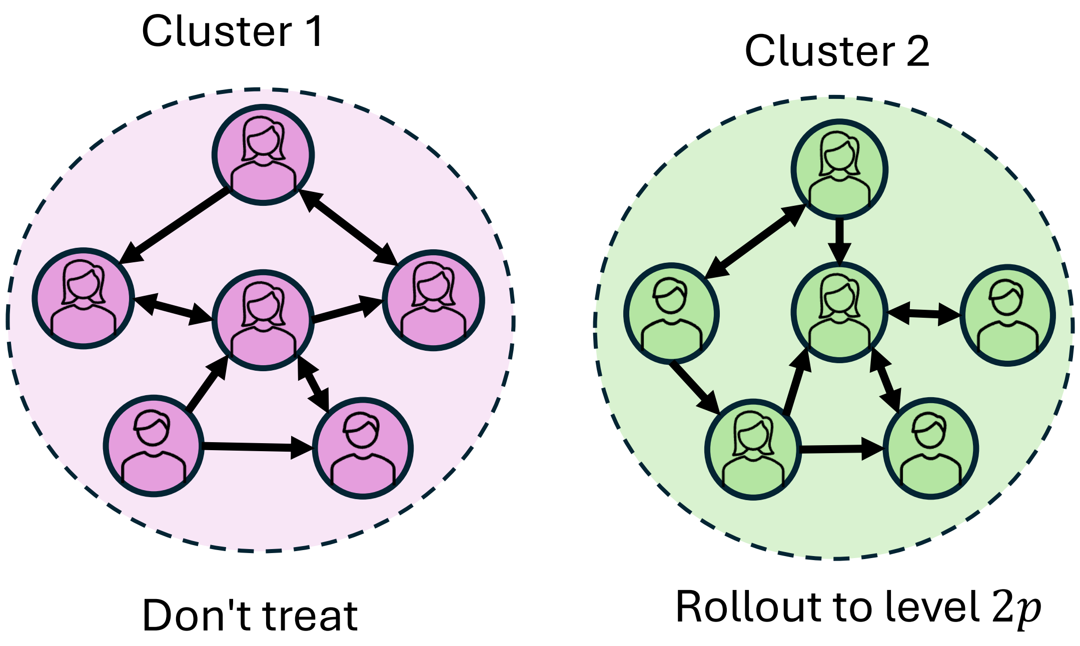


$F(x) = \frac{1}{2}(F_1(x) + F_2(x))$ $\quad$ $\mathrm{TTE} = \frac{1}{2}(\mathrm{TTE}_1 + \mathrm{TTE}_2)$

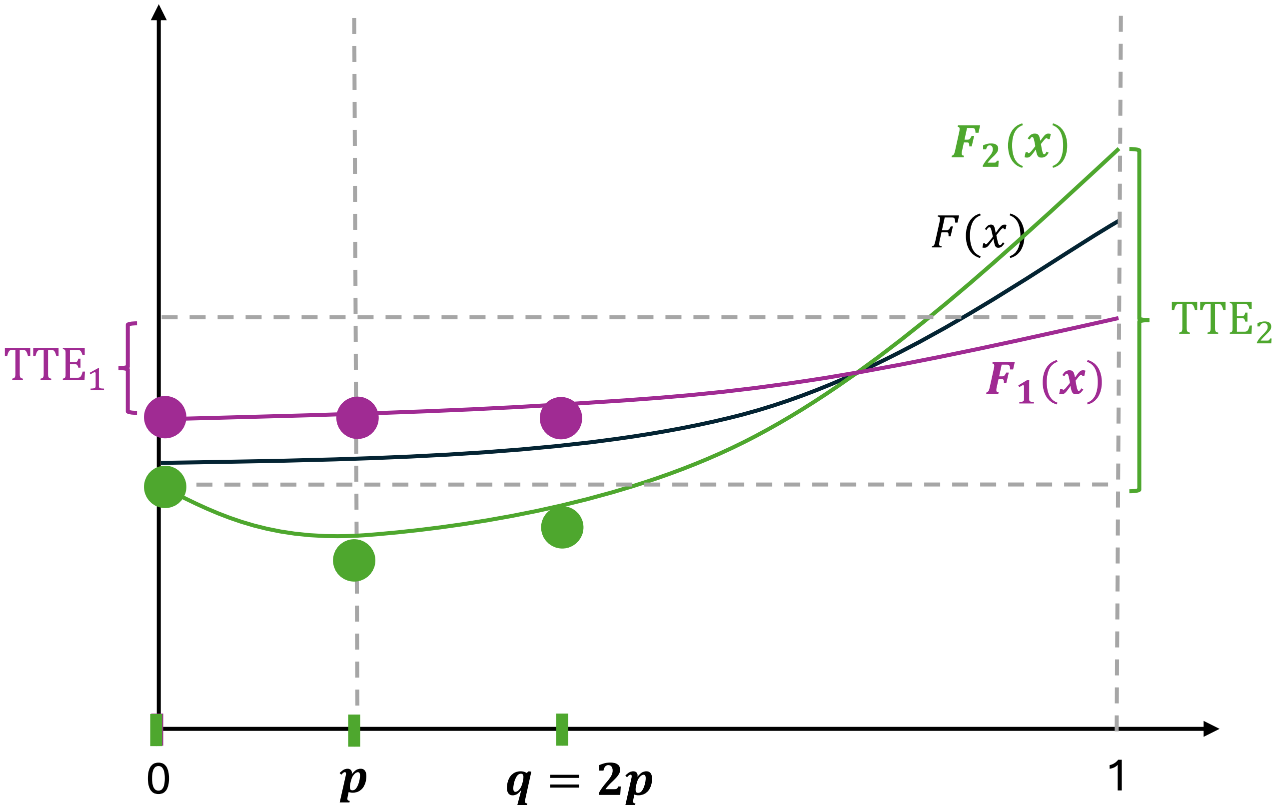


Subsampling (Motivation)

Imagine our interference graph splits into two, equal-sized disjoint clusters:

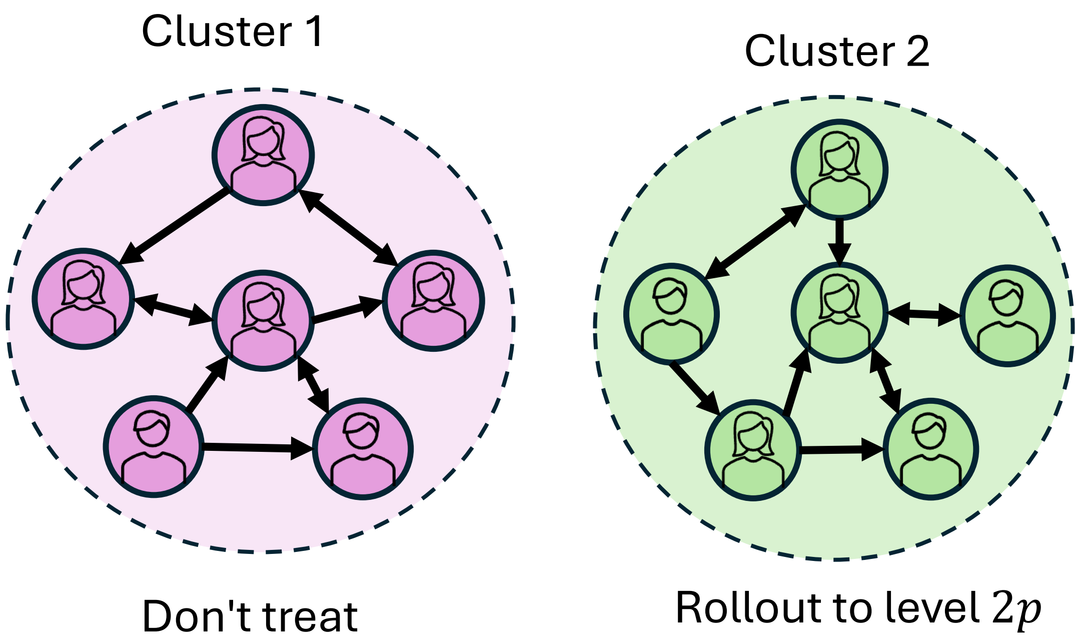


$F(x) = \frac{1}{2}(F_1(x) + F_2(x))$ $\qquad$ $\text{TTE} = \frac{1}{2}(\text{TTE}_1 + \text{TTE}_2)$



Subsampling (Motivation)

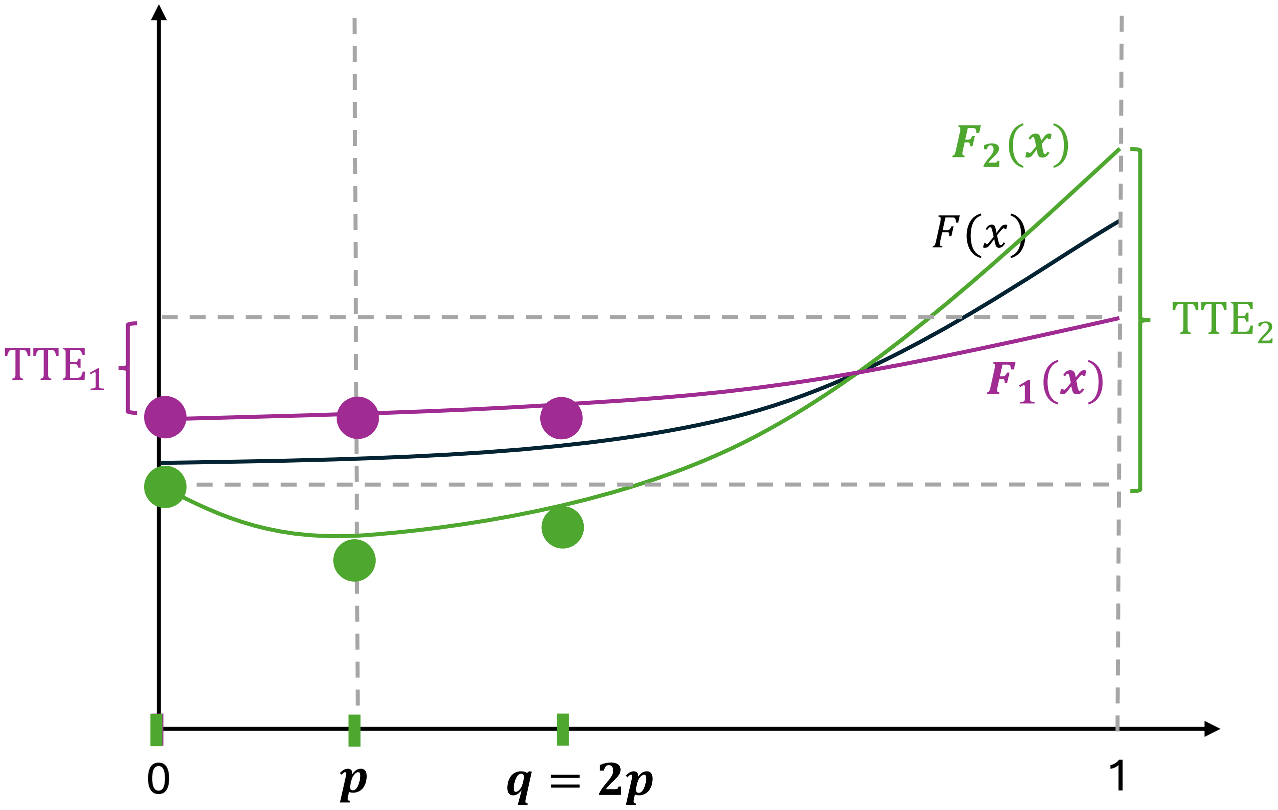
Imagine our interference graph splits into two, equal-sized disjoint clusters:



$\mathbb{E}[\widehat{\text{TTE}}_{\text{PI}} \mid \text{cluster assignments}] = 0 + \frac{1}{2}\ \text{TTE}_2$

$$\widehat{\text{TTE}} = 2\,\widehat{\text{TTE}}_{\text{PI}}$$

$F(x) = \frac{1}{2}(F_1(x) + F_2(x))$ $\text{TTE} = \frac{1}{2}(\text{TTE}_1 + \text{TTE}_2)$



Two-Stage Rollout Design

For some chosen parameter q with $p \leq q \leq 1$:

Stage 1: Choose a set of experimental units $\mathcal{U}$, including each individual with in $\mathcal{U}$ with marginal probability $\frac{p}{q}$

Stage 2: Run a staggered rollout experiment:

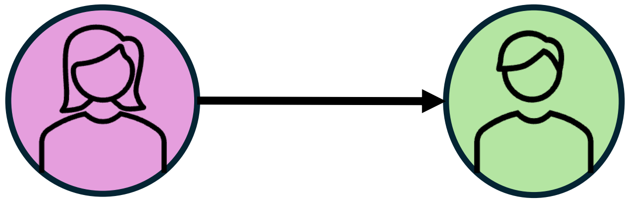
- Treat a q-fraction of individuals in $\mathcal{U}$
- Deterministically don't treat anyone outside of $\mathcal{U}$

Estimate: $\widehat{\text{TTE}}_{2-\text{Stage}} = \frac{q}{p} * (\widehat{\text{TTE}}_{\text{PI}}$ with budget $q)$

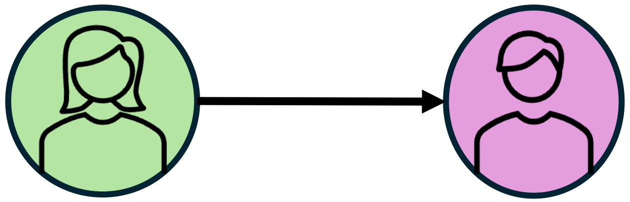
$\mathcal{U}$-Crossing edges

Edges between $\mathcal{U}$ and $[n]\backslash\mathcal{U}$ make the picture "fuzzy"

- Some neighbors of individuals in $\mathcal{U}$ will be untreatable, q overestimates the treatment fraction



- Some neighbors of individuals in $[n]\backslash\mathcal{U}$ will be treated, their estimated treatment effect is not 0

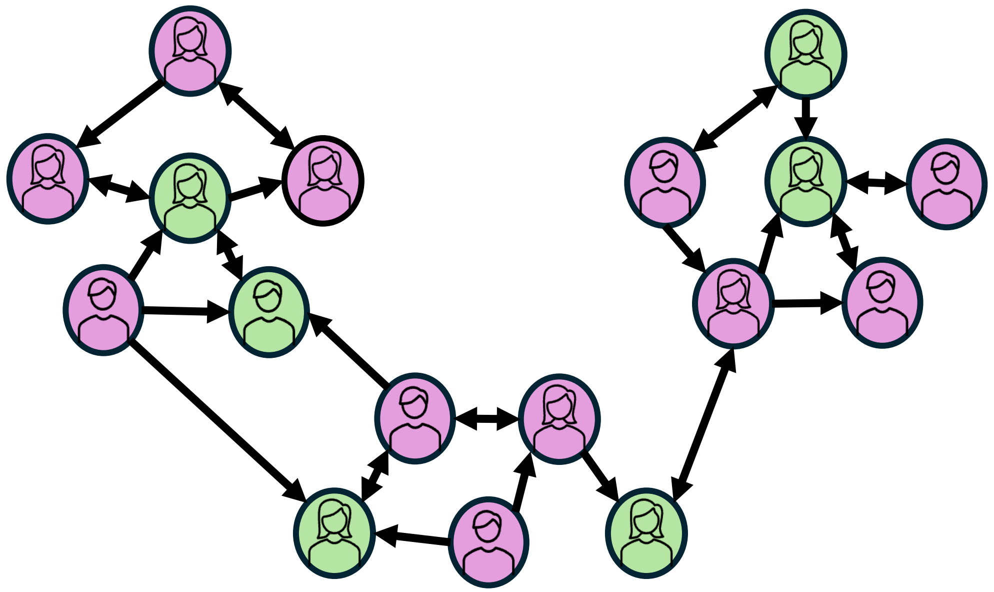


Crossing Edges Contribute Bias $\approx \; (\sum \text{crossing effects}) * \left(1 - \left(\frac{p}{q}\right)^{\beta-1}\right)$

Selecting $\mathcal{U}$

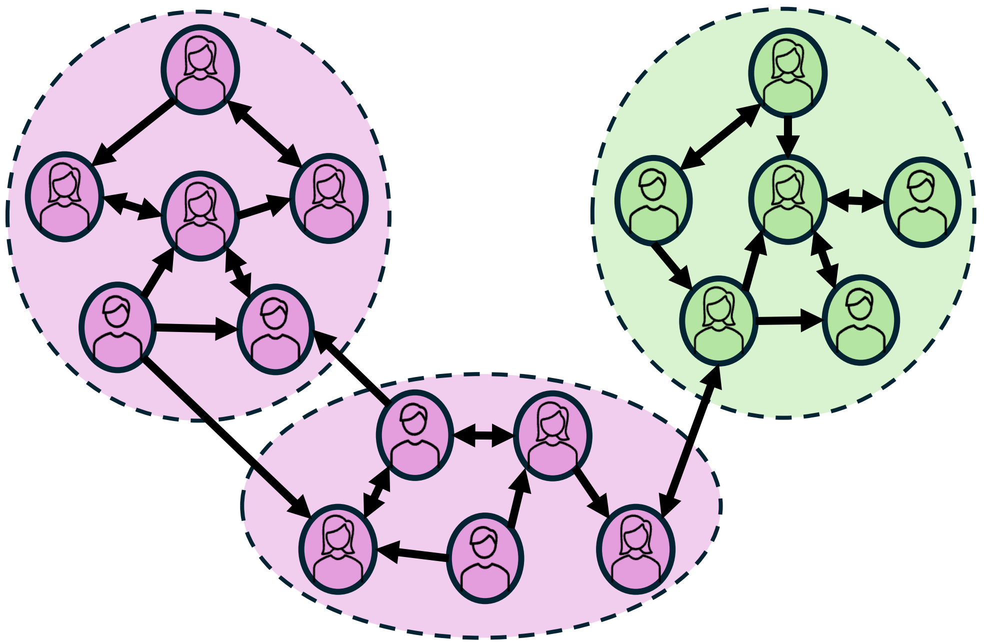
No network knowledge:

CRD over units



Network/Covariate knowledge:

CRD over clusters



Variance

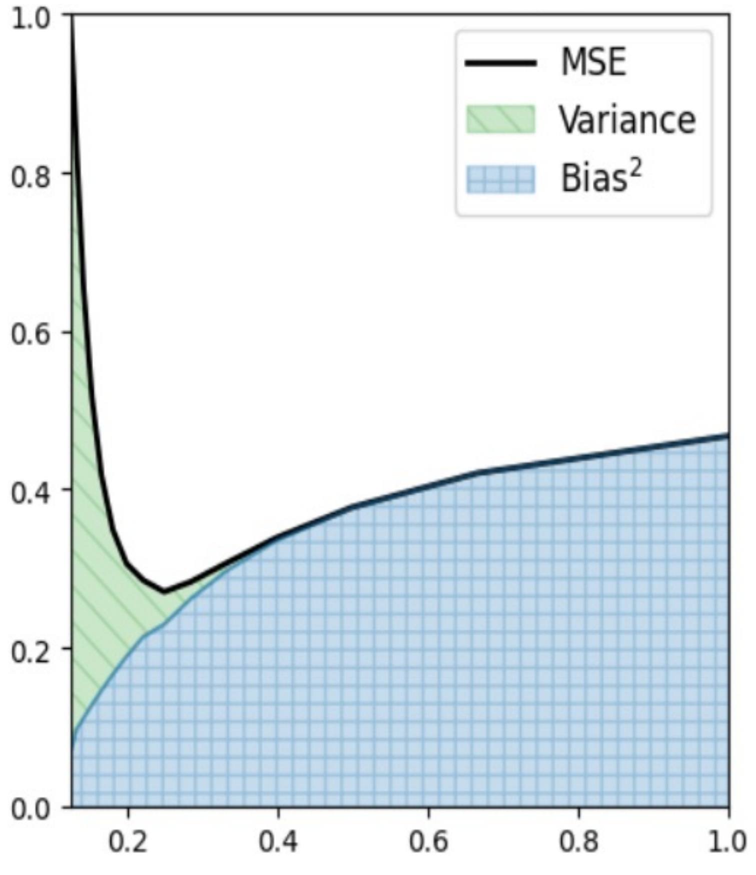
If we sample $\mathcal{U}$ with a CRD design over n_c equal-sized cluster, then $\widehat{\text{TTE}}_{2-\text{stage}}$ has variance bounded by:

$$\underbrace{\frac{q^3\beta^2Y^2d^3}{p^2n}\left(\frac{\beta}{q}\right)^{2\beta}}_{\substack{\text{Extrapolation} \\ \text{Decreases with } q}} + \underbrace{\frac{q-p}{p(n_c-1)}\widehat{\text{Var}}_\pi\left(\bar{L}_\pi\right)}_{\substack{\text{Cluster Variability} \\ \text{Increases with } q}} + \underbrace{\mathbb{I}(q>p)\frac{2d^2Y}{n_c}C(\delta(\Pi))}_{\substack{\text{Crossing Effects} \\ \text{Increases with } q}}$$

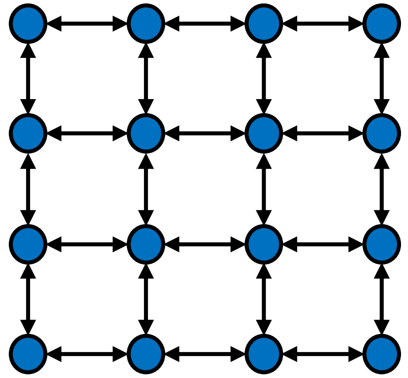
$\bar{L}_\pi$ = Average outgoing treatment effect of unit in cluster π

$C\big(\delta(\Pi)\big)$ = sum of effects including individuals in both $\mathcal{U}$ and its complement

Varying q



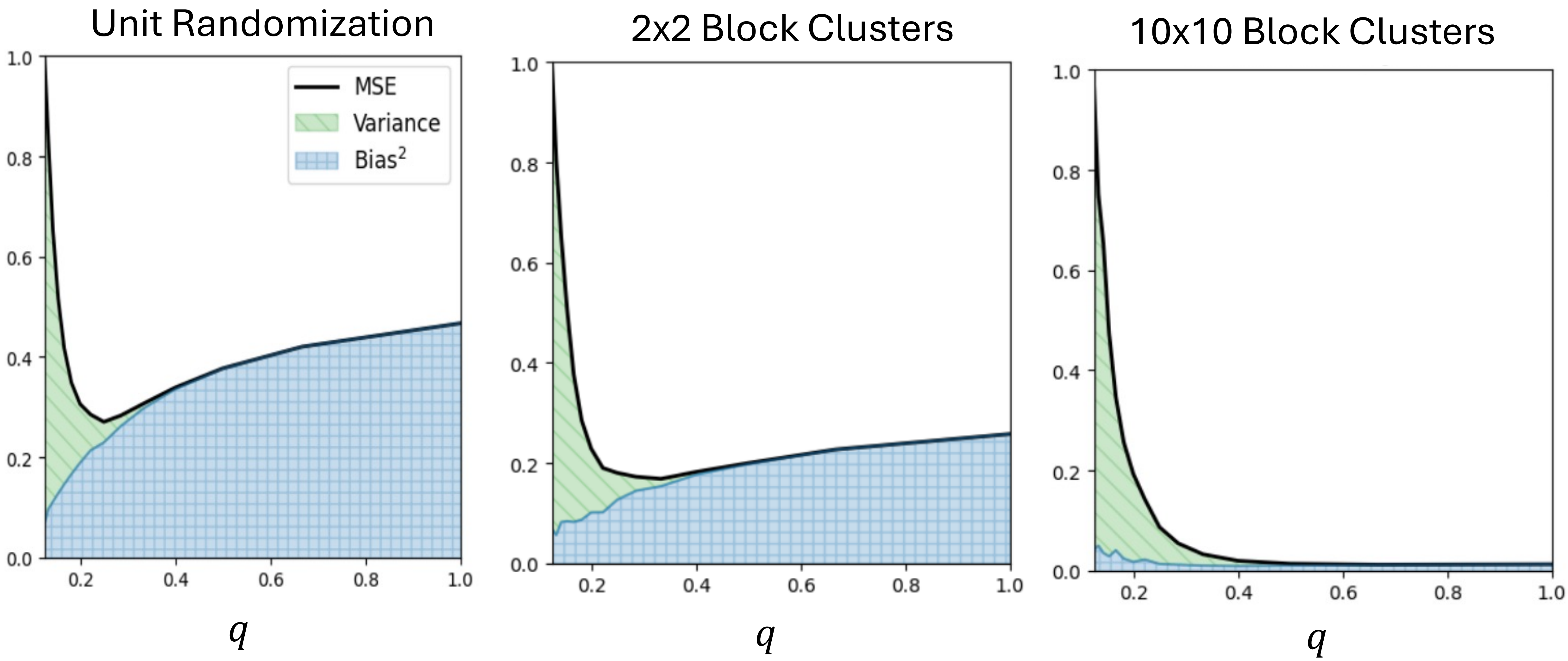
q



- 100 x 100 Lattice Graph
- Synthetic Potential Outcomes from Ugander & Yin
  - Homophily
  - Heterogenous Effect Scaling
- Treatment budget $p = 0.1$
- Model order $\beta = 3$
- **Singleton Clusters** (Unit Randomization)

Ugander, J. and Yin, H. (2023). Randomized graph cluster randomization. Journal of Causal Inference, 11(1):20220014.

Correlating $\mathcal{U}$ via clustering



Clustering in Real-World Networks

Network of 14,436 DVDs available on Amazon

- Connected to frequent co-purchases (average degree = 6)
- Annotated with ~13.2 out of 13,591 category labels (genre, actors, setting, etc.)

Two Ways to Cluster (METIS):

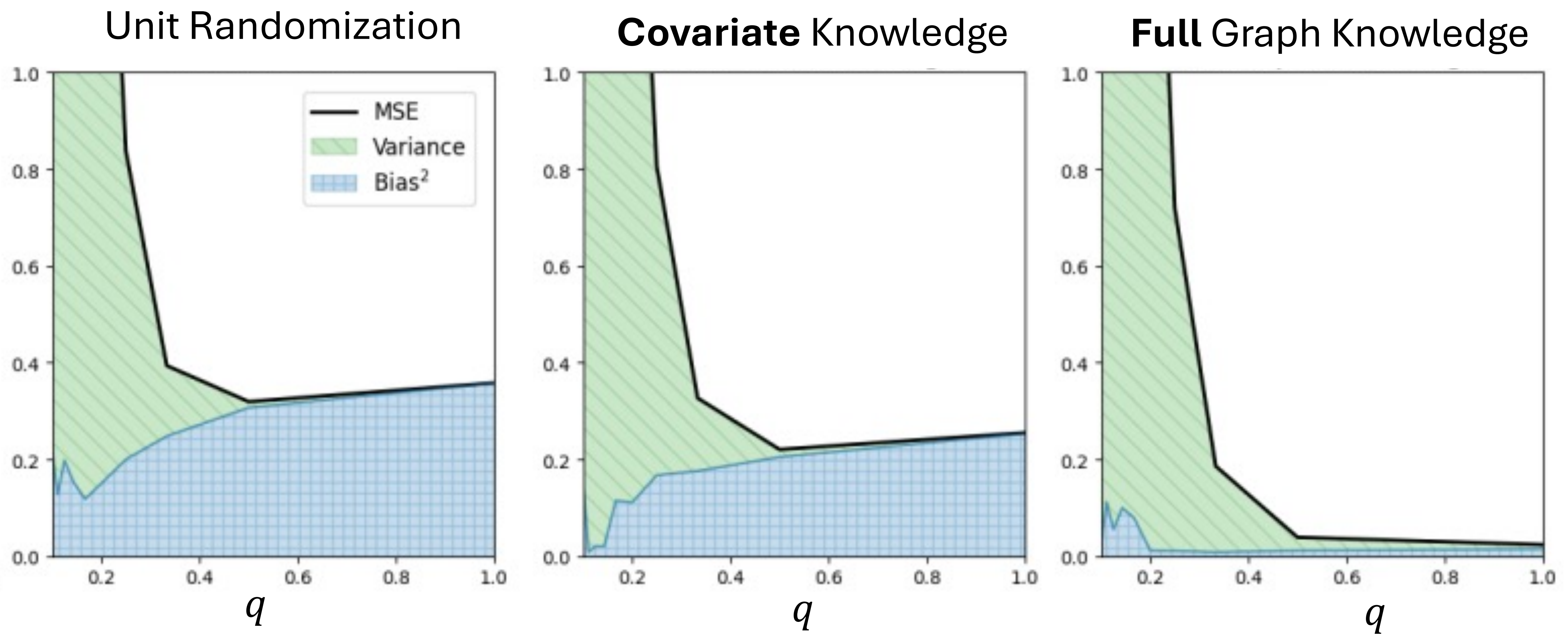
Full Knowledge: Ground truth edges

Covariate Knowledge: Using weighted feature graph

Cluster	$\widehat{\mathrm{Var}}(\bar{L}_{\pi})$	$C(\delta(\Pi))$	Cuts
Full	0.2488	0.1258	7670
Covariate	0.0426	0.5436	41243

Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007a). The dynamics of viral marketing. ACM Transactions on the Web (TWEB), 1(1):5–es.

Clustering in Real-World Networks



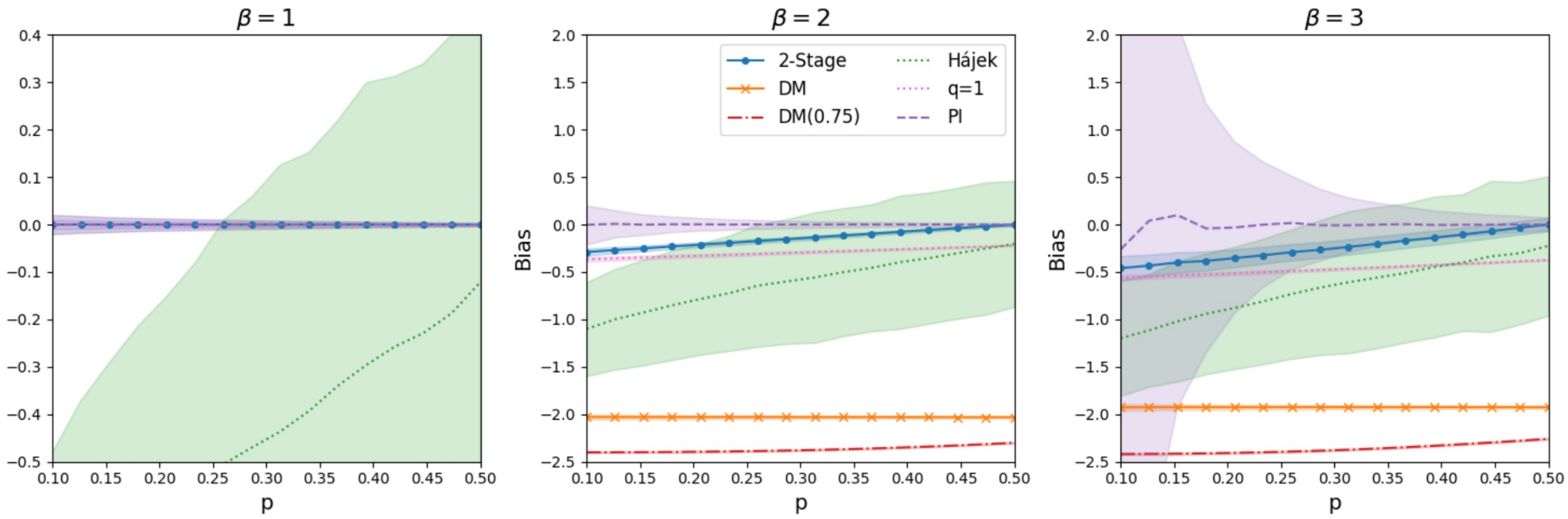
Comparing to Other Estimators

Difference in Means Estimators (Orange + Red) have high bias

Hajek (Green) and Horvitz-Thompson (Not shown) have high variance

(1-Stage) Interpolation (Purple) is unbiased, but has high variance for $\beta > 1$

2-Stage Estimator (Blue: q=0.5, Pink: q=1) *without clustering* has small bias, variance



Main Takeaways

More Details in our paper:

- β-order interactions
  - Rich framework for modeling interference
  - Rollout experiments recast estimation as polynomial fitting

- 2-Stage Interpolation
  - Vanilla interpolation estimators have high extrapolation variance
  - Subsampling reduces variance, incurs some bias

- Clustered Designs
  - Correlating $\mathcal{U}$ reduces bias from crossing edges
  - Homophily introduces trade-off in variance