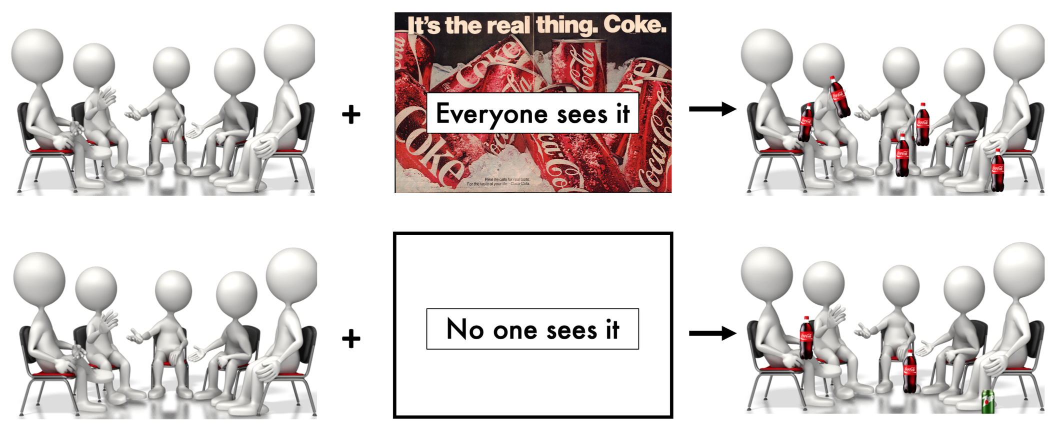
Two-Stage Rollout Designs with Clustering for Causal Inference under Network Interference

Mayleen Cortez-Rodriguez [1] Matthew Eichhorn [2] Christina Lee Yu [3]

[1]Center for Applied Mathematics [2]Computer Science [3]Operations Research and Information Engineering

The Problem

- Company runs experiment to estimate value of ad campaign
- **Total Treatment Effect (TTE)**: average change in sales when everyone versus no one sees the ad



- **Network Interference:** Word-of-mouth spreads ad's message beyond direct ad viewers
- Interference violates SUTVA, biasing classic estimators

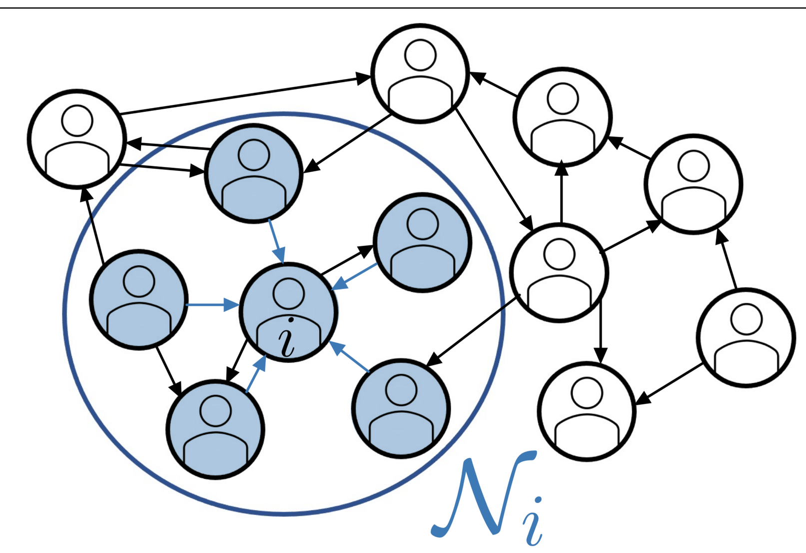
Formalizing the Problem

Population $[n] := \{1, \ldots, n\}$

Treatments $\mathbf{z} \in \{0,1\}^n$

Outcomes $Y_i(\mathbf{z})\colon \{0,1\}^n \to \mathbb{R}$

Neighborhood Interference:
$Y_i(\mathbf{z})$ depends on treatments of i's *neighbors* $\mathcal{N}_i$ w.r.t. interference graph, $d = \max_i |\mathcal{N}_i|$



β-**Order Interactions**: Only small subsets of *treated* neighbors affect i's outcome

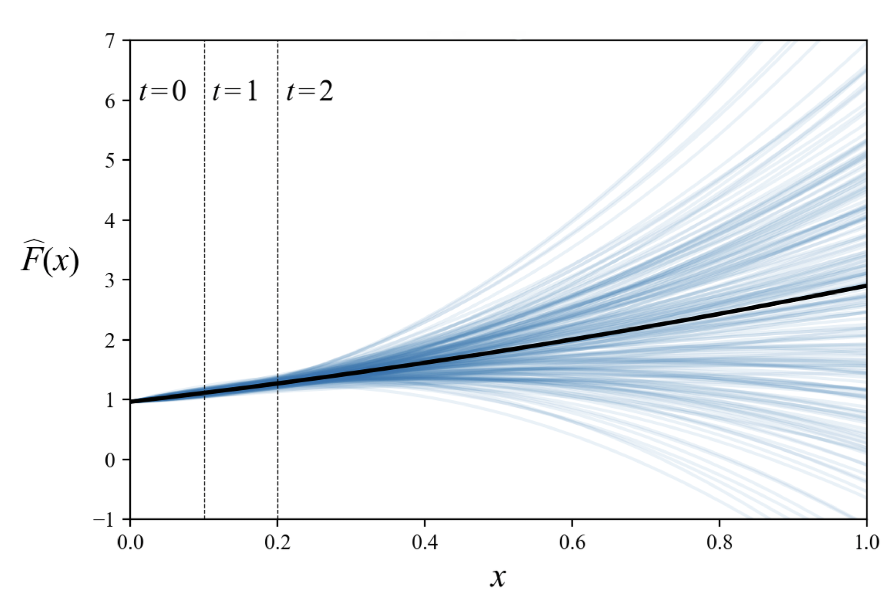
$$Y_i(\mathbf{z}) = \sum_{\mathcal{S} \in \mathcal{S}_i^\beta} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j \quad \Rightarrow \quad \text{TTE} = \frac{1}{n}\sum_{i=1}^{n} \sum_{\mathcal{S} \in \mathcal{S}_i^\beta \setminus \varnothing} c_{i,\mathcal{S}}\,, \qquad \mathcal{S}_i^\beta := \{\mathcal{S} \subseteq \mathcal{N}_i \;:\; |\mathcal{S}| \le \beta\}$$

Past Approach [1]: Bernoulli Rollout Design

- $F(p) = \mathbb{E}_{\mathbf{z}}\left[\frac{1}{n}\sum_{i=1}^n Y_i(\mathbf{z})\right]$ is β-degree polynomial, note $\text{TTE} = F(1) - F(0)$
- Staggered rollout design: in each time step t, tpn/β individuals randomly assigned to treatment
- This gives $\beta+1$ samples of F; we can estimate TTE with Lagrange interpolation

This estimator:

- ✓ Is unbiased
- ✓ Does not require knowledge of the interference network
- ✓ Outperforms baseline estimators
- ✗ Has high variance when $\beta > 1$, p small due to extrapolation



Research Objective

Develop a design/estimator pair that:

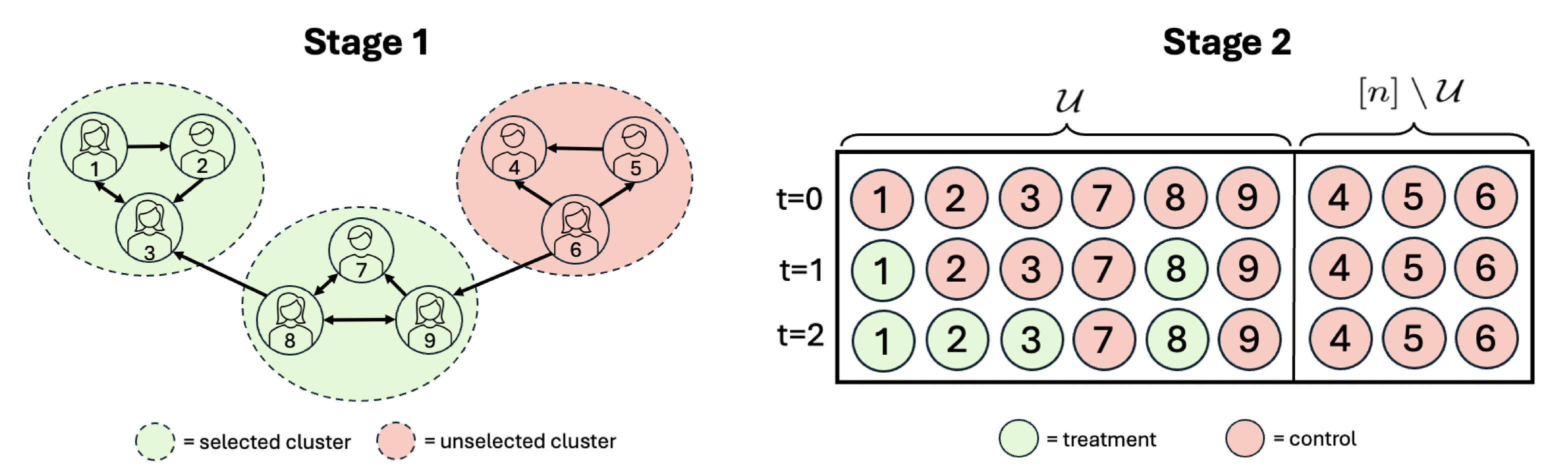
- Improves performance (over [1]) when $\beta > 1$ and treatment budget p is small
- Does not require full knowledge of the interference network, but can use network information to improve performance

Two-Stage Clustered Rollout Design

Idea: Artificially "increase" treatment budget p by running experiment on subpopulation, treating a greater proportion $q > p$ of units

Stage 1: Partition network into n_c clusters. Include clusters in *experimental units* $\mathcal{U}$ with probability $\frac{p}{q}$

Stage 2: Do rollout experiment on $\mathcal{U}$ with max treatment fraction q



2-Stage Estimator:

$$\widehat{\text{TTE}} := \frac{q}{np}\sum_{i=1}^{n}\sum_{t=0}^{\beta} h_{t,q} \cdot Y_i(\mathbf{z}^t), \qquad h_{t,q} = \prod_{\substack{s=0 \\ s \ne t}}^{\beta} \frac{\beta/q - s}{t-s} - \prod_{\substack{s=0 \\ s \ne t}}^{\beta} \frac{-s}{t-s}$$

Performance of the Two-Stage Estimator

Bias bounded by the **cut effect**, the total impact of edges crossing between clusters:

$$C(\delta(\Pi)) := \frac{1}{n}\sum_{i \in [n]} \sum_{\mathcal{S} \in \mathcal{S}_i^\beta \setminus \varnothing} c_{i,\mathcal{S}} \cdot \mathbb{I}(|\Pi(\mathcal{S})| \ge 2), \qquad \Pi(\mathcal{S}) \text{ is set of clusters containing units from } \mathcal{S}$$

- Cut effect is 0 when $\beta = 1$ or there are no crossing edges

Variance bounded above by:

$$\underbrace{\frac{d^3\beta^{2\beta}Y_{\max}^2}{np^2q^{2\beta}}}_{\substack{\text{Extrapolation,} \\ \text{Goes away if } q=1}} + \underbrace{\frac{q-p}{pn_c} \cdot \widehat{\text{Var}}(\bar{L}_\pi)}_{\text{Covariate imbalance}} + \underbrace{\frac{d^2 Y_{\max}}{n_c} \cdot C(\delta(\Pi))}_{\substack{\text{Crossing edges} \\ \text{Goes away if } q=p}},$$

where $\widehat{\text{Var}}(\bar{L}_\pi)$ is empirical variance of average treatment effect of clusters and $Y_{\max}$ bound on outcomes

Insights

- Cut effect tells us to reduce bias by reducing number of cut edges
- $\widehat{\text{Var}}(\bar{L}_\pi)$ tells us to reduce variance by increasing covariate balance
- If there is homophily, there may be a tension in these two clustering objectives
- Clustering on edges may reduce bias but increase variance
- Clustering to target covariate balance may increase bias and reduce variance

Simulation Setup

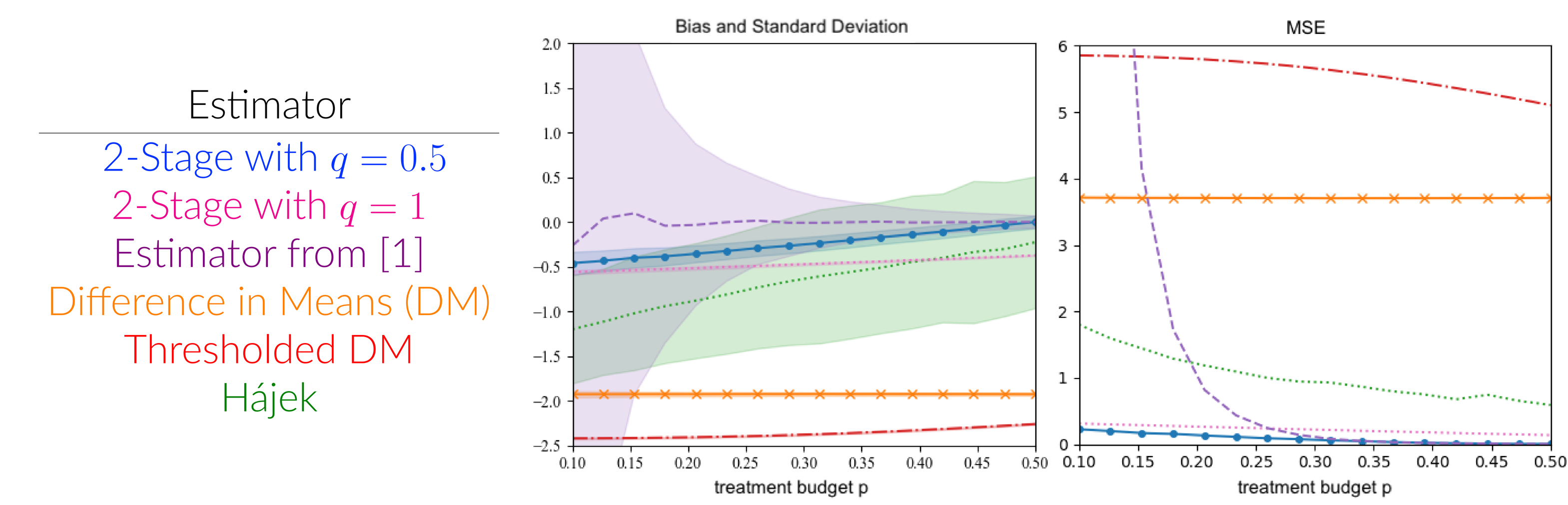
Network:

- Dataset [3] of $n = 19{,}828$ Amazon DVD product listings
- Directed edges from each DVD to five frequent co-purchases ($1 \le |\mathcal{N}_i| \le 247$)
- Each DVD has subset of ≈ 13 out of 13,591 category labels (genre, actors, setting, etc.)

Potential Outcomes: Model from [4], generalized to β-order interactions, incorporates homophily & degree correlated outcomes

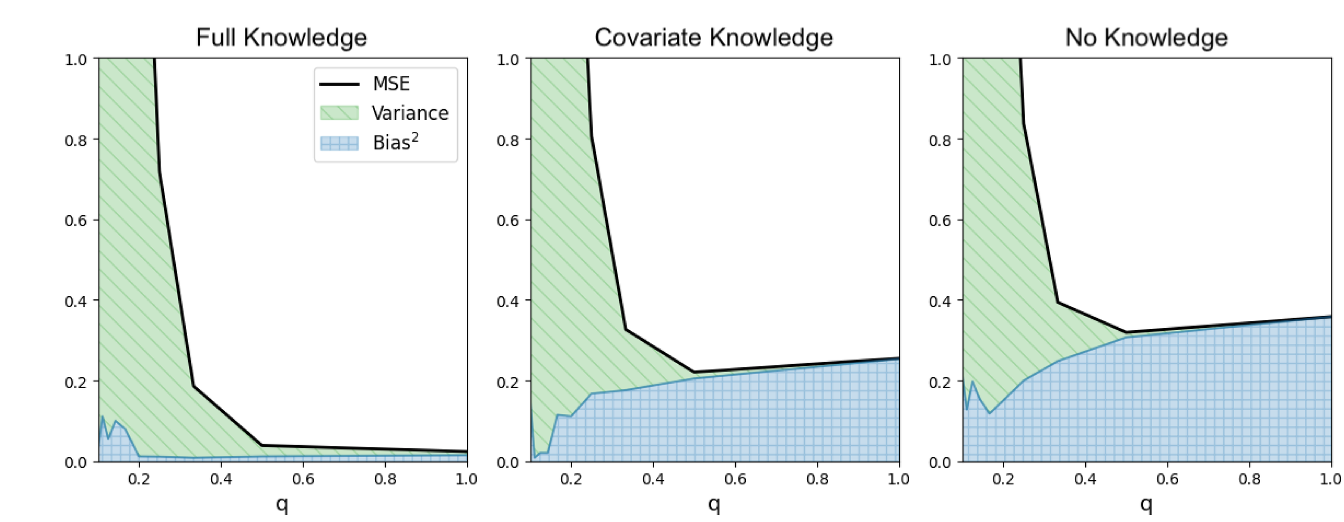
Experimental Results

Comparing performance of different estimators ($\beta = 3$):



- Thresholded DM and Hájek the only estimators requiring full network knowledge
- The estimator from [1] is the only unbiased estimator

Comparing performance of 2-stage approach under different levels of network knowledge:



- No knowledge means 2-Stage design with clusters of size 1

Insights:

- Clustering with full network knowledge achieves best overall performance
- 2-stage approach may still reduce MSE (versus single-stage) even without network knowledge

References

[1] Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. Staggered rollout designs enable causal inference under interference without network knowledge. *Advances in Neural Information Processing Systems*, 35:7437–7449, 2022.

[2] Mayleen Cortez-Rodriguez, Matthew Eichhorn, and Christina Lee Yu. Combining rollout designs and clustering for causal inference under low-order interference. 2024.

[3] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5–es, 2007.

[4] Johan Ugander and Hao Yin. Randomized graph cluster randomization. *Journal of Causal Inference*, 11(1):20220014, 2023.