Clustered Rollout Designs for Causal Inference with Network Interference

Cornell University

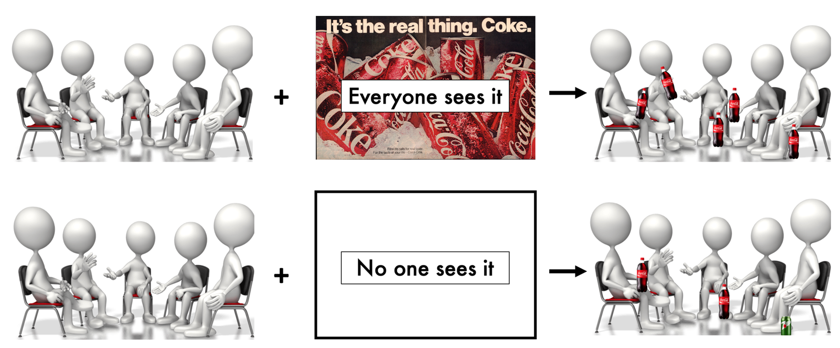
Mayleen Cortez-Rodriguez [1] Matthew Eichhorn [1] Christina Lee Yu [2]

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

Cornell University

The Problem

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



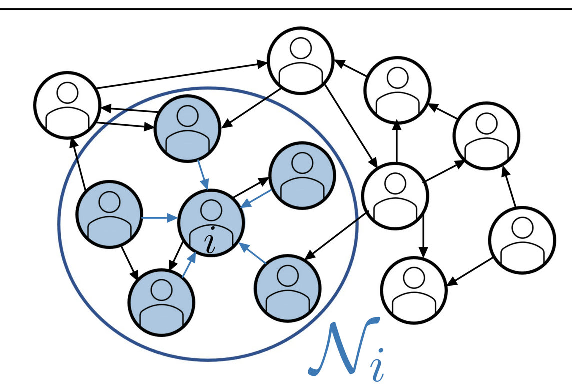
- **Interference:** Word-of-mouth spreads ad's message to others
- Interference violates SUTVA, biasing classic estimators
- Interference structure may be **unknown**

Formalizing the Problem

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

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

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



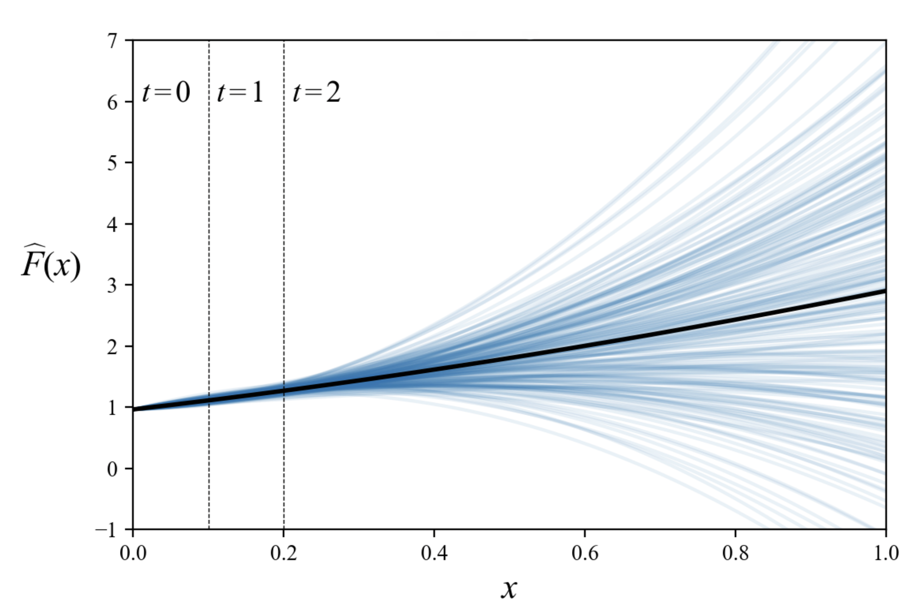
Neighborhood Interference: $Y_i(\mathbf{z})$ depends only on treatments of i's *neighbors* $\mathcal{N}_i$ w.r.t. (unknown) interference graph, $d = \max_i |\mathcal{N}_i|$

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

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

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
- Staggered rollout design gives $\beta+1$ samples of F
- Estimate $\mathrm{TTE} = F(1) - F(0)$ 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 an experimental design/estimator that:

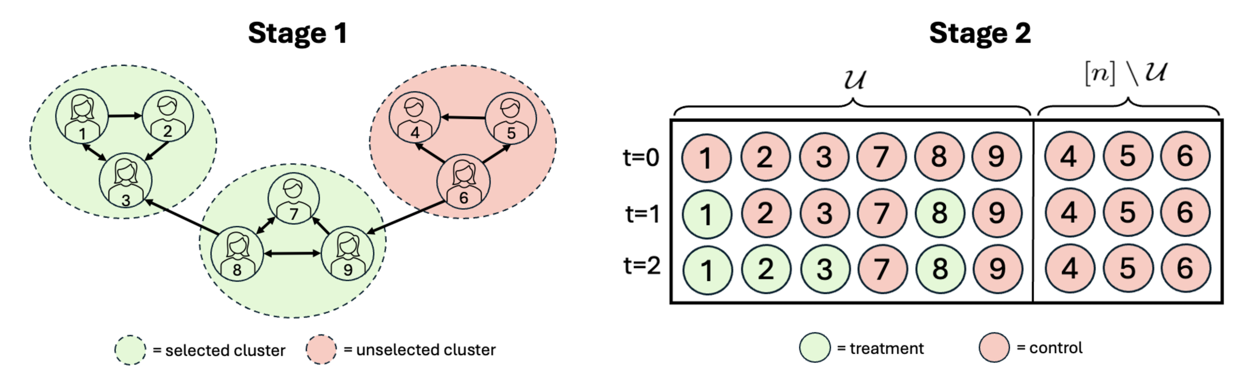
- Requires no knowledge of the interference network
- Has improved performance over [1] when $\beta > 1$ and treatment budget p is small
- Can use network knowledge to improve performance

Two-Stage Clustered Rollout Design

Idea: Run a rollout experiment on only a subpopulation, where we have the budget to treat a greater proportion $q > p$ of individuals.

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

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

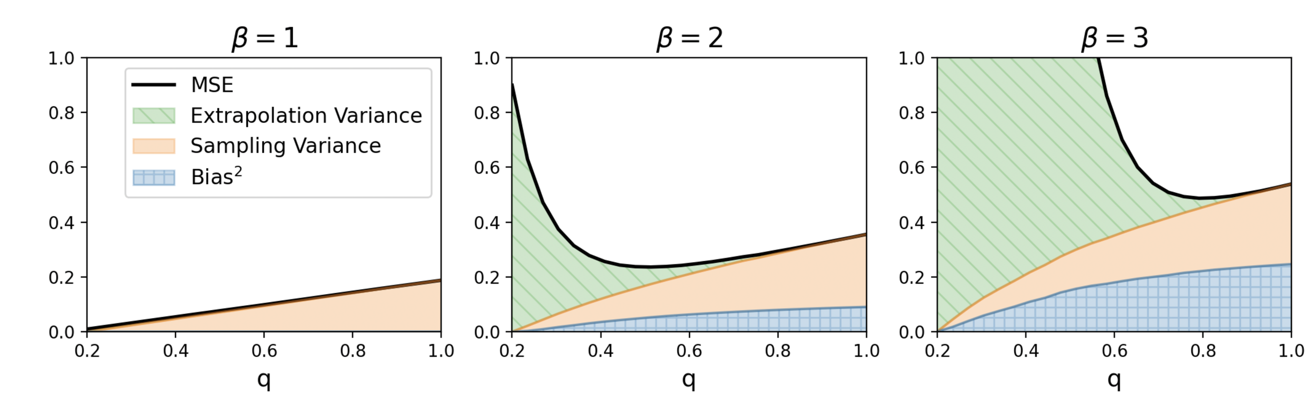


2-Stage Estimator:

$$\widehat{\mathrm{TTE}} := \frac{q}{np}\sum_{i=1}^{n}\sum_{t=0}^{\beta}\Big(\ell_t(1) - \ell_t(0)\Big)\cdot Y_i(\mathbf{z}^t), \qquad \ell_t(x) = \prod_{\substack{s=0 \\ s\ne t}}^{\beta} \frac{\beta x - qs}{qt - qs}$$

Performance of the Two-Stage Estimator

These plots visualize the MSE (black line) of the Two-Stage Estimator on a Stochastic Block Model network for three values of β.



Shading distinguishes three components of the MSE:

(Squared) Bias: $$\mathbb{E}\Big[\widehat{\mathrm{TTE}}\Big] - \mathrm{TTE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\mathcal{S}\in\mathcal{S}_i^\beta\setminus\varnothing} c_{i,\mathcal{S}}\left[\left(\frac{p}{q}\right)^{|\Pi(\mathcal{S})|-1} - 1\right]$$

Sampling Variance: $$\operatorname*{Var}_{\mathcal{U}}\Big(\mathbb{E}_{\mathbf{z}}\Big[\widehat{\mathrm{TTE}}\Big]\Big) = O\Big(q\cdot\max_{\pi\in\Pi}|\pi|\cdot\frac{d^2}{np}\Big)$$

Extrapolation Variance: $$\mathbb{E}_{\mathcal{U}}\Big[\operatorname*{Var}_{\mathbf{z}}\Big(\widehat{\mathrm{TTE}}\Big)\Big] = O\Big(\frac{1}{q^{2(\beta-1)}}\cdot\frac{d^2\beta^{2(\beta+1)}}{np^2}\Big)$$

Main Takeaways:

- When $\beta = 1$, 2-stage estimator is unbiased, clustering increases (sampling) variance
- When $\beta > 1$, bias and sampling variance increase slowly with q, while extrapolation variance sharply decreases with q
- **Two-stage design can significantly reduce MSE**

Experimental Results

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:

$$Y_i(\mathbf{z}) = Y_i(\mathbf{0})\cdot\left(1 + \delta z_i + \sum_{k=1}^{\beta}\gamma_k\cdot\binom{d_i}{k}^{-1}\sum_{\substack{\mathcal{S}\in\mathcal{S}_i^\beta \\ |\mathcal{S}|=k}}\prod_{j\in\mathcal{S}} z_j\right),$$

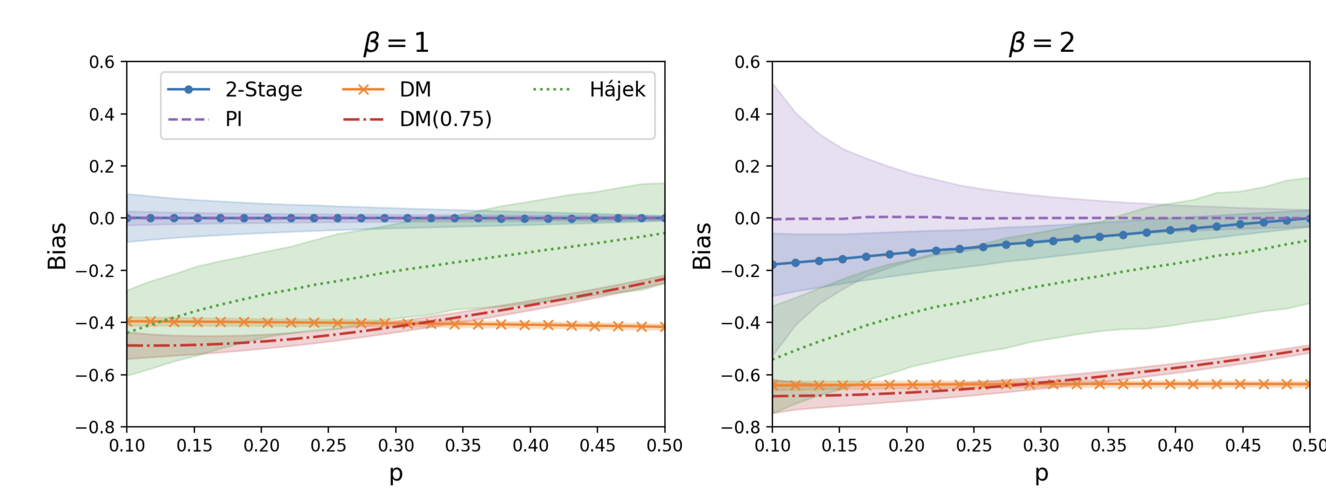
$$Y_i(\mathbf{0}) = \Big(a + b\cdot h_i + \varepsilon_i\Big)\cdot\frac{d_i}{d}$$

- Incorporates homophily (h_i) & degree (d_i) correlated responses

Estimators:

Estimator	Unbiased	Graph Knowledge?
2-Stage Interpolation	No	No
Bernoulli Interpolation	Yes	No
Difference in Means	No	No
Thresholded DM	No	Yes
Hájek	No	Yes

Visualizing bias & standard deviation, varying treatment budget p



Future/Ongoing Work

- Incorporating time-varying dynamics
- Analyzing model misspecification
- Understanding value of additional measurements in the rollout

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.