NEURAL INFORMATION PROCESSING SYSTEMS

Optimized Covariance Design for A/B Test on Social Network

Experiment Design under Interference

Lecturer: Qianyi Chen
Tsinghua University

Introduction

Background: A/B Test

- AB test is the gold standard for modern online platform to support data-driven decision making.
- It's widely adopted by social platform such as LinkedIn, Wechat, to validate new features of product, e.g., new algorithm, new UI.
- Even the 1% relative improvement is very valuable for such large platforms.
- The demand for A/B test increases rapidly. LinkedIn and Wechat usually launch more than 2,000 new experiments weekly.

Background: Complexity of A/B Test

- The number of parallel experiments is limited! (resource restriction)
- The classical experiment design scales badly for modern scenario.
- The social interaction and/or supply-demand balance induces complex dependency among users, which challenges the analysis of bias and variance of causal estimator.

Background: Experimentation on Network

- Network interference happens when people on social network interact with each other, and the influence (on concerned metric) of treatment is propagated along edges.
- Platform is concerned with global average treatment effect (GATE), whose estimation is blurred by severe bias brought by interference.

Position of Our Work

- Interference type: interference conducted by social network

- Design : treatment allocation (contrasting post-treatment techniques, such as regression adjustment)

- Regime: intensity of interference is comparable to direct treatment effect.

Limitations of Existing Literatures

- Most existing works focus on variance reduction, while bias is also very important, even dominate variance, as exposed in our simulation.

- There are a variety of variance bounds of different estimators, while seldom of them can directly instruct experiment design.

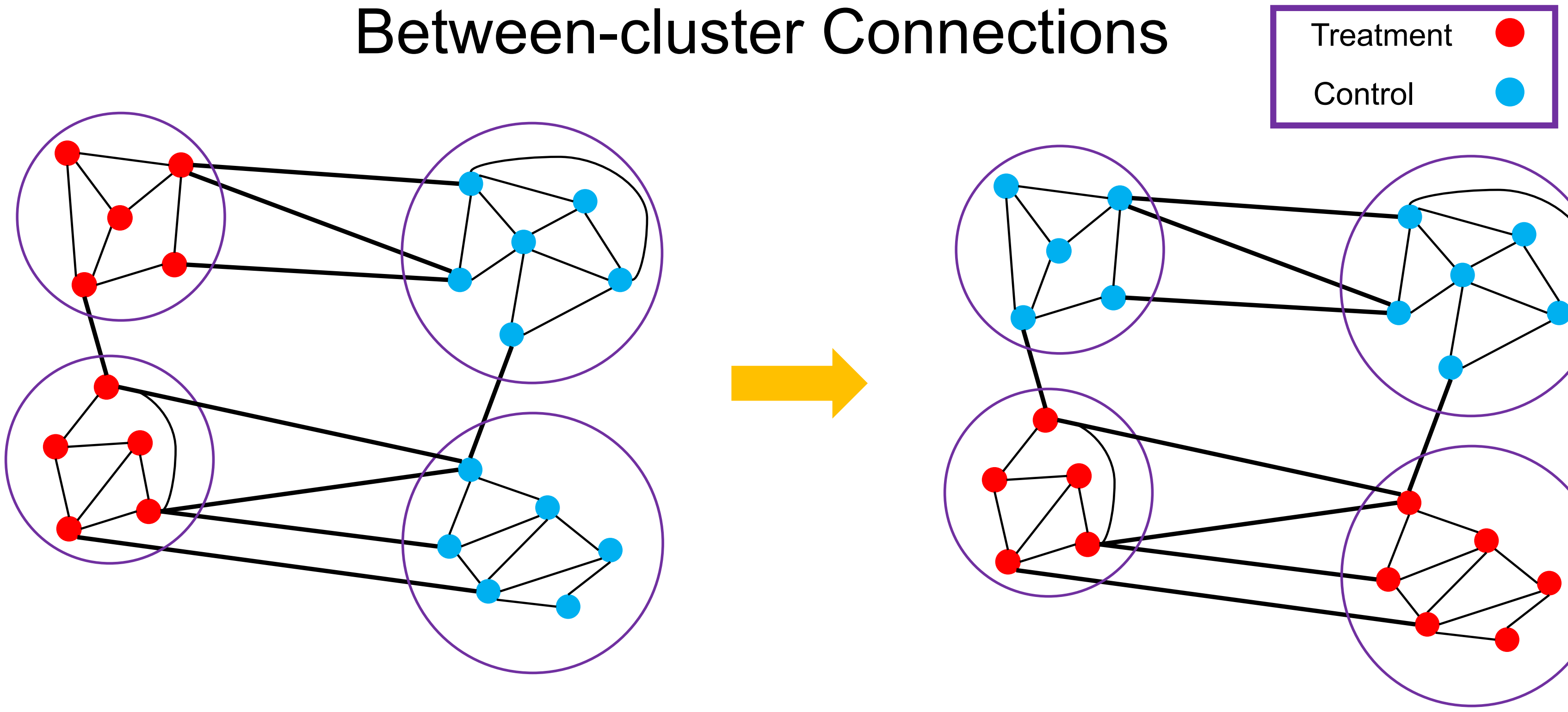
- Many experiment designs are concerned with mathematical programming (such as SDP, MILP) that scale badly for social platform.

Traits of our Optimized Covariance Design

- We're concerned with minimization of a tight MSE upper bound that consider bias and variance in a meanwhile.
- We derive an optimizable bound on MSE under a potential outcome model that enables covariance matrix of treatment vector to be decision variables.
- We propose a sampling procedure and a projected gradient descent algorithm that supports efficient optimization.

Adaptive to Between-cluster Connection

Treatment Covariance Adaptive to Between-cluster Connections



Basic Setting

- We consider binary treatment vector

$$\boldsymbol{z} = (z_1, z_2, \ldots, z_n) \in \{0, 1\}^n$$

- The estimand is GATE

$$\tau := \frac{1}{n} \sum_{i \in [n]} (Y_i(\mathbf{1}) - Y_i(\mathbf{0}))$$

- We consider graph cluster randomization, and cluster-level treatment vector is

$$\boldsymbol{t} = (t_1, t_2, \ldots, t_K) \in \{0, 1\}^K$$

Basic Setting

- We consider balanced cluster-level randomization

$$\mathbb{P}\left(t_k = 1 \mid \mathcal{G}\right) = \frac{1}{2} \quad \mathbb{E}\left[z_i\right] = \frac{1}{2}$$

- We consider standard HT estimator (without exposure indicator)

$$\hat{\tau} = \frac{1}{n} \sum_{i \in [n]} \left(\left(\frac{z_i}{\mathbb{E}\left[z_i\right]} - \frac{(1 - z_i)}{\mathbb{E}\left[1 - z_i\right]} \right) Y_i(\boldsymbol{z}) \right)$$

- We consider following linear potential outcome model

$$Y_i(\boldsymbol{z}) = \alpha_i + \beta_i z_i + \gamma \sum_{j \in N_i} z_j$$

Bias and Variance Analysis

Bias of HT Estimator

- Firstly, we define a core term in our methodology, which characterizes the connections between/within clusters. Here S_k is the k-th cluster

$$\boldsymbol{C}_{ij} = |\{(u,v) : (u,v) \in \mathcal{E}, u \in S_i, v \in S_j\}|$$

- Now we can present the bias of HT estimator under our model

$$\mathbb{E}[\hat{\tau}] - \tau = \frac{\gamma}{n}\left(4\,\mathrm{trace}(\boldsymbol{C}\,\mathrm{Cov}[\boldsymbol{t}]) - \sum_{i,j \in [K]} \boldsymbol{C}_{ij}\right)$$

- This bias formula implies
  - Only connections between clusters can contribute to bias.
  - Only positive correlation can reduce bias.

Variance of HT Estimator

- To derive a clean variance, we must introduce an assumption on base level α_i, which is we know all base levels in advance.

- This assumption is reasonable for social platform since they collect concerned metrics constantly, and it remove the giant influence of α_i in variance (since $\alpha_i \gg \beta_i$ in such experiments)

- Based on it, we can derive the expression of variance.

$$\mathrm{Var}[\hat{\tau}] = \frac{4}{n^2}\left(\mathrm{trace}\left(\boldsymbol{h}\boldsymbol{h}^T\,\mathrm{Cov}[\boldsymbol{t}]\right) + 4\gamma\,\mathrm{Cov}\left[\boldsymbol{h}^T\boldsymbol{t}, \boldsymbol{t}^T\boldsymbol{C}\boldsymbol{t}\right] + 4\gamma^2\,\mathrm{Var}\left[\boldsymbol{t}^T\boldsymbol{C}\boldsymbol{t}\right]\right)$$

Methodology

Bypass Parameter Estimation

- The expression of variance can't be optimized directly without knowing interference intensity γ in advance.
- We introduce following comparability assumption that restricts our scope to the scene that interference is comparable to direct treatment effect

Assumption 3 (Comparability between Direct Treatment Effect and Interference) *Given potential outcome model in equation (6), we assume there exists a constant $\omega > 0$ such that*

$$|\boldsymbol{h}_k| \leq \omega\gamma(\sum_{i \in S_k} d_i) \tag{13}$$

holds for each cluster $k \in [K]$.

Bypass Parameter Estimation

- Now we can construct a variance bound that depends on experiment design only through covariance matrix of treatment vector.
- Moreover, this bound is well-crafted and allow us to bypass the estimation on γ: if we're concerned with minimize this bound, γ^2 is a common multiplier in squared bias and variance bound!

Proposition 3 (Variance Bound) *The variance of the standard HT estimator has following upper bound,*

$$\mathrm{Var}[\hat{\tau}] \leq \frac{8\gamma^2(\omega^2+4)}{n^2} \operatorname{trace}\left(\boldsymbol{d}\boldsymbol{d}^T(\mathrm{Cov}[\boldsymbol{t}] + \frac{1}{4}\mathbf{1}\mathbf{1}^T)\right) \tag{14}$$

where $\mathbf{d}$ *is the vector* $(\sum_{i \in S_k} d_i)_{k=1}^K$.

Enable Sampling Following Optimized Covariance

- Before we finish the formulation of optimizing the covariance matrix, we should guarantee two points
  - The covariance matrix is legal for multi-variate Bernoulli distribution.
  - We can sample treatment vector that's subject to such covariance.

- To realize it, we introduce the Grothendieck's identity and a Cholesky-based parameterization, and the covariance matrix is parameterized as

$$X(R) = \frac{\arcsin\left(RR^T\right)}{2\pi}$$

Optimization Issues

- Through this parameterization, the constraints is simplified significantly.

$$\begin{aligned} \min_{R} \quad & M(R) = B(X(R))^2 + \bar{V}_\omega(X(R)) \\ \text{s.t.} \quad & (RR^T)_{i,j} \in [-1, 1] \quad \forall i \neq j,\ i, j \in [K] \\ & (RR^T)_{i,i} = 1 \qquad\qquad \forall i \in [K] \end{aligned}$$

- We verify that row-normalization is a projection to feasible domain, and propose a projected gradient descent algorithm for the optimization.

- After optimization, we can sample directly from desired distribution

$$\boldsymbol{t} = \frac{\mathbf{1} + \operatorname{sgn}\left(R^* \mathcal{N}\left(\mathbf{0}, I_K\right)\right)}{2}$$

Simulation Result

- Our optimized covariance design present significant improvement on both statistical metrics (bias, variance, MSE), and computational efficiency.

Table 2: The average bias, standard deviation and MSE of HT estimator under multiplicative model

gamma		0.5			1.0			2.0	
metric	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
method									
Ber	-0.365	0.348	0.255	-0.736	0.394	0.698	-1.475	0.493	2.421
CR	-0.368	0.235	0.191	-0.744	0.274	0.629	-1.477	0.336	2.297
ReAR	-0.402	0.178	0.194	-0.809	0.174	0.685	-1.548	0.226	2.450
PSR	-0.366	0.134	0.152	-0.738	0.153	0.569	-1.479	0.192	2.227
IBR	-0.369	0.155	0.161	-0.737	0.178	0.576	-1.484	0.221	2.252
IBR-p	-0.368	0.163	0.163	-0.739	0.185	0.581	-1.482	0.232	2.252
OCD	-0.258	0.040	**0.069**	-0.517	0.050	**0.271**	-1.034	0.054	**1.073**