Minimax-Optimal Experimental Design for Network Interference on Pseudo-Random Graphs

Shuangping Li

Shuangning Li

Yale University

LUX ET VERITAS

CHICAGO BOOTH

The University of Chicago Booth School of Business

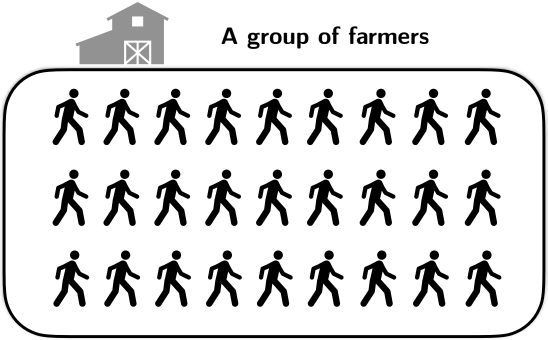
Minimax-Optimal Experimental Design for Network Interference on Pseudo-Random Graphs

A motivating example:

- ? What is interference?
- ? How does interference complicate causal inference?
- ? Why do we need better experiments?

Weather insurance

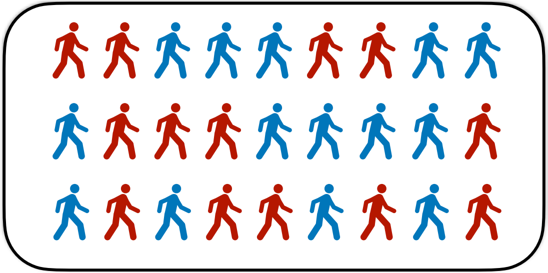
Example: Farmers' Insurance Decisions



[Cai, de Janvry, & Sadoulet, 2015]

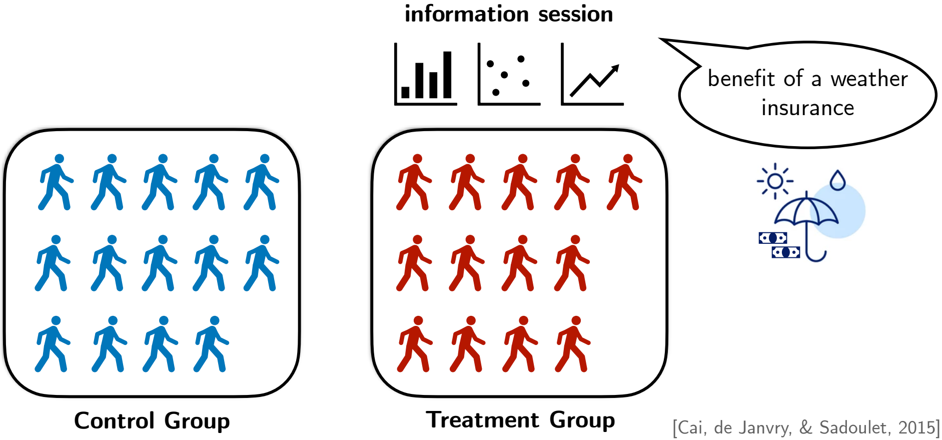
Example: Farmers' Insurance Decisions

A group of farmers



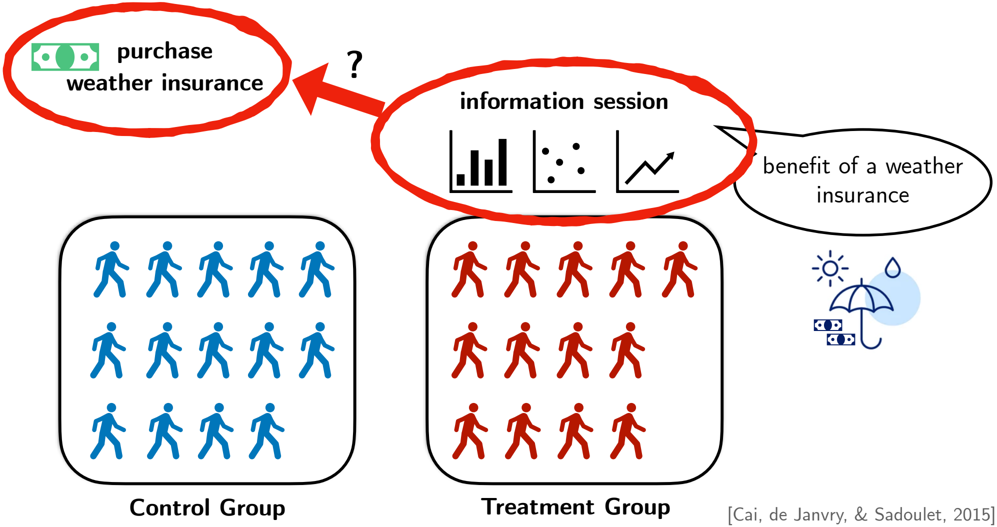
[Cai, de Janvry, & Sadoulet, 2015]

Example: Farmers’ Insurance Decisions



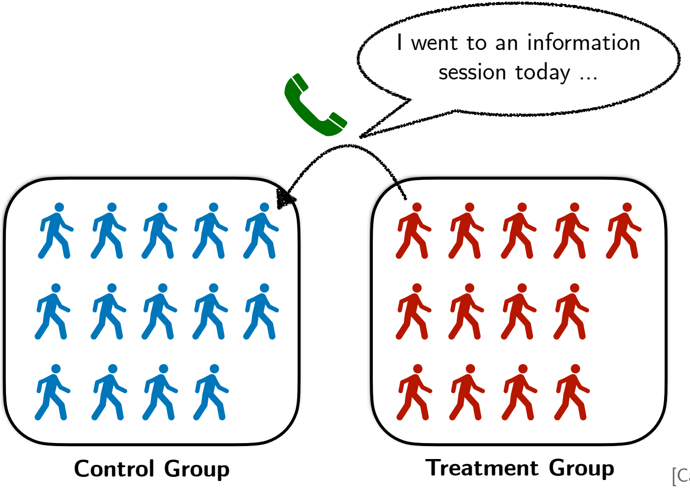
[Cai, de Janvry, & Sadoulet, 2015]

Example: Farmers’ Insurance Decisions



[Cai, de Janvry, & Sadoulet, 2015]

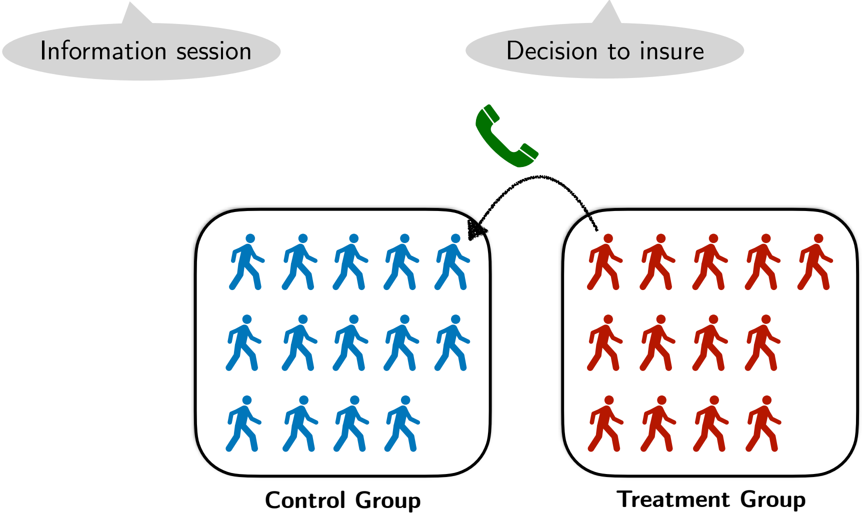
Example: Farmers' Insurance Decisions



[Cai, de Janvry, & Sadoulet, 2015]

Example: Farmers' Insurance Decisions

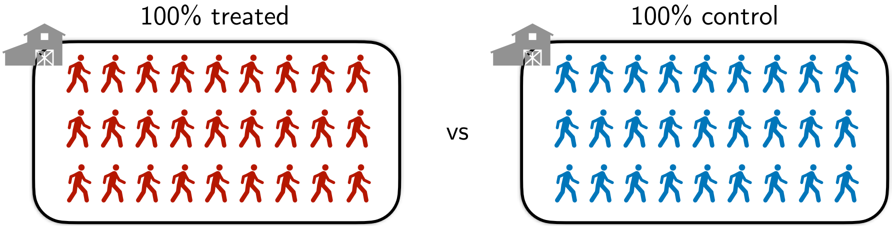
Treatment assigned to one unit may affect outcomes for **other** units—**Interference**.



[Cai, de Janvry, & Sadoulet, 2015]

Example: Farmers' Insurance Decisions

Goal: average outcome (everyone treated) - average outcome (everyone in control group)



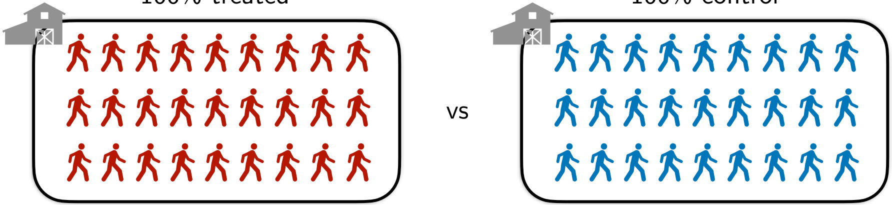
[Cai, de Janvry, & Sadoulet, 2015]

How does interference complicate causal inference?

Goal: average outcome (everyone treated) - average outcome (everyone in control group)

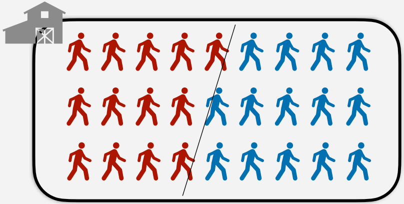
100% treated

100% control



vs

Can only observe:



50% treated

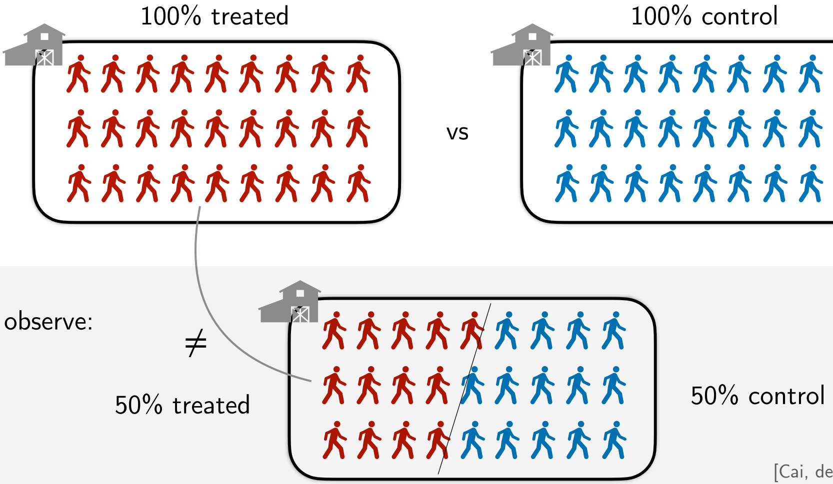
50% control

[Cai, de Janvry, & Sadoulet, 2015]

How does interference complicate causal inference?

Goal: average outcome (everyone treated) - average outcome (everyone in control group)

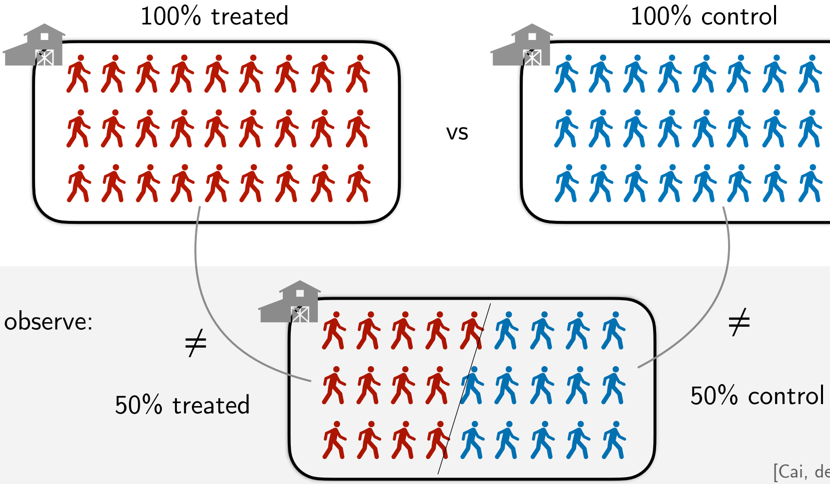
Can only observe:



[Cai, de Janvry, & Sadoulet, 2015]

How does interference complicate causal inference?

Goal: average outcome (everyone treated) - average outcome (everyone in control group)

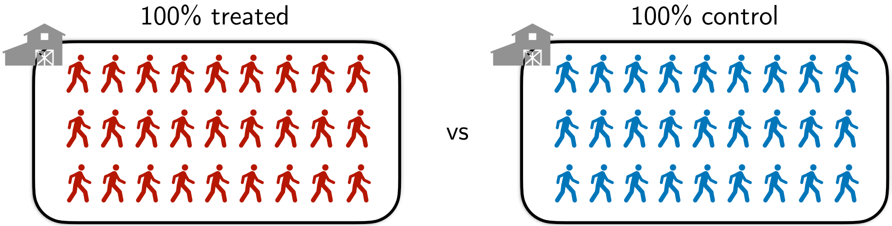


Can only observe:

[Cai, de Janvry, & Sadoulet, 2015]

How does interference complicate causal inference?

Goal: average outcome (everyone treated) - average outcome (everyone in control group)



Difference-in-means estimator

average outcome (treated units) - average outcome (control units)

The difference-in-means estimator is biased!

[Cai, de Janvry, & Sadoulet, 2015]

Network Interference Model

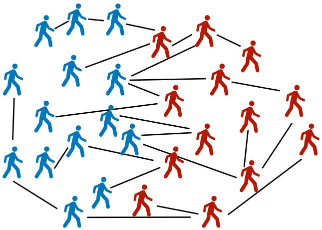
Treatment assigned to one unit may affect outcomes for **other** units—**Interference**.

Information session

Decision to insure

Survey

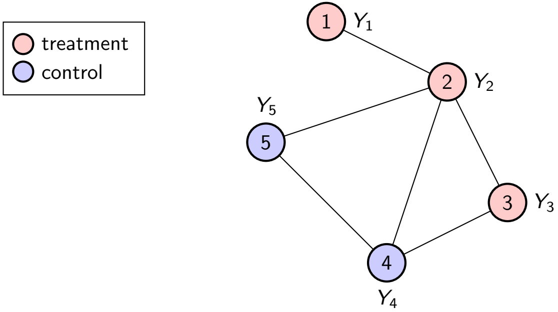
✦ List your close friends, within or outside your village, with whom you most frequently discuss rice production or financial issues.



[Cai, de Janvry, & Sadoulet, 2015]

Network Interference Model

- Subjects indexed $i = 1, \ldots, n$.
- Treatment $W_i \in \{0, 1\}$.
- Outcome $Y_i \in \mathbb{R}$.
- Graph with edge set $\{E_{ij}\}_{i,j=1}^{n}$:
  - Y_i may depend on W_j only if there is an edge from i to j.
  - Potential outcomes $Y_i(\cdot)$ with $Y_i = Y_i(W_i, W_{\mathcal{N}_i})$.



Global Average Treatment Effect

- Goal: estimate the global average treatment effect (GATE):

$$\text{GATE} = \frac{1}{n} \sum_i \left[Y_i(\vec{1}) - Y_i(\vec{0}) \right].$$

An Unbiased Estimator: The Horvitz–Thompson Estimator

- Define

$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right),$$

  where T_i is the indicator that unit i and all of its neighbors are treated, C_i is the indicator that unit i and all of its neighbors are in control.
  - Formally, $T_i = W_i \prod_{j\in\mathcal{N}_i} W_j$ and $C_i = (1 - W_i)\prod_{j\in\mathcal{N}_i}(1 - W_j)$.
- This estimator is unbiased for the GATE under any randomized design.

Does a Bernoulli Experiment Work?

- Define

$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right),$$

where T_i is the indicator that unit i and all of its neighbors are treated, C_i is the indicator that unit i and all of its neighbors are in control.

  – Formally, $T_i = W_i \prod_{j \in \mathcal{N}_i} W_j$ and $C_i = (1 - W_i)\prod_{j \in \mathcal{N}_i}(1 - W_j)$.

- In a Bernoulli experiment where $W_i \sim \text{Bern}(\pi)$ independently,

$$\mathbb{E}\left[T_i\right] = \pi^{|\mathcal{N}_i|+1}.$$

The denominator decays **exponentially** with the number of neighbors.

Does a Bernoulli Experiment Work?

- Define

$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right),$$

  where T_i is the indicator that unit i and all of its neighbors are treated, C_i is the indicator that unit i and all of its neighbors are in control.
  - Formally, $T_i = W_i \prod_{j\in\mathcal{N}_i} W_j$ and $C_i = (1 - W_i)\prod_{j\in\mathcal{N}_i}(1 - W_j)$.
- In a Bernoulli experiment where $W_i \sim \mathrm{Bern}(\pi)$ independently,

$$\mathbb{E}\left[T_i\right] = \pi^{|\mathcal{N}_i|+1}.$$

  The denominator decays **exponentially** with the number of neighbors. Not ideal!
- Can we design better experiments?

Goal of This Work

What experimental design minimizes the worst-case variance of the HT estimator?

- An experimental design $\mathcal{E}$ is a joint distribution over the treatment vector $W = (W_1, \ldots, W_n)^\top$.
- Let

$$\mathcal{Y}_n = \left\{ \{Y_i(\cdot)\}_{i=1}^n : \sum_{i=1}^n [Y_i(w_i, w_{\mathcal{N}_i})]^2 \leq n \text{ for all } \vec{w} \in \{0,1\}^n \right\}$$

  be the class of potential outcomes whose two-norm, under every treatment assignment, is at most $\sqrt{n}$.
- The goal is to find $\mathcal{E}$ that minimizes

$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \mathrm{Var}_{W \sim \mathcal{E}} \left[\hat{\tau}_{\mathrm{HT}}(W) \right].$$

Two Competing Considerations

$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right).$$

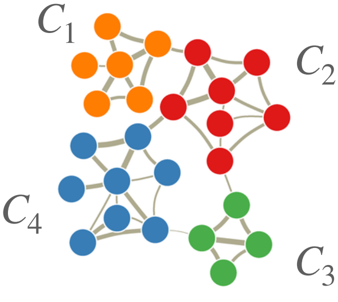
For a good experimental design:

- We want to avoid small denominators. Therefore, treatment assignments of neighboring units should occur together with reasonably high probability.
- We do not want to induce excessive dependence between T_i and T_j across units.
- There is a fundamental tension between these two goals: independence is easiest to achieve under Bernoulli designs, but such designs lead to exponentially small denominators.

Prior Approaches in the Literature

Cluster randomization [Ugander, Karrer, Backstrom, & Kleinberg, 2013; Ugander & Yin, 2023]

- Let the vertices be partitioned into K clusters $C_1, C_2, \ldots, C_K$
  - Define $C(\cdot) : \{1, \ldots, N\} \to \{1, \ldots, K\}$ as mapping vertex indices to cluster indices.
  - C_j refers to cluster j by its index.
  - $C(i)$ refers to the cluster containing vertex i.

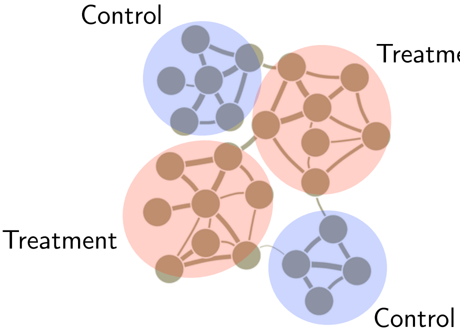


Prior Approaches in the Literature

Cluster randomization [Ugander, Karrer, Backstrom, & Kleinberg, 2013; Ugander & Yin, 2023]

- Treatments are assigned at the cluster level.
- Each cluster C_j is assigned a treatment $Z_j \sim \text{Bern}(0.5)$.
- The treatments assigned to vertices are those assigned to their clusters:

$$W_i = Z_{C(i)}.$$



Prior Approaches in the Literature

Cluster randomization [Ugander, Karrer, Backstrom, & Kleinberg, 2013; Ugander & Yin, 2023]

- Choosing the cluster smartly yields:

$$\sup_{|Y_i(\cdot)|\le 1} \operatorname{Var}\left[\hat{\tau}_{\mathrm{HT}}(W)\right] = \mathcal{O}\left(\lambda(G^2)\, d_{\max}(G)\, \kappa^3/n\right),$$

  where

  - $\lambda(G^2)$ is the largest eigenvalue of the adjacency matrix of the two-hop graph G^2,
  - $d_{\max}(G)$ is the maximum degree of G,
  - κ is the restricted growth coefficient of the graph.

- Since κ is typically on the order of $d_{\max}(G)$, this result yields a rate of approximately $\mathcal{O}\left(\lambda(G^2)\, d_{\max}(G)^4/n\right)$.

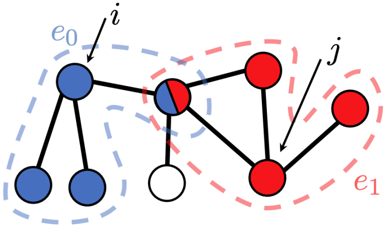
Prior Approaches in the Literature

Conflict graph design [Kandiros, Pipis, Daskalakis, & Harshaw, 2024]

- Using a modified version of the HT estimator, the authors show that this design achieves the improved rate

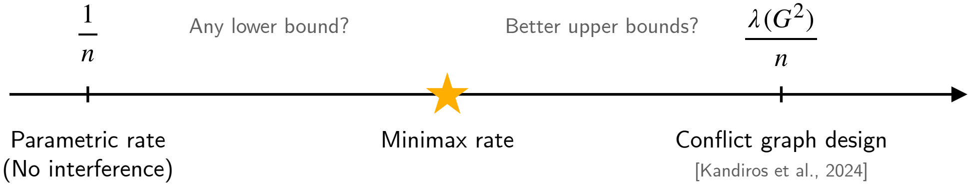
$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \text{Var}\left[\hat{\tau}_{\widetilde{\text{HT}}}(W)\right] = \mathcal{O}\big(\lambda(G^2)/n\big) .$$

- Represents a factor-of-d_{max}^4 improvement over the graph cluster randomization in typical settings where $\kappa = \Theta(d_{\text{max}})$.



Two Questions We Need to Answer

- Lower bound: Establish a lower bound on the worst-case variance.
- Upper bound: Improve existing upper bounds through better experimental design.
- Goal: Match the lower and upper bounds, thereby characterizing the optimal rate.



A Lower Bound

Theorem (A Lower Bound)

For any experimental design $\mathcal{E}$,

$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \mathrm{Var}_{W\sim\mathcal{E}}\left[\hat{\tau}_{\mathrm{HT}}(W)\right] = \tilde{\Omega}\left(\frac{d_{\max}(G)}{n}\right), \tag{1}$$

where $d_{\max}(G)$ is the maximum degree of G.

- Under any experimental design, there exists a potential outcome model in $\mathcal{Y}_n$ for which the variance of the HT estimator scales as $\widetilde{\Omega}(d_{\max}(G)/n)$.

Lower Bound: Star Graph Intuition

- Consider a star graph where the center node influences all leaf nodes.
- The GATE consists of a treatment component $\frac{1}{n}\sum_{i=1}^{n} Y_i(\vec{1})$ and a control component $\frac{1}{n}\sum_{i=1}^{n} Y_i(\vec{0})$.
- Accurate estimation requires information about both components.
- There are essentially two worlds: one where the center node is treated and the other where it is not. The design only reveals one of these worlds in any given draw of W.
- This structural limitation yields a near-constant lower bound on the variance.

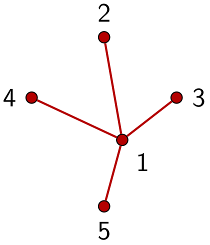


Figure: Star subgraph H.

Lower Bound: Extension to General Graphs

- The star-graph intuition extends to general graphs.
- Consider the node with the largest degree and its induced neighborhood.
- Assign nonzero potential outcomes only to this subgraph.
- This subgraph behaves like a star graph.
- The resulting lower bound scales with $d_{\max}(G)$ rather than n.
- This establishes the worst-case variance rate.

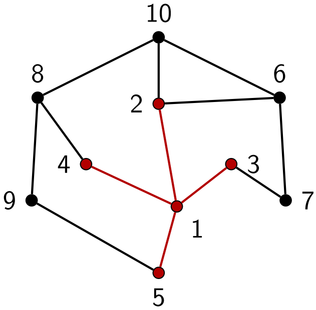


Figure: Graph G with a star-like subgraph highlighted.

Upper Bound (Ongoing)

- Recall that
$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right).$$
- *As a first attempt*, we focus on
$$\hat{\tau}_{\mathrm{HT}}^{(1)}(W) = \frac{1}{n}\sum_{i=1}^{n}\frac{T_i Y_i}{\mathbb{E}[T_i]}.$$
This is an unbiased estimator of $\frac{1}{n}\sum_{i=1}^{n} Y_i(\vec{1})$. For simplicity, we refer to this estimator as $\mathrm{HT}^{(1)}$.
- We consider the scenario where $\mathbb{P}(W_i = 1) \leq 1/2$ to maintain problem complexity.

A Simple Design that Protects Denominators

Proposed design

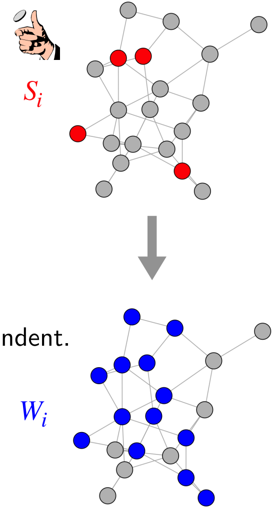
- Sample $S_i \sim \mathrm{Bern}(q)$ independently.
- Set

$$W_i = \max_{j \in N_i \cup \{i\}} S_j.$$

- Interpretation: if $S_i = 1$, then i and all of its neighbors are treated.

Why it helps:

- *Not too small denominators:* $\mathbb{E}[T_i] \geq q$ by construction.
- *Limited covariance:* dependence is induced locally, but the S_i are independent.
- *Simplicity:* easy to implement and analyze.



Upper Bound on Pseudo-Random Graphs

Theorem (An upper bound for pseudo-random graphs)

Assume $|\mathcal{N}_i \cap \mathcal{N}_j| \leq |\mathcal{N}_i|/(\log n)^2$ for any i, j. Consider the design above with $q = 1/(2d_{\max}(G))$. Then

$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \mathrm{Var}_{W \sim \mathcal{E}} \left[\hat{\tau}_{\mathrm{HT}}^{(1)}(W) \right] = \widetilde{\mathcal{O}}\left(\frac{d_{\max}(G)}{n} \right).$$

- This is the variance of the $\mathrm{HT}^{(1)}$ estimator, not the full HT estimator.
- The bound matches the lower bound up to polylogarithmic factors.
- The codegree condition holds for many standard random graph models.
  - *Erdős–Rényi graphs* $G(n, p)$ with $p = n^{-\epsilon}$ for any fixed $\epsilon \in (0, 1)$.
  - *Stochastic block models* with expected degrees of order $n^{1-\epsilon}$.
  - *Graphon models* with expected degrees of order $n^{1-\epsilon}$.

Ongoing Direction I: Beyond Controlled Codegree

Current limitation:

- Theorem (Upper bound) relies on a controlled codegree condition.
- Key technical step: showing $\mathrm{Cov}\left[T_i, T_j\right]$ is small.

Intuition:

- $T_i = 1$ occurs either when:
  1. $S_i = 1$, or
  2. $\mathcal{N}_i$ is covered by neighbors with $S_j = 1$.
- When codegree is small, the second event has low probability, keeping covariance small.
- This argument can fail when codegree is large.

Self-Normalization of the HT Estimator

Key observation:

- Even when T_i and T_j are highly correlated, the HT estimator may self-normalize.

Extreme example:

$$\hat{\tau}_{\mathrm{HT}}(W) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\mathbb{E}[T_i]} - \frac{C_i Y_i}{\mathbb{E}[C_i]}\right)$$

Twice the original size

- If i and j have identical neighborhoods,

$$\mathcal{N}_i \cup \{i\} = \mathcal{N}_j \cup \{j\},$$

  then T_i and T_j are perfectly correlated.
- However, $\mathbb{E}[T_i]$ and $\mathbb{E}[T_j]$ are also substantially larger.
- The increased HT denominators offset the increased correlation.

Question:

- Can we analyze regimes between small codegree and perfect overlap?

Ongoing Direction II: Adaptive Designs Based on Graph Structure

Motivating observations:

- If codegree is large, then $\lambda(G^2)$ is close to $d_{\max}(G)$.
- In this regime, the conflict graph design achieves a rate $\lambda(G^2)/n$, which is near-optimal.
- If codegree is small, our proposed design achieves the optimal rate.

Question:

- Should we allow the design to select different assignment strategies depending on local graph features?
- Structural quantities such as codegree may guide the choice of design.

Ongoing Direction III: From $\text{HT}^{(1)}$ to the HT Estimator

- Currently, the $\text{HT}^{(1)}$ estimator includes only one of the two components in the HT estimator.
- Can we combine it with the $\text{HT}^{(0)}$ counterpart and its corresponding design to obtain the full HT estimator?

Summary

- We study experimental design for estimating the global average treatment effect under network interference.
- The Horvitz–Thompson estimator is unbiased under any design, but its variance depends critically on the assignment strategy.
- We establish a minimax lower bound showing that, for any design,

$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \mathrm{Var}_{W\sim\mathcal{E}}\left[\hat{\tau}_{\mathrm{HT}}(W)\right] \gtrsim \frac{d_{\max}(G)}{n}.$$

- We propose a simple, locally dependent design that protects the denominators in $\mathrm{HT}^{(1)}$.
- For pseudo-random graphs with controlled codegree, this design achieves the minimax rate for the $\mathrm{HT}^{(1)}$ estimator up to polylogarithmic factors.

Summary

- We study experimental design for estimating the global average treatment effect under network interference.
- The Horvitz–Thompson estimator is unbiased under any design, but its variance depends critically on the assignment strategy.
- We establish a minimax lower bound showing that, for any design,

$$\sup_{\{Y_i(\cdot)\}_{i=1}^n \in \mathcal{Y}_n} \mathrm{Var}_{W\sim\mathcal{E}}\left[\hat{\tau}_{\mathrm{HT}}(W)\right] \gtrsim \frac{d_{\max}(G)}{n}.$$

- We propose a simple, locally dependent design that protects the denominators in $\mathrm{HT}^{(1)}$.
- For pseudo-random graphs with controlled codegree, this design achieves the minimax rate for the $\mathrm{HT}^{(1)}$ estimator up to polylogarithmic factors.