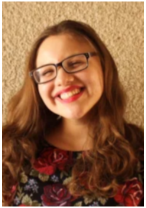
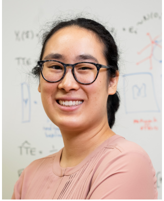
Simple yet Efficient Estimators for Network Causal Inference Even When the Network is Unknown

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu

May 24, 2022





Our Goal

Predict the effect of exposing an entire population to a treatment

- How will a national advertising campaign influence sales?
- How will a new website feature impact user engagement?
- Will proposed public health measures mitigate the spread of a disease?

https://www.cbia.com/resources/coronavirus/reopen-connecticut/covid-19-poster-slow-the-spread/

Total Treatment Effect

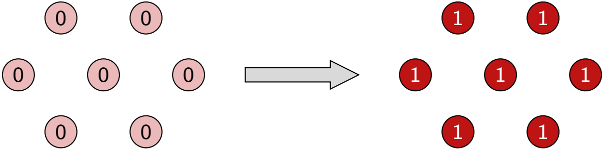
Population: Individuals $\{1, \ldots, n\}$

Treatment: $\mathbf{z} \in \{0,1\}^n$, $1 =$ treated , $0 =$ untreated

Outcomes: Each individual i has function $Y_i : \{0,1\}^n \to \mathbb{R}$

The *Total Treatment Effect* is the average effect on an individual's outcome if entire population is treated.

$$TTE = \frac{1}{n} \sum_{i=1}^{n} \Big(Y_i(\mathbf{1}) - Y_i(\mathbf{0}) \Big).$$



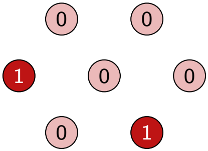
Estimating TTE

We wish to estimate TTE after treating a small fraction of individuals

- Experimental trials can be costly or resource intensive
- Treated individuals could see permanent worsened outcome

Classical Approach:

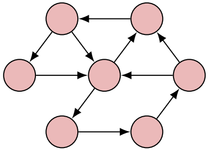
- Treat a random subset of individuals, and scale appropriately
- Difference of means estimator
- Relies on SUTVA: each Y_i is a function only of z_i

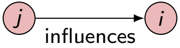


Incorporating Network Effects

Real-world applications include interactions between individuals.

- People purchase goods based on recommendations from friends
- Engagement on a social platform is affected by how engaged one's connections are
- Public health policy in one community affects disease transmission to neighboring communities





Network effects violate SUTVA, introduce bias into classical estimators.

Potential Outcomes with Network Effects

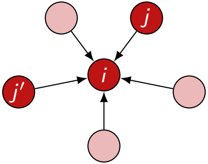
Neighborhood Interference:

Y_i is a function of $\{z_j : j \in \mathcal{N}_i\}$

Low Degree (β):

Subsets $\mathcal{S} \subseteq \mathcal{N}_i$ of treated neighbors ($|\mathcal{S}| \leq \beta$) realize effects $c_{i,\mathcal{S}}$ in Y_i

$$Y_i = c_{i,\varnothing} + \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ 1 \leq |\mathcal{S}| \leq \beta}} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j$$



$\beta = 1$: Linear Heterogeneous Outcomes Model
$\beta = |\mathcal{N}_i|$: Completely General

Techniques for Network Causal Inference

Existing Approaches:

Difference of Means: Biased

OLS: Works well for uniform treatment effects $c_{i,\mathcal{S}}$, otherwise biased

Horwitz-Thompson/Hájek: Unbiased, but require treatment / non-treatment of entire neighborhoods (clustering helps, but is costly)

Our Contributions:

Known Network Structure: Unbiased estimator that performs well under *Bernoulli randomized design*

Unknown Network Structure: Unbiased estimator under *staggered rollout Bernoulli design* based on polynomial extrapolation

Graph-Aware Estimator

$$\widehat{TTE} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{T}\subseteq\mathcal{N}_i \\ |\mathcal{T}|\leq\beta}} f_{|\mathcal{T}|} \prod_{k\in\mathcal{T}} \left(\frac{z_k}{p} - \frac{1-z_k}{1-p}\right), \qquad f_{|\mathcal{T}|} = (1-p)^{|\mathcal{T}|} - (-p)^{|\mathcal{T}|}$$

Theorem

$\widehat{TTE}$ is unbiased with variance $O\left(\frac{d^2 Y_{\max}^2}{n} \max\left(\left(\frac{2\beta}{p}\right)^{2\beta}, \left(\frac{d}{p(1-p)}\right)^{\beta}\right)\right)$.

Ingredients for Unbiasedness:

- Lots of linearity of expectation
- Identity $\mathbb{E}\left[\prod_{j\in\mathcal{S}} z_j \prod_{k\in\mathcal{T}} \left(\frac{z_k}{p} - \frac{1-z_k}{1-p}\right)\right] = \mathbb{I}(\mathcal{T}\subseteq\mathcal{S}) \cdot p^{|\mathcal{S}|-|\mathcal{T}|}$
- Binomial theorem (using $f_{|\mathcal{T}|}$s) to remove $c_{i,\varnothing}$

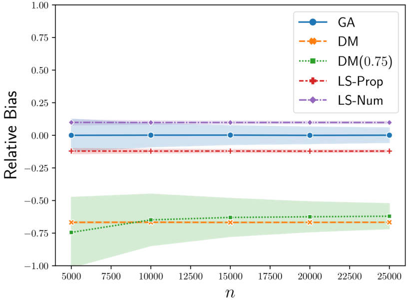
Graph-Aware Estimator in Practice

Erdős-Rényi directed networks with varying n, edge probability $\frac{10}{n}$

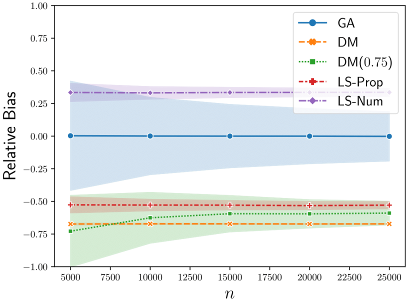
Randomly sampled $\{c_{i,\mathcal{S}}\}$: ~33% magnitude from direct effects $\{c_{i,\{i\}}\}$

Overall treatment probability $p = 0.2$

Linear



Quadratic



Graph-Agnostic Estimator

Staggered-Rollout: Assign treatment to individuals over $\beta+1$ stages: $\mathbf{z}^0 \preceq \mathbf{z}^1 \preceq \ldots \preceq \mathbf{z}^\beta$, with $\mathbf{z}^t \sim \mathrm{Bern}(p_t)$

Measure outcomes $Y_i(\mathbf{z}^t)$ after each stage

Recast estimation as polynomial extrapolation problem:

Learn $F(p) = \mathop{\mathbb{E}}_{\mathbf{z}\sim\mathrm{Bern}(p)}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\right]$ to calculate $\mathrm{TTE} = F(1) - F(0)$

Estimator based on Lagrange interpolation:

$$\widehat{TTE} := \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{\beta}\Big(\ell_t(1) - \ell_t(0)\Big)\cdot Y_i(\mathbf{z}^t), \qquad \ell_t(p) = \prod_{\substack{s=0 \\ s\neq t}}^{\beta} \frac{p-p_s}{p_t-p_s}$$

Theorem

When $p_t = \frac{tp}{\beta}$, $\widehat{TTE}$ is unbiased with variance $O\left(\frac{d^2\beta^2}{n}\cdot Y_{\max}^2\cdot\left(\frac{\beta}{p}\right)^{2\beta}\right)$.

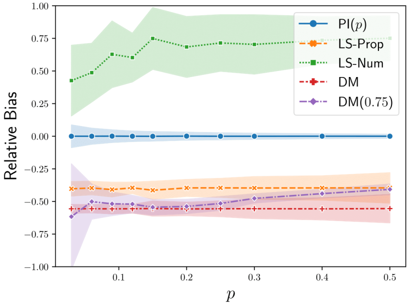
Graph-Agnostic Estimator in Practice

Random directed network on 5000 individuals using configuration model

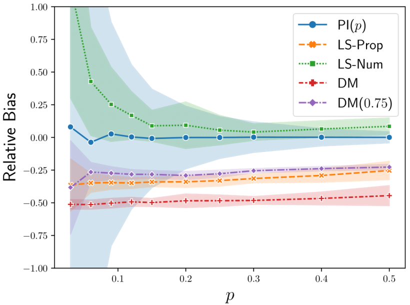
- In-degrees distributed according to power law ($\alpha = 2.5$)
- Out-degrees are consistent

Randomly sampled $\{c_{i,\mathcal{S}}\}$: ∼44% magnitude from direct effects $\{c_{i,\{i\}}\}$

Linear



Quadratic



Thank You!