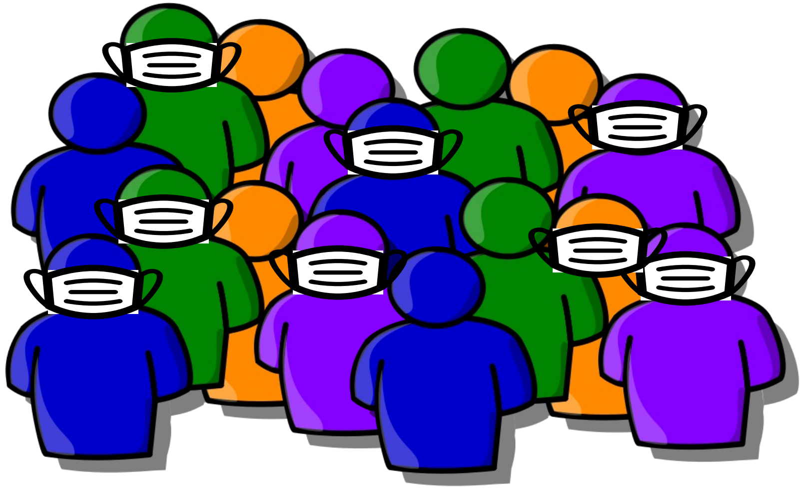
Efficiently Exploiting Model Structure in Network Causal Inference with and without Knowledge of the Network

Christina Lee Yu
Cornell University

Joint work with Edoardo Airoldi, Christian Borgs, Jennifer Chayes, Mayleen Cortez, and Matthew Eichhorn

Q: How to estimate effect of mask-wearing?

observed mask wearing behavior



measure health outcomes

Naïve: estimate average difference in health outcomes between mask wearing and not mask wearing.

Issue 1: Confounding – ppl who choose to wear masks have different susceptibility and preexisting conditions

Issue 2: Bias – Individual outcomes are affected by actions of others

Q: Evaluate social media newsfeed algorithm?

Naïve: Randomly assign small fraction of users to new algorithm and compare performance

Issue: Individuals behavior on social media platform have influence on each other

Causal Inference Setup

- Want to evaluate "effect" of a proposed treatment vs control on a population of size n which is connected via some network
- Potential Outcomes function specifies all hypothetical outcomes

$$Y: \{0,1\}^n \to \mathbb{R}^n$$

- Causal estimand: difference between outcomes under diff treatments

$$\text{Total Treatment Effect (TTE)} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i(\mathbf{1}) - Y_i(\mathbf{0})\right)$$

- Challenge: we only observe data from a single treatment vector

Key Assumptions

- Classical work assumes Stable Unit Treatment Value Assumption (SUTVA), individual's outcome only depends on individual's own treatment

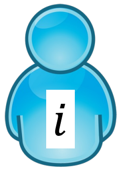
$$Y_i: \{0,1\} \to \mathbb{R}$$

- We assume neighborhood interference, an individual's outcome only depends on treatment of its neighborhood

$$Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}') \text{ if } z_{\mathcal{N}_i} = z'_{\mathcal{N}_i}, \text{ i.e. } Y_i: \{0,1\}^{\mathcal{N}_i} \to \mathbb{R}$$

- Assume treatments are assigned randomly to avoid confounding
  "randomized design" refers to the distribution of the treatment vector $\boldsymbol{z}$

A/B Testing Experiment under SUTVA



Randomly assign

A ↙ ↘ B

$z_i = 1$ treatment ↓ $Y_i(1)$

$z_i = 0$ control ↓ $Y_i(0)$

Assumes i's outcome only depends on z_i
"Stable Unit Treatment Value Assumption" (SUTVA)

Estimand: Total Treatment Effect

$$\text{TTE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i(\mathbf{1}) - Y_i(\mathbf{0})) = \frac{1}{n}\sum_{i\in[n]} Y_i(1) \; - \; \frac{1}{n}\sum_{i\in[n]} Y_i(0)$$

Difference in Means Estimator

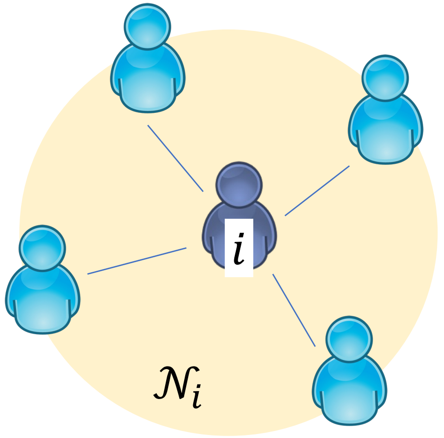
A B

$$\widehat{\text{TTE}} = \frac{1}{|S_A|}\sum_{i\in S_A} Y_i(z_i) - \frac{1}{|S_B|}\sum_{i\in S_B} Y_i(z_i)$$

Relies on SUTVA and randomization

Challenge under Network Interference

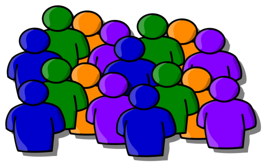
Neighborhood Interference



$Y_i(z_{\mathcal{N}_i})$ denotes i's outcome

Total Treatment Effect Estimand

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

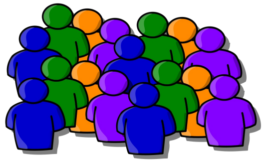


+

treatment

↓

$\{Y_i(\mathbf{1})\}$



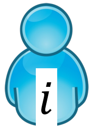
+

control

↓

$\{Y_i(\mathbf{0})\}$

Data collected from A/B test



Randomly assign

treatment ↓ $Y_i(z_{\mathcal{N}_i})$

control ↓ $Y_i(z_{\mathcal{N}_i})$

We may not observe $Y_i(\mathbf{1})$ or $Y_i(\mathbf{0})$!

Naively using difference in means estimator can incur significant bias.

How do we use data collected from A/B test to estimate total treatment effect under interference?

Simple first attempt

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

- Bernoulli design – each $i \in [n]$ assigned to treatment indep w/prob p
- Horvitz-Thompson estimator

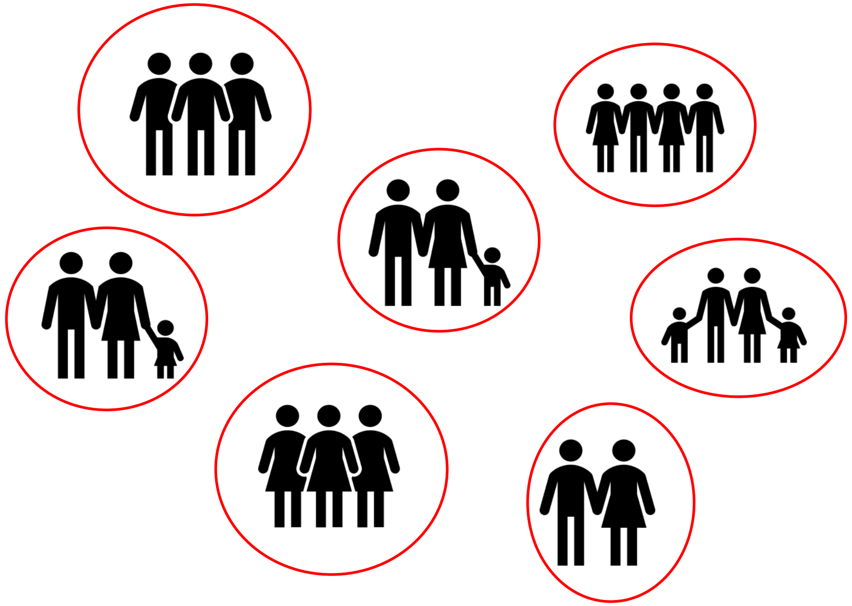
$$\widehat{TTE} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\left(\frac{\mathbb{I}(\mathbf{z}_{\mathcal{N}_i\cup\{i\}} = \mathbf{1})}{\mathbb{P}(\mathbf{z}_{\mathcal{N}_i\cup\{i\}} = \mathbf{1})} - \frac{\mathbb{I}(\mathbf{z}_{\mathcal{N}_i\cup\{i\}} = \mathbf{0})}{\mathbb{P}(\mathbf{z}_{\mathcal{N}_i\cup\{i\}} = \mathbf{0})}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\left(\prod_{j\in\mathcal{N}_i\cup\{i\}} \frac{z_j}{p_j} - \prod_{j\in\mathcal{N}_i\cup\{i\}} \frac{1-z_j}{1-p_j}\right)$$

- Variance under Bernoulli design is $O\left(\frac{Y_{\max}^2 d^2}{np^d}\right)$
- Can we do better?

Brief Literature Review

- “nonparametric” approaches – focus on designing clever designs
  - Depends heavily on graph structure (clusterable)



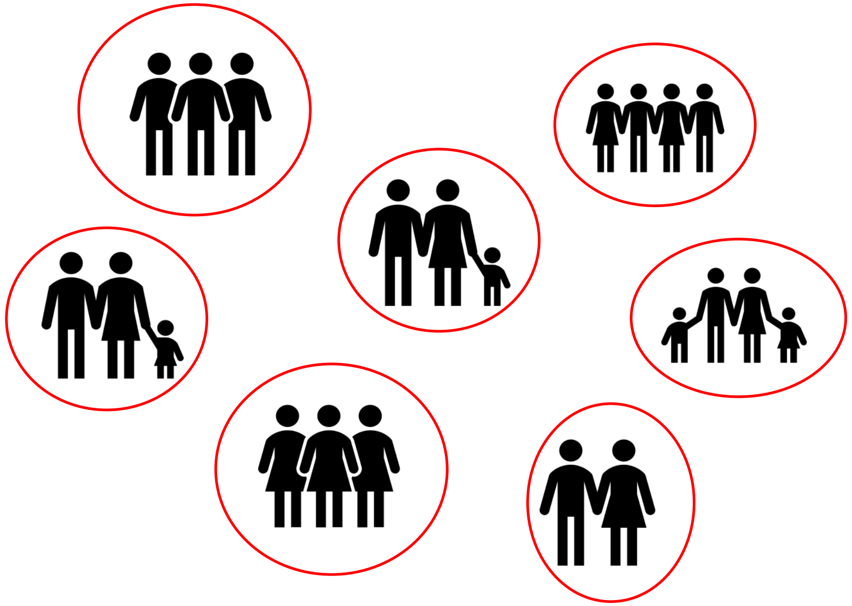
(a) fully disconnected [Sobel06][Rosenbaum07] [HudgensHalloran08][TchetgenVanderWeele12] and more

If randomize over each of the C clusters independently, treating with prob p,

Horvitz-Thompson has variance $O\left(\frac{Y_{\max}^2}{Cp}\right)$

Brief Literature Review

- "nonparametric" approaches – focus on designing clever designs
  - Depends heavily on graph structure (clusterable)
  - Computationally complex randomized designs



(a) fully disconnected [Sobel06][Rosenbaum07] [HudgensHalloran08][TchetgenVanderWeele12] and more

(a) 3-net clustering for restricted-growth graphs [GuiXuBhasinHan15] [EcklesKarrerUgander17] [UganderKarrerBackstromKleinberg13] [Ugander and Yin 20]

Brief Literature Review

- "nonparametric" approaches – focus on designing clever designs
  - Depends heavily on graph structure (clusterable)
  - Computationally complex randomized designs

Variance of Horvitz-Thompson under randomized clustered randomized design for κ-restricted growth graphs is $O\left(\frac{Y_{\max}^2 \kappa^4 d^2}{np}\right)$ [Ugander and Yin 20]

(a) 3-net clustering for restricted-growth graphs
[GuiXuBhasinHan15] [EcklesKarrerUgander17] [UganderKarrerBackstromKleinberg13] [Ugander and Yin 20]

Brief Literature Review

- “nonparametric” approaches – focus on designing clever designs
  - Depends heavily on graph structure (clusterable)
  - Computationally complex randomized designs
- “parametric” approaches – utilizes regression/ML
  - Linear model with respect to known features [ToulisKao13] [GuiXuBhasinHan15] [BasseAiroldi15] [Cai2015] [Parker2016] [Chin2019]

Brief Literature Review

- “nonparametric” approaches – focus on designing clever designs
  - Depends heavily on graph structure (clusterable)
  - Computationally complex randomized designs
- “parametric” approaches – utilizes regression/ML
  - Requires more data than parameters to fully identify model
  - Assumes anonymous interference, imposes strong symmetries in model
  - Fragile to model misspecification, but fewer requirements on randomization
- All previous solutions require (approx) knowledge of network!!
- In nonparametric setting, how can we exploit model structure?

Neighborhood Interference

- We can express the potential outcomes as a polynomial in $\boldsymbol{z}$

$$Y_i(\mathbf{z}) = \sum_{\mathcal{S} \subseteq \mathcal{N}_i} a_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j \prod_{k \in \mathcal{N}_i \setminus \mathcal{S}} (1 - z_k)$$

- Polynomial degree bounded by $|\mathcal{N}_i|$, but what if it were smaller?
- Degree β polynomial potential outcomes model

$$Y_i(\mathbf{z}) = \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \leq \beta}} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j$$

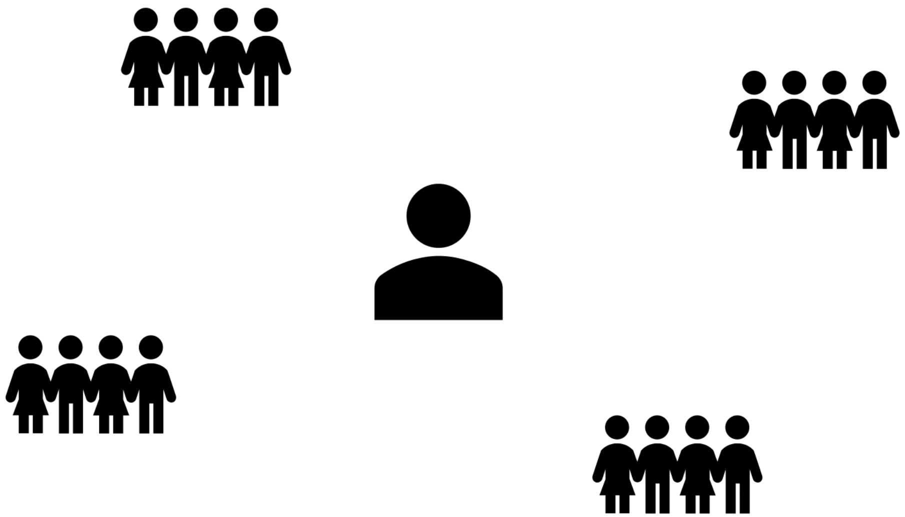
- # parameters scales as nd^{β}, much greater than # observations

Low Order Interactions / Low Degree Polynomial

- Degree β polynomial potential outcomes model

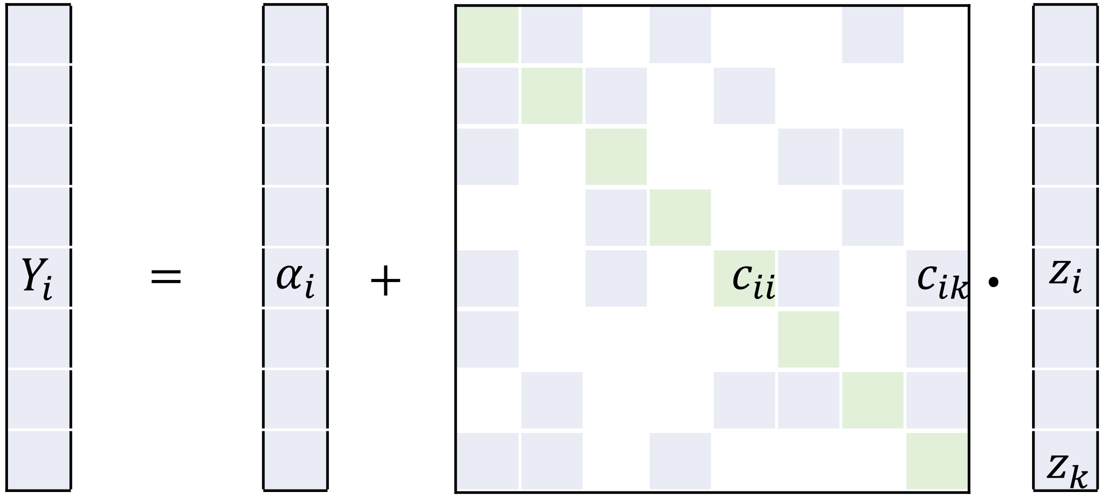
$$Y_i(\mathbf{z}) = \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \leq \beta}} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j$$

- Example: if network effects are additive across subcommunities but could have arbitrary interactions within sub-communities, then degree would be at most max size of subcommunity

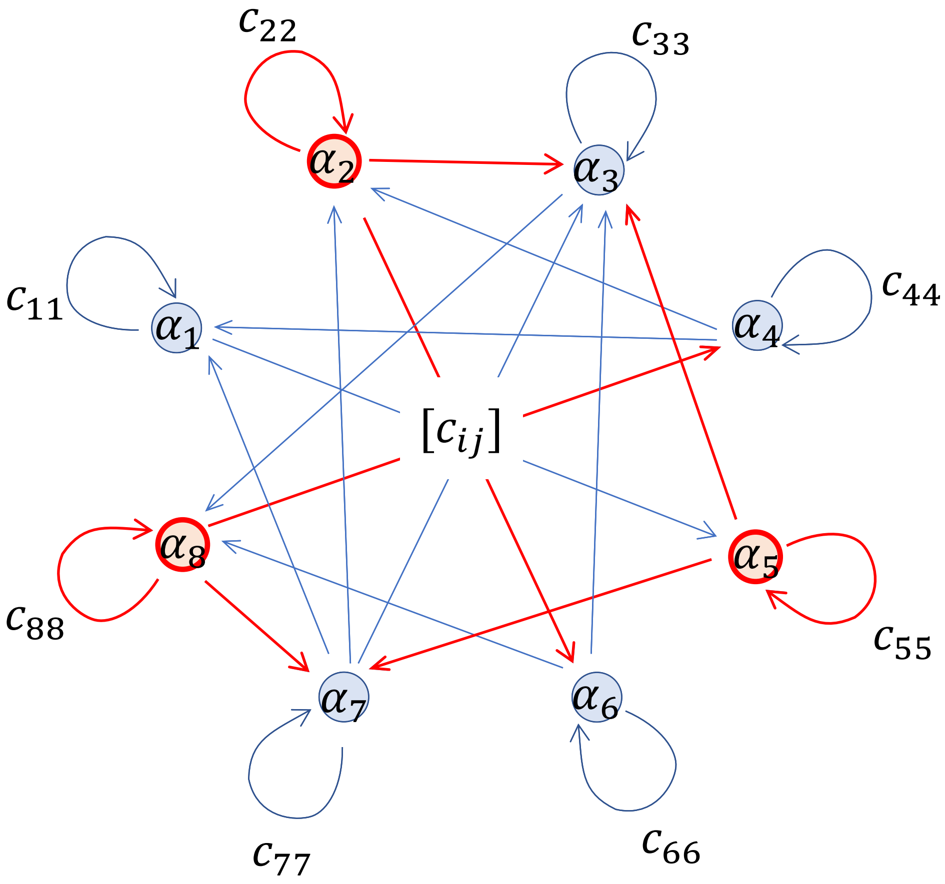


Heterogeneous Additive Network Effects

$$Y_i\left(z_i, z_{\mathcal{N}_i}\right) = \overbrace{\alpha_i}^{c_{i,\emptyset}} + c_{ii} z_i + \overbrace{\sum_{k \in \mathcal{N}_i} c_{ik} z_k}^{\text{Additive network effects}}$$



$$[\boldsymbol{Y}] = [\boldsymbol{\alpha}] + [\boldsymbol{c}] \cdot [\boldsymbol{z}]$$



Heterogeneous Additive Network Effects

$$Y_i(z_i, z_{\mathcal{N}_i}) = \alpha_i + c_{ii}z_i + \sum_{k\in\mathcal{N}_i} c_{ik}z_k$$

- Allows for full heterogeneity in α_i, c_{ii}, c_{ik}, can be positive or negative
- More parameters (2n + #edges) than possible measurements (n)
- Can easily add mean zero independent measurement noise

Exploiting Low Order Interaction

- Present new estimators whose performance is characterized by new complexity measure, i.e. polynomial degree β

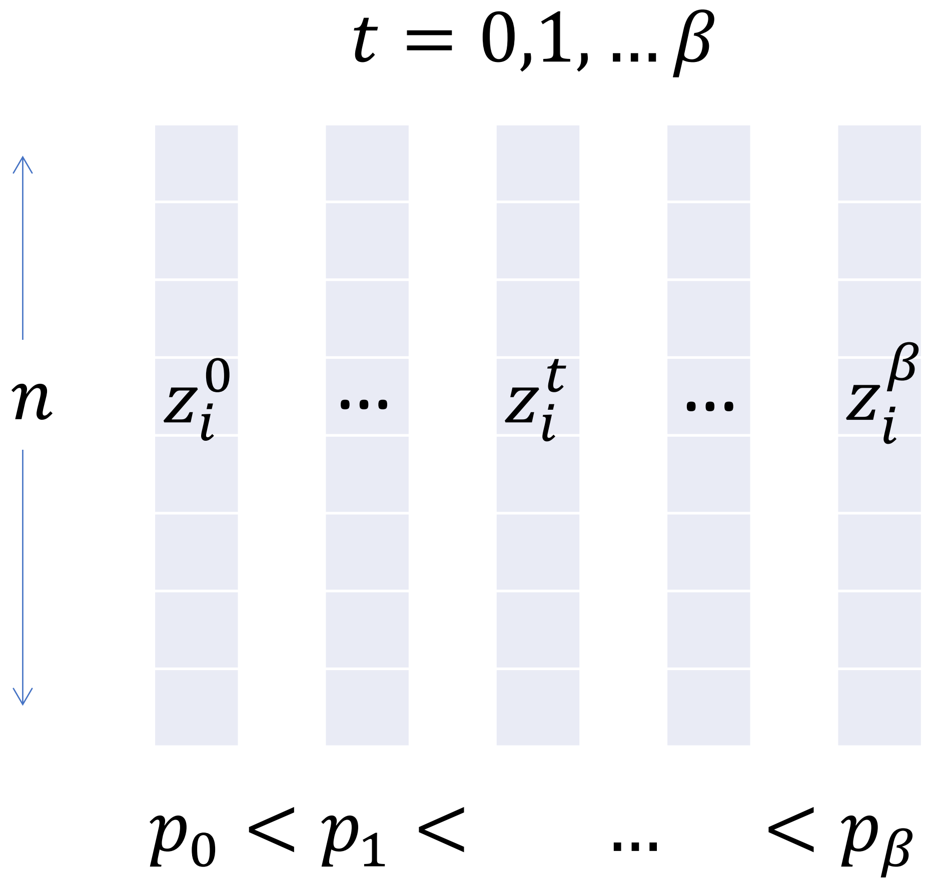
(1) Staggered Rollout Design – richer experimental setup

Enables graph agnostic estimators, but requires uniform treatment probabilities

(2) Bernoulli Design – classical experimental setup

Allows nonuniform treatment, but requires knowledge of graph

Staggered Rollout Bernoulli Design



- Treatments are nested, s.t. $z_i^t < z_i^{t'}$ for $t < t'$
- $\mathbb{E}[z_i^t] = p_t$
- For all $i \neq j, t, t'$, $z_i^t \perp\!\!\!\perp z_j^{t'}$
- Already used in practice (e.g. tech industry, healthcare/medicine)

Result

- If potential outcomes model has polynomial degree β …

$$Y_i(\mathbf{z}) = \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \leq \beta}} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j$$

- Without knowledge of the network, we propose an estimator which is unbiased under a staggered rollout Bernoulli design with variance

$$O\left(\frac{Y_{\max}^2 d^2 \beta^2}{n} \left(\frac{\beta}{p}\right)^{2\beta}\right)$$

- The ability to take multiple measurements (e.g. via a staggered rollout design) enables estimation of TTE without knowledge of the network

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge." Arxiv:2205.14552, 2022.

Reduction to Polynomial Interpolation

- Under a β degree polynomial potential outcomes model and a Bernoulli(p) randomized design, the average outcome is a β degree polynomial with respect to p

$$f(p) = \mathbb{E}\left[\frac{1}{n}\sum_i Y_i(\mathbf{z})\right] = \frac{1}{n}\sum_i \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\leq\beta}} c_{i,\mathcal{S}} \cdot p^{|\mathcal{S}|}$$

- As $TTE = f(1) - f(0)$, problem reduces to polynomial interpolation if we can observe at least $\beta + 1$ different treatment fractions, as could be implemented in a staggered rollout design

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge." Arxiv:2205.14552, 2022.

Result (specialized to linear setting)

Given knowledge of average baselines $\bar{\alpha}$, for any randomized design such that $\mathbb{E}[z_i] = p$ for all $i \in [n]$, the following simple estimator

$$\widehat{TTE} = \frac{1}{p}\left(\frac{1}{n}\sum_{i\in[n]} Y_i(\boldsymbol{z}) - \overbrace{\frac{1}{n}\sum_{i\in[n]} \alpha_i}^{\bar{\alpha}}\right)$$

is an unbiased estimator for any network under the heterogeneous linear outcomes model. Under Bernoulli design, $\text{Var}[\widehat{\text{TTE}}] = O\left(\frac{L^2}{pn}\right)$.

Does not require knowledge of the network!!

"Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure". Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. Arxiv:2205.12803, 2022.

Intuition in linear setting

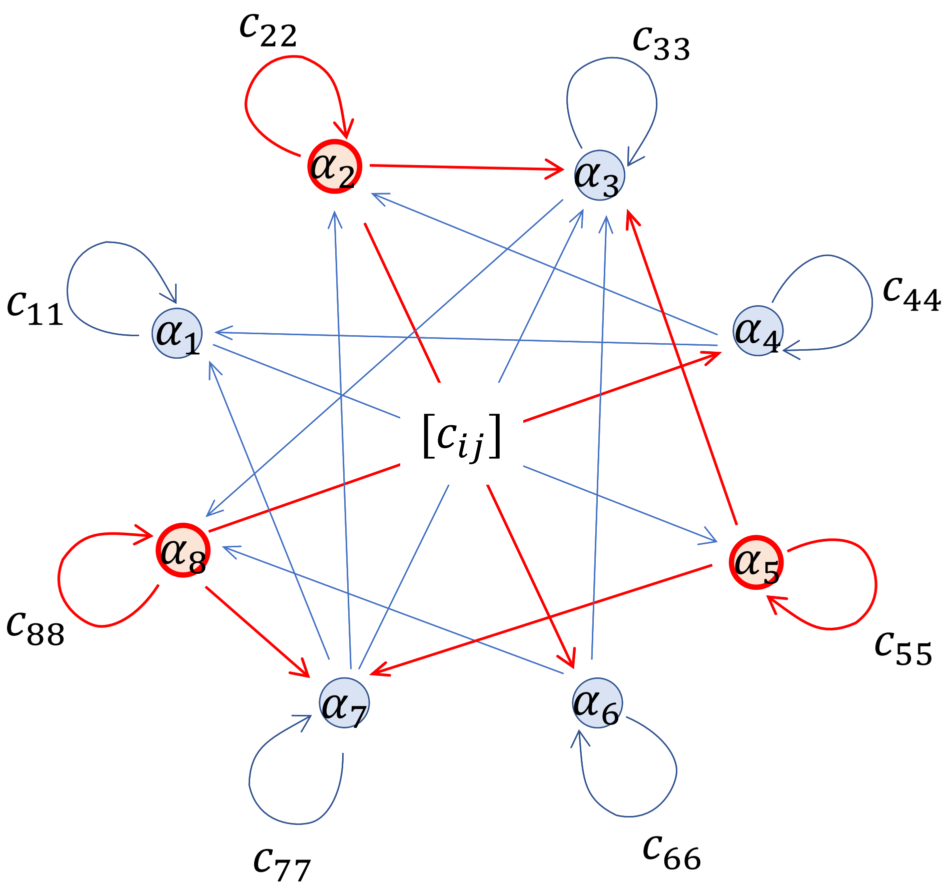
- Total treatment effect equals sum of weighted edges

$$TTE = \frac{1}{n}\sum_i \sum_{k\in[n]} c_{ik}$$

- Treating an individual "activates" its outgoing edges
- Estimator is sum of activated edges

$$\widehat{TTE} = \frac{1}{pn}\sum_{i\in[n]} Y_i(z) - \frac{1}{pn}\sum_{i\in[n]} \alpha_i$$

$$= \frac{1}{pn}\sum_{i\in[n]} \underbrace{\left(\sum_{k\in[n]} c_{ki}\right)}_{\text{"influence" } L_i} z_i$$



"Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure". Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. Arxiv:2205.12803, 2022.

Intuition in linear setting

$$\widehat{TTE} = \frac{1}{pn}\sum_{i\in[n]}\overbrace{\left(\sum_{k\in[n]} c_{ki}\right)}^{L_i} z_i$$

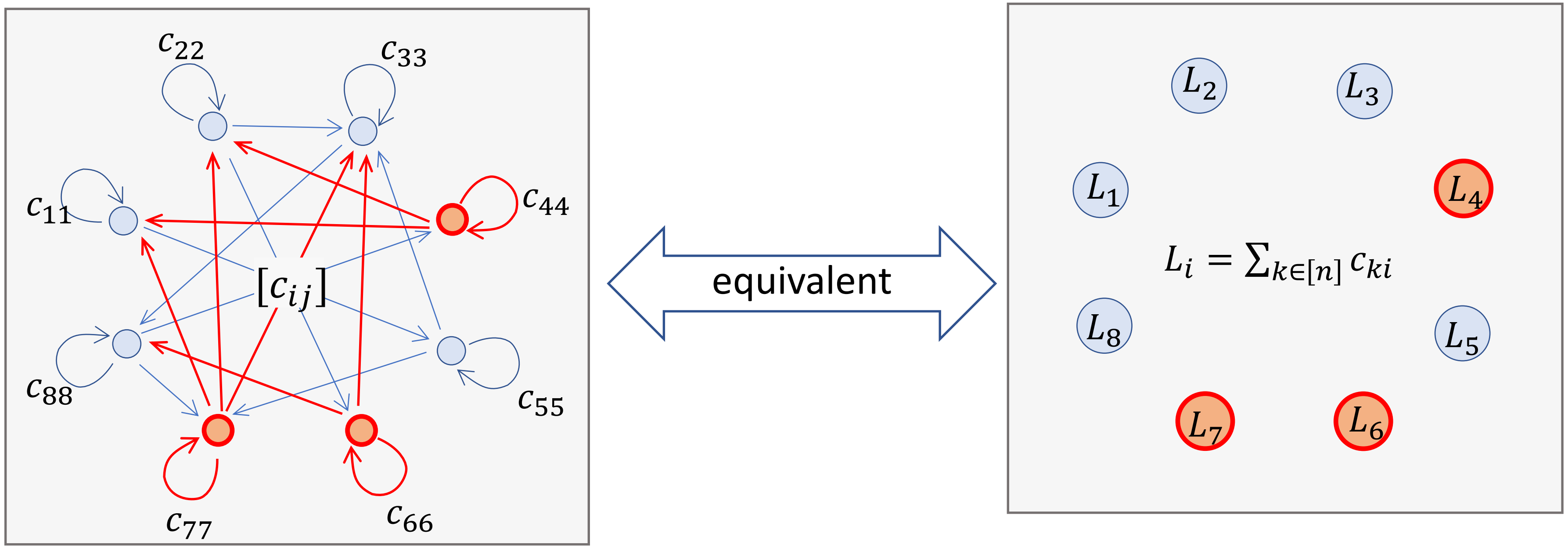
$$TTE = \frac{1}{n}\sum_{i}\left(\sum_{k\in[n]} c_{ik}\right)$$

equivalent

$$\widehat{TTE} = \frac{1}{pn}\sum_{i\in[n]} L_i z_i$$

$$TTE = \frac{1}{n}\sum_{i\in[n]} L_i$$

Given baseline estimates, network causal inference is as easy as estimating population mean!



"Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure". Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. Arxiv:2205.12803, 2022.

Intuition in linear setting

$$\widehat{TTE} = \frac{1}{pn}\sum_{i\in[n]}\overbrace{\left(\sum_{k\in[n]} c_{ki}\right)}^{L_i} z_i$$

$$TTE = \frac{1}{n}\sum_i\left(\sum_{k\in[n]} c_{ik}\right)$$

⟺ equivalent

$$\widehat{TTE} = \frac{1}{pn}\sum_{i\in[n]} L_i z_i$$

$$TTE = \frac{1}{n}\sum_{i\in[n]} L_i$$

Given baseline estimates, network causal inference is as easy as estimating population mean!

- Easy to show unbiasedness, i.e. $\mathbb{E}\left[\widehat{TTE}\right] = TTE$
- Easy to show low variance under simple designs, e.g. for Bernoulli design

$$\mathrm{Var}\left[\widehat{TTE}\right] = \frac{1-p}{pn}\left(\frac{1}{n}\sum_{i\in[n]} L_i^2\right) \approx \frac{\overline{L^2}}{pn}$$

- Approach + guarantees allows for fully dense network
- We can also analyze other randomized designs beyond Bernoulli

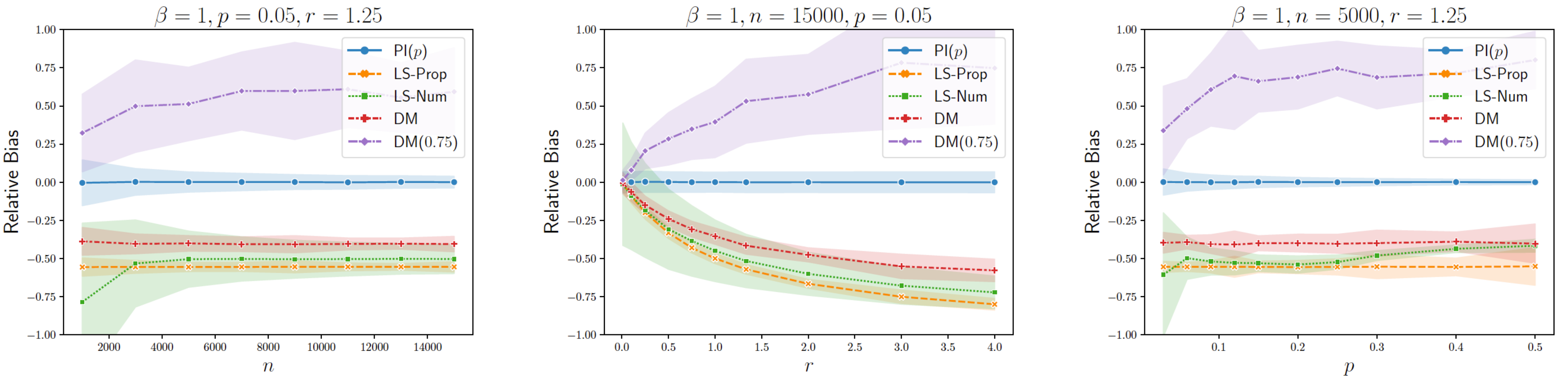
"Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure".
Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. Arxiv:2205.12803, 2022.

Experiment

- Directed graph with power law in-degrees, uniform out-degrees
- β-degree model with weights sampled such that r is average ratio of network effect to direct effect
- Bernoulli staggered rollout over β stages
- Compared our polynomial interpolation estimator against ordinary least squares (OLS) regressing on the fraction or number of neighbors treated, as well as difference in means estimators
- Compare performance when varying population size, ratio of network:direct effects, and overall treatment budget

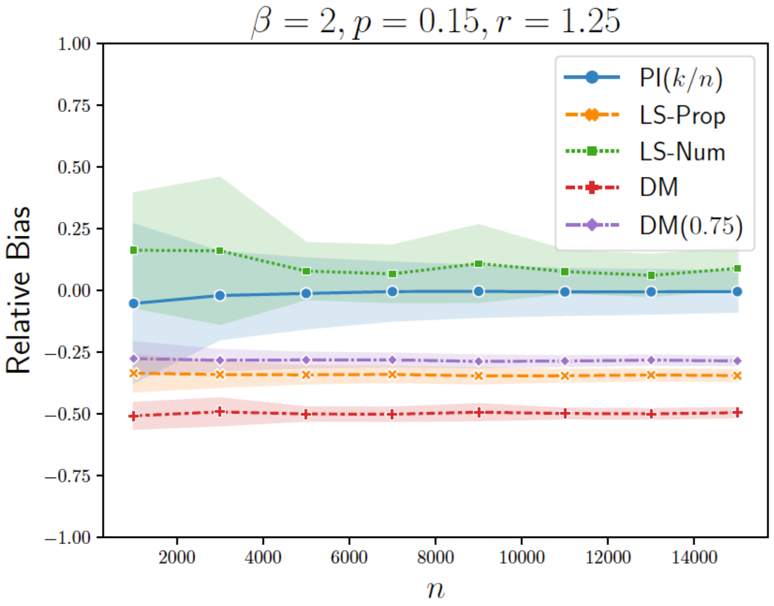
Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge." Arxiv:2205.14552, 2022.

Experiment

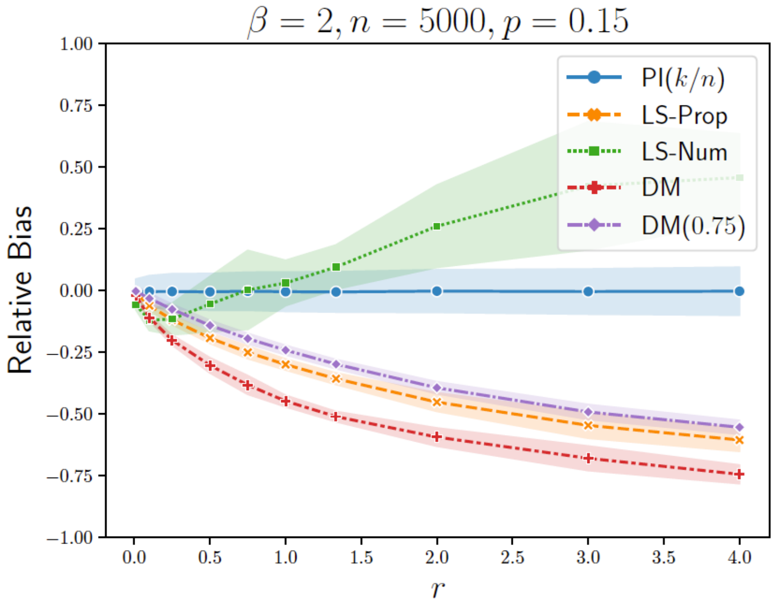


(a) Varying population size (b) Varying direct:indirect effects (c) Varying treatment budget

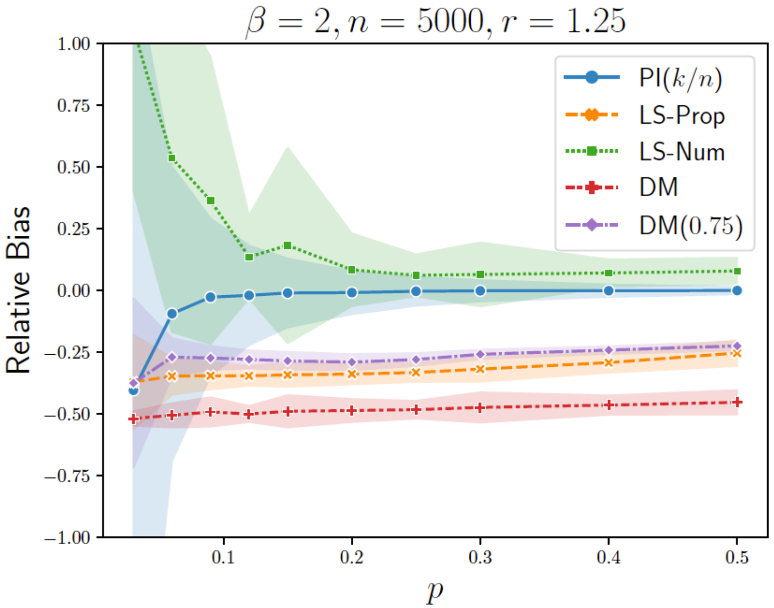
Experiment



(a) Varying population size



(b) Varying direct:indirect effects



(c) Varying treatment budget

Estimators for Low Degree Polynomial Models

- No longer limited to data collected from $\{i: z_{\mathcal{N}_i} = \mathbf{1}\}, \{j: z_{\mathcal{N}_j} = \mathbf{0}\}$
- Cannot use standard regression / ML algorithms

(1) Staggered Rollout Design – richer experimental setup

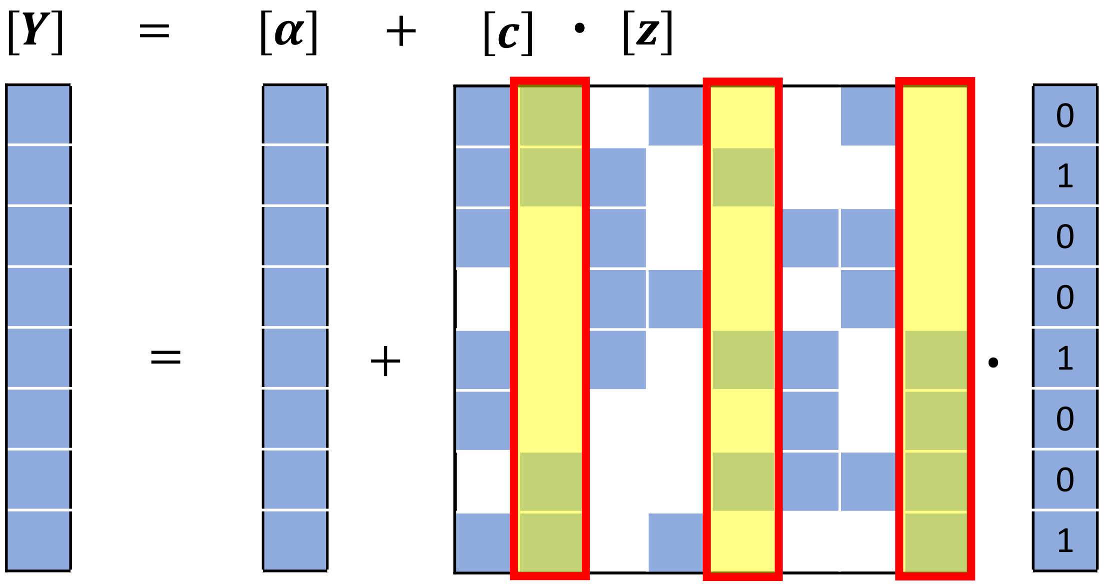
Enables graph agnostic estimators, but requires uniform treatment probabilities

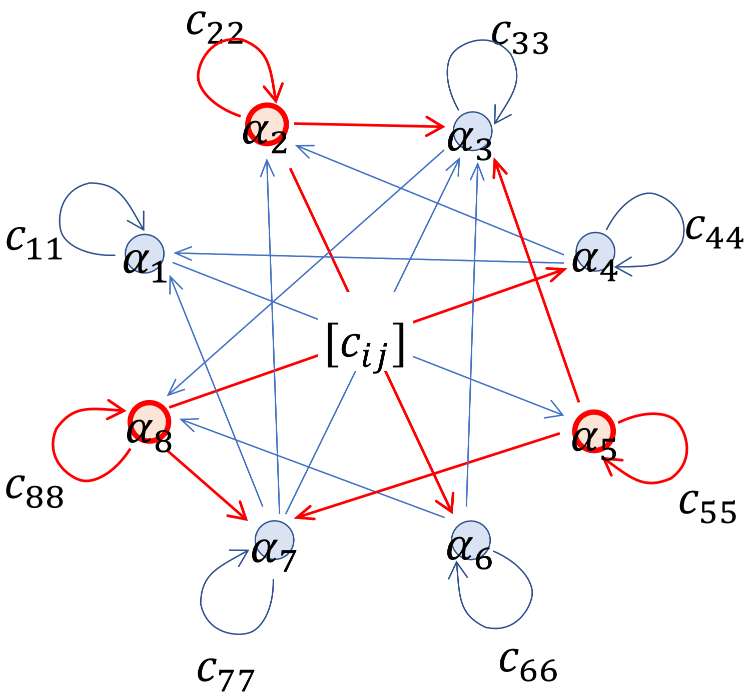
(2) Bernoulli Design – classical experimental setup

Allows nonuniform treatment, but requires knowledge of graph

Linear Model + Bernoulli Design

- Nonuniform Bernoulli design, j treated with prob p_j independently
- Treatments act on outgoing edges, i.e. selects columns of matrix
- Measurements are on incoming edges, i.e. row sums of selected cols





Linear Model + Bernoulli Design

- Nonuniform Bernoulli design, j treated with prob p_j independently

Sum is nonstandard

$$\widehat{TTE} = \frac{1}{n} \sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{j \in \mathcal{N}_i} \left(\frac{z_j}{p_j} - \frac{1 - z_j}{1 - p_j} \right)$$

Similar to inverse probability weights (IPW)

- Estimator is unbiased and $\mathrm{Var}[\widehat{TTE}] = O\left(\frac{Y_{\max}^2 d^3}{np(1-p)} \right)$

"Exploiting neighborhood interference with low order interactions under unit randomized design."
Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. ArXiv:2208.05553, 2022.

Polynomial Model + Bernoulli Design

- Nonuniform Bernoulli design, j treated with prob p_j independently
- Structured Neighborhood Interference - Polynomial Estimator (SNIPE)

$$\widehat{TTE} = \frac{1}{n} \sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \leq \beta}} g(\mathcal{S}) \prod_{j \in \mathcal{S}} \left(\frac{z_j}{p_j} - \frac{1 - z_j}{1 - p_j} \right)$$

$$g(\mathcal{S}) = \prod_{s \in \mathcal{S}} (1 - p_s) - \prod_{s \in \mathcal{S}} (-p_s)$$

Coefficients come from solving for unbiasedness conditions

- Estimator equivalent to Horvitz-Thompson if $\beta > |\mathcal{N}_i|$

"Exploiting neighborhood interference with low order interactions under unit randomized design."
Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. ArXiv:2208.05553, 2022.

Polynomial Model + Bernoulli Design

- Nonuniform Bernoulli design, j treated with prob p_j independently
- Structured Neighborhood Interference - Polynomial Estimator (SNIPE)

$$\widehat{TTE} = \frac{1}{n} \sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{S} \subseteq \mathcal{N}_i \\ |\mathcal{S}| \leq \beta}} g(\mathcal{S}) \prod_{j \in \mathcal{S}} \left(\frac{z_j}{p_j} - \frac{1 - z_j}{1 - p_j} \right)$$

- Given knowledge of local neighborhoods, the SNIPE(β) estimator is unbiased under nonuniform Bernoulli design with variance

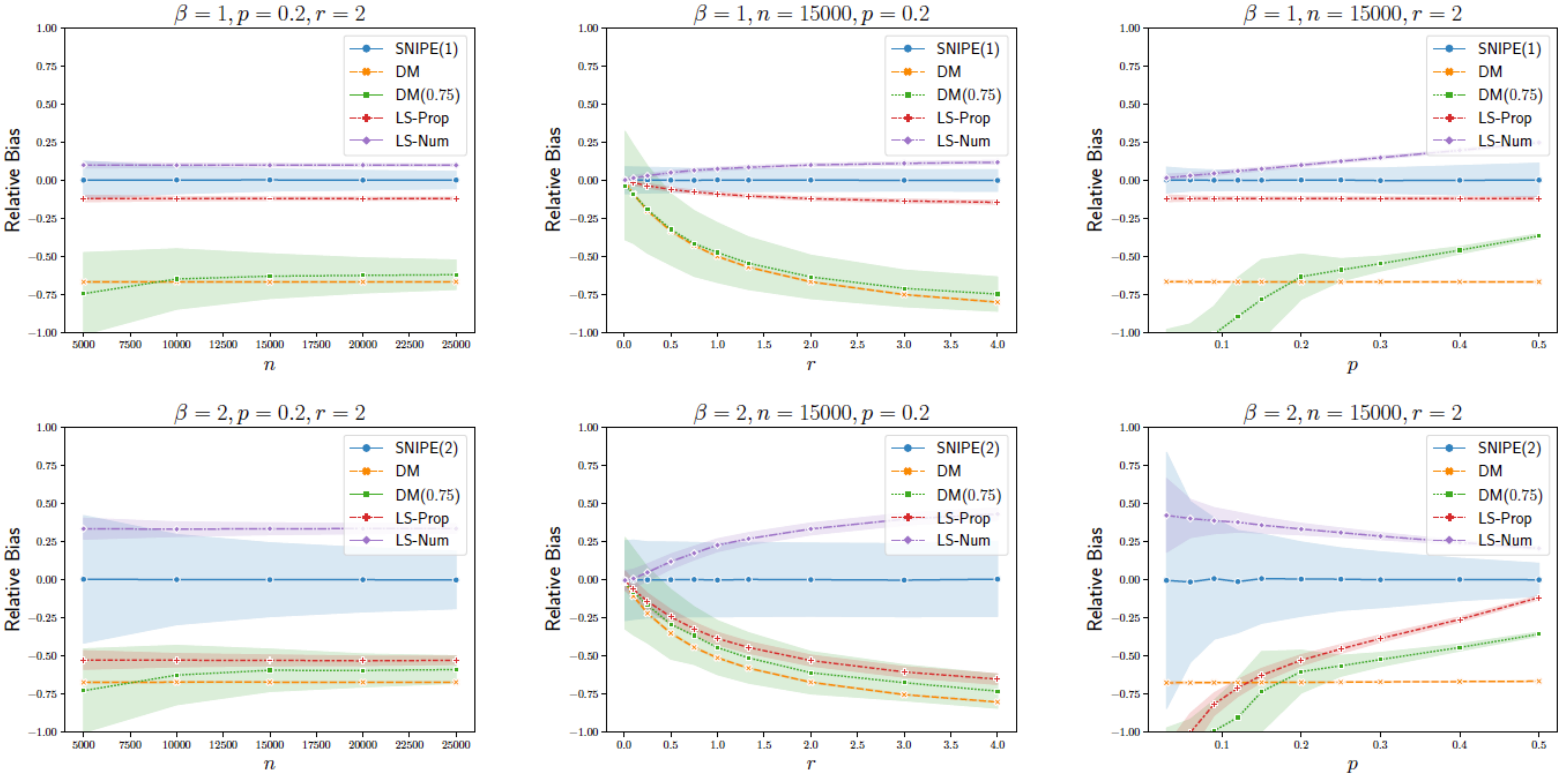
$$O\left(\frac{Y_{\max}^2 d^2}{n} \cdot \left(\frac{d^2}{p(1-p)} \right)^{\beta} \right)$$

"Exploiting neighborhood interference with low order interactions under unit randomized design."
Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. ArXiv:2208.05553, 2022.

Experiment

- Erdos-Renyi graph with constant expected degree of 10
- β-degree model with weights sampled such that r is average ratio of network effect to direct effect
- Bernoulli(p) randomized design
- Compared our SNIPE(β) estimator against ordinary least squares (OLS) regressing on the fraction or number of neighbors treated, as well as difference in means estimators
- Compare performance when varying population size, ratio of network:direct effects, and overall treatment budget

"Exploiting neighborhood interference with low order interactions under unit randomized design."
Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. ArXiv:2208.05553, 2022.



(a) Varying population size (b) Varying direct:indirect effects (c) Varying treatment budget

Conclusion

- New estimators that exploit structure of potential outcomes model to estimate total treatment effect under network interference
- Complexity characterized by polynomial degree of model
- Fully graph agnostic if allowed staggered rollout designs
- Can be extended to unconfounded observational settings

Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. "Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure ". Arxiv:2205.12803, 2022.

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge." Arxiv:2205.14552, 2022.

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Exploiting neighborhood interference with low order interactions under unit randomized design." Arxiv:2208.05553, 2022.

Many Open Questions ….

- Construct estimator for general randomized designs
- Optimizing randomized design using graph structure
- Beyond neighborhood models
- Optimally incorporate covariates
- Handling unobserved confounders
- Computing valid/efficient confidence intervals

Christina Lee Yu, Edo Airoldi, Christian Borgs, and Jennifer Chayes. "Estimating Total Treatment Effect in Randomized Experiments with Unknown Network Structure ". Arxiv:2205.12803, 2022.

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge." Arxiv:2205.14552, 2022.

Mayleen Cortez, Matthew Eichhorn, Christina Lee Yu. "Exploiting neighborhood interference with low order interactions under unit randomized design." Arxiv:2208.05553, 2022.