Causal clustering: design of cluster experiments under network interference

Davide Viviano (Harvard University)
w/ Lihua Lei, Guido Imbens, Brian Karrer, Okke Schrijvers, Liang Shi

November, 2023

Example

Ex1 Informing users on platforms for voting (Bond et al., 2012)

Ex2 Informing farmers exposed to environmental disasters to increase insurance take-up in rural China (Cai et al., 2015)

Example

Ex1 Informing users on platforms for voting (Bond et al., 2012)

Ex2 Informing farmers exposed to environmental disasters to increase insurance take-up in rural China (Cai et al., 2015)

Q: Effect of scaling up the treatment on the pop? (overall effect)

$\Rightarrow$ Challenge: information spillovers betw/ friends

Example

Ex1 Informing users on platforms for voting (Bond et al., 2012)

Ex2 Informing farmers exposed to environmental disasters to increase insurance take-up in rural China (Cai et al., 2015)

Q: Effect of scaling up the treatment on the pop? (overall effect)

$\Rightarrow$ Challenge: information spillovers betw/ friends

$\Rightarrow$ Cluster designs (Karrer et al., 2022; Duflo and Banerjee, 2018)

(i) Researchers partition the population into subgroups (e.g., villages)

(ii) They randomize treatments at the cluster level

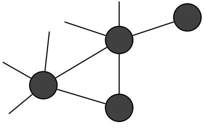
(iii) Estimator as difference in means betw/ treated and controls

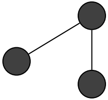
Cluster designs to capture overall effects

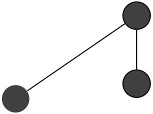
- Many independent clusters to capture overall effects
  - ⇒ Means difference betw/ treated and controls unbiased for overall effect

Cluster designs to capture overall effects

- Many independent clusters to capture overall effects
  - $\Rightarrow$ Means difference betw/ treated and controls unbiased for overall effect



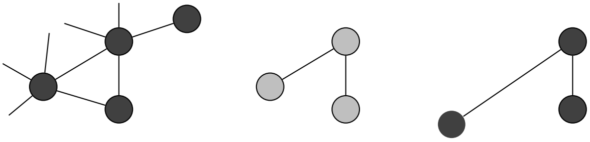




Cluster 1 Cluster 2 Cluster 3

Cluster designs to capture overall effects

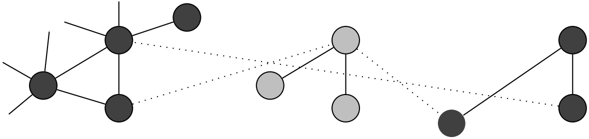
- Many independent clusters to capture overall effects
  - ⇒ Means difference betw/ treated and controls unbiased for overall effect



Cluster 1 Cluster 2 Cluster 3

Cluster designs to capture overall effects

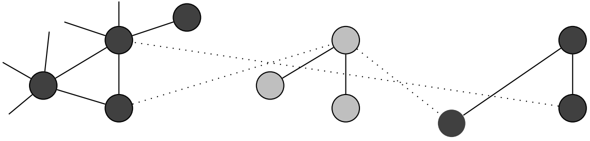
- Many independent clusters to capture overall effects
  - ⇒ Means difference betw/ treated and controls unbiased for overall effect



Cluster 1 Cluster 2 Cluster 3

Cluster designs to capture overall effects

- Many independent clusters to capture overall effects
  - ⇒ Means difference betw/ treated and controls unbiased for overall effect



Cluster 1 Cluster 2 Cluster 3

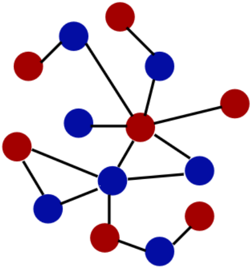
In practice...

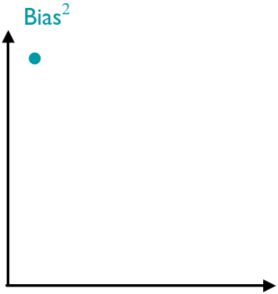
⇒ Not many clusters available (e.g., only one network)

⇒ Uncertainty over spillover (and direct) effects

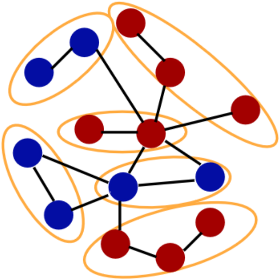
⇒ Trade-offs in the choice and number of the clusters

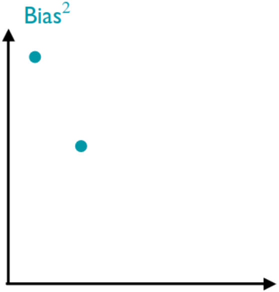
Bias-variance tradeoff



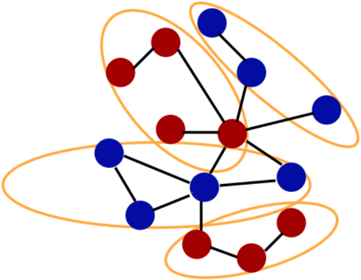


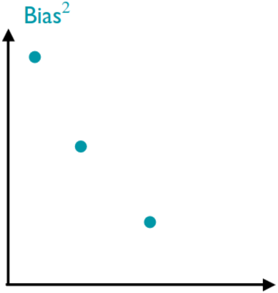
Bias-variance tradeoff



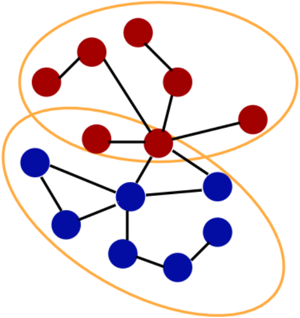


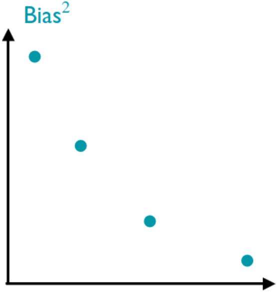
Bias-variance tradeoff



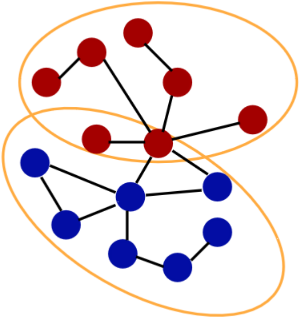


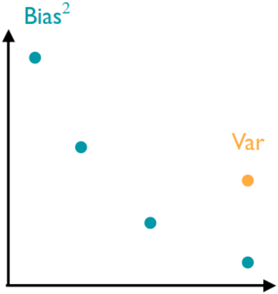
Bias-variance tradeoff



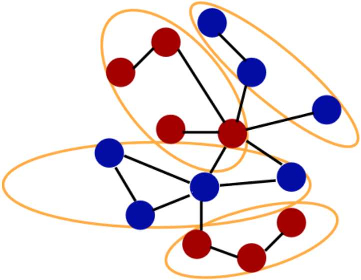


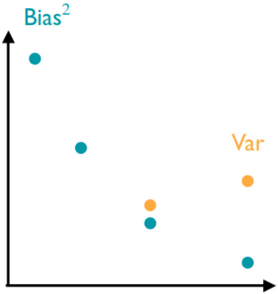
Bias-variance tradeoff



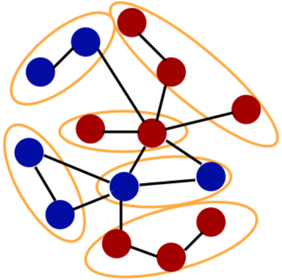


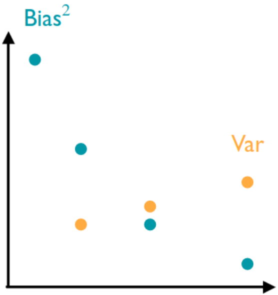
Bias-variance tradeoff





Bias-variance tradeoff





This paper: clustering for global effects estimation

- When should you run a cluster experiment?
  - Novel characterization of worst-case bias and variance
  - Comparison with Bernoulli (*i.i.d.*) design with worst-case MSE

This paper: clustering for global effects estimation

- When should you run a cluster experiment?
  - Novel characterization of worst-case bias and variance
  - Comparison with Bernoulli ($i.i.d.$) design with worst-case MSE
- How should you design the clusters?
  - MSE-Optimal clustering as simple optimization (sd program)
  - Application to unique data from Facebook graph

This paper: clustering for global effects estimation

- When should you run a cluster experiment?
  - Novel characterization of worst-case bias and variance
  - Comparison with Bernoulli (*i.i.d.*) design with worst-case MSE
- How should you design the clusters?
  - MSE-Optimal clustering as simple optimization (sd program)
  - Application to unique data from Facebook graph
- Conceptual contribution: clustering for treatment effect estimation
  - Provide intuitive “cluster purity” measure as bias justification
- Theoretical contribution: model spillovers through local asymptotics
  - Variance only depends on average cluster size variation

Literature

- Experimental design with spillovers [Baird et al., 2018; Basse and Airoldi, 2018; Johari et al. (2020); Viviano (2020); Pouget-Abadie et al. (2019, 2023) ...]
- Inference under interference [Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); Liu and Wager, 2022]
- Inference with clusters [Abadie et al., 2021; Ibragimov and Mueller, 2016; Canay et al., 2017]

Literature

- Experimental design with spillovers [Baird et al., 2018; Basse and Airoldi, 2018; Johari et al. (2020); Viviano (2020); Pouget-Abadie et al. (2019, 2023) ...]
- Inference under interference [Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); Liu and Wager, 2022]
- Inference with clusters [Abadie et al., 2021; Ibragimov and Mueller, 2016; Canay et al., 2017]

  ⇒ None study the problem of optimal cluster designs

Literature

- Experimental design with spillovers [Baird et al., 2018; Basse and Airoldi, 2018; Johari et al. (2020); Viviano (2020); Pouget-Abadie et al. (2019, 2023) ...]
- Inference under interference [Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); Liu and Wager, 2022]
- Inference with clusters [Abadie et al., 2021; Ibragimov and Mueller, 2016; Canay et al., 2017]

  $\Rightarrow$ None study the problem of optimal cluster designs

- Graph clustering: rate optimal cluster size for ε-clustering [Ugander et al., 2013; Eckles et al., 2017; Leung, 2022; Faridani and Niehaus, 2022]

  $\Rightarrow$ Here optimal clustering method + novel bias/variance characterization

Literature

- Experimental design with spillovers [Baird et al., 2018; Basse and Airoldi, 2018; Johari et al. (2020); Viviano (2020); Pouget-Abadie et al. (2019, 2023) ...]
- Inference under interference [Hudgens and Halloran (2008); Aronow and Samii (2017); Leung (2020); Athey et al. (2018); Savje et al. (2021); Liu and Wager, 2022]
- Inference with clusters [Abadie et al., 2021; Ibragimov and Mueller, 2016; Canay et al., 2017]

  $\Rightarrow$ None study the problem of optimal cluster designs

- Graph clustering: rate optimal cluster size for ε-clustering [Ugander et al., 2013; Eckles et al., 2017; Leung, 2022; Faridani and Niehaus, 2022]

  $\Rightarrow$ Here optimal clustering method + novel bias/variance characterization
- Clustering in stat [Von Luxburg, 2007; Newman 2013; Lei et al. (2020), et al.]

  $\Rightarrow$ Here different focus on clustering for causal inference

Content

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n
  - Cluster of unit i : $c(i) \in \{1, \cdots, K_n\}$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n
  - Cluster of unit i : $c(i) \in \{1, \cdots, K_n\}$
  - Cluster level treatment

$$D_i = \tilde{D}_{c(i)} \sim \mathrm{Bern}(1/2)$$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$

- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n
  - Cluster of unit i : $c(i) \in \{1, \cdots, K_n\}$
  - Cluster level treatment

$$D_i = \tilde{D}_{c(i)} \sim \mathrm{Bern}(1/2)$$

  - n_k: number of units in cluster k

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n
  - Cluster of unit i : $c(i) \in \{1, \cdots, K_n\}$
  - Cluster level treatment

$$D_i = \tilde{D}_{c(i)} \sim \mathrm{Bern}(1/2)$$

  - n_k: number of units in cluster k
- Estimand and difference in means estimator:

$$\tau_n = \frac{1}{n}\sum_{i=1}^{n}\Big[Y_i(\mathbf{1}) - Y_i(\mathbf{0})\Big],$$

Setup

- Finite pop: n units connected under adjacency matrix A (observed)
  - Potential outcomes: $Y_i(\mathbf{d}), \mathbf{d} \in \{0,1\}^n$
  - Set of neighbors: $\mathcal{N}_i = \{j : A_{i,j} = 1\}$
  - Local interference: $Y_i(\mathbf{d}) = \mu_i(\mathbf{d}_i, \mathbf{d}_{\mathcal{N}_i}), \mu_i \in \mathcal{M}$
- Cluster experiment: for cluster design $\mathcal{C}_n$
  - Number of clusters: K_n
  - Cluster of unit i : $c(i) \in \{1, \cdots, K_n\}$
  - Cluster level treatment

$$D_i = \tilde{D}_{c(i)} \sim \mathrm{Bern}(1/2)$$

  - n_k: number of units in cluster k
- Estimand and difference in means estimator:

$$\tau_n = \frac{1}{n} \sum_{i=1}^{n} \Big[Y_i(\mathbf{1}) - Y_i(\mathbf{0}) \Big], \quad \hat{\tau}_n = \frac{2}{n} \sum_{i=1}^{n} \Big[Y_i D_i - Y_i(1 - D_i) \Big]$$

Model on spillover effects

- Class of models: for $\mathbf{d}' \in \{\mathbf{1}, \mathbf{0}\}$

$$\sup_{\mu_i(1,\cdot)\in\mathcal{M}_1} \Big|\mu_i(1,\mathbf{d}) - \mu_i(1,\mathbf{d}')\Big| = \frac{\bar{\phi}_n}{|\mathcal{N}_i|} \sum_{i\in\mathcal{N}_i} \Big|\mathbf{d}'_k - \mathbf{d}_k\Big|$$

Model on spillover effects

- Class of models: for $\mathbf{d}' \in \{\mathbf{1}, \mathbf{0}\}$

$$\sup_{\mu_i(1,\cdot)\in\mathcal{M}_1} \left|\mu_i(1,\mathbf{d}) - \mu_i(1,\mathbf{d}')\right| = \frac{\bar{\phi}_n}{|\mathcal{N}_i|}\sum_{i\in\mathcal{N}_i}\left|\mathbf{d}'_k - \mathbf{d}_k\right|$$

Asm Consider local asymptotics $\bar{\phi}_n = o(1)$:

$\Rightarrow$ signal to noise is of order smaller than $n^{1/2}$
$\Rightarrow$ $\bar{\phi}_n \gg n^{-1/2}$ "moderately large" spillover effects
$\Rightarrow$ $\bar{\phi}_n \leq n^{-1/2}$ "small spillover effects

Model on spillover effects

- Class of models: for $\mathbf{d}' \in \{\mathbf{1}, \mathbf{0}\}$

$$\sup_{\mu_i(1,\cdot)\in\mathcal{M}_1} \Big|\mu_i(1,\mathbf{d}) - \mu_i(1,\mathbf{d}')\Big| = \frac{\bar{\phi}_n}{|\mathcal{N}_i|} \sum_{i\in\mathcal{N}_i} \Big|\mathbf{d}'_k - \mathbf{d}_k\Big|$$

Asm Consider local asymptotics $\bar{\phi}_n = o(1)$:

$\Rightarrow$ signal to noise is of order smaller than $n^{1/2}$

$\Rightarrow$ $\bar{\phi}_n \gg n^{-1/2}$ "moderately large" spillover effects

$\Rightarrow$ $\bar{\phi}_n \leq n^{-1/2}$ "small spillover effects

E.g. Write the overall effect as $\tau_n = \Delta_n + \bar{\phi}_n$, for $\Delta_n, \bar{\phi}_n = o(1)$.

E.g. Information experiments [Karrer et al. (2021)]

Content

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2 = \frac{\bar{\phi}_n}{n}$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left| \mathcal{N}_i \bigcap \left\{ j : c(i) \neq c(j) \right\} \right|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n\right] - \tau_n\right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left|\mathcal{N}_i \bigcap \left\{j : c(i) \neq c(j)\right\}\right|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left| \mathcal{N}_i \bigcap \left\{ j : c(i) \neq c(j) \right\} \right|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n \mathcal{N}^2_{n,\max} = o(1)$, $\bar{\psi} = \sup_i \max\{\mu_i(\mathbf{1})^2, \mu_i(\mathbf{0})^2\}$, then

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left| \mathcal{N}_i \bigcap \left\{ j : c(i) \neq c(j) \right\} \right|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n \mathcal{N}_{n,\max}^2 = o(1)$, $\bar{\psi} = \sup_i \max\{\mu_i(\mathbf{1})^2, \mu_i(\mathbf{0})^2\}$, then

$$\sup_{\mu \in \mathcal{M}} \mathrm{Var}\left(\hat{\tau} | \mathbf{A}, \mathcal{C}_n \right) =$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu\in\mathcal{M}}\Big(\mathbb{E}\Big[\hat{\tau}_n|\mathbf{A},\mathcal{C}_n\Big]-\tau_n\Big)^2=\frac{\bar{\phi}_n}{n}\sum_{i=1}^{n}\frac{1}{|\mathcal{N}_i|}\underbrace{\Big|\mathcal{N}_i\bigcap\Big\{j:c(i)\neq c(j)\Big\}\Big|}_{\#i\text{'s friends not in }i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n\mathcal{N}_{n,\max}^2=o(1)$, $\bar{\psi}=\sup_i\max\{\mu_i(\mathbf{1})^2,\mu_i(\mathbf{0})^2\}$, then

$$\sup_{\mu\in\mathcal{M}}\mathrm{Var}\Big(\hat{\tau}|\mathbf{A},\mathcal{C}_n\Big)=\underbrace{\bar{\psi}\sum_{k=1}^{K_n}\frac{n_k^2}{n^2}}_{\text{Variation in clusters' size}}$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n \right] - \tau_n \right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left| \mathcal{N}_i \bigcap \left\{ j : c(i) \neq c(j) \right\} \right|}_{\# i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n \mathcal{N}_{n,\max}^2 = o(1)$, $\bar{\psi} = \sup_i \max\{\mu_i(\mathbf{1})^2, \mu_i(\mathbf{0})^2\}$, then

$$\sup_{\mu \in \mathcal{M}} \mathrm{Var}\left(\hat{\tau} | \mathbf{A}, \mathcal{C}_n \right) = \underbrace{\bar{\psi} \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}}_{\text{Variation in clusters' size}} + o(1/K_n)$$

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \left(\mathbb{E}\left[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n\right] - \tau_n\right)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\left|\mathcal{N}_i \bigcap \left\{j : c(i) \neq c(j)\right\}\right|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n \mathcal{N}_{n,\max}^2 = o(1)$, $\bar{\psi} = \sup_i \max\{\mu_i(\mathbf{1})^2, \mu_i(\mathbf{0})^2\}$, then

$$\sup_{\mu \in \mathcal{M}} \mathrm{Var}\left(\hat{\tau} | \mathbf{A}, \mathcal{C}_n\right) = \underbrace{\bar{\psi} \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}}_{\text{Variation in clusters' size}} + o(1/K_n)$$

$\Rightarrow$ More clusters' size variation imply smaller variance

Worst-case variance/bias?

Lem Worst-case outcome model justifies cluster purity:

$$\sup_{\mu \in \mathcal{M}} \Big(\mathbb{E}\Big[\hat{\tau}_n | \mathbf{A}, \mathcal{C}_n\Big] - \tau_n\Big)^2 = \frac{\bar{\phi}_n}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|} \underbrace{\Big|\mathcal{N}_i \bigcap \Big\{j : c(i) \neq c(j)\Big\}\Big|}_{\#i\text{'s friends not in } i\text{'s cluster}}$$

$\Rightarrow$ Clusters' purity justified as estimator's bias!

Lem Let $\bar{\phi}_n \mathcal{N}_{n,\max}^2 = o(1)$, $\bar{\psi} = \sup_i \max\{\mu_i(\mathbf{1})^2, \mu_i(\mathbf{0})^2\}$, then

$$\sup_{\mu \in \mathcal{M}} \mathrm{Var}\Big(\hat{\tau} | \mathbf{A}, \mathcal{C}_n\Big) = \underbrace{\bar{\psi} \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}}_{\text{Variation in clusters' size}} + o(1/K_n)$$

$\Rightarrow$ More clusters' size variation imply smaller variance

Key Bound (i, j) covariances with $\bar{\phi}_n$ for (i, j) not in same cluster

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0$$

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

$\Rightarrow$ Exact threshold in the paper (rule of thumb $\sqrt{K_n}\bar{\phi}_n > 2.3$)

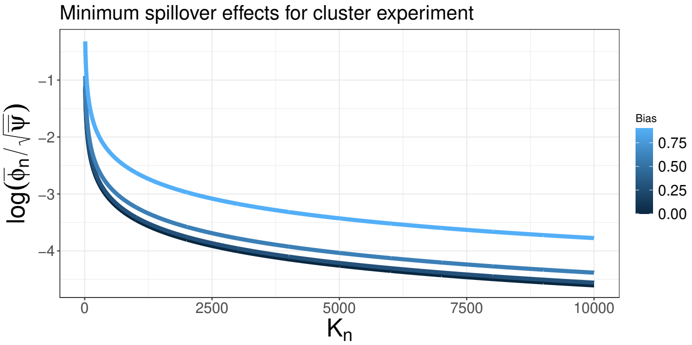
Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

$\Rightarrow$ Exact threshold in the paper (rule of thumb $\sqrt{K_n}\bar{\phi}_n > 2.3$)



Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

Description	Mathematical Formulation	Implication
Spillovers are small and # of clusters is small	$\sqrt{K_n}\bar{\phi}_n = o(1)$	Run Bernoulli design

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

Description	Mathematical Formulation	Implication
Spillovers are non-negligible and # of clusters is large	$\sqrt{K_n}\bar{\phi}_n \to \infty$	Run Cluster design

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

Description	Mathematical Formulation	Implication
Spillovers are small and # of clusters is very large	$K_n \propto n, \bar{\phi}_n \propto n^{-1/3}$	Run Cluster design

Clustering vs Bernoulli design

Thm For $\mathcal{C}_{0,n}$ denoting a Bernoulli design, under regularities

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \geq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to 0$$

$$\lim_{n\to\infty} \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_n) - \sup_{\mu} \mathrm{MSE}_{\mu}(\mathcal{C}_0) \leq 0 \text{ if } \sqrt{K_n}\bar{\phi}_n \to \infty$$

Description	Mathematical Formulation	Implication
Spillovers are not that small and # of clusters is very small	$K_n = \mathcal{O}(1), \bar{\phi}_n = o(1)$	Run Bernoulli design

Content

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto$$

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|\mathcal{N}_i|}\Big|j \in \mathcal{N}_i : c(i) \neq c(j)\Big|\Big)^2 +$$

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|\mathcal{N}_i|}\Big|j \in \mathcal{N}_i : c(i) \neq c(j)\Big|\Big)^2 + \xi_n \sum_{k=1}^{K_n}\frac{n_k^2}{n^2}$$

Objective function for optimal clustering

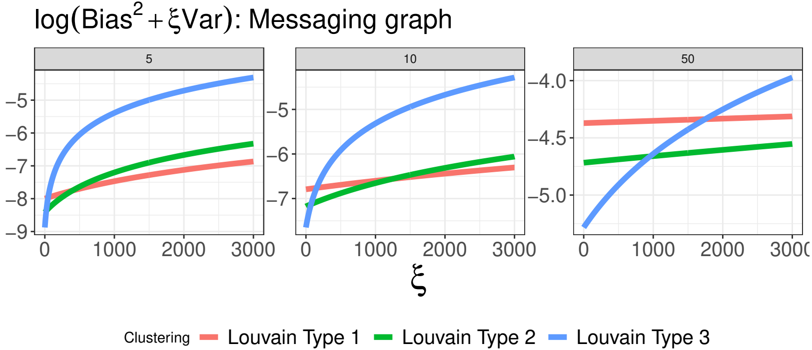
Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \Big(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|}\Big|j \in \mathcal{N}_i : c(i) \neq c(j)\Big|\Big)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \Big(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\mathcal{N}_i|}\Big|j \in \mathcal{N}_i : c(i) \neq c(j)\Big|\Big)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$



Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \mathrm{Bias}_n(\mathcal{C}_n)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \mathrm{Bias}_n(\mathcal{C}_n)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

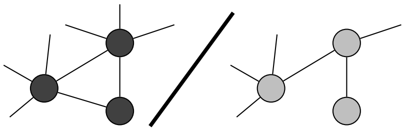
Idea Focus on (surrogate) $|\mathrm{Bias}_n(\mathcal{C}_n)| + \xi_n' \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}$

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \mathrm{Bias}_n(\mathcal{C}_n)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

Idea Focus on (surrogate) $|\mathrm{Bias}_n(\mathcal{C}_n)| + \xi_n' \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}$

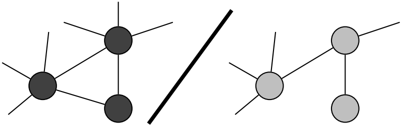


Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \mathrm{Bias}_n(\mathcal{C}_n)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

Idea Focus on (surrogate) $|\mathrm{Bias}_n(\mathcal{C}_n)| + \xi_n' \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}$



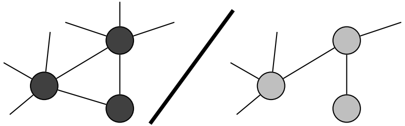
Thm Let $\mathbf{M}_{c,K}^{(i,k)} = 1\{c(i) = k\}$.

Objective function for optimal clustering

Thm For any clustering $\mathcal{C}_n$, for $\bar{\phi}_n = o(1), \xi_n = \frac{\bar{\psi}}{\bar{\phi}_n^2}$,

$$\sup_{\mu} \mathrm{MSE}_n(\mathcal{C}) \propto \mathrm{Bias}_n(\mathcal{C}_n)^2 + \xi_n \sum_{k=1}^{K_n} \frac{n_k^2}{n^2} + o(1/K_n)$$

Idea Focus on (surrogate) $|\mathrm{Bias}_n(\mathcal{C}_n)| + \xi_n' \sum_{k=1}^{K_n} \frac{n_k^2}{n^2}$



Thm Let $\mathbf{M}_{c,K}^{(i,k)} = 1\{c(i) = k\}$. For $K_n = K$, optimization is equivalent to

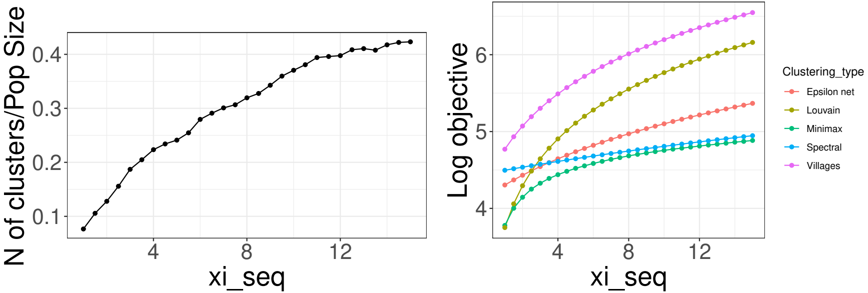
$$\max_c \mathrm{tr}\Big((\mathbf{L} - \frac{\xi_n'}{n}\mathbf{1}_n\mathbf{1}_n)\mathbf{M}_{c,K}\mathbf{M}_{c,K}^\top\Big), \quad \mathbf{L} = \mathrm{diag}(\mathbf{A}\mathbf{1}_n)^{-1}\mathbf{A}$$

Implementation

- Choose $n\xi_n'$ (e.g., 3.29)
- Solution for sequence of K_n
- Trace-optimization solved via SDP $\mathbf{M}_{c,K}\mathbf{M}_{c,K}^\top$ + k-means
- Compute best K_n and clustering by comparing objectives

Implementation

- Choose $n\xi_n'$ (e.g., 3.29)
- Solution for sequence of K_n
- Trace-optimization solved via SDP $\mathbf{M}_{c,K}\mathbf{M}_{c,K}^{\top}$ + k-means
- Compute best K_n and clustering by comparing objectives



Content

Conclusions

- Novel *causal* clustering algorithm motivated by ATE estimation
- Bias/Variance motivate measures of cluster purity/variation
- Comparisons Cluster vs Bernoulli design
- Algorithm to compute optimal clustering
- Application using data from Meta (and field experiment)

Conclusions

- Novel *causal* clustering algorithm motivated by ATE estimation
- Bias/Variance motivate measures of cluster purity/variation
- Comparisons Cluster vs Bernoulli design
- Algorithm to compute optimal clustering
- Application using data from Meta (and field experiment)

Thanks very much, questions? :)

Worst-case variance

Worst case variance depends on

$$\text{Var}\left(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\right) = \frac{4}{n^2} \sum_{i,j} \underbrace{\text{Cov}\left(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\right)}_{=\sigma_{i,j}}$$

hmm... and now?

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2} \sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2} \sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends,

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2} \sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big)$$

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) \leq \underbrace{\bar{\phi}_n/n^2}_{\text{Bias}} \times$$

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i,j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) \leq \underbrace{\bar{\phi}_n/n^2}_{\text{Bias}} \times \mathcal{O}\Big(\underbrace{n \times \mathcal{N}_{n,\max}^2}_{\#\text{ 2nd degree}} \times$$

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) \leq \underbrace{\bar{\phi}_n/n^2}_{\text{Bias}} \times \mathcal{O}\Big(\underbrace{n \times \mathcal{N}^2_{n,\max}}_{\#\text{ 2nd degree}} \times \underbrace{\frac{\bar{\gamma} n}{K_n}}_{\text{largest cluster size}}\Big) +$$

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) = \frac{4}{n^2} \sum_{i,j} \underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i - 1), \mu_j(\mathbf{D})(2D_j - 1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i, j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A}, \mathcal{C}_n\Big) \leq \underbrace{\bar{\phi}_n/n^2}_{\text{Bias}} \times \mathcal{O}\Big(\underbrace{n \times \mathcal{N}^2_{n,\max}}_{\#\text{ 2nd degree}} \times \underbrace{\frac{\bar{\gamma} n}{K_n}}_{\text{largest cluster size}}\Big) + \underbrace{\frac{\bar{\psi}}{n^2} \sum_{i=1}^{n} |c(i)|}_{\text{ppl in same cluster}}$$

Worst-case variance

Worst case variance depends on

$$\mathrm{Var}\Big(\hat{\tau}|\mathbf{A},\mathcal{C}_n\Big) = \frac{4}{n^2}\sum_{i,j}\underbrace{\mathrm{Cov}\Big(\mu_i(\mathbf{D})(2D_i-1),\mu_j(\mathbf{D})(2D_j-1)\Big)}_{=\sigma_{i,j}}$$

hmm... and now?

Lem $|\sigma_{i,j}| = \mathcal{O}\Big(\mathrm{Bias}_n(\mathcal{C}_n)\Big)$ if (i,j) are (a) not friends or friends of friends, (b) in different clusters, and (c) j is not in a cluster with a friend of i

$$\Rightarrow \mathrm{Var}\Big(\hat{\tau}|\mathbf{A},\mathcal{C}_n\Big) \leq \underbrace{\bar{\phi}_n/n^2}_{\text{Bias}} \times \mathcal{O}\Big(\underbrace{n\times\mathcal{N}^2_{n,\max}}_{\#\text{ 2nd degree}} \times \underbrace{\frac{\bar{\gamma}n}{K_n}}_{\text{largest cluster size}}\Big) + \underbrace{\frac{\bar{\psi}}{n^2}\sum_{i=1}^{n}|c(i)|}_{\text{ppl in same cluster}}$$

$$= o(1/K_n) + \bar{\psi}\sum_{k=1}^{K_n}\frac{n_k^2}{n^2}$$