Approaches to Interference in Networked Experiments

Yuan Yuan

MIT IDSS

December 14, 2018

Outline

- A motivating example
- Aronow & Samii (2017), "*Estimating Average Causal Effects Under Interference Between Units*".
- Ugander, Karrer, Backstrom, & Kleinberg (2013), "*Graph Cluster Randomization: Network Exposure to Multiple Universes*".
- Eckles, Karrer & Ugander (2017), "*Design and analysis of experiments in networks: Reducing bias from interference*".
- Aral (2016), "*Networked Experiments*".

A motivating example

The goal: We want estimate the treatment effect of a policy.
The problem: We cannot observe two counterfactual worlds simultaneously!

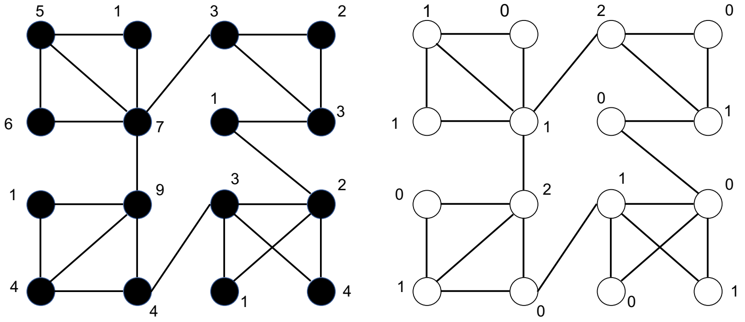


Figure 1: The goal is to compute the difference between a world where everyone is treated and and the other world where everyone is not treated.

A motivating example

One solution: Randomly assign individuals to treatment or control.

The problem: Strong network effects may exist.

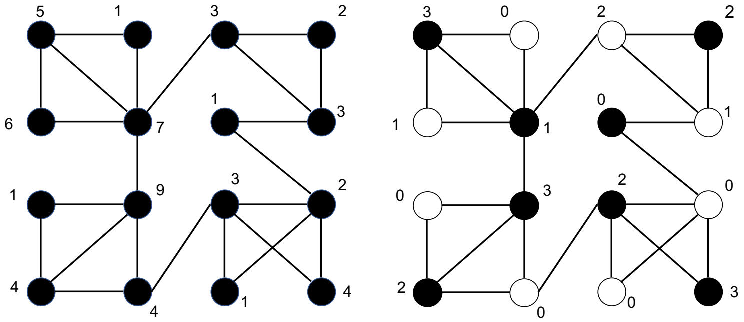


Figure 2: The problem of Bernoulli randomization.

A motivating example

The other solution: Graph cluster randomization!
Result: Getting closer to the truth.

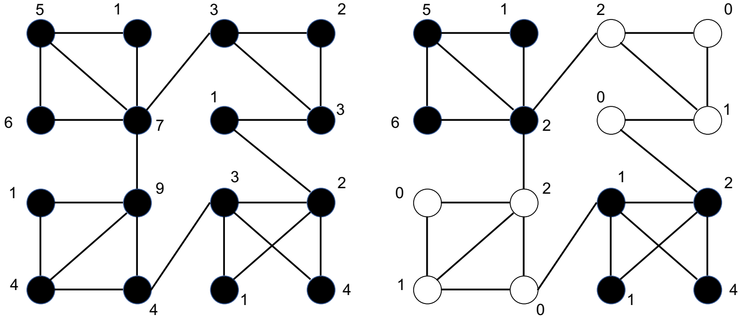


Figure 3: An example of graph cluster randomization.

Aronow & Samii (2017)

Key message

Proposing a framework for networked experiments

1. Treatment assignment
   – probability distribution for treatment vector $\mathbb{P}(\mathbf{Z})$
2. Exposure mapping
   – from treatment assignment vector $\mathbf{z}$ to **exposure**, for $i = 1, 2, ..., N$
3. Estimands
   – e.g., direct or indirect treatment effects

Treatment assignment

Treatment assignment

- $\mathbf{Z} = (Z_1, ..., Z_N)$: random vector
- $\mathbf{z} = (z_1, ..., z_N)$: treatment assignment vector
- $z_i \in \{1, ..., M\}$; $M = 2$ for binary treatment
- Ω: support for $\mathbf{Z}$; the size could be as large as M^N
- We can design $\mathbb{P}(\mathbf{Z})$!

Example

- Bernoulli randomization: $(Z_i \overset{\text{i.i.d.}}{\sim} Bern(p))$
- Graph cluster randomization: imagine when we have only 2 clusters...

SUTVA

The stable unit treatment value assumption (SUTVA)

- No spillover effect
- The treatment of j, Z_j, does not affect the outcome of i Y_i; or $Y_i(\mathbf{Z}) = Y_i'(Z_i)$.
- It does not hold in many cases: vaccines, advertisements, ...

Use exposures to address violation of SUTVA

Exposure: The exposure of j, D_j, does not affect the outcome of i Y_i; or $Y_i(\mathbf{Z}) = Y_i''(D_i)$ (Conditions 1 & 2).

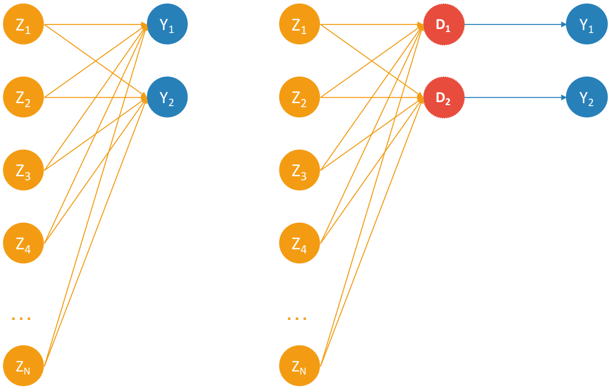


Figure 4: The causal diagrams when exposure mappings are included.

Exposure mapping

Exposure mapping: for i, $D_i = f(\mathbf{z}, \theta_i)$
Examples for θ_i:

- Index for i and $f(\mathbf{z}, \theta_i) = z_i$
- ith row of adjacency matrix (only consider i's neighbors)
- Group index for individual i (e.g., groups have different susceptibility to neighbors' treatments)
- Any heterogeneous effects...

Q: Are those exposure mappings realistic?

Some notations for exposure mapping

- $f(\mathbf{z}, \theta_i) \in \{d_1, ..., d_K\}$
- $\pi_i(d_k) = \mathbb{P}(f(\mathbf{z}, \theta_i) = d_k)$
- $\pi_{ij}(d_k) = \mathbb{P}(f(\mathbf{z}, \theta_i) = d_k, f(\mathbf{z}, \theta_j) = d_k)$
- Prop. 3.1: Approximate $\pi_i(d_k)$ and $\pi_{ij}(d_k)$ by R random replications of $\mathbf{z}$ ($|\Omega|$ may be too large).

Estimating y

Horvitz-Thompson estimator:

$$\hat{y}_{HT}^{T}(d_k) = \sum_{i=1}^{N} \mathbb{1}(D_i = d_k) \frac{Y_i}{\pi_i(d_k)} \tag{1}$$

which is unbiased (Lemma 4.1):

$$\mathbb{E}[\hat{y}_{HT}^{T}(d_k)] = \sum_{i=1}^{N} \mathbb{E}[\mathbb{1}(D_i = d_k)] \frac{y_i(d_k)}{\pi_i(d_k)} = \sum_{i=1}^{N} y_i(d_k)$$

$$\begin{aligned} \mathrm{Var}[\hat{y}_{HT}^{T}(d_k)] = \sum_{i=1}^{N} \pi_i(d_k)[1 - \pi_i(d_k)][\frac{y_i(d_k)}{\pi_i(d_k)}]^2 + \\ \sum_{i=1}^{N} \sum_{j \neq i} [\pi_{ij}(d_k) - \pi_i(d_k)\pi_j(d_k)] \frac{y_i(d_k)}{\pi_i(d_k)} \frac{y_j(d_k)}{\pi_j(d_k)} \end{aligned} \tag{2}$$

Issue (solved by conservative estimators in Section 5):

- $y_i(d_k) y_j(d_k)$ is not observed if $\pi_{ij}(d_k) = 0$, and thus unidentified.

Estimating τ

Following estimators for y:

$$\hat{\tau}_{HT}(d_k, d_l) = \frac{1}{N}[y_i(d_k) - y_i(d_l)] \tag{3}$$

$$\mathbb{E}[\hat{\tau}_{HT}(d_k, d_l)] = \frac{1}{N}[y_i(d_k) - y_i(d_l)] \tag{4}$$

$$\begin{aligned}\text{Var}[\hat{\tau}_{HT}(d_k, d_l)] = \frac{1}{N^2}\{&\text{Var}[y_{HT}^T(d_k)] + \text{Var}[y_{HT}^T(d_l)] \\ &-2\text{Cov}[y_{HT}^T(d_k), y_{HT}^T(d_l)]\}\end{aligned} \tag{5}$$

$$\begin{aligned}\text{Cov}[\hat{y}_{HT}^T(d_k), \hat{y}_{HT}^T(d_l)] = \sum_{i=1}^{N}\sum_{j\neq i}\frac{y_i(d_k)}{\pi_i(d_k)}\frac{y_j(d_l)}{\pi_j(d_l)}[&\pi_{ij}(d_k, d_l) - \\ &\pi_i(d_k)\pi_j(d_l)] - \sum_{i=1}^{N} y_i(d_k)y_i(d_l)\end{aligned} \tag{6}$$

Issue (solved by conservative estimators in Section 5):

- $y_i(d_k)y_i(d_l)$ is never observed, and thus unidentified.

Variance estimators

Horvitz and Thompson estimator:

$$\hat{\text{Var}}[\hat{y}_{HT}^{T}(d_k)] = \sum_{i=1}^{N} \mathbb{1}(D_i = d_k)[1 - (D_i = d_k)][\frac{Y_i}{\pi_i(d_k)}]^2 +$$

$$\sum_{i=1}^{N} \sum_{j \neq i} \frac{\mathbb{1}(D_i = d_k)\mathbb{1}(D_j = d_k)}{\pi_{ij}(d_k)}[\pi_{ij}(d_k) - \pi_i(d_k)\pi_j(d_k)]\frac{Y_i Y_j}{\pi_i(d_k)\pi_j(d_k)} \tag{7}$$

According to the equation above, we derive $\hat{\text{Var}}[\hat{\tau}_{HT}^{T}(d_k, d_l)]$, and show the variance estimator is conservative (Prop. 5.6):

$$\mathbb{E}[\hat{\text{Var}}[\hat{\tau}_{HT}^{T}(d_k)]] \geq \text{Var}[\hat{\tau}_{HT}^{T}(d_k)]$$

Quick summary for Sections 6-8

Section 6

- Consistency and confidence intervals (under some conditions)

Section 7

- Covariate adjustment
- Hájek estimation (alternative to Horvitz–Thompson; (slightly) biased; small variance)

Section 8

- Misspecification (one exposure corresponds to multiple outcomes–Prop 8.1 weight the outcomes by their probability)

Evaluation: Simulation

Network structure on AddHealth.
1/10 individuals are randomly treated.

$$f(\mathbf{z}, \theta_i) = \begin{cases} d_{11} \text{ (Direct + Indirect Exposure) :} & z_i \mathbf{I}(\mathbf{z}'\theta_i > 0) = 1, \\ d_{10} \text{ (Isolated Direct Exposure) :} & z_i \mathbf{I}(\mathbf{z}'\theta_i = 0) = 1, \\ d_{01} \text{ (Indirect Exposure) :} & (1 - z_i)\mathbf{I}(\mathbf{z}'\theta_i > 0) = 1, \\ d_{00} \text{ (No Exposure) :} & (1 - z_i)\mathbf{I}(\mathbf{z}'\theta_i = 0) = 1, \end{cases}$$

$$y_i(d_{11}) = 2y_i(d_{00})$$

$$y_i(d_{10}) = 1.5y_i(d_{00})$$

$$y_i(d_{01}) = 1.25y_i(d_{00})$$

Exposure conditions and results

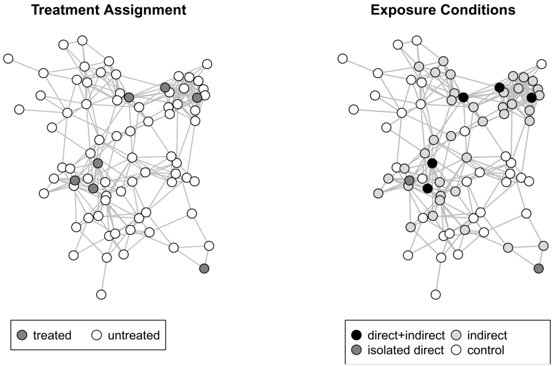


FIG. 1. *Illustration of a treatment assignment (left) and then treatment-induced exposures (right) for one of the school classes in the study. Each dot is a student, and each line represents an undirected friendship tie.*

Results (Table 1): the proposed estimators are unbiased; OLS is biased...

Evaluation: Field experiment

1. 28 of 56 schools were randomly selected to host the anti-conflict program, via block randomization ($s_i = 1$).
2. Within all schools, a group of between 40 to 64 students were nonrandomly selected as eligible to participate in the program.
3. Within each school hosting the program, half of the eligible students were then block randomized to participate in the program, with blocking on gender, grade, and a measure of network closure ($z_i = 1$).

$$f(\mathbf{z}, \theta_i) = \begin{cases} d_{111} \text{ (Direct + Indirect Exposure)}: & z_i \mathbf{I}(\mathbf{z}'\theta_i > 0) s_i = 1, \\ d_{101} \text{ (Isolated Direct Exposure)}: & z_i \mathbf{I}(\mathbf{z}'\theta_i = 0) s_i = 1, \\ d_{011} \text{ (Indirect Exposure)}: & (1 - z_i) \mathbf{I}(\mathbf{z}'\theta_i > 0) s_i = 1, \\ d_{001} \text{ (School Exposure)}: & (1 - z_i) \mathbf{I}(\mathbf{z}'\theta_i = 0) s_i = 1, \\ d_{000} \text{ (No Exposure)}: & (1 - s_i) = 1. \end{cases}$$

Results: HT, Hajek, and WLS with covariates have similar estimations for effects of exposures.

Discussion questions

- For the experiments: Are the exposure assumptions realistic?
- How can we effectively test the estimators in real-world data?
- Is it really necessary to assume a small number of exposure conditions (can we just parametrize exposure mapping $f(\mathbf{z}, \theta_i)$)?

Ugander et al. (2013)

Motivations

The goal is to estimate

$$\tau(\mathbf{Z} = \vec{1}, \mathbf{Z} = \vec{0}) = \frac{1}{N} \sum_{i=1}^{N} [y_i(\mathbf{Z} = \vec{1}) - y_i(\mathbf{Z} = \vec{0})] \qquad (8)$$

but $y_i(\mathbf{Z} = \vec{1})$ and $y_i(\mathbf{Z} = \vec{0})$ cannot be observed at the same time.
Solutions?

- **Analysis step**: Make assumptions about exposure conditions: **neighborhood/core exposure**!
- **Design step**: Design a proper treatment assignment policy ($\mathbb{P}(\mathbf{Z})$): **graph cluster randomization**! Nodes in a cluster receive the same assignment.

Network exposure: exposure assumptions

Use exposures σ_i^1, σ_i^0 to approximate $\mathbf{Z} = \vec{1}$ and $\mathbf{Z} = \vec{0}$.

$$y_i(\mathbf{Z} = \vec{1}) = y_i(\mathbf{Z} = \mathbf{z_1}), \text{ for } \mathbf{z_1} \in \sigma_i^1;$$

$$y_i(\mathbf{Z} = \vec{0}) = y_i(\mathbf{Z} = \mathbf{z_0}), \text{ for } \mathbf{z_0} \in \sigma_i^0.$$

Neighborhood exposure

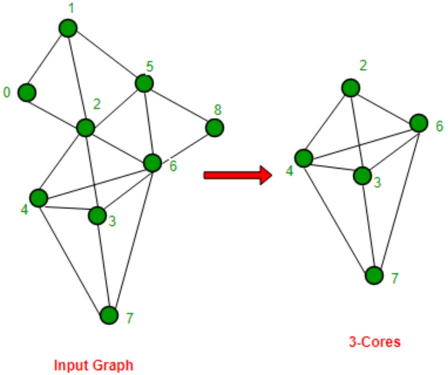
Neighborhood exposure (only consider i's immediate graph neighborhood, $\mathcal{N}(i)$)

- Full neighborhood exposure
  - $\sigma_i^1 = \{\mathbf{z} : z_i = 1 \text{ and } z_j = 1, \forall j \in \mathcal{N}(i)\}$
  - $\sigma_i^0 = \{\mathbf{z} : z_i = 0 \text{ and } z_j = 0, \forall j \in \mathcal{N}(i)\}$
- Absolute k-neighborhood exposure
  - $\sigma_i^1 = \{\mathbf{z} : z_i = 1 \text{ and } \sum_{j \in \mathcal{N}(i)} z_j \geq k\}$
  - $\sigma_i^0 = \{\mathbf{z} : z_i = 0 \text{ and } \sum_{j \in \mathcal{N}(i)} 1 - z_j \geq k\}$
- Fractional q-neighborhood exposure
  - $\sigma_i^1 = \{\mathbf{z} : z_i = 1 \text{ and } \sum_{j \in \mathcal{N}(i)} z_j \geq q|\mathcal{N}(i)|\}$
  - $\sigma_i^0 = \{\mathbf{z} : z_i = 0 \text{ and } \sum_{j \in \mathcal{N}(i)} 1 - z_j \geq q|\mathcal{N}(i)|\}$

1. $\sigma_i^1 \cup \sigma_i^0 \neq \Omega$: losing statistical power!
2. We may introduce bias if exposure conditions are chosen inappropriately!

k-core and fractional q-core

k-core: maximal subgraph where every node has a degree $\geq k$[1]



Fractional q-core: maximal subgraph where every node has a degree $\geq q|\mathcal{N}(i)|$

[1] Image source: `www.geeksforgeeks.org/find-k-cores-graph/`

Core exposure

- Component exposure: nodes in the same component receive the same treatment
- Absolute k-core exposure: neighbors in the same k-core graph receive the same treatment
- Fractional q-core exposure: neighbors in the same fractional q-core graph receive the same treatment

Stronger requirements than associated notwork exposures!
Q: how practical are these core exposure conditions?

Exposure probability

Consider cluster randomization: for each cluster, draw $\sim Bern(p)$
Non-trivial to compute $\mathbb{P}(Z \in \sigma_i^1)$ and $\mathbb{P}(Z \in \sigma_i^0)$ for absolute and fractional neighborhood exposure conditions

- Dynamic programming

k-core and fractional q-core? (core $\Rightarrow$ neighborhood)

- $\mathbb{P}(\mathbf{Z} \in \sigma_i^x | k\text{-core}) \leq \mathbb{P}(\mathbf{Z} \in \sigma_i^x | k\text{-nhood})$
- $\mathbb{P}(\mathbf{Z} \in \sigma_i^x | \text{frac } q\text{-core}) \leq \mathbb{P}(\mathbf{Z} \in \sigma_i^x | \text{frac } q\text{-nhood})$

Estimators

Horvitz-Thompson estimator

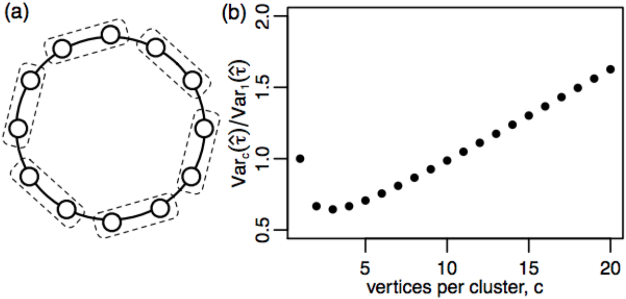
- the same as Aronow & Samii (2017) [pp. 13]
- $D_i \in \{\sigma_i^1, \sigma_i^0\}$

(Prop 3.3) The variance of the Horvitz-Thompson estimator under graph cluster randomization is ($\mathcal{O}(1/N)$) if

- Maximum degree $\mathcal{O}(1)$: no "hub".
- Cluster size $\mathcal{O}(1)$: $\mathcal{O}(N)$ clusters.

Application: Cycles

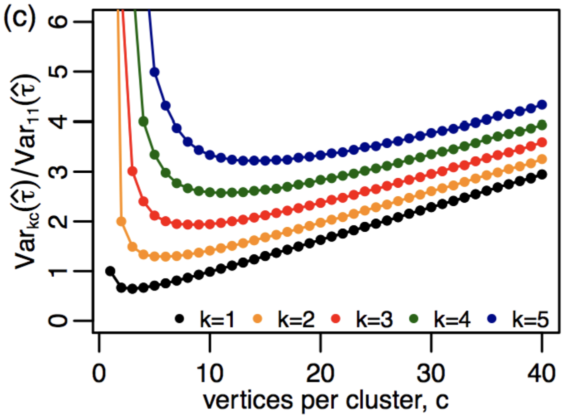
Actual treatment effect: $Y_i(\sigma_i^1) = \bar{Y}$, $Y_i(\sigma_i^0) = 0$ and full neighborhood exposure.



Trade-off: probability of full neighborhood exposure vs dependency between nodes (# clusters)

k-power Cycle Graphs

"Watts-Strogatz model without rewiring"–connecting to nodes with a distance $\leq k$.



Q: Why does the variance increase with respect to k?
(#full-neighborhood exposure)

Clustering Restricted Growth Graphs

Restricted growth graph:

- $|B_{r+1}(v)| \leq \kappa |B_r(v)|$.
- $|B_r(v)|$ is the set of nodes whose distance with v no greater than r.

Clustering algorithm (ϵ-net):

- Initially all vertices are unmarked.
- While there are unmarked vertices, in step j find an arbitrary unmarked vertex v, selecting v to be vertex v_j and marking all vertices in $B_2(v_j)$.
- Suppose k such vertices are defined, and let $S = \{v_1, v_2, ..., v_k\}$.
- For every vertex w of G, assign w to the closest vertex $v_i \in S$, breaking ties consistently (e.g. in order of lowest index).
- For every v_j, let C_j be the set of all vertices assigned to v_j.

Variance bounds

Lower bound for Bernoulli randomization:

PROPOSITION 4.3. *The variance of the HT estimator under full neighborhood exposure for vertex randomization of a graph with n vertices is lower bounded by an exponential function in the degree d of the graph,* $\text{Var}[\hat{\tau}(Z)] \geq O(1/n)(p^{-(d+1)} + (1-p)^{-(d+1)} - 1)$.

Upper bound for the proposed graph cluster randomization:

PROPOSITION 4.4. *The variance of the HT estimator under full, q-fractional, or k-absolute neighborhood exposure for a 3-net cluster randomization of a restricted-growth graph is upper bounded by a function linear in the degree d of the graph.*

Discussion:

- Recall $\mathcal{O}(1)$ maximum cluster size is a condition for $\mathcal{O}(1/n)$ variance (Prop. 3.3). The maximum cluster size is κ^3 (Prop. 4.2)!

Discussion questions

- How would neighborhood exposure assumptions affect our results?
- What conditions are used to show the variance bounds? Are they realistic?
- How realistic are restricted-growth graphs (think about Facebook friendship network)?
- Why not use community detection methods or other methods?

Eckles et al. (2017)

Key message

Motivations

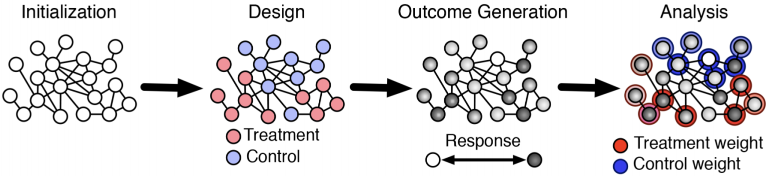
- "Evaluate methods for designing and analyzing randomized experiments under minimal, realistic assumptions compatible with broad interference."
- "Aim to reduce bias and possibly overall error in estimates of average effects of a global treatment."

Again, the goal is to estimate

$$\tau(\mathbf{Z} = \mathbf{z_1}, \mathbf{Z} = \mathbf{z_0}) = \frac{1}{N} \sum_{i=1}^{N} [y_i(\mathbf{Z} = \mathbf{z_1}) - y_i(\mathbf{Z} = \mathbf{z_0})] \tag{9}$$

Prototype: $\mathbf{z_1} = \vec{1}$ and $\mathbf{z_1} = \vec{0}$.

Model of experiments in networks



- **Initiation**: generates the graph and vertex characteristics
- **Design**: randomization
- **Outcome generation**: observe/simulate behavior
- **Analysis**: estimators

Initiation and Design

Initiation:

- Network formation
- Vertex characteristics and prior behaviors

Design (focused on cluster randomization):

- Global community detection methods (do not distinguish small clusters)
- Local clustering methods (e.g., 3-net clustering)
- Observed community membership (villages)

Outcome generation and observation

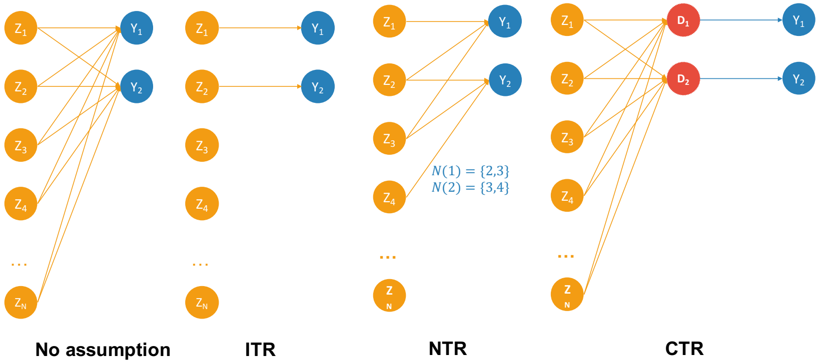
$Y_i = f_i(\mathbf{z}, \boldsymbol{\theta})$:

- Very similar to the exposure mapping in Aronow & Samii (2017)...
- But this function directly maps to outcomes rather than exposures!

Treatment response assumptions (Manski, 2013)

- Individualistic treatment response (ITR): SUTVA!
- Constant treatment response (CTR): equivalent to exposure assumption (effective treatments)
- Neighborhood treatment response (NTR): outcome Y only depends on z_i, and z_j for $j \in \mathcal{N}(i)$

Illustration of response models



Q: Where do k-core and fractional q-core neighborhood exposure assumptions belong?

Implausibility of tractable treatment response assumptions

Issue:

- For example, even NTR is often not realistic (time, peers' behavior, 2-hop influence, ...)

Dynamic model with discrete time steps

$$y_{i,t} = h_{i,t}(\mathbf{z}, \mathbf{y}_{i \cup \mathcal{N}(i), t-1}, \boldsymbol{\theta}) : \tag{10}$$

- $\mathbf{z}$: assignment vector
- $\mathbf{y}_{i \cup \mathcal{N}(i), t-1}$: the behavior of i and i's immediate neighbors at the last time step
- $\boldsymbol{\theta}$: characteristics of all individuals (including network structure)

Implausibility of tractable treatment response assumptions

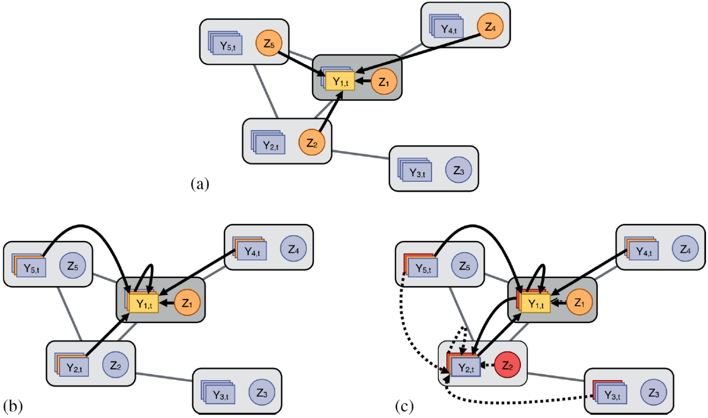


Figure 5: (a) Neighborhood treatment response (**NTR**). (b) Egos' behavior are affected by the behaviors of **peers** at time $t-1$. (c) Egos' behavior are affected by the behaviors of **peers of peers** at time $t-2$.

Linear-in-means model

$$Y_{i,j}^{*} = \alpha + \beta Z_i + \gamma \frac{A_i' Y_{t-1}}{k_i} + U_{i,t}$$
$$Y_{i,j} = a(Y_{i,j}*)$$

where A_i is the i row of adjacency matrix A; k_i is the degree of i; $a(x)$ can be binary ($a(x) = 1[x \geq 0]$) or continuous ($a(x) = x$).

Bias reduction through design

Thm. 2.1 in one sentence: Graph cluster randomization can reduce bias!

Assumptions:

1. Linear outcome model: $\mathbb{E}[Y_i(\mathbf{z})] = a_i + \sum_j B_{ij} z_j$
2. Monotonicity: $Y_i(\mathbf{z})$ is monotonically increasing in $\mathbf{z}$ (or $B_{ij} \geq 0$).

Results:

Any graph cluster randomization has a bias smaller than Bernoulli randomization.

$$\text{Relative bias} = \tau^{gcr}_{\text{ITR}}(1,0)/\tau(1,0) - 1 = \frac{\sum_{i=1}^N \sum_{j=1}^N B_{ij} 1[C(i) = C(j)]}{\sum_{i=1}^N \sum_{j=1}^N B_{ij}} - 1$$

Q: Are those assumptions realistic enough?

Q: Derivation of Eq. (7)?

Bias reduction through analysis

Thm 2.3 in one sentence: We can reduce bias by making more restrict treatment response assumptions!

Def 2.2 Treatment response A is more restrictive than treatment response B if $g_i^A(\mathbf{z}) = g_i^A(\mathbf{z}')$ implies $g_i^B(\mathbf{z}) = g_i^B(\mathbf{z}')$.

- Fractional neighborhood treatment response (FNTR) is more restrictive than ITR and NTR
- NTR is more restrictive than ITR

Q: Why don't we make the most restrictive assumptions to minimize bias?

bias-variance trade-off?

Estimators

Sample means

- $\hat{y}_S^T(d_k) = \sum_{i=1}^N \mathbb{1}(D_i = d_k)\frac{Y_i}{\sum_{i=1}^N \mathbb{1}(D_i=d_k)}$
- Biased
- Smaller variacne

Horvitz-Thompson estimator

- $\hat{y}_{HT}^T(d_k) = \sum_{i=1}^N \mathbb{1}(D_i = d_k)\frac{Y_i}{\pi_i(d_k)}$
- Unbiased
- Very large variance

Hajek

- $\hat{y}_{Hajek}^T(d_k) = (\sum_{i=1}^N \frac{\mathbb{1}(D_i=d_k)}{\pi_i(d_k)})^{-1} \sum_{i=1}^N \frac{Y_i \mathbb{1}(D_i=d_k)}{\pi_i(d_k)}$
- Slightly biased
- Smaller variance

Experiments: Setup

Setting:

- Small-world network ($N = 1000$; degree$= 10$; with rewiring probability $\{0, 0.01, 0.1, 0.5, 1\}$)
- Degree-corrected block model ($N = 1000$, 10 communities, within community edges: $\{0.2,\ 0.5,\ 0.8\}$; log-normal degree distribution)

Outcome generation

$$Y^*_{i,t} = -1.5 + \beta Z_i + \gamma \frac{A'_i Y_{i,t-1}}{k_i} + U_{i,t}, \qquad Y_{i,t} = 1\{Y^*_{i,t} > 0\}$$

Design: ϵ-net graph clustering
Analysis: ITR + sample mean; FNTR + sample mean; FNTR + Hajek.

Experiments: Results (Bias)

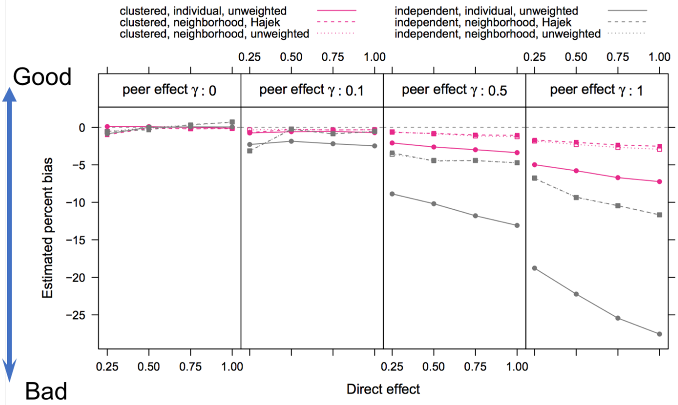


FIG 4. *Relative bias in ATE estimates for different assignment procedures, exposure models, and estimation methods. The most striking differences are between the assignment procedures, though the neighborhood exposure model also reduces bias (at the cost of increased variance — see Figure 5). Relative bias is not defined when the true value is zero, so we exclude simulations with the direct effect* $\beta = 0$. *For all networks, the rewiring probability was* $p_{\mathrm{rw}} = 0.01$.

Experiments: Results (Bias+Variance=RMSE)

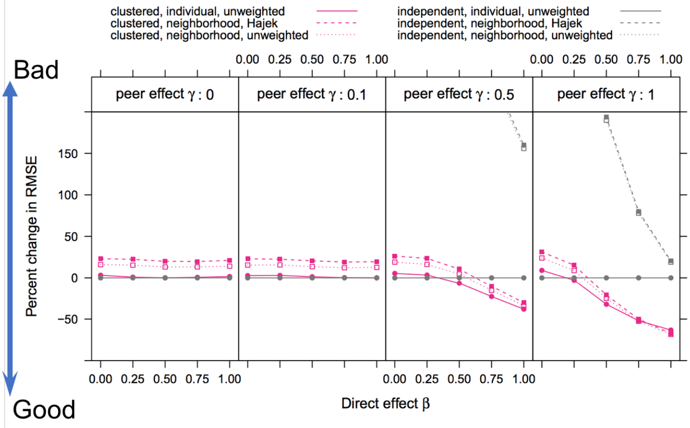


FIG 5. *Percent change in root-mean-squared-error (RMSE) compared with independent assignment with the simple difference-in-means estimator. Using the neighborhood condition with independent assignment results in large increases in variance: for the two smaller values of γ, this produces an almost 400% increase in RMSE. For this reason, the y-axis is limited to not show these cases. Rewiring probability* $p_{\mathrm{rw}} = 0.01$.

Q: Why do graph cluster randomization sometimes have a worse RMSE?

Experiments: Results of Degree-corrected block model

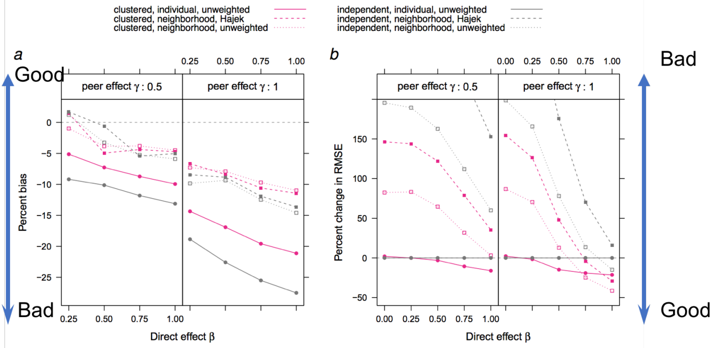


FIG 7. *Relative bias (a) and change in RMSE (b) in ATE estimates for different assignment procedures, exposure models, and estimation method, using the degree-corrected block model with community proportion* $p_{comm} = 0.8$. *Analysis using the exposure model provides additional bias reduction over using graph cluster randomization only — with a cost in variance.*

Why the performance is worse than that in small-world networks?
More high-degree nodes, lower clustering coefficients?

Discussion

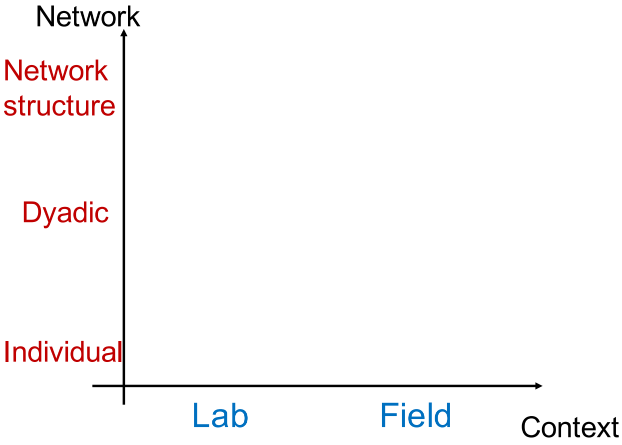
1. Can we test those methods on real-world data sets?
2. How we can check if our randomization methods and response assumptions are reasonable choices?

Aral (2016)

Key decisions in networked experiments

- Design
  - Setting
  - Sampling
  - Randomization
  - Assignment
- Analysis
  - Modeling
  - Inference
  - Estimation

Setting


Network
Network
structure
Dyadic
Individual
Lab
Field
Context

Sampling

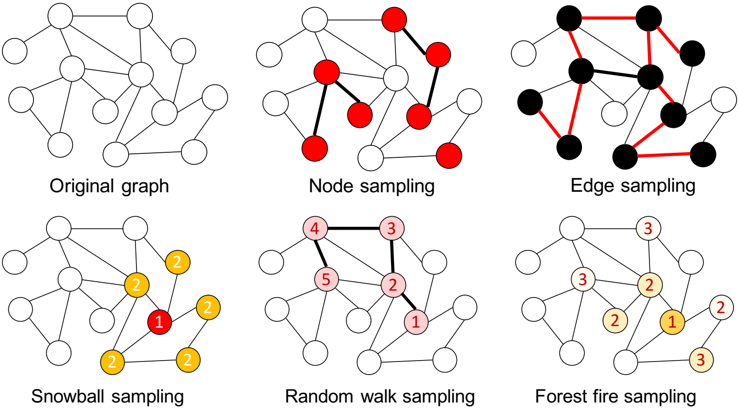
Goals:

- Estimate node or edge attributes
- Collect representative subgraphs
- Collect paths that reliably reproduce diffusion events

Sampling methods:

- Node & edge
- Snowball
- Random walk
- Forest fire
- Respondent driven

Sampling methods

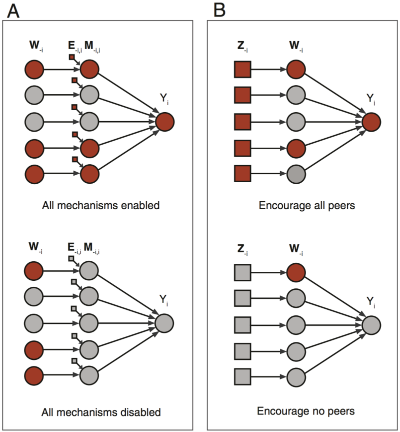


Q: Where can we apply these sampling methods?

Q: Is there selection bias?

Randomization

Mechanism designs vs peer encouragement designs[2]



Structural designs (network structure) and setting designs (context; e.g., randomize incentives)

[2] Eckles et al. (2016) Estimating peer effects in networks with peer encouragement designs

Treatment

Network autocorrelation

- e.g., graph cluster randomization

Sequencing

- e.g., global treatment impact vs statistical power (Aral and Walker (2011))

Key decisions in networked experiments

- Design
  - Setting
  - Sampling
  - Randomization
  - Assignment
- Analysis
  - Modeling
  - Inference
  - Estimation

Modeling

Exposure/Response assumptions

- CTR (exposure), ITR (SUTVA), NTR (neighborhood exposure), etc.

Model assumptions

- e.g., linear-in-means model

Inference

Estimands

- Direct causal effects: $Y_i(\mathbf{z}_i = 1, \mathbf{z}_{-i}) - Y_i(\mathbf{z}_i = 0, \mathbf{z}_{-i})$
- Indirect causal effects: $Y_i(\mathbf{z}_i = 0, \mathbf{z}_{-i}) - Y_i(\mathbf{z}_i = 0, \mathbf{z}_{-i} = \mathbf{0})$
- Total causal effects: $Y_i(\mathbf{z}_i = 1, \mathbf{z}_{-i} = \mathbf{1}) - Y_i(\mathbf{z}_i = 0, \mathbf{z}_{-i} = \mathbf{0})$

Estimation

- Model specification
  - Average treatment effects, heterogeneous treatment effects, more sophisticated model (hazard)
- Interference
  - Inference strategies (e.g., exposure assumptions; challenge: how to validate?)
  - Design strategies (treatment cluster vs treatment separating[3])
- Estimators
  - Horvitz-Thompson, Hajek estimators..

[3] Separate treated nodes (make their distance long.)

Summary and future directions

Summary

- Design: setting, sampling, randomization, and assignment
- Analysis: Modeling (assumptions), inference, and estimation

Future directions

- Adaptive treatment assignment (sequential)
- Novel randomization techniques
- Linking online treatments to offline responses (e.g., voting, HIV)
- Experimental validation of observational methods (propensity score matching)

Outline

- Aronow & Samii (2017): A theoretical framework for network interference.
- Ugander, Karrer, Backstrom, & Kleinberg (2013): A good example to address network interference.
- Eckles, Karrer & Ugander (2017): Theoretical explains for cluster randomization and exposure assumptions.
- Aral (2016): A high-level framework for networked experiments and examples