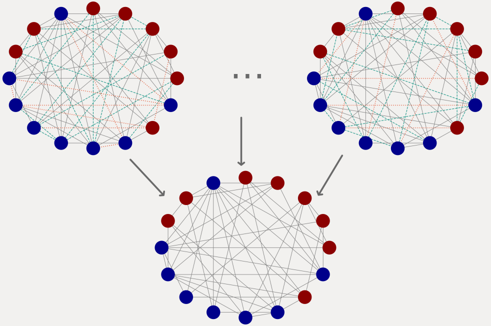
Estimating Causal Effects Using Proxy Interference Networks

Bar Weinstein
Daniel Nevo

Statistics & OR
Tel Aviv University

IDSAI 2025



Field experiement in 56 middle-schools.

Field experiement in 56 middle-schools.

1. *Goal:* reduce student conflicts through education.

Illustrative Example – Paluck et al. (2016)

Field experiement in 56 middle-schools.

1. *Goal:* reduce student conflicts through education.
2. *Challenge:* students are influenced by their peers.

Illustrative Example – Paluck et al. (2016)

Field experiement in 56 middle-schools.

1. *Goal:* reduce student conflicts through education.
2. *Challenge:* students are influenced by their peers.
3. *Data:* Multiple friendship networks.
   - Who do you spend time with?
   - Who is your best friend?
   - Measured at pre and post.
   - Self-reported.

Illustrative Example – Paluck et al. (2016)

Field experiement in 56 middle-schools.

1. *Goal:* reduce student conflicts through education.
2. *Challenge:* students are influenced by their peers.
3. *Data:* Multiple friendship networks.
   - Who do you spend time with?
   - Who is your best friend?
   - Measured at pre and post.
   - Self-reported.

- Which of the networks, if any, is the true network? which to choose?

Illustrative Example – Paluck et al. (2016)

Field experiement in 56 middle-schools.

1. *Goal:* reduce student conflicts through education.
2. *Challenge:* students are influenced by their peers.
3. *Data:* Multiple friendship networks.
    - Who do you spend time with?
    - Who is your best friend?
    - Measured at pre and post.
    - Self-reported.

- Which of the networks, if any, is the true network? which to choose?
- **Objective:** Estimate the intervention effects using the proxy networks.

- **Causal Inference.** Effect of treatment on an outcome.

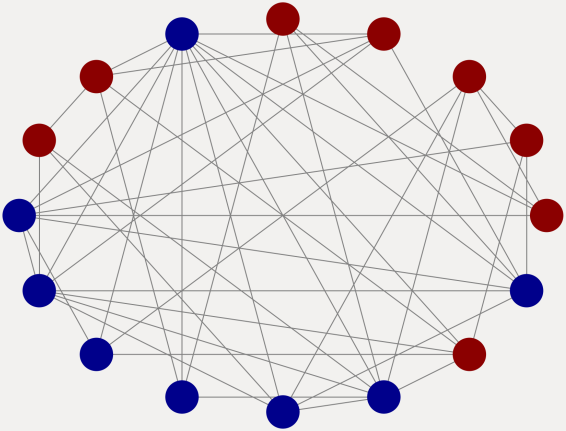
Background

- **Causal Inference.** Effect of treatment on an outcome.
- **Interference.** Treatment of one unit affect the outcomes of others.
  - Spillovers, peer effects, contagion, etc.

Background

- **Causal Inference.** Effect of treatment on an outcome.
- **Interference.** Treatment of one unit affect the outcomes of others.
  - Spillovers, peer effects, contagion, etc.
- Treatments spreads through a **network**.
  - Nodes: units. Edges: pairwise interactions.

- **Causal Inference.** Effect of treatment on an outcome.
- **Interference.** Treatment of one unit affect the outcomes of others.
  - Spillovers, peer effects, contagion, etc.
- Treatments spreads through a **network**.
  - Nodes: units. Edges: pairwise interactions.
- Examples:
  - *Social networks.* Information & behavior spread.
  - *Public health.* transmisstion of infectious diseases or addictive drugs.
  - A/B testing in marketplaces.



The Challenge

- Randomized experiment in networked population.

The Challenge

- Randomized experiment in networked population.
- Accurately measuring social networks is challenging.

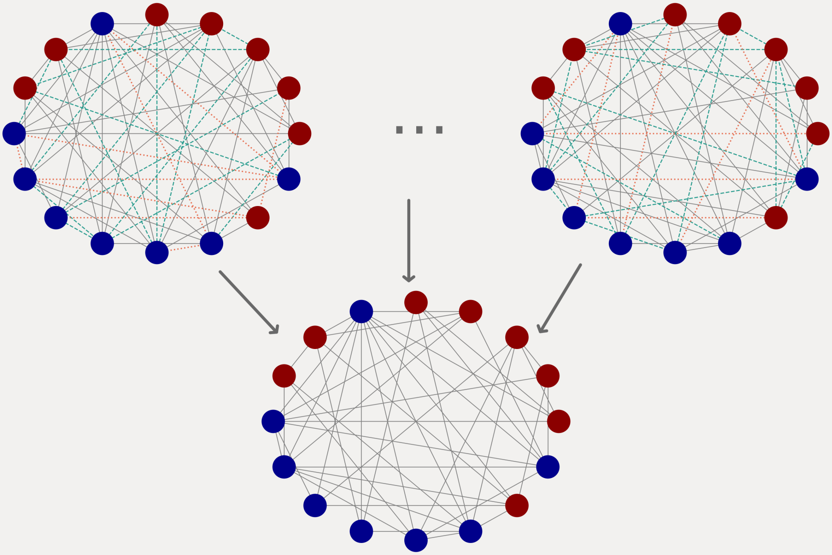
The Challenge

- Randomized experiment in networked population.
- Accurately measuring social networks is challenging.
- We observe only proxy measurements of the true network.
  - Measurements error.
  - Multiple sources of data.
  - Multilayer networks.
- True network remains latent.

The Challenge

- Randomized experiment in networked population.
- Accurately measuring social networks is challenging.
- We observe only proxy measurements of the true network.
    - Measurements error.
    - Multiple sources of data.
    - Multilayer networks.
- True network remains latent.

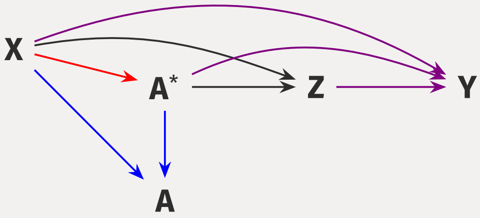
How can we estimate causal effects using proxy networks?



Formal Setup

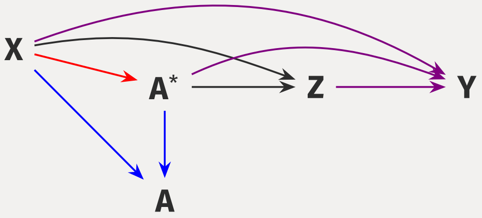
- Finite population $i \in \{1, \dots, N\}$.
- Treatments: $\mathbf{Z} \in \{0, 1\}^N$.
- Outcomes: $\mathbf{Y} \in \mathbb{R}^N$.
- Covariates/features: $\mathbf{X}$.
- True interference network: $\mathbf{A}^* \in \{0, 1\}^{N \times N}$.
- Proxy networks: $\mathbf{A} = \left(\mathbf{A}^1, \dots, \mathbf{A}^B\right)$.

- Population-level directed acyclic graph (DAG):



Structural Causal Model

- Population-level directed acyclic graph (DAG):



- Requires probabilistic models:
  1. *True network.* $p\left(\mathbf{A}^* | \mathbf{X}, \theta\right)$.
  2. *Proxy networks.* $p\left(\mathbf{A} | \mathbf{A}^*, \mathbf{X}, \gamma\right)$.
  3. *Outcomes.* $p\left(\mathbf{Y} | \mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \eta\right)$.

Causal Estimands

Impact of hypothetical treatments assignment policies.

Impact of hypothetical treatments assignment policies.

1. **Static.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = \mathbf{z}), \mathbf{X}, \mathbf{A}^*\right]$.
    - Treating all ($\mathbf{z} = \mathbf{1}$) vs none ($\mathbf{z} = \mathbf{0}$).

Causal Estimands

Impact of hypothetical treatments assignment policies.

1. **Static.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = \mathbf{z}), \mathbf{X}, \mathbf{A}^*\right]$.
    - Treating all ($\mathbf{z} = \mathbf{1}$) vs none ($\mathbf{z} = \mathbf{0}$).
2. **Dynamic.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = h(\mathbf{X}, \mathbf{A}^*)), \mathbf{X}, \mathbf{A}^*\right]$.
    - Treating units with specific features, e.g., above certain age.

Impact of hypothetical treatments assignment policies.

1. **Static.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = \mathbf{z}), \mathbf{X}, \mathbf{A}^*\right]$.
    - Treating all ($\mathbf{z} = \mathbf{1}$) vs none ($\mathbf{z} = \mathbf{0}$).
2. **Dynamic.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = h(\mathbf{X}, \mathbf{A}^*)), \mathbf{X}, \mathbf{A}^*\right]$.
    - Treating units with specific features, e.g., above certain age.
3. **Stochastic.** $\mathbb{E}_{\pi_\alpha(\mathbf{Z})}\mathbb{E}\left[Y_i | do(\mathbf{Z}), \mathbf{X}, \mathbf{A}^*\right]$.
    - Randomly treating α_1 vs α_0 percent of units.

- *Data:* $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. *Latent:* $\mathbf{A}^*$ and models' parameters (η, γ, θ).

- *Data:* $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. *Latent:* $\mathbf{A}^*$ and models' parameters (η, γ, θ).
- Posterior distribution:

$$\begin{aligned} p(\mathbf{A}^*, \eta, \gamma, \theta | \mathbf{O}) \propto\ & p(\mathbf{Y}|\mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \eta)p(\eta) \\ & \times p(\mathbf{A}|\mathbf{A}^*, \mathbf{X}, \gamma)p(\gamma) \\ & \times p(\mathbf{A}^*|\mathbf{X}, \theta)p(\theta). \end{aligned}$$

- *Data:* $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. *Latent:* $\mathbf{A}^*$ and models' parameters (η, γ, θ).
- Posterior distribution:

$$
\begin{aligned}
p(\mathbf{A}^*, \eta, \gamma, \theta | \mathbf{O}) \propto\ & p(\mathbf{Y} | \mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \eta) p(\eta) \\
& \times p(\mathbf{A} | \mathbf{A}^*, \mathbf{X}, \gamma) p(\gamma) \\
& \times p(\mathbf{A}^* | \mathbf{X}, \theta) p(\theta).
\end{aligned}
$$

- Mixed space. Discrete $(\mathbf{A}^*)$ have $O(N^2)$ terms!

- *Data:* $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. *Latent:* $\mathbf{A}^*$ and models' parameters (η, γ, θ).
- Posterior distribution:

$$\begin{aligned} p(\mathbf{A}^*, \eta, \gamma, \theta | \mathbf{O}) \propto\ & p(\mathbf{Y}|\mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \eta) p(\eta) \\ & \times p(\mathbf{A}|\mathbf{A}^*, \mathbf{X}, \gamma) p(\gamma) \\ & \times p(\mathbf{A}^*|\mathbf{X}, \theta) p(\theta). \end{aligned}$$

- Mixed space. Discrete $(\mathbf{A}^*)$ have $O(N^2)$ terms!
- Propose two sampling schemes:
    1. *Modularization.* "break" the posterior into smaller, more manageable parts.
    2. *Gibbs sampling.* Sample discrete with *Local Informed Proposals*.

Simulations

- $N = 500$ units. True network is measured with error. Outcomes from MRF.

Simulations

- N = 500 units. True network is measured with error. Outcomes from MRF.

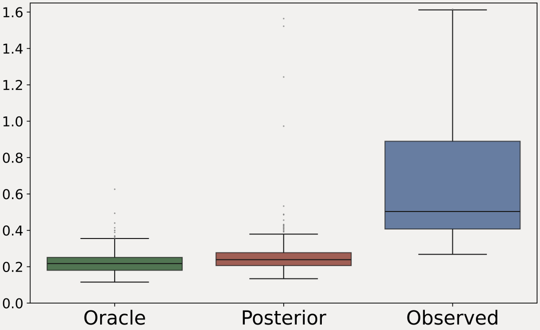


Figure 1: MAPE (↓) of stochastic estimand.

- N = 500 units. True network is measured with error. Outcomes from MRF.

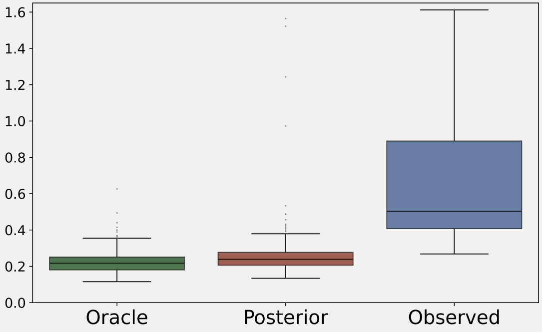


Figure 1: MAPE (↓) of stochastic estimand.

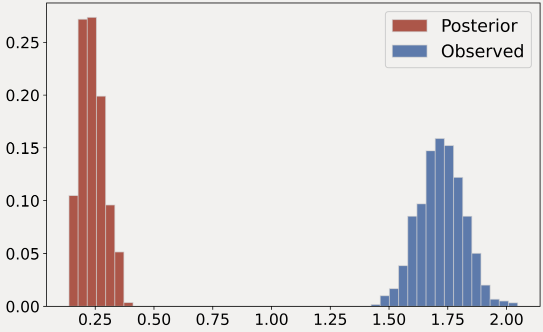
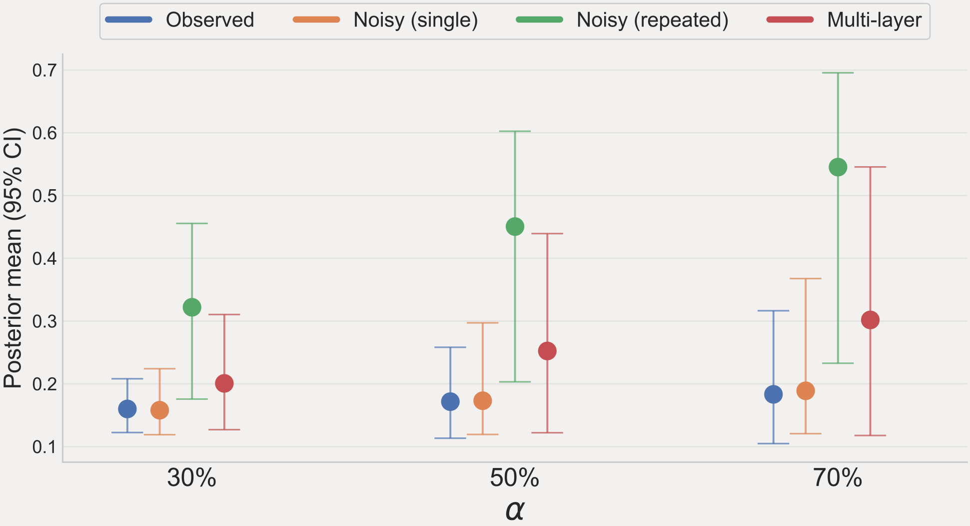


Figure 2: MAE (←) of network statistics.

Data Analysis - Paluck et al. (2016)

- Outcome is indicator of anti-conflict behavior.
- Stochastic estimand.
- Four available networks.
  - Use one as ”Observed” true network.
  - Analysis using different combination of the four proxy networks.

Data Analysis - Paluck et al. (2016)



Key Takeaways

- Network interference is common $\Rightarrow$ implications for many A/B tests.

Key Takeaways

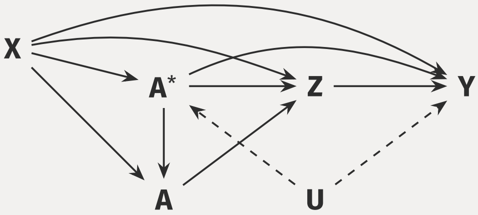
- Network interference is common $\Rightarrow$ implications for many A/B tests.
- Correctly measuring social relations is often impossible.
- Can estimate causal effects using proxy networks.

Key Takeaways

- Network interference is common $\Rightarrow$ implications for many A/B tests.
- Correctly measuring social relations is often impossible.
- Can estimate causal effects using proxy networks.
- Bayesian framework for inference. Computation is challenging.

Thank You!

Appendix – DAG



- **X**: covariates.
- **A***: true network.
- **A**: proxy networks.
- **Z**: treatments.
- **Y**: outcomes.
- **U**: unobserved confounders.

Appendix – SCM

$$
\begin{aligned}
\mathbf{U}_i &= f_U\left(\varepsilon_{U_i}\right), && i \in [N] \\
\mathbf{X}_i &= f_X\left(\varepsilon_{X_i}\right), && i \in [N] \\
\mathbf{A}^* &= f_{A^*}\left(\mathbf{X}, \mathbf{U}, \varepsilon_{A^*}\right) && \\
\mathbf{A}^b &= f_{A^b}\left(\mathbf{A}^*, \mathbf{X}, \varepsilon_{A^b}\right), && b = 1, \dots, B \\
Z_i &= f_Z\left(\mathbf{A}^*, \mathbf{A}, \mathbf{X}_i, \mathbf{X}_{-i}, \varepsilon_{Z_i}\right) && i \in [N] \\
Y_i &= f_Y\left(Z_i, \mathbf{X}_i, \mathbf{Z}_{-i}, \mathbf{X}_{-i}, \mathbf{A}^*, \mathbf{U}_i, \varepsilon_{Y_i}\right) && i \in [N],
\end{aligned}
$$

often simplified to

$$
\begin{aligned}
Z_i &= f_Z\left(\mathbf{X}_i, \phi_2(\mathbf{X}_{-i}, \mathbf{A}^*), \phi_{3,i}(\mathbf{A}^*), \varepsilon_{Z_i}\right) && i \in [N] \\
Y_i &= f_Y\left(Z_i, \mathbf{X}_i, \phi_1(\mathbf{Z}_{-i}, \mathbf{A}^*), \phi_2(\mathbf{X}_{-i}, \mathbf{A}^*), \phi_{3,i}(\mathbf{A}^*), \mathbf{U}_i, \varepsilon_{Y_i}\right) && i \in [N].
\end{aligned}
$$

Appendix – Examples of proxy network models

1. **Random noise.** $p(\mathbf{A}|\mathbf{A}^*, \mathbf{X}, \gamma) = \prod_{i>j} \Pr(A_{ij}|A^*_{ij}, \gamma)$ where $\Pr(A_{ij} = 1|A^*_{ij} = k, \gamma) = \gamma_k,\ k = 0, 1$, corresponding to false positive rate when $k = 0$ and true positive rate when $k = 1$.

2. **Edge censoring.** Each unit report maximal C friends, that is, report only $\min(C, d_i)$ friends. The probability that $A_{ij} = 1$ given that $A^*_{ij} = 1$ can be modeled by
$$\Pr(A_{ij} = 1|A^*_{ij} = 1, \gamma) = 1 - \big(1 - \min(1, C/d_i)\big)\big(1 - \min(1, C/d_j)\big)$$

3. **Repeated measures.** Each proxy modeled some thing like 1., and we have joint over all B proxies.

4. **Multilayer networks.** Each layer is a proxy network, and we have a model $p(\mathbf{A}^b|\mathbf{A}^*, \mathbf{X}, \gamma_b)$ for each proxy. For example, we have conditional independence across proxies but hierarchical structure for the parameters.

Appendix – Sampling from the posterior

Bayesian Modularization. Note that

$$p(\eta, \theta, \gamma, \mathbf{A}^* | \mathbf{O}) \propto \underbrace{p(\eta | \mathbf{Y}, \mathbf{Z}, \mathbf{A}^*, \mathbf{X})}_{\text{Outcome module}} \underbrace{p(\theta, \gamma, \mathbf{A}^* | \mathbf{A}, \mathbf{X})}_{\text{Network module}} \underbrace{p(\mathbf{Y} | \mathbf{Z}, \mathbf{A}^*, \mathbf{X})}_{\text{Feedback term}} .$$

Remove feedback by sampling from *cut-posterior*:

$$p_{cut}(\eta, \theta, \gamma, \mathbf{A}^* | \mathbf{O}) \propto p(\eta | \mathbf{Y}, \mathbf{Z}, \mathbf{A}^*, \mathbf{X}) p(\theta, \gamma, \mathbf{A}^* | \mathbf{A}, \mathbf{X}).$$

Sample from network module by first sample θ, γ from

$$p(\theta, \gamma | \mathbf{A}, \mathbf{X}) \propto p(\theta) p(\gamma) \sum_{\mathbf{A}^*} p(\mathbf{A} | \mathbf{A}^*, \mathbf{X}, \gamma) p(\mathbf{A}^* | \mathbf{X}, \theta)$$

and then sample $\mathbf{A}^*$ from

$$p(\mathbf{A}^* | \mathbf{A}, \mathbf{X}, \theta, \gamma) \propto \frac{p(\mathbf{A} | \mathbf{A}^*, \mathbf{X}, \gamma) p(\mathbf{A}^* | \mathbf{X}, \theta)}{\sum_{\mathbf{A}^*} p(\mathbf{A} | \mathbf{A}^*, \mathbf{X}, \gamma) p(\mathbf{A}^* | \mathbf{X}, \theta)}$$

Appendix – Sampling from the posteior 2

Gibbs sampling. Given $\mathbf{A}^*$, sampling (η, θ, γ) is easy and can be done with any gradient-based sampler, such as MALA or HMC.

Updating $\mathbf{A}^*$ given the rest is challenging as the space is huge.

Can be done with *Local Informed Proposals*: “flip” edges with probaility proportional to likelihood ratios.

Can be even more efficient by approximating likelihood ratios with gradients.

Block Gibbs is therefore the following for multiple iterations.

- Update $\mathbf{A}^*$ given η, θ, γ.
- Update (η, θ, γ) given $\mathbf{A}^*$.

Appendix – Simulation DGP

$$
\begin{aligned}
X_{1,i} &\sim N(0, 3^2) \\
X_{2,i} &\sim Ber(0.1) \\
\mathbf{U}_i &\sim N(0, I_2) \\
\Pr(A^*_{ij} = 1 | \mathbf{X}, \mathbf{U}, \theta) &= expit\Big(- 2 + 1.5\tilde{X}_{2,ij} - \|\mathbf{U}_i - \mathbf{U}_j\| \Big) \\
\Pr(A_{ij} = 1 | A^*_{ij}, \mathbf{X}, \gamma) &= expit\Big(A^*_{ij} 1.1 + (1 - A^*_{ij})(0.2 - \tilde{X}_{1,ij} + \tilde{X}_{2,ij}) \Big) \\
Z_i &\sim Ber(0.5) \\
\psi &\sim N\Big(0, \tilde{Q}^{-1}\Big) \\
\Pr(Y_i = 1 | \cdot) &= expit\Big(- 1 - 0.25 X_{1,i} + 3 Z_i + 3\phi_1(\mathbf{Z}_{-i}, \mathbf{A}^*_i) + 1.5\psi_i \Big)
\end{aligned}
$$

Appendix – Data Analysis

- Noisy (single) – ST network measured with differential noise.
- Noisy (repeated) – ST pre and post measured with differential noise.
- Multi-layer – ST and BT networks are function of latent predecssor $\mathbf{A}^*$.

Degree distribution in observed networks: