Estimating Causal Effects Using Proxies of a Latent Interference Network

Bar Weinstein
Tel Aviv University

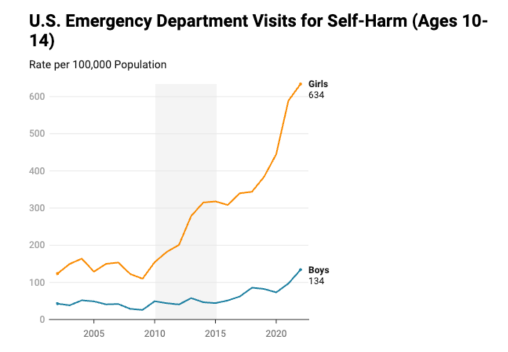
TAU Statistics Seminar
January 6, 2026

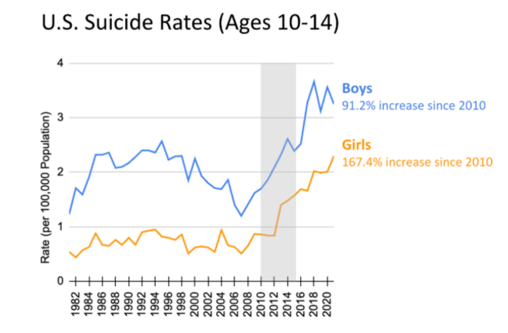
Talk Outline

- Causal inference under **interference**.
- **Network interference structure:** measurements, misspecification, and proxies.

Talk Outline

- Causal inference under **interference**.
- **Network interference structure:** measurements, misspecification, and proxies.
- Research conducted during my PhD (supervised by Daniel Nevo):
    1. Impact of network misspecificaiton and network-misspecification-robust estimation.
    2. **Estimation with proxy networks** (focus of today's talk).
    3. Sensitivity analysis for contamination in egocentric-network RCTs.


U.S. Emergency Department Visits for Self-Harm (Ages 10-14)
Rate per 100,000 Population
600
500
400
300
200
100
0
2005
2010
2015
2020
Girls
634
Boys
134


U.S. Suicide Rates (Ages 10-14)
Rate (per 100,000 Population)
4
3
2
1
0
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
Boys
91.2% increase since 2010
Girls
167.4% increase since 2010

- Are these increases related to the rise of social media?
  - iPhone in 2007; Instagram in 2010.
  - Main thesis of Jonathan Haidt (*The Anxious Generation, 2024*)

- Are these increases related to the rise of social media?
  - iPhone in 2007; Instagram in 2010.
  - Main thesis of Jonathan Haidt (*The Anxious Generation, 2024*)
- Recent RCTs of social media restriction had minimal effects
  - Treatment: Social media usage restriction
  - Outcomes: Various mental health and social behavior measures
  - Estimand: ATE
  - Recent commentary by Fulton (2025).

- Are these increases related to the rise of social media?
  - iPhone in 2007; Instagram in 2010.
  - Main thesis of Jonathan Haidt (*The Anxious Generation, 2024*)
- Recent RCTs of social media restriction had minimal effects
  - Treatment: Social media usage restriction
  - Outcomes: Various mental health and social behavior measures
  - Estimand: ATE
  - Recent commentary by Fulton (2025).
- However, social media usage among friends may affect the outcomes of an individual $\Rightarrow$ **Interference!**
- ATE is irrelevant and other estimands are needed.

What can we do?

- Cluster-randomized trial with cluster-level social media restriction

What can we do?

- Cluster-randomized trial with cluster-level social media restriction
- Network-randomized trial
  - Estimands can be something like the average effect if $p\%$ of individuals are treated.
  - Requires correct measurement of the interference network.

The Problem

- **Interference:** treatment of one unit affects the outcome of another unit.
- To estimate effects under interference, we typically need to know **who interferes with whom**

The Problem

- **Interference:** treatment of one unit affects the outcome of another unit.
- To estimate effects under interference, we typically need to know **who interferes with whom**
- **The Challenge:** Measuring networks of social interaction is a formidable task!

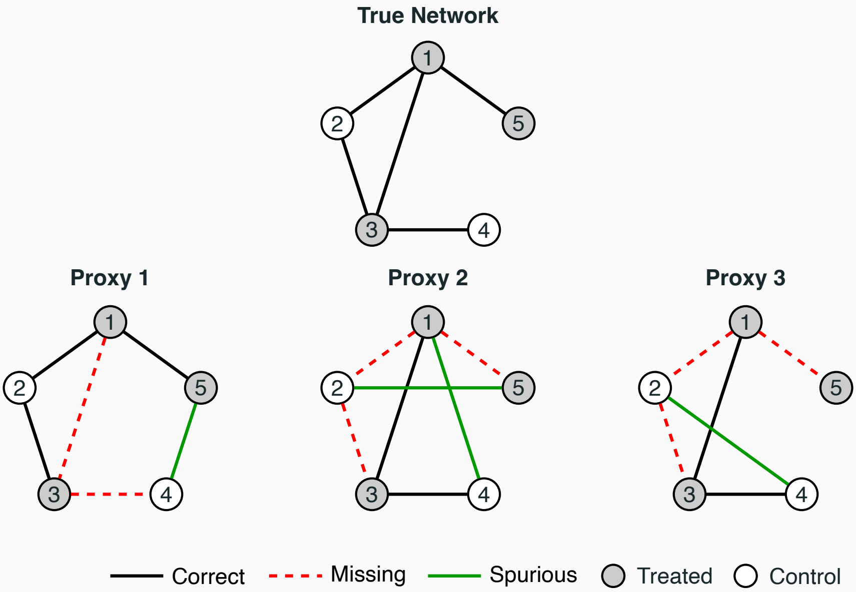
The Problem

- **Interference:** treatment of one unit affects the outcome of another unit.
- To estimate effects under interference, we typically need to know **who interferes with whom**
- **The Challenge:** Measuring networks of social interaction is a formidable task!
- At best, we only observe **proxies** of the true network
  - Networks measured with error
  - Networks from multiple sources
  - Multilayer networks

The Problem

- **Interference:** treatment of one unit affects the outcome of another unit.
- To estimate effects under interference, we typically need to know **who interferes with whom**
- **The Challenge:** Measuring networks of social interaction is a formidable task!
- At best, we only observe **proxies** of the true network
  - Networks measured with error
  - Networks from multiple sources
  - Multilayer networks

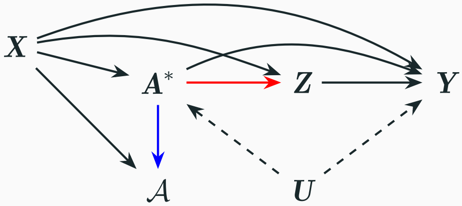
How can we estimate causal effects when we only observe proxies of the interference network?

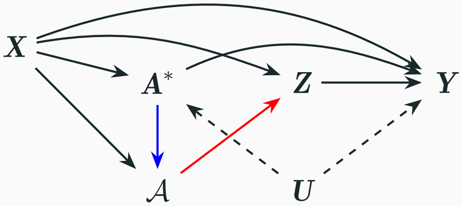

True Network
1
2
3
4
5
Proxy 1
Proxy 2
Proxy 3
Correct
Missing
Spurious
Treated
Control

Setup

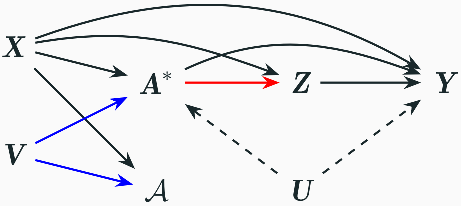
- Finite population of N units indexed by $i \in [N] = \{1, 2, \ldots, N\}$.
- Covariates $\boldsymbol{X}_i$.
- Treatments $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_N)$.
- Outcomes $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$.
- Latent interference network $\boldsymbol{A}^*$; assume undirected and unweighted for simplicity.
- Observed proxies of the interference network $\mathcal{A} = \{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_B\}$.
  - Each $\boldsymbol{A}_b$ is an $N \times N$ adjacency matrix.
- Observed data: $\boldsymbol{\mathcal{D}} = (\boldsymbol{Y}, \boldsymbol{Z}, \mathcal{A}, \boldsymbol{X})$.

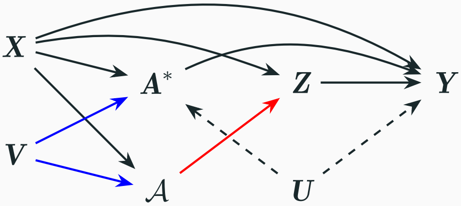
DAGs (Causal Proxies)





DAGs (Non-Causal Proxies)





Structural Causal Models

- Each DAG corresponds to a structural causal model (SCM).
- Each variable is a function of its parents and an exogenous error term.
  - Not necessarily iid.

Structural Causal Models

- Each DAG corresponds to a structural causal model (SCM).
- Each variable is a function of its parents and an exogenous error term.
    - Not necessarily iid.
- Outcomes generated as

$$Y_i = f_Y\big(Z_i, \boldsymbol{X}_i, \boldsymbol{Z}_{-i}, \boldsymbol{X}_{-i}, \boldsymbol{A}^*, \boldsymbol{U}_i, \varepsilon_{Y_i}\big) \quad i \in [N]$$

Structural Causal Models

- Each DAG corresponds to a structural causal model (SCM).
- Each variable is a function of its parents and an exogenous error term.
  - Not necessarily iid.
- Outcomes generated as

$$Y_i = f_Y\big(Z_i, \boldsymbol{X}_i, \boldsymbol{Z}_{-i}, \boldsymbol{X}_{-i}, \boldsymbol{A}^*, \boldsymbol{U}_i, \varepsilon_{Y_i}\big) \quad i \in [N]$$

- Function of high-dimensional inputs; In practice summarized via

$$Y_i = f_Y\big(Z_i, \boldsymbol{X}_i, \phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}^*), \phi_2(\boldsymbol{X}_{-i}, \boldsymbol{A}^*), \phi_{3,i}(\boldsymbol{A}^*), \boldsymbol{U}_i, \varepsilon_{Y_i}\big), \;\; i \in [N]$$

  - ϕ_1 – Exposure mapping (e.g., proportion of treated neighbors).
  - ϕ_2 – Summarizing covariates (e.g., average age of neighbors).
  - $\phi_{3,i}$ – Network summary statistics (e.g., global centrality measures).

Estimands

- Effects of hypothetical interventions on the population-level treatments.
- Define $Y_i(\boldsymbol{z})$ as outcomes under the intervention do$(\boldsymbol{Z} = \boldsymbol{z})$.

Estimands

- Effects of hypothetical interventions on the population-level treatments.
- Define $Y_i(\boldsymbol{z})$ as outcomes under the intervention $\mathrm{do}(\boldsymbol{Z} = \boldsymbol{z})$.
- Treatment assignment policies:
  1. **Static policy.** Set $\boldsymbol{Z}$ to some fixed $\boldsymbol{z}$ (e.g., all or none treated).
  2. **Dynamic policy.** Set $\boldsymbol{Z}$ as a function of covaraites $\boldsymbol{z} = h(\boldsymbol{X})$.
  3. **Stochastic policy.** $\mathrm{do}(\boldsymbol{Z} = \boldsymbol{z})$ with probability $\pi_\alpha(\boldsymbol{z})$.

Estimands

- Effects of hypothetical interventions on the population-level treatments.
- Define $Y_i(\boldsymbol{z})$ as outcomes under the intervention do$(\boldsymbol{Z} = \boldsymbol{z})$.
- Treatment assignment policies:
  1. **Static policy.** Set Z to some fixed $\boldsymbol{z}$ (e.g., all or none treated).
  2. **Dynamic policy.** Set Z as a function of covaraites $\boldsymbol{z} = h(\boldsymbol{X})$.
  3. **Stochastic policy.** do$(\boldsymbol{Z} = \boldsymbol{z})$ with probability $\pi_\alpha(\boldsymbol{z})$.
- Estimands are contrasts of

$$N^{-1} \sum_{i=1}^{N} \mathbb{E}[\, Y_i(\boldsymbol{z}) \mid \boldsymbol{A}^*]$$

for different policies (in stochastic case, marginalize over $\pi_\alpha(\boldsymbol{z})$).

Estimands

- Effects of hypothetical interventions on the population-level treatments.
- Define $Y_i(\boldsymbol{z})$ as outcomes under the intervention $\mathrm{do}(\boldsymbol{Z} = \boldsymbol{z})$.
- Treatment assignment policies:
  1. **Static policy.** Set $\boldsymbol{Z}$ to some fixed $\boldsymbol{z}$ (e.g., all or none treated).
  2. **Dynamic policy.** Set $\boldsymbol{Z}$ as a function of covaraites $\boldsymbol{z} = h(\boldsymbol{X})$.
  3. **Stochastic policy.** $\mathrm{do}(\boldsymbol{Z} = \boldsymbol{z})$ with probability $\pi_\alpha(\boldsymbol{z})$.
- Estimands are contrasts of

$$N^{-1} \sum_{i=1}^{N} \mathbb{E}[\, Y_i(\boldsymbol{z}) \mid \boldsymbol{A}^*]$$

  for different policies (in stochastic case, marginalize over $\pi_\alpha(\boldsymbol{z})$).
- For most DAGs shown, had we observed $\boldsymbol{A}^*$:

$$\mathbb{E}[\, Y_i(\boldsymbol{z}) \mid \boldsymbol{A}^*] = \mathbb{E}[\, Y_i \mid \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}].$$

Types of Proxy Networks

- **Measurement errors.** Each proxy is a noisy measurement of A^*.

Types of Proxy Networks

- **Measurement errors.** Each proxy is a noisy measurement of A^*.
- **Multiple sources.** Proxies from different sources.
  - Online networks, surveys, biological measurements, spatial data …

Types of Proxy Networks

- **Measurement errors.** Each proxy is a noisy measurement of A^*.
- **Multiple sources.** Proxies from different sources.
  - Online networks, surveys, biological measurements, spatial data …
- **Multilayer networks.** Each proxy represents a different type of relational structure (e.g., friendship, family, colleagues, etc.)

$$\begin{tikzcd} & \boldsymbol{A}^* \\ \boldsymbol{V} \arrow[ur] \arrow[r] \arrow[dr] & \boldsymbol{A}_1 \\ & \boldsymbol{A}_2 \end{tikzcd}$$

$$\begin{tikzcd} & \boldsymbol{A}_1 \\ \boldsymbol{A}^* \arrow[ur] \arrow[dr] & \\ & \boldsymbol{A}_2 \end{tikzcd}$$

Parametric Identifiability

- $\boldsymbol{\eta}$ parameterize the **outcomes** model $p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\eta})$.
- $\boldsymbol{\beta}_Z$ parameterize the **treatments** model $p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\beta}_Z)$.
- $\boldsymbol{\theta}$ parameterize the **latent network** model $p(\boldsymbol{A}^* \mid \boldsymbol{X}; \boldsymbol{\theta})$.
- γ parameterize the **proxy networks** model $p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}; \gamma)$.
- Let $\Theta = (\boldsymbol{\eta}, \boldsymbol{\beta}_Z, \boldsymbol{\theta}, \gamma)$ be the full parameter vector.

Parametric Identifiability

- $\boldsymbol{\eta}$ parameterize the **outcomes** model $p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\eta})$.
- $\boldsymbol{\beta}_Z$ parameterize the **treatments** model $p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\beta}_Z)$.
- $\boldsymbol{\theta}$ parameterize the **latent network** model $p(\boldsymbol{A}^* \mid \boldsymbol{X}; \boldsymbol{\theta})$.
- γ parameterize the **proxy networks** model $p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}; \gamma)$.
- Let $\boldsymbol{\Theta} = (\boldsymbol{\eta}, \boldsymbol{\beta}_Z, \boldsymbol{\theta}, \gamma)$ be the full parameter vector.
- The complete-data likelihood is (causal proxies + observational study)

$$p(\boldsymbol{\mathcal{D}}, \boldsymbol{A}^* \mid \boldsymbol{\Theta}) = p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}}) p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\gamma}) p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}).$$

- The observed-data likelihood is $p(\boldsymbol{\mathcal{D}} \mid \boldsymbol{\Theta}) = \sum_{\boldsymbol{A}^*} p(\boldsymbol{\mathcal{D}}, \boldsymbol{A}^* \mid \boldsymbol{\Theta})$.

Parametric Identifiability

- $\boldsymbol{\eta}$ parameterize the **outcomes** model $p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\eta})$.
- $\boldsymbol{\beta}_Z$ parameterize the **treatments** model $p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}; \boldsymbol{\beta}_Z)$.
- $\boldsymbol{\theta}$ parameterize the **latent network** model $p(\boldsymbol{A}^* \mid \boldsymbol{X}; \boldsymbol{\theta})$.
- γ parameterize the **proxy networks** model $p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}; \gamma)$.
- Let $\boldsymbol{\Theta} = (\boldsymbol{\eta}, \boldsymbol{\beta}_Z, \boldsymbol{\theta}, \gamma)$ be the full parameter vector.
- The complete-data likelihood is (causal proxies + observational study)

$$p(\boldsymbol{\mathcal{D}}, \boldsymbol{A}^* \mid \boldsymbol{\Theta}) = p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}}) p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}).$$

- The observed-data likelihood is $p(\boldsymbol{\mathcal{D}} \mid \boldsymbol{\Theta}) = \sum_{\boldsymbol{A}^*} p(\boldsymbol{\mathcal{D}}, \boldsymbol{A}^* \mid \boldsymbol{\Theta})$.
- **Under what conditions and models is Θ identifiable?**

Parametric Identifiability

- $(\boldsymbol{Y}, \boldsymbol{Z}) \perp\!\!\!\perp \mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}$.
  - Under causal proxies + observational study DAG without $\boldsymbol{U}$.
- Observed correlations between $\mathcal{A}$ and $\boldsymbol{Y}, \boldsymbol{Z}$ are mediated by $\boldsymbol{A}^*$.

Parametric Identifiability

- $(\boldsymbol{Y}, \boldsymbol{Z}) \perp\!\!\!\perp \mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}$.
  - Under causal proxies + observational study DAG without $\boldsymbol{U}$.
- Observed correlations between $\mathcal{A}$ and $\boldsymbol{Y}, \boldsymbol{Z}$ are mediated by $\boldsymbol{A}^*$.
- **Strategy:** Construct a system of moment equations
  - e.g., $\text{Cov}(\boldsymbol{Y}, \mathcal{A} \mid \boldsymbol{X}), \text{Cov}(\boldsymbol{Z}, \mathcal{A} \mid \boldsymbol{X}), \text{Cov}(\boldsymbol{A}_1, \boldsymbol{A}_2 \mid \boldsymbol{X})$
- If the moment map $\mathcal{M}(\boldsymbol{\Theta})$ is injective, $\boldsymbol{\Theta}$ is globally identifiable (Rothenberg, 1971).

Parametric Identifiability (Example)

$$
\begin{aligned}
A^*_{ij} &\sim \mathsf{Ber}(\theta), \\
A_{ij} \mid A^*_{ij} = k &\sim \mathsf{Ber}(\gamma_k),\ \ k = 0, 1, \\
Z_i &\sim \mathsf{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A^*_{ij} Z_j + \varepsilon_i
\end{aligned}
$$

Parametric Identifiability (Example)

$$
\begin{aligned}
A^*_{ij} &\sim \mathsf{Ber}(\theta), \\
A_{ij} \mid A^*_{ij} = k &\sim \mathsf{Ber}(\gamma_k), \;\; k = 0, 1, \\
Z_i &\sim \mathsf{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A^*_{ij} Z_j + \varepsilon_i
\end{aligned}
$$

- **Base Case:** Four unknowns $(\theta, \gamma_0, \gamma_1, \eta_2)$, but only three distinct equations $\mathbb{E}[Y_i], \Pr(A_{ij} = 1), \mathsf{Cov}(Y_i, d_i^{obs}) \Rightarrow$ **not identifiable.**
- We can't distnguish between dense networks with high error and sparse networks with low error.

Parametric Identifiability (Example)

$$
\begin{aligned}
A^*_{ij} &\sim \text{Ber}(\theta), \\
A_{ij} \mid A^*_{ij} = k &\sim \text{Ber}(\gamma_k),\ \ k = 0, 1, \\
Z_i &\sim \text{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A^*_{ij} Z_j + \varepsilon_i
\end{aligned}
$$

- **Global Identification is achieved** in any of these extensions:

Parametric Identifiability (Example)

$$\begin{aligned} A^*_{ij} &\sim \text{Ber}(\theta), \\ A_{ij} \mid A^*_{ij} = k &\sim \text{Ber}(\gamma_k), \;\; k = 0, 1, \\ Z_i &\sim \text{Ber}(p_z), \\ Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A^*_{ij} Z_j + \varepsilon_i \end{aligned}$$

- **Global Identification is achieved** in any of these extensions:
    1. If γ_0 is known or estimated from validation data.

Parametric Identifiability (Example)

$$
\begin{aligned}
A^*_{ij} &\sim \text{Ber}(\theta), \\
A_{ij} \mid A^*_{ij} = k &\sim \text{Ber}(\gamma_k), \;\; k = 0, 1, \\
Z_i &\sim \text{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A^*_{ij} Z_j + \varepsilon_i
\end{aligned}
$$

- **Global Identification is achieved** in any of these extensions:
  1. If γ_0 is known or estimated from validation data.
  2. **Multiple proxies:** If two proxies $\boldsymbol{A}_1, \boldsymbol{A}_2$ are available.

Parametric Identifiability (Example)

$$
\begin{aligned}
A_{ij}^* &\sim \text{Ber}(\theta), \\
A_{ij} \mid A_{ij}^* = k &\sim \text{Ber}(\gamma_k), \;\; k = 0, 1, \\
Z_i &\sim \text{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A_{ij}^* Z_j + \varepsilon_i
\end{aligned}
$$

- **Global Identification is achieved** in any of these extensions:
  1. If γ_0 is known or estimated from validation data.
  2. **Multiple proxies:** If two proxies $\boldsymbol{A}_1, \boldsymbol{A}_2$ are available.
  3. **Network-Dependent $\boldsymbol{Z}$:** For example, if $\Pr(Z_i = 1 \mid \boldsymbol{A}^*) \propto d_i^*$.

Parametric Identifiability (Example)

$$
\begin{aligned}
A_{ij}^* &\sim \text{Ber}(\theta), \\
A_{ij} \mid A_{ij}^* = k &\sim \text{Ber}(\gamma_k), \;\; k = 0, 1, \\
Z_i &\sim \text{Ber}(p_z), \\
Y_i &= \eta_1 Z_i + \eta_2 \sum_{j \neq i} A_{ij}^* Z_j + \varepsilon_i
\end{aligned}
$$

- **Global Identification is achieved** in any of these extensions:
  1. If γ_0 is known or estimated from validation data.
  2. **Multiple proxies:** If two proxies $\boldsymbol{A}_1, \boldsymbol{A}_2$ are available.
  3. **Network-Dependent $\boldsymbol{Z}$:** For example, if $\Pr(Z_i = 1 \mid \boldsymbol{A}^*) \propto d_i^*$.
  4. **Network-correlated $\boldsymbol{Y}$:** For example, if $Cov(\varepsilon_i, \varepsilon_j) = \rho A_{ij}^*$.

Bayesian Estimation

- Assume prior independence. The posterior (causal proxies + observational study) is

$$
\begin{aligned}
p(\boldsymbol{\Theta}, \boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}) \propto\ & p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\
& \times p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\mathbf{Z}}) p(\boldsymbol{\beta}_{\mathbf{Z}}) \\
& \times p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\gamma) \\
& \times p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).
\end{aligned}
$$

- Mixed space of continuous ($\boldsymbol{\Theta}$) and discrete ($\boldsymbol{A}^*$) parameters.

Bayesian Estimation

- Assume prior independence. The posterior (causal proxies + observational study) is

$$\begin{aligned} p(\boldsymbol{\Theta}, \boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}) \propto\ & p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\ & \times p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}}) p(\boldsymbol{\beta}_{\boldsymbol{Z}}) \\ & \times p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\gamma) \\ & \times p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}). \end{aligned}$$

- Mixed space of continuous ($\boldsymbol{\Theta}$) and discrete ($\boldsymbol{A}^*$) parameters.
- SOTA is to marginalize out $\boldsymbol{A}^*$ – not possible here!

Bayesian Estimation

- Assume prior independence. The posterior (causal proxies + observational study) is

$$
\begin{aligned}
p(\boldsymbol{\Theta}, \boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}) \propto\; & p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\
& \times p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}}) p(\boldsymbol{\beta}_{\boldsymbol{Z}}) \\
& \times p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\gamma) \\
& \times p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).
\end{aligned}
$$

- Mixed space of continuous ($\boldsymbol{\Theta}$) and discrete ($\boldsymbol{A}^*$) parameters.
- SOTA is to marginalize out $\boldsymbol{A}^*$ – not possible here!
- Possible solutions:
  - Continuous relaxations of $\boldsymbol{A}^*$.
  - Mixed Hamiltonian Monte Carlo.
  - Random-walk Metropolis within Gibbs.

Bayesian Estimation

- Assume prior independence. The posterior (causal proxies + observational study) is

$$
\begin{aligned}
p(\boldsymbol{\Theta}, \boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}) \propto\ & p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta}) p(\boldsymbol{\eta}) \\
& \times p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}}) p(\boldsymbol{\beta}_{\boldsymbol{Z}}) \\
& \times p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\gamma) \\
& \times p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}).
\end{aligned}
$$

- Mixed space of continuous ($\boldsymbol{\Theta}$) and discrete ($\boldsymbol{A}^*$) parameters.
- SOTA is to marginalize out $\boldsymbol{A}^*$ – not possible here!
- Possible solutions:
  - Continuous relaxations of $\boldsymbol{A}^*$.
  - Mixed Hamiltonian Monte Carlo.
  - Random-walk Metropolis within Gibbs.
- **Our solution:** Block Gibbs with Locally Informed Proposals for $\boldsymbol{A}^*$.

Locally Informed Proposals

- Updating $\boldsymbol{A}^*$ state corresponds to 'flipping' some entries ($A^*_{ij} = 1 \rightarrow A^*_{ij} = 0$ or $A^*_{ij} = 0 \rightarrow A^*_{ij} = 1$)
- Random-walk proposals flips $\Rightarrow$ high rejection rate and slow mixing.

Locally Informed Proposals

- Updating $\boldsymbol{A}^*$ state corresponds to 'flipping' some entries ($A^*_{ij} = 1 \to A^*_{ij} = 0$ or $A^*_{ij} = 0 \to A^*_{ij} = 1$)
- Random-walk proposals flips $\Rightarrow$ high rejection rate and slow mixing.
- **Solution:** use the conditional posterior to guide proposals:

$$p(\boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}, \boldsymbol{\Theta}) \propto \underbrace{p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta})}_{\text{Outcomes}} \underbrace{p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}})}_{\text{Treatments}} \underbrace{p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma)}_{\text{Proxies}} \underbrace{p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})}_{\text{Prior}}.$$

Locally Informed Proposals

- Updating $\boldsymbol{A}^*$ state corresponds to 'flipping' some entries ($A^*_{ij} = 1 \to A^*_{ij} = 0$ or $A^*_{ij} = 0 \to A^*_{ij} = 1$)
- Random-walk proposals flips $\Rightarrow$ high rejection rate and slow mixing.
- **Solution:** use the conditional posterior to guide proposals:

$$p(\boldsymbol{A}^* \mid \boldsymbol{\mathcal{D}}, \boldsymbol{\Theta}) \propto \underbrace{p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\eta})}_{\text{Outcomes}} \underbrace{p(\boldsymbol{Z} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\beta}_{\boldsymbol{Z}})}_{\text{Treatments}} \underbrace{p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma)}_{\text{Proxies}} \underbrace{p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})}_{\text{Prior}} .$$

- Construct proposal distribution that biases flips towards high posterior probability states.
- **Information Triangulation:** each module contributes to the proposal.

Locally Informed Proposals

- The LIP for updating $\boldsymbol{A}_t^*$ to $\boldsymbol{A}_{t+1}^*$ is with structure (Zanella, 2020):

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto g\left(\frac{p(\boldsymbol{A}_{t+1}^* \mid \cdot)}{p(\boldsymbol{A}_t^* \mid \cdot)}\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big),$$

Locally Informed Proposals

- The LIP for updating $\boldsymbol{A}_t^*$ to $\boldsymbol{A}_{t+1}^*$ is with structure (Zanella, 2020):

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto g\left(\frac{p(\boldsymbol{A}_{t+1}^* \mid \cdot)}{p(\boldsymbol{A}_t^* \mid \cdot)}\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big),$$

- $H(\boldsymbol{A}_t^*)$ is Hamming ball of size 1 (i.e., all states that differ in one entry).

Locally Informed Proposals

- The LIP for updating $\boldsymbol{A}_t^*$ to $\boldsymbol{A}_{t+1}^*$ is with structure (Zanella, 2020):

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto g\left(\frac{p(\boldsymbol{A}_{t+1}^* \mid \cdot)}{p(\boldsymbol{A}_t^* \mid \cdot)}\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big),$$

- $H(\boldsymbol{A}_t^*)$ is Hamming ball of size 1 (i.e., all states that differ in one entry).
- $g(\cdot)$ satisfy the balancing condition $g(a) = ag(1/a)$ for locally reversible with respect to $p(\boldsymbol{A}^* \mid \cdot)$.
- e.g., $g(a) = \sqrt{a}$ or $g(a) = \frac{a}{1+a}$.

Locally Informed Proposals

- The LIP for updating $\boldsymbol{A}_t^*$ to $\boldsymbol{A}_{t+1}^*$ is with structure (Zanella, 2020):

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto g\left(\frac{p(\boldsymbol{A}_{t+1}^* \mid \cdot)}{p(\boldsymbol{A}_t^* \mid \cdot)}\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big),$$

- $H(\boldsymbol{A}_t^*)$ is Hamming ball of size 1 (i.e., all states that differ in one entry).
- $g(\cdot)$ satisfy the balancing condition $g(a) = ag(1/a)$ for locally reversible with respect to $p(\boldsymbol{A}^* \mid \cdot)$.
- e.g., $g(a) = \sqrt{a}$ or $g(a) = \frac{a}{1+a}$.
- Asymptotic efficent and good mixing (Zanella, 2020).

Locally Informed Proposals

- Let $\Delta(\boldsymbol{A}_{t+1}^*, \boldsymbol{A}_t^* \mid \cdot) = \log(p(\boldsymbol{A}_{t+1}^* \mid \cdot)) - \log(p(\boldsymbol{A}_t^* \mid \cdot))$
- Assume that $g(a) = \sqrt{a}$. Then,

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto \exp\left(\frac{1}{2}\Delta(\boldsymbol{A}_{t+1}^*, \boldsymbol{A}_t^* \mid \cdot)\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big)$$

Locally Informed Proposals

- Let $\Delta(\boldsymbol{A}_{t+1}^*, \boldsymbol{A}_t^* \mid \cdot) = \log(p(\boldsymbol{A}_{t+1}^* \mid \cdot)) - \log(p(\boldsymbol{A}_t^* \mid \cdot))$
- Assume that $g(a) = \sqrt{a}$. Then,

$$Q(\boldsymbol{A}_{t+1}^* \mid \boldsymbol{A}_t^*, \cdot) \propto \exp\left(\frac{1}{2}\Delta(\boldsymbol{A}_{t+1}^*, \boldsymbol{A}_t^* \mid \cdot)\right)\mathbb{I}\big(\boldsymbol{A}_{t+1}^* \in H(\boldsymbol{A}_t^*)\big)$$

- Equivalent to sample an entry with probability Softmax$\left(\frac{1}{2}\Delta(ij, t \mid \cdot)\right)$ and flip its value.
- Entails computing all $\Delta(ij, t \mid \cdot)$ differences which is $\mathcal{O}(N^2)$ operations.

Locally Informed Proposals

- Let $\Delta(\boldsymbol{A}^*_{t+1}, \boldsymbol{A}^*_t \mid \cdot) = \log(p(\boldsymbol{A}^*_{t+1} \mid \cdot)) - \log(p(\boldsymbol{A}^*_t \mid \cdot))$
- Assume that $g(a) = \sqrt{a}$. Then,

$$Q(\boldsymbol{A}^*_{t+1} \mid \boldsymbol{A}^*_t, \cdot) \propto \exp\left(\frac{1}{2}\Delta(\boldsymbol{A}^*_{t+1}, \boldsymbol{A}^*_t \mid \cdot)\right)\mathbb{I}\left(\boldsymbol{A}^*_{t+1} \in H(\boldsymbol{A}^*_t)\right)$$

- Equivalent to sample an entry with probability Softmax$\left(\frac{1}{2}\Delta(ij, t \mid \cdot)\right)$ and flip its value.
- Entails computing all $\Delta(ij, t \mid \cdot)$ differences which is $\mathcal{O}(N^2)$ operations.
- Gradient-based approximations using First-order Taylor series (Grathwohl et al., 2021):

$$\widetilde{\Delta}(ij, t \mid \cdot) = -(2A^*_{t,ij} - 1)\frac{\partial \log(p(\boldsymbol{A}^*_t \mid \cdot))}{\partial A^*_{t,ij}} \approx \Delta(ij, t \mid \cdot)$$

- Single gradient evaluation and implemented via automatic differentiation.

Algorithm A Single $\boldsymbol{A}^*$ Update with LIP

Input: Data $\boldsymbol{\mathcal{D}}$, continuous parameters $\boldsymbol{\Theta}$, current state $\boldsymbol{A}_t^*$, log-posterior $\log p(\boldsymbol{A}_t^* \mid \cdot)$, number of entries to update $K \geq 1$.

1: Compute differences $\widetilde{\boldsymbol{\Delta}}(\boldsymbol{A}_t^* \mid \cdot) \equiv \widetilde{\boldsymbol{\Delta}}(\boldsymbol{A}_t^* \mid \boldsymbol{\mathcal{D}}, \boldsymbol{\Theta})$.

2: Compute $q(ij \mid \boldsymbol{A}_t^*, \cdot) = \text{Softmax}\left(\frac{1}{2}\widetilde{\boldsymbol{\Delta}}(\boldsymbol{A}_t^* \mid \cdot)\right)$.

3: Sample without replacement K edges $\mathcal{I}$ using Gumbel-Max trick with probabilities $q(ij \mid \boldsymbol{A}_t^*, \cdot)$.

4: Compute *forward* probability $q(\mathcal{I} \mid \boldsymbol{A}_t^*, \cdot) = \prod_{ij \in \mathcal{I}} q(ij \mid \boldsymbol{A}_t^*, \cdot)$.

5: $\boldsymbol{A}_{t+1}^* \leftarrow \texttt{FlipEntries}(\boldsymbol{A}_t^*, \mathcal{I})$.

6: Compute *backward* probability $q(\mathcal{I} \mid \boldsymbol{A}_{t+1}^*, \cdot) = \prod_{ij \in \mathcal{I}} q(ij \mid \boldsymbol{A}_{t+1}^*, \cdot)$.

7: Accept with probability

$$\min\left(\exp\left(\Delta(\boldsymbol{A}_{t+1}^*, \boldsymbol{A}_t^* \mid \cdot)\right) \frac{q(\mathcal{I} \mid \boldsymbol{A}_{t+1}^*, \cdot)}{q(\mathcal{I} \mid \boldsymbol{A}_t^*, \cdot)}, 1\right).$$

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).
- **Initialization strategy:** Sample from cut-posterior followed by LIP refinement.

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).
- **Initialization strategy:** Sample from cut-posterior followed by LIP refinement.
    1. Sample from $p(\boldsymbol{\theta}, \gamma \mid \mathcal{A}, \boldsymbol{X}) \propto p(\boldsymbol{\theta})p(\gamma) \sum_{\boldsymbol{A}^*} p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma)p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})$, and save the posterior means $\hat{\boldsymbol{\theta}}, \hat{\gamma}$.

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).
- **Initialization strategy:** Sample from cut-posterior followed by LIP refinement.
    1. Sample from $p(\boldsymbol{\theta}, \gamma \mid \mathcal{A}, \boldsymbol{X}) \propto p(\boldsymbol{\theta})p(\gamma)\sum_{\boldsymbol{A}^*} p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma)p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})$, and save the posterior means $\hat{\boldsymbol{\theta}}, \hat{\gamma}$.
    2. Sample $\boldsymbol{A}^*$ from $p(\boldsymbol{A}^* \mid \mathcal{A}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}, \hat{\gamma})$ and select $\boldsymbol{A}_0^*$ as the network with the highest conditional probability.

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).
- **Initialization strategy:** Sample from cut-posterior followed by LIP refinement.
  1. Sample from $p(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid \mathcal{A}, \boldsymbol{X}) \propto p(\boldsymbol{\theta})p(\boldsymbol{\gamma}) \sum_{\boldsymbol{A}^*} p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{\gamma})p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})$, and save the posterior means $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}}$.
  2. Sample $\boldsymbol{A}^*$ from $p(\boldsymbol{A}^* \mid \mathcal{A}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})$ and select $\boldsymbol{A}_0^*$ as the network with the highest conditional probability.
  3. Estimate the outcome model parameters $\boldsymbol{\eta}$ with $\boldsymbol{A}_0^*$ as a fixed covariate, yielding initial estimates $\boldsymbol{\Theta}_0$.

Block Gibbs Sampler

- Alternate between updating $\boldsymbol{A}^*$ with LIP and $\boldsymbol{\Theta}$ with any continuous kernel (HMC, NUTS, etc.)
- We perform L updates of $\boldsymbol{A}^*$ per iteration (maximum of $L \times K$ total flips per iteration).
- **Initialization strategy:** Sample from cut-posterior followed by LIP refinement.
  1. Sample from $p(\boldsymbol{\theta}, \gamma \mid \mathcal{A}, \boldsymbol{X}) \propto p(\boldsymbol{\theta}) p(\gamma) \sum_{\boldsymbol{A}^*} p(\mathcal{A} \mid \boldsymbol{A}^*, \boldsymbol{X}, \gamma) p(\boldsymbol{A}^* \mid \boldsymbol{X}, \boldsymbol{\theta})$, and save the posterior means $\hat{\boldsymbol{\theta}}, \hat{\gamma}$.
  2. Sample $\boldsymbol{A}^*$ from $p(\boldsymbol{A}^* \mid \mathcal{A}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}, \hat{\gamma})$ and select $\boldsymbol{A}_0^*$ as the network with the highest conditional probability.
  3. Estimate the outcome model parameters $\boldsymbol{\eta}$ with $\boldsymbol{A}_0^*$ as a fixed covariate, yielding initial estimates $\boldsymbol{\Theta}_0$.
  4. Refine $\boldsymbol{A}_0^*$ by performing a short sequence of LIP updates given $\boldsymbol{\Theta}_0$.

Numerical Illustration

- Simulated data with $N = 500$ units.
- Latent network $\Pr(A^*_{ij} = 1) = \text{expit}(\boldsymbol{\theta}^\top \boldsymbol{X})$ with edge-level covariates.
- Single or two proxy networks from differential measurement error models.
- Treatments assigned as $\Pr(Z_i = 1) \propto d^*_i$.

Numerical Illustration

- Simulated data with $N = 500$ units.
- Latent network $\Pr(A^*_{ij} = 1) = \text{expit}(\boldsymbol{\theta}^\top \boldsymbol{X})$ with edge-level covariates.
- Single or two proxy networks from differential measurement error models.
- Treatments assigned as $\Pr(Z_i = 1) \propto d^*_i$.
- Outcomes generated as $Y_i = \boldsymbol{\mu} + \varepsilon_i$
  - $\mu_i = \eta_1 Z_i + \eta_2 \phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}^*) + \boldsymbol{\eta}'_x \boldsymbol{X}_i$
  - $\phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}^*) = \sum_{j \neq i} w_j Z_j A^*_{ij}$ with $w_j = \frac{d^*_j}{N-1}$

Numerical Illustration

- Simulated data with $N = 500$ units.
- Latent network $\Pr(A^*_{ij} = 1) = \text{expit}(\boldsymbol{\theta}^\top \boldsymbol{X})$ with edge-level covariates.
- Single or two proxy networks from differential measurement error models.
- Treatments assigned as $\Pr(Z_i = 1) \propto d^*_i$.
- Outcomes generated as $Y_i = \boldsymbol{\mu} + \varepsilon_i$
  - $\mu_i = \eta_1 Z_i + \eta_2 \phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}^*) + \boldsymbol{\eta}'_x \boldsymbol{X}_i$
  - $\phi_1(\boldsymbol{Z}_{-i}, \boldsymbol{A}^*) = \sum_{j \neq i} w_j Z_j A^*_{ij}$ with $w_j = \frac{d^*_j}{N-1}$
- **Estimands:**
  1. Dynamic policy assigning treatment through two thresholds of one covariate.
  2. Total Treatment Effect: treating all versus none of the units.

Numerical Illustration

- **Baseline methods:**
  1. **True:** using true network A^* (“True”).
  2. **Naive:** using one proxy network A_1 as if it were true (“Obs.”).

Numerical Illustration

- **Baseline methods:**
  1. **True:** using true network $\boldsymbol{A}^*$ ("True").
  2. **Naive:** using one proxy network $\boldsymbol{A}_1$ as if it were true ("Obs.").
  3. **Continuous Relaxation** of $\boldsymbol{A}^*_{ij}$ into $[0, 1]$ using the CONCRETE distribution (Maddison et al., 2017).

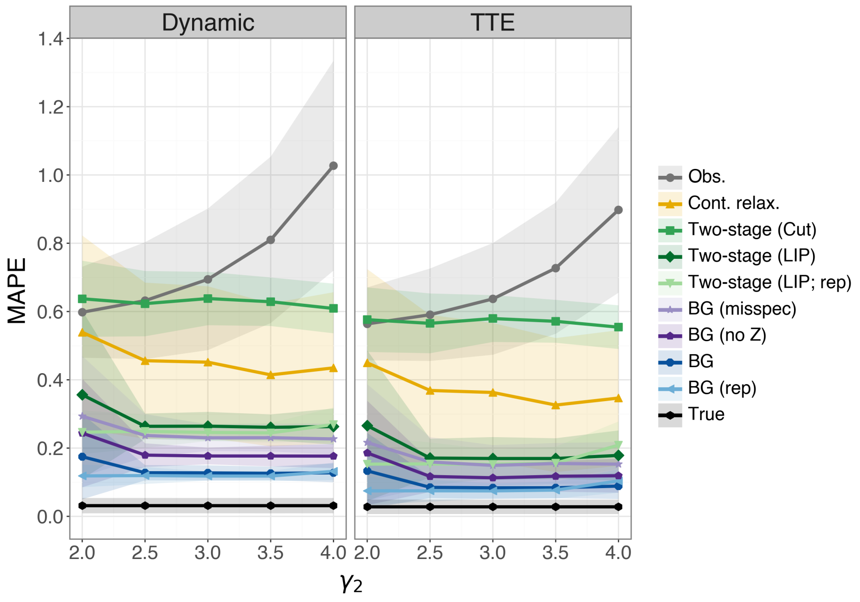
Numerical Illustration

- **Baseline methods:**
  1. **True:** using true network $\boldsymbol{A}^*$ ("True").
  2. **Naive:** using one proxy network $\boldsymbol{A}_1$ as if it were true ("Obs.").
  3. **Continuous Relaxation** of $\boldsymbol{A}^*_{ij}$ into $[0, 1]$ using the CONCRETE distribution (Maddison et al., 2017).
  4. **Two-stage samplers:** estimate $\boldsymbol{A}^*$ from proxies $\mathcal{A}$ and plug-in into outcome model.
     - Without LIP refinement ("Cut")
     - With LIP refinement with one ("LIP") or two proxies ("LIP; rep").

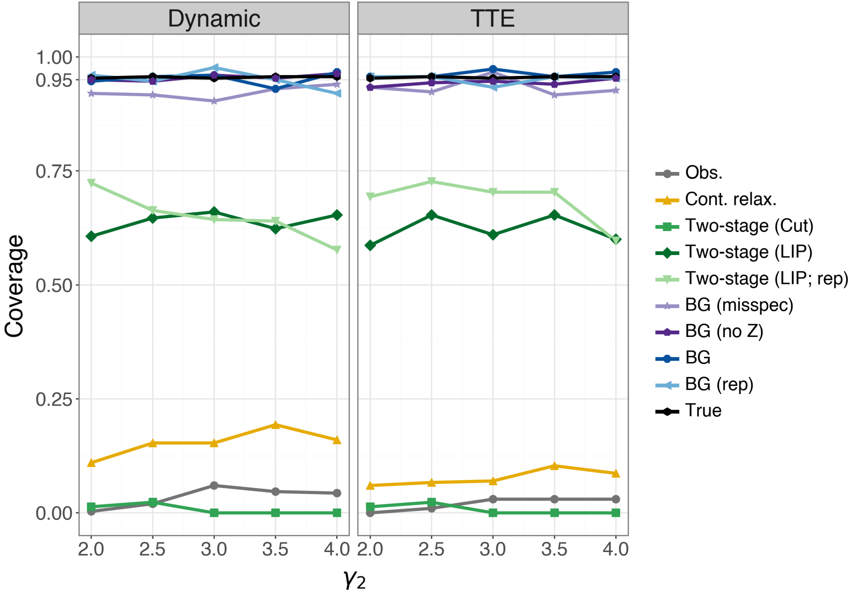
Numerical Illustration

- **Baseline methods:**
  1. **True:** using true network $\boldsymbol{A}^*$ ("True").
  2. **Naive:** using one proxy network $\boldsymbol{A}_1$ as if it were true ("Obs.").
  3. **Continuous Relaxation** of $\boldsymbol{A}^*_{ij}$ into $[0, 1]$ using the CONCRETE distribution (Maddison et al., 2017).
  4. **Two-stage samplers:** estimate $\boldsymbol{A}^*$ from proxies $\mathcal{A}$ and plug-in into outcome model.
     - Without LIP refinement ("Cut")
     - With LIP refinement with one ("LIP") or two proxies ("LIP; rep").
  5. **Block Gibbs** (with $K = L = 1$ updates per iteration):
     - With one ("BG") or two proxies ("BG (rep)").
     - Without the treatment model ("BG (no Z)").
     - With misspecification due to omitted covariate ("BG (misspec)").

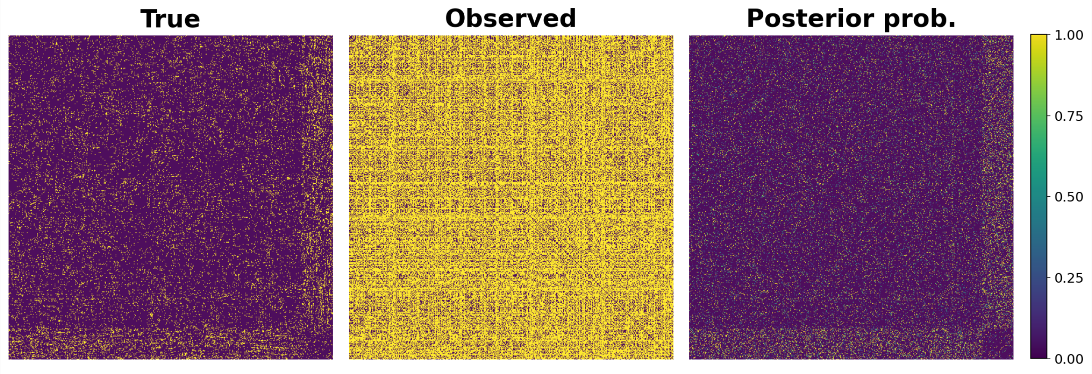
$$MAPE = N^{-1}\sum_{i=1}^{N}\left|\frac{\hat{\tau}_i-\tau_i}{\tau_i}\right|$$



Empirical coverage of 95% CIs



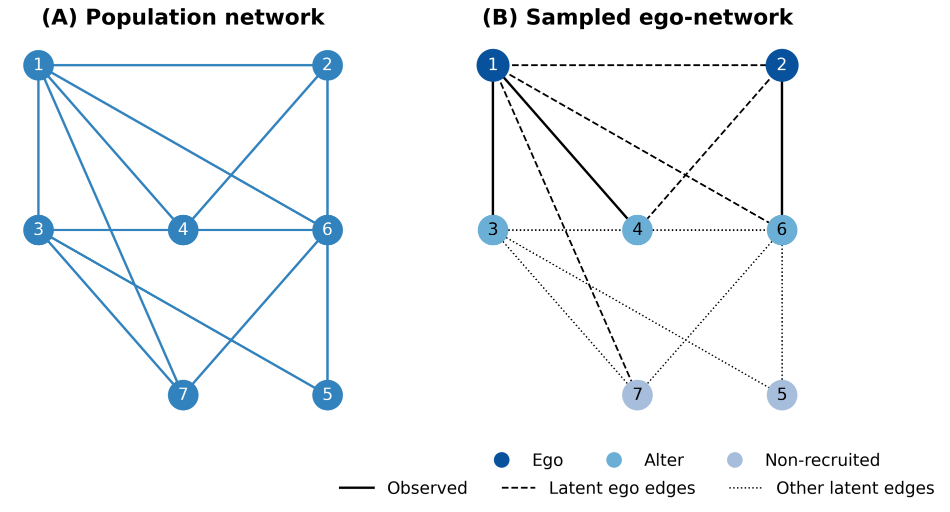
Network reconstruction in one iteration



Extensions and Future Work

- Estimands of intervention on the interference network structure.
- DOE with latent or uncertain interference network.
- Dimension of latent network is $\mathcal{O}(N^2)$ – feasible for small to moderate networks.
- Robustness to model misspecification.

Egocentric-Network Randomized Trials



Thank You!

- BW and DN. Causal Inference with Misspecified Network Interference Structure. *Biometrics*, 2026. Discussion paper.
- BW and DN. Bayesian Estimation of Causal Effects Using Proxies of a Latent Interference Network. R&R at *JMLR*.
- BW and DN. Sensitivity Analysis for Contamination in Egocentric-Network Randomized Trials with Interference. *To be submitted*.

DAG given A^*

$$
\begin{tikzcd}
Z_2 \arrow[r] \arrow[dr] & Y_{21} \arrow[r] \arrow[dr] & Y_{22} \\
Z_1 \arrow[r] \arrow[ur] & Y_{11} \arrow[r] \arrow[ur] & Y_{12}
\end{tikzcd}
$$

Parameteric Identifiability – Local

We can write the observed-data Fisher Information Matrix as

$$\mathscr{I}_{obs}(\boldsymbol{\Theta}) = \mathscr{I}_{comp}(\boldsymbol{\Theta}) - \mathscr{I}_{miss}(\boldsymbol{\Theta}).$$

- $\boldsymbol{\Theta}$ is locally identifiable at $\boldsymbol{\Theta}_0$ if $\mathscr{I}_{obs}(\boldsymbol{\Theta}_0)$ nonsingular.
- That will be the case if $\mathscr{I}_{obs}(\boldsymbol{\Theta})$ is strictly block-diagnoally dominant matrix.
- Intuitively, identification is stronger when the data modules (outcomes, treatments, proxies) depend on the latent network $\boldsymbol{A}^*$ through *distinct* structural features.

Parametric Identifiability (Example)

Base Case:

- $\mathbb{E}[Z_i] = p_z$
- $\mathbb{E}[Y_i \mid Z_i = z] = \eta_1 z + \eta_2(N-1)p_z\theta$
- $\mathbb{E}[Y_i] = \eta_1 p_z + \eta_2(N-1)p_z\theta$
- $\mu_A = \Pr(A_{ij} = 1) = \gamma_1\theta + \gamma_0(1-\theta)$
- Cov$(Y_i, d_i^{obs}) = \eta_2(N-1)p_z\theta(\gamma_1 - \mu_A)$

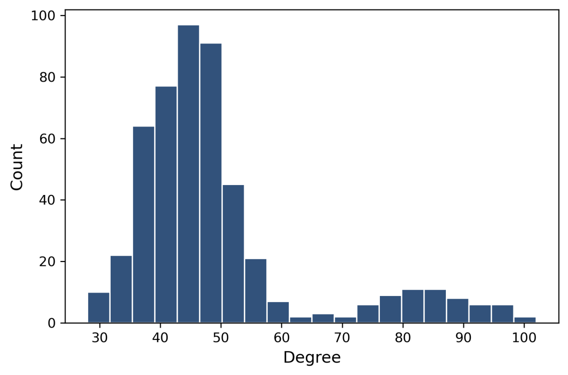
1. **Two proxies:** $m_{AD} = \frac{2}{N(N-1)}\sum_{i<j}|A_{1,ij} - A_{2,ij}|$ and $\mathbb{E}[m_{AD}] = 2\left[\gamma_1(1-\gamma_1)\theta + \gamma_0(1-\gamma_0)(1-\theta)\right]$
2. **Network-dependent *Z*:** $\mathbb{E}[Z_iZ_j] = \frac{p_z^2}{(N-1)^2}\left[\frac{1}{\theta} + ((N-1)^2 - 1)\right]$ and $Cov(Z_i, d_i^{obs}) = p_z(1-\theta)(\gamma_1 - \gamma_0)$
3. **Network-correlated *Y*:** $\mathbb{E}[Y_iY_j \mid A_{ij} = 1] - \mathbb{E}[Y_iY_j \mid A_{ij} = 0] = \left[\frac{\gamma_1\theta}{\mu_A} - \frac{(1-\gamma_1)\theta}{1-\mu_A}\right]\left[2\eta_2 p_z(\mathbb{E}[Y_i] - \eta_2\theta p_z) + \eta_2^2 p_z^2 + \rho.\right]$.

Cut-posterior

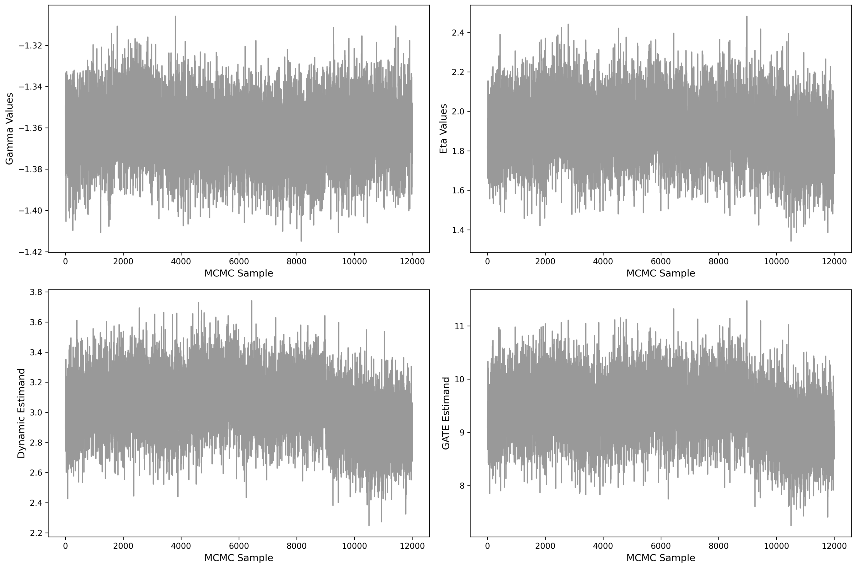
$$p(\boldsymbol{\eta}, \boldsymbol{\theta}, \gamma, \boldsymbol{A}^* \mid \mathcal{D}) \propto \underbrace{p(\boldsymbol{\eta} \mid \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Outcome module}} \underbrace{p(\boldsymbol{\theta}, \gamma, \boldsymbol{A}^* \mid \mathcal{A}, \boldsymbol{X})}_{\text{Network module}} \underbrace{p(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{A}^*, \boldsymbol{X})}_{\text{Feedback term}}.$$

Numerical Illustration

Degree distribution of the A^* in one iteration



Traceplot of key parameters and estimands



Numerical Illustration – Full Details

$$
\begin{aligned}
X_{1,i} \sim N(0,1), &\quad X_{2,i} \sim Ber(0.1) \\
\Pr(A^*_{ij} = 1 \mid \boldsymbol{X}, \boldsymbol{\theta}) &= s\big(-2 + \widetilde{X}_{2,ij} \big) \\
\Pr(A_{ij} = 1 \mid A^*_{ij}, \boldsymbol{X}, \boldsymbol{\gamma}) &= s\big(A^*_{ij}\gamma_0 + (1 - A^*_{ij})(\gamma_1 + \gamma_2 \widetilde{X}_{1,ij} + \gamma_3 \widetilde{X}_{2,ij}) \big) \\
\Pr(Z_i = 1) &= p_z \frac{d^*_i}{N-1} \\
\mu_i &= -1 + 3Z_i + 2\phi_1 - 0.5X_{1,i} \\
Y_i &= \mu_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,1)
\end{aligned}
$$

where we set $\gamma_0 =$ logit$(0.95) - \gamma_2/2$, $\gamma_1 =$ logit$(0.05) + \gamma_2/2$, and took $\gamma_2 \in \{2, 2.5, 3, 3.5, 4\}$. Edge-level covariates are defined as $\widetilde{X}_{1,ij} = \big|X_{1,i} - X_{1,j}\big|$, capturing the covariate distance, and $\widetilde{X}_{2,ij} = \mathbb{I}\{X_{2,i} + X_{2,j} = 1\}$, an indicator of exactly one of the two units have $X_2 = 1$.

Numerical Illustration – Priors and Estimands

We assume $N(0, 3^2)$ priors for all coefficients and $\sigma \sim$ HalfNormal$(0, 3^2)$.

The dynamic estimand is $\tau(h_{c_1}, h_{c_2}) = \mu(h_{c_1}) - \mu(h_{c_2})$, with thresholds $c_1 = 0.75$ and $c_2 = 1.5$. The static policy is the Total Treatment Effects (TTE), $\tau(\mathbf{1}, \mathbf{0})$, comparing the scenarios where all units are treated versus none.

Semi-synthetic Data Application

- Multilayer network dataset collected by Magnani et al. (2013), which originally contained five layers describing social connections among N = 61 employees at Aarhus University's Department of Computer Science. We remove the co-authorship layer.
- we sequentially designated each layer as the latent interference network $\boldsymbol{A}^*$, treating the remaining three layers as observed proxies $\mathcal{A}$. The joint distribution of the four layers was modeled using LSM with shared bivariate latent positions $\boldsymbol{V}$

$$\Pr(A_{b,ij} = 1 \mid \boldsymbol{V}, \boldsymbol{\gamma}) = s\left(\gamma_{b,0} - e^{\gamma_{b,1}} \|\boldsymbol{V}_i - \boldsymbol{V}_j\|_2\right),$$

  where $\gamma_{b,0}$ controls the baseline probability of edge creation in layer b and $\gamma_{b,1}$ controls how latent distances affects edge probabilities.
- We assumed $\boldsymbol{V}_i \sim N_2(\mathbf{0}, I)$, and hierarchical priors for the layer-specific parameters

$$\gamma_{b,j} \sim N(\mu_j, \sigma_j^2),\ \ \mu_j \sim N(0, 3^2),\ \ \sigma_j \sim \text{HalfNormal}(0, 3^2),\ \ j = 0, 1.$$

Semi-synthetic Data Application

Method	*Facebook*	*Leisure*	*Lunch*	*Work*
True	0.070 (0.045)	0.104 (0.080)	0.055 (0.041)	0.040 (0.024)
BG	0.132 (0.059)	0.157 (0.095)	0.111 (0.049)	0.096 (0.036)
Two-stage	0.397 (0.099)	0.187 (0.059)	0.169 (0.044)	0.287 (0.073)
OR	0.783 (0.198)	0.419 (0.229)	0.272 (0.057)	0.320 (0.065)
AND	0.501 (0.093)	0.233 (0.061)	0.425 (0.072)	0.324 (0.055)

Table: Mean (SD) of MAPE values across 300 iterations of the TTE estimand by layer and estimation method.

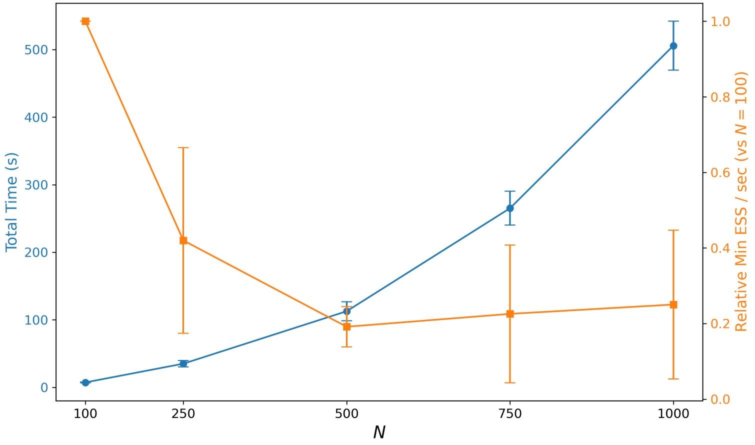


Figure: Scaling and mixing of the BG across 10 iterations in one setup. For varying sample sizes N, the figure shows the average ($\pm$ SD) of wall-clock time (in seconds) to obtain 1.2×10^4 MCMC samples (left y-axis, blue), and the relative minimal effective sample size (ESS) per second compared to $N = 100$ (right y-axis, orange). Clock-wall times were obtained on a PC equipped with a 12-core Apple Silicon (ARM64) processor and 48 GB of RAM, running macOS (Darwin Kernel 24.6.0).