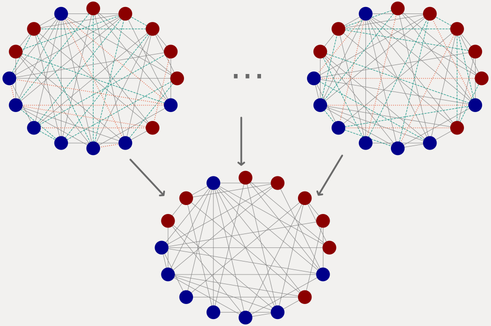
# Estimating Causal Effects Using Proxy Interference Networks

Bar Weinstein
Daniel Nevo

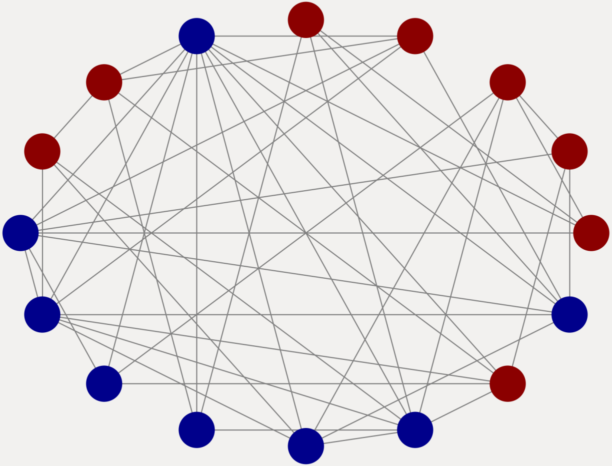
Statistics & OR
Tel Aviv University

IDSAI 2025



- **Causal Inference.** Estimate the effect of treatment on an outcome.
- **Interference.** Treatment of one unit affect the outcomes of others.
- Treatments spreads through a network.
  - Nodes: units; Edges: magntiude of pairwise interfernce.

- **Causal Inference.** Estimate the effect of treatment on an outcome.
- **Interference.** Treatment of one unit affect the outcomes of others.
- Treatments spreads through a network.
  - Nodes: units; Edges: magntiude of pairwise interfernce.
- Examples:
  - *Social networks.* Transmission of information, behavior, encouragements, etc.
  - *Epidemiology.* Mitigating spread of infectious diseases or addictive drugs.
  - A/B testing in marketplaces.



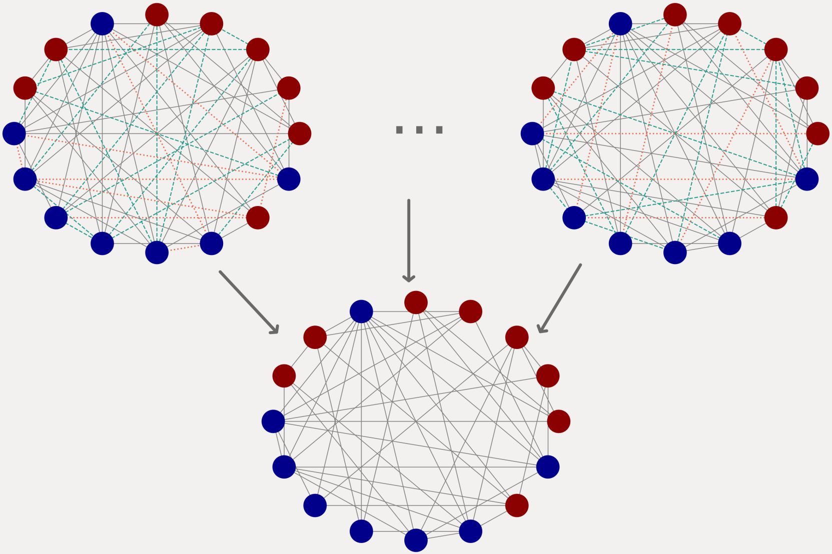
## The Challenge

- 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

- 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?**



# Illustrative Example – Paluck et al. (2016)

- Field experiement in 56 middle-schools.
- Study how anti-conflict education spread through social networks.
- Measured social networks using self-reported friendships.
  - Bi-layer networks: frequently interacted and best friends.
  - Measured at pre- and post-intervention period.

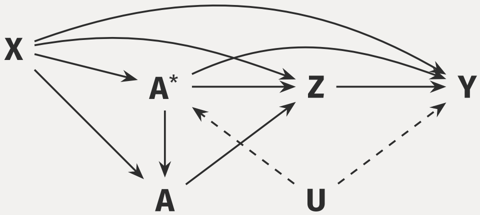
# Illustrative Example – Paluck et al. (2016)

- Field experiement in 56 middle-schools.
- Study how anti-conflict education spread through social networks.
- Measured social networks using self-reported friendships.
  - Bi-layer networks: frequently interacted and best friends.
  - Measured at pre- and post-intervention period.
- Which of the networks, if any, is the true network?
- **Objective:** Estimate the intervention effects using the proxy networks.

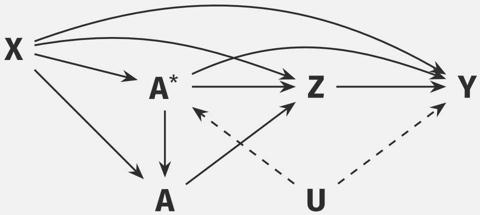
# Formal Setup

- 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):



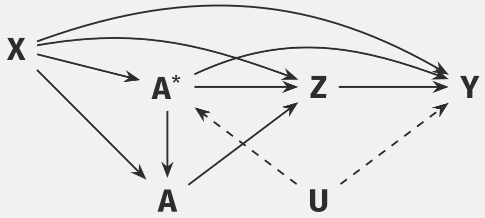
- Population-level directed acyclic graph (DAG):




- 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.* $p\left(\mathbf{Y}|\mathbf{Z},\mathbf{A}^*,\mathbf{X},\eta\right)$.

# 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.* $p\left(\mathbf{Y} | \mathbf{Z}, \mathbf{A}^*, \mathbf{X}, \eta\right)$.

- Outcome often simplified to $p\left(Y_i | Z_i, \mathbf{X}_i, \phi_1(\mathbf{Z}_{-i}, \mathbf{A}^*), \phi_2(\mathbf{X}_{-i}, \mathbf{A}^*), \phi_{3,i}(\mathbf{A}^*)\right)$.

# Causal Estimands

Causal effects are the impact of hypothetical interventions on $\mathbf{Z}$.

Can be viewed as population-level treatment assignment policies.

Causal effects are the impact of hypothetical interventions on $\mathbf{Z}$.
Can be viewed as population-level treatment assignment policies.

1. **Static.** $\mathbb{E}\left[Y_i | do(\mathbf{Z} = \mathbf{z}), \mathbf{X}, \mathbf{A}^*\right]$.
   - Treating all ($\mathbf{z} = 1$) versus none ($\mathbf{z} = 0$).
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.** $\mathbb{E}_{\pi_\alpha(\mathbf{Z})}\mathbb{E}\left[Y_i | do(\mathbf{Z}), \mathbf{X}, \mathbf{A}^*\right]$.
   - Expected impact of randomly treating $\alpha_1$ vs $\alpha_0$ percent of units.

# Bayesian Inference

- Observed data $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. Latent variables $(\mathbf{A}^*, \eta, \gamma, \theta)$.

- Observed data $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. Latent variables $(\mathbf{A}^*, \eta, \gamma, \theta)$.
- Posterior distribution:

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

- Observed data $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. Latent variables $(\mathbf{A}^*, \eta, \gamma, \theta)$.
- 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 of continuous $(\eta, \gamma, \theta)$ and discrete $\mathbf{A}^*$. Discrete have $O(N^2)$ terms!

- Observed data $\mathbf{O} = (\mathbf{Y}, \mathbf{Z}, \mathbf{X}, \mathbf{A})$. Latent variables $(\mathbf{A}^*, \eta, \gamma, \theta)$.
- 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 of continuous $(\eta, \gamma, \theta)$ and discrete $\mathbf{A}^*$. Discrete have $O(N^2)$ terms!
- Propose two sampling schemes:
    1. *Modularization.* "break" the posterior into smaller, more manageable parts.
    2. *Gibbs.* Iterate between continuous and discrete. Sample discrete with *Local Informed Proposals.*

# Simulations

Two figures: MAPE of estimated treatment effects and MAE of exposure mapping.

Results of Paluck et al. (2016) analysis.


**Figure 1:** Figure caption.

| | Heading 1 | Heading 2 |
|---|---|---|
| Row 1 | $v_{11}$ | $v_{12}$ |
| Row 2 | $v_{21}$ | $v_{22}$ |
| Row 3 | $v_{31}$ | $v_{32}$ |

**Table 1:** Table caption.

Thanks for using **Focus**!

# References

Donald E. Knuth.
**Computer programming as an art.**
*Commun. ACM*, pages 667–673, 1974.

Donald E. Knuth.
**Two notes on notation.**
*Amer. Math. Monthly*, 99:403–422, 1992.

Leslie Lamport.
***LaTeX: A Document Preparation System.***
Pearson Education India, 1994.

# Appendix

More math details.