Estimating network-mediated causal effects via spectral embeddings

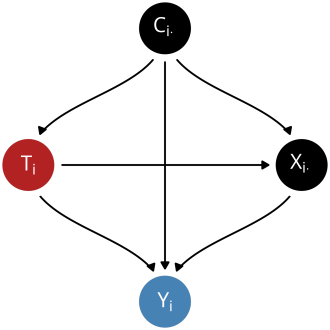
Alex Hayes, Mark Fredrickson & Keith Levin

2023-08-09 @ JSM 2023
Recent Developments in Causal Inference

Department of Statistics, University of Wisconsin-Madison

Causal mediation

Treatment	T_i	$\in \{0, 1\}$
Outcome	Y_i	$\in \mathbb{R}$
Mediators	$X_{i\cdot}$	$\in \mathbb{R}^{1 \times d}$
Confounders	$C_{i\cdot}$	$\in \mathbb{R}^{1 \times p}$



Decompose effect of T_i on Y_i:

1. Effect operating along $T_i \rightarrow Y_i$ path (direct)
2. Effect operating along $T_i \rightarrow X_{i\cdot} \rightarrow Y_i$ path (indirect)

Decomposing the average treatment effect

$Y_i(t)$: counterfactual value of Y_i when T_i is set to t. The average treatment effect is

$$\Psi_{\mathrm{ate}} = \mathbb{E}[Y_i(t) - Y_i(t^*)]$$

which decomposes into the natural direct effect (not operating through $X_{i\cdot}$) and natural indirect effect (operating through $X_{i\cdot}$)

$$\begin{aligned}
\Psi_{\mathrm{ate}} &= \Psi_{\mathrm{nde}} + \Psi_{\mathrm{nie}} \\
\Psi_{\mathrm{nde}} &= \mathbb{E}[Y_i(t, X_{i\cdot}(t^*)) - Y_i(t^*, X_{i\cdot}(t^*))] \\
\Psi_{\mathrm{nie}} &= \mathbb{E}[Y_i(t, X_{i\cdot}(t)) - Y_i(t, X_{i\cdot}(t^*))]
\end{aligned}$$

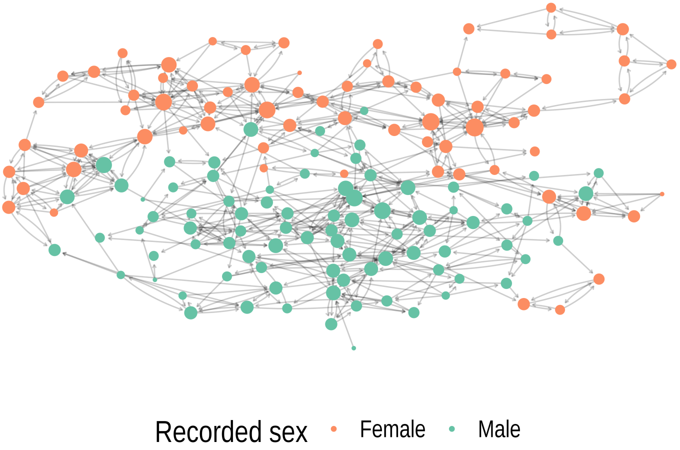
This talk: social networks as mediators

In particular: friends groups in social networks as mediators

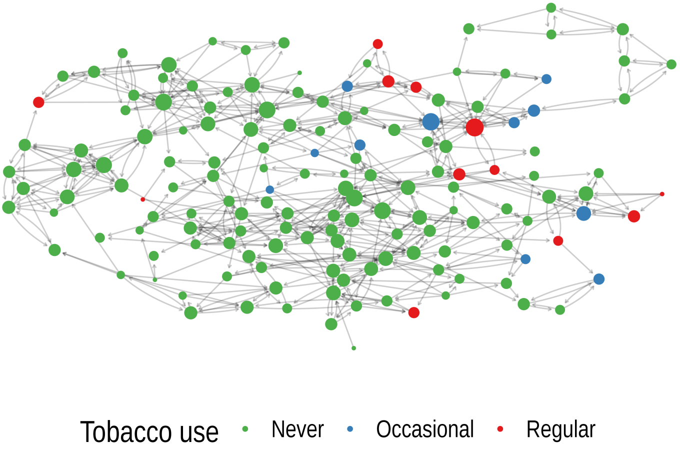
Motivating example: friend groups mediate the effect of sex on smoking in an adolescent social network

- Girls smoke than boys
- Girls and boys in different friend groups
- Smoking varies with friend group

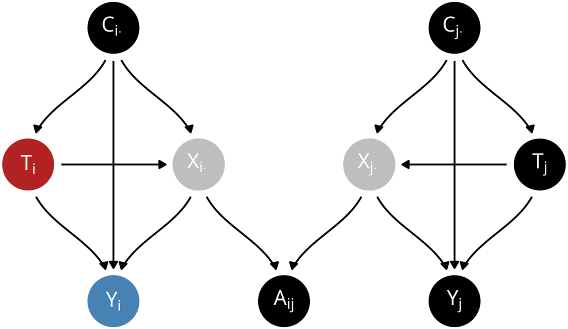
Teenage Friends and Lifestyle Study, Glasgow, 1996



Teenage Friends and Lifestyle Study, Glasgow, 1996



Mediation on a social network



Adjacency matrix	A	$\in \mathbb{R}^{n \times n}$
Edge $i \sim j$	A_{ij}	$\in \mathbb{R}$
Treatment	T_i	$\in \{0, 1\}$
Outcome	Y_i	$\in \mathbb{R}$
Confounders	$C_{i\cdot}$	$\in \mathbb{R}^{1 \times p}$
Friend group	$X_{i\cdot}$	$\in \mathbb{R}^{1 \times d}$

The challenge: the friend groups $X_{i\cdot}$ are not observed!

Solution: estimate friend groups $X_{i\cdot}$ using community models!

We use a semi-parametric estimator that accommodates:

- Stochastic blockmodels
- Degree-corrected stochastic blockmodels
- Mixed-membership stochastic blockmodels
- Overlapping stochastic blockmodels
- Random dot product graphs
- Etc

Intuition: stochastic blockmodels

d "blocks" or communities

$X_{i\cdot} \in \{0,1\}^d$ one-hot indicator of node i's block

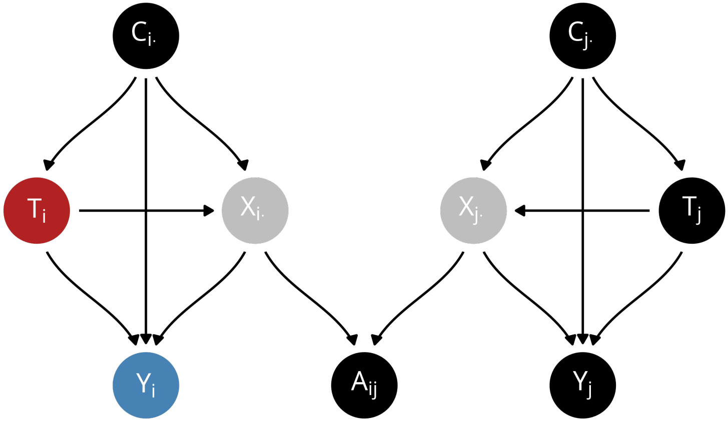
X is latent (i.e. unobserved)

$B \in [0,1]^{d\times d}$ inter-block edge probabilities

Friendships depend on group memberships and B

$\mathbb{P}\left(A_{ij} = 1 \,\middle|\, X\right) = X_{i\cdot} B X_{j\cdot}^T$

Returning to the structural causal model for a moment



Mediation estimators require X, which is unknown

We will estimate X and plug our estimate into standard estimators!

Definition (ASE)
Given a network A, the $\widehat{d}$-dimensional adjacency spectral embedding of A is

$$\widehat{X} = \widehat{U}\widehat{S}^{1/2}$$

where $\widehat{U}\widehat{S}\widehat{U}^T$ is the rank-$\widehat{d}$ truncated singular value decomposition of A.

Note that the analyst must specify $\widehat{d}$

Semi-parametric regression estimators

Under the assumption that:

$$\underbrace{\mathbb{E}[Y_i \mid T_i, C_{i\cdot}, X_{i\cdot}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\beta_\mathsf{t}}_{\mathbb{R}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\beta_\mathsf{c}}_{\mathbb{R}^p} + \underbrace{X_{i\cdot}}_{\mathbb{R}^{1\times d}} \underbrace{\beta_\mathsf{x}}_{\mathbb{R}^d},$$

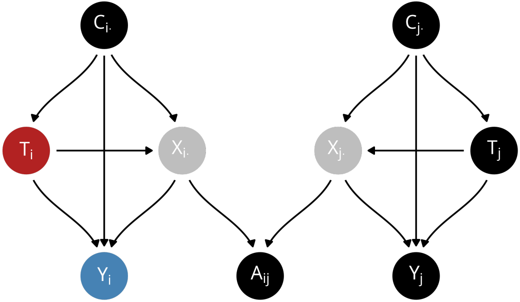
$$\underbrace{\mathbb{E}[X_{i\cdot} \mid T_i, C_{i\cdot}]}_{\mathbb{R}^{1\times d}} = \underbrace{\theta_0}_{\mathbb{R}^{1\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\theta_\mathsf{t}}_{\mathbb{R}^{1\times d}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\Theta_\mathsf{c}}_{\mathbb{R}^{p\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\Theta_\mathsf{tc}}_{\mathbb{R}^{p\times d}}.$$

Then:

$$\Psi_{\mathrm{nde}}(t, t^*) = (t - t^*)\,\beta_\mathsf{t}$$

$$\Psi_{\mathrm{nie}}(t, t^*) = (t - t^*)\,\theta_\mathsf{t}\,\beta_\mathsf{x} + (t - t^*)\,\mu_c\,\Theta_\mathsf{tc}\,\beta_\mathsf{x}.$$

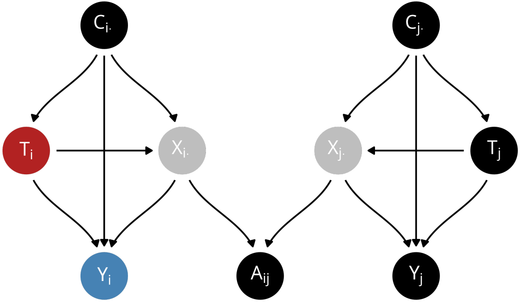
A regression model for friend group membership



Idea: interventions T_i can cause community membership $X_{i\cdot}$

$$\underbrace{\mathbb{E}[X_{i\cdot} \mid T_i, C_{i\cdot}]}_{\mathbb{R}^{1\times d}} = \underbrace{\theta_0}_{\mathbb{R}^{1\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\theta_{\mathrm{t}}}_{\mathbb{R}^{1\times d}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\Theta_{\mathrm{c}}}_{\mathbb{R}^{p\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\Theta_{\mathrm{tc}}}_{\mathbb{R}^{p\times d}}.$$

A regression model for outcomes



Idea: community membership $X_{i\cdot}$ can cause outcomes Y_i

$$\underbrace{\mathbb{E}[Y_i \mid T_i, C_{i\cdot}, X_{i\cdot}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\beta_{\mathrm{t}}}_{\mathbb{R}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p}} \underbrace{\beta_{\mathrm{c}}}_{\mathbb{R}^p} + \underbrace{X_{i\cdot}}_{\mathbb{R}^{1\times d}} \underbrace{\beta_{\mathrm{x}}}_{\mathbb{R}^d}$$

$\widehat{X}$ can be plugged in for X just fine

Let $\widehat{D} = \begin{bmatrix} 1 & T & C & \widehat{X} \end{bmatrix} \in \mathbb{R}^{n \times (2+p+d)}$ and
$L = \begin{bmatrix} 1 & T & C & T \cdot C \end{bmatrix} \in \mathbb{R}^{n \times (2p+2)}$.

We estimate β_{w} and β_{x} via ordinary least squares as follows

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_{\mathrm{t}} \\ \widehat{\beta}_{\mathrm{c}} \\ \widehat{\beta}_{\mathrm{x}} \end{bmatrix} = \left(\widehat{D}^T \widehat{D}\right)^{-1} \widehat{D}^T Y.$$

Similarly, we estimate Θ via ordinary least squares as

$$\widehat{\Theta} = \left(L^T L\right)^{-1} L^T \widehat{X}.$$

Causal estimators

To estimate Ψ_{nde} and Ψ_{nie} in our semi-parametric setting, we combine regression coefficients from the network regression models:

$$\widehat{\Psi}_{\mathrm{cde}} = \widehat{\Psi}_{\mathrm{nde}} = (t - t^*)\,\widehat{\beta}_{\mathsf{t}} \qquad \text{and}$$
$$\widehat{\Psi}_{\mathrm{nie}} = (t - t^*)\,\widehat{\theta}_{\mathsf{t}}\,\widehat{\beta}_{\mathsf{x}} + (t - t^*) \cdot \widehat{\mu}_c \cdot \widehat{\Theta}_{\mathsf{tc}}\,\widehat{\beta}_{\mathsf{x}}.$$

It's standard to fit two regressions and multiply coefficients to estimate an indirect effect like this (VanderWeele and Vansteelandt, 2014).

Theorem (Regression coefficients are asymptotically normal)

Under some mild assumptions, there is an unknown orthogonal matrix Q such that

$$\sqrt{n}\,\widehat{\Sigma}_{\beta}^{-1/2}\begin{pmatrix}\widehat{\beta}_w - \beta_w \\ Q\,\widehat{\beta}_x - \beta_x\end{pmatrix} \to \mathcal{N}(0, I_d), \textit{ and}$$

$$\sqrt{n}\,\widehat{\Sigma}_{\mathrm{vec}(\Theta)}^{-1/2}\left(\mathrm{vec}\left(\widehat{\Theta}\,Q^T\right) - \mathrm{vec}(\Theta)\right) \to \mathcal{N}\left(0, I_{pd}\right).$$

where $\widehat{\Sigma}_{\mathrm{vec}(\Theta)}^{-1/2}$ and $\widehat{\Sigma}_{\beta}^{-1/2}$ are the typical heteroscedasticity robust covariance estimators, with $\widehat{X}$ plugged in for X.

Corollary

Theorem (Causal estimators are asymptotically normal)

Under the same statistical assumptions as before, plus mediating homophily,

$$\sqrt{n\,\widehat{\sigma}^2_{\mathrm{nde}}}\left(\widehat{\Psi}_{\mathrm{nde}} - \Psi_{\mathrm{nde}}\right) \to \mathcal{N}(0,1), \text{ and}$$
$$\sqrt{n\,\widehat{\sigma}^2_{\mathrm{nie}}}\left(\widehat{\Psi}_{\mathrm{nie}} - \Psi_{\mathrm{nie}}\right) \to \mathcal{N}(0,1).$$

where $\widehat{\sigma}^2_{\mathrm{nde}}$ and $\widehat{\sigma}^2_{\mathrm{nie}}$ are rather unfriendly variance estimators derived via the delta method and the previous theorem.

Thank you! Questions?

Read the manuscript at
https://arxiv.org/abs/2212.12041

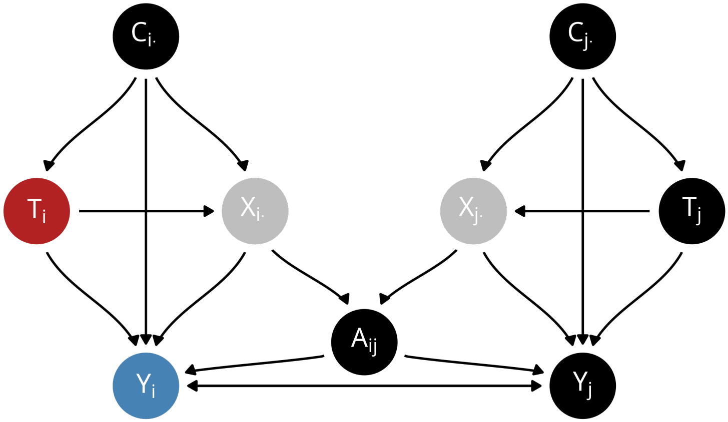
R package netmediate

Stay in touch

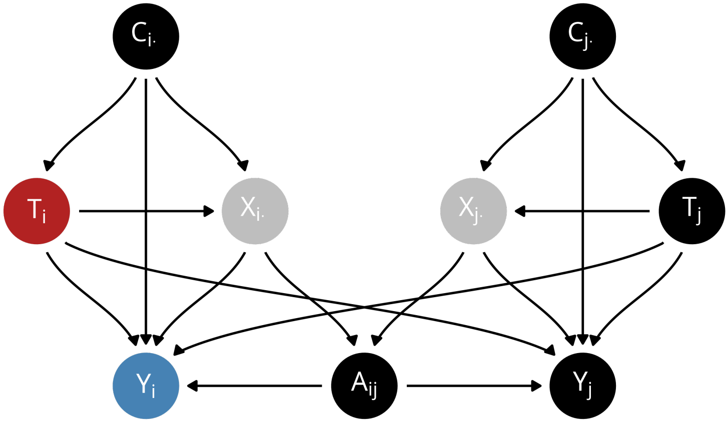
- @alexpghayes
- alex.hayes@wisc.edu
- https://www.alexpghayes.com
- https://github.com/alexpghayes

I'm looking for a post-doc, say hi if this work interests you!

Disambiguation: contagion ($Y_j \to Y_i$) is not allowed



Disambiguation: interference ($T_j \to Y_i$) is not allowed



More on interference and contagion

Interference and contagion effects are allowed so long as they happen in the latent space. Suppose

$$\mathbb{E}[Y_i \mid W_{i\cdot}, X_{i\cdot}] = W_{i\cdot}\beta_{\mathsf{w}} + X_{i\cdot}\beta'_{\mathsf{x}} + \delta_{\mathsf{y}} \sum_j X_{i\cdot}^T X_{j\cdot} Y_j$$

This latent space contagion model is a special parametric case of the regression outcome model (take $\beta_{\mathsf{x}} = \beta'_{\mathsf{x}} + X^T Y \delta_{\mathsf{y}}$).

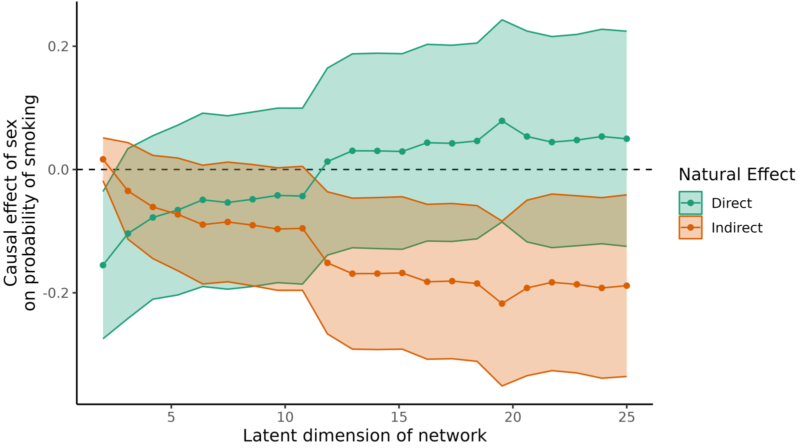
Semi-parametric network model

Let $A \in \mathbb{R}^{n \times n}$ be a random symmetric matrix, such as the adjacency matrix of an undirected graph. Let $P = \mathbb{E}[A \,|\, X] = XX^T$ be the expectation of A conditional on $X \in \mathbb{R}^{n \times d}$, which has independent and identically distributed rows $X_{1\cdot}, \ldots, X_{n\cdot}$. That is, P has $\operatorname{rank}(P) = d$ and is positive semi-definite with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0 = \lambda_{d+1} = \cdots = \lambda_n$. Conditional on X, the upper-triangular elements of $A - P$ are independent (ν_n, b_n)-sub-gamma random variables.

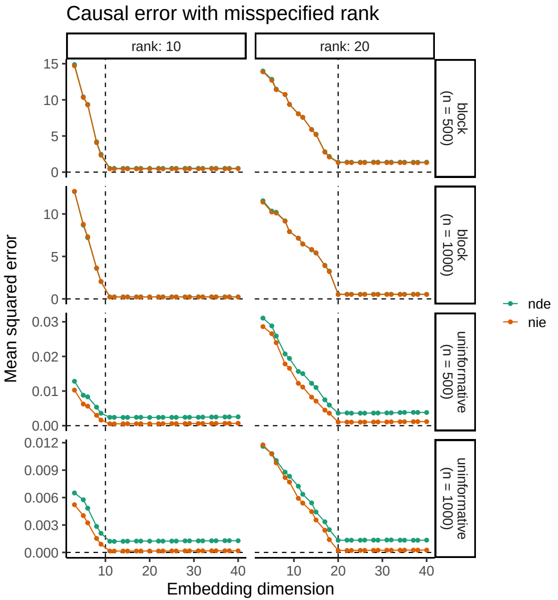
Semi-parametric network model: identification of X

$P = XX^T = (XQ)(XQ)^T$ for any $d \times d$ orthogonal matrix Q, the latent positions X are only identifiable up to an orthogonal transformation.

Choosing $\widehat{d}$: do a multiverse analysis



Choosing $\widehat{d}$: overestimating the embedding dimension is fine



Identifying assumptions

The random variables $(Y_i, Y_i(t,x), X_{i\cdot}, X_{i\cdot}(t), C_{i\cdot}, T_i)$ are independent over $i \in [n]$ and obey the following three properties.

1. Consistency:
   if $T_i = t$, then $X_{i\cdot}(t) = X_{i\cdot}$ with probability 1, and
   if $T_i = t$ and $X_{i\cdot} = x$, then $Y_i(t,x) = Y_i$ with probability 1
2. Sequential ignorability:
$$\{Y_i(t^*,x), X_{i\cdot}(t)\} \perp\!\!\!\perp T_i \mid C_{i\cdot} \quad \text{and} \quad \{Y_i(t^*,x)\} \perp\!\!\!\perp X_{i\cdot} \mid T_i = t, C_{i\cdot}$$
3. Positivity:
$$\mathbb{P}(x \mid T_i, C_{i\cdot}) > 0 \text{ for each } x \in \operatorname{supp}(X_{i\cdot})$$
$$\mathbb{P}(t \mid C_{i\cdot}) > 0 \text{ for each } t \in \operatorname{supp}(T_i)$$

Interventions allowed

Provided that controls $C_{i\cdot}$ are sufficiently informative about group membership $X_{i\cdot}$, treatment T_i is allowed to cause:

- Changes in popularity within a group
- Movement to a new friend group
- Becoming a member of a new friend group while remaining in current friend group
- Friendships becoming more or less likely between distinct friend groups
- Combinations of the above

See Appendix of manuscript for details.

References

VanderWeele, T. and S. Vansteelandt (2014, January). Mediation Analysis with Multiple Mediators. Epidemiologic methods 2(1), 95–115.