Estimating network-mediated causal effects via spectral embeddings

Alex Hayes

2022-10-14

University of Wisconsin-Madison

This is joint work

Figure 1: Keith Levin, assistant professor in statistics

Figure 2: You, our new collaborator??

Motivating example

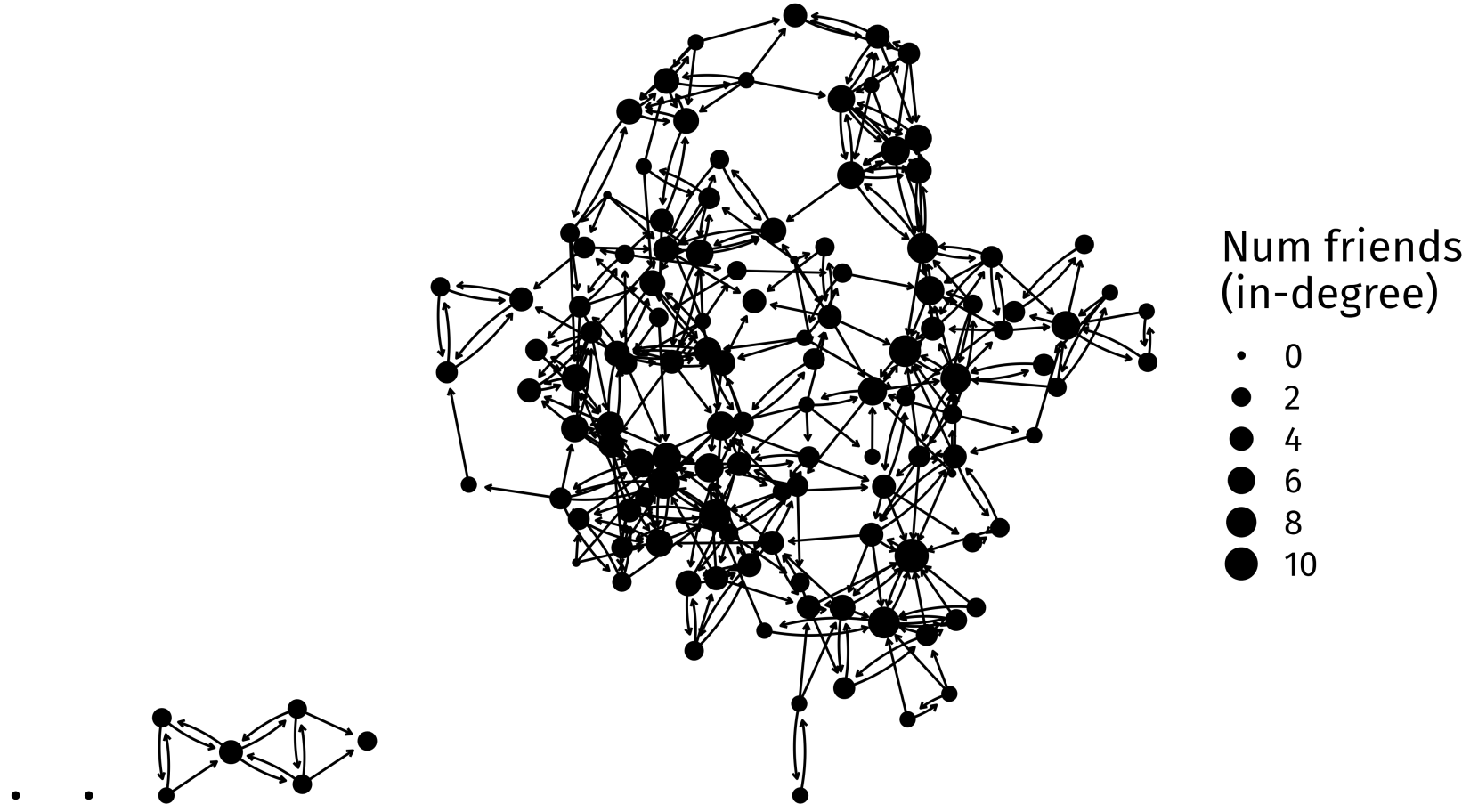
Does access to additional spending money cause adolescents to smoke, drink, and get high more often?

Data

- Social network of 129 adolescents in secondary school in Glasgow, measured in 1995
- Spending money available to each adolescent
- Smoking (tobacco & marijuana) and drinking behaviors
- Many potential controls

Many behaviors are social and depend on network structure

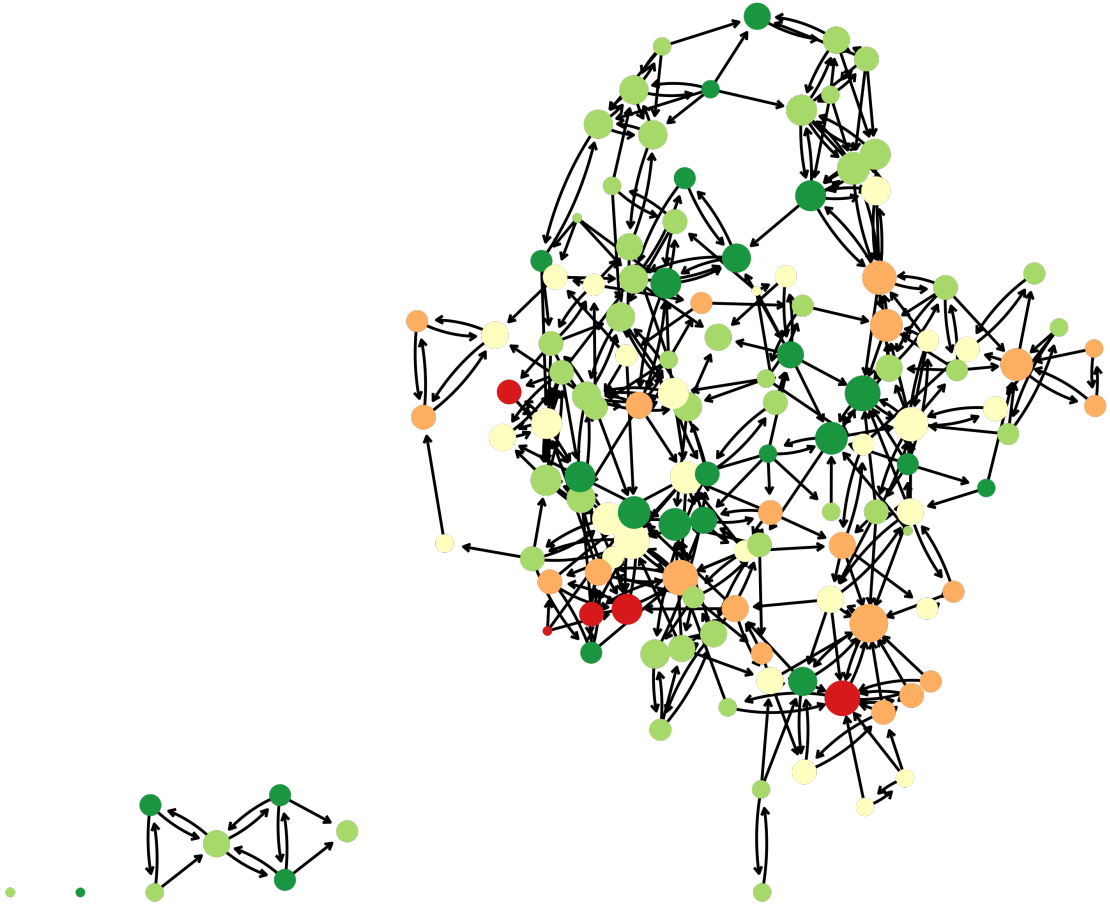
Each node represents one adolescent



Source: Teenage Friends and Lifestyle Study, s129 dataset, 1995

How to account for the social nature of drinking?

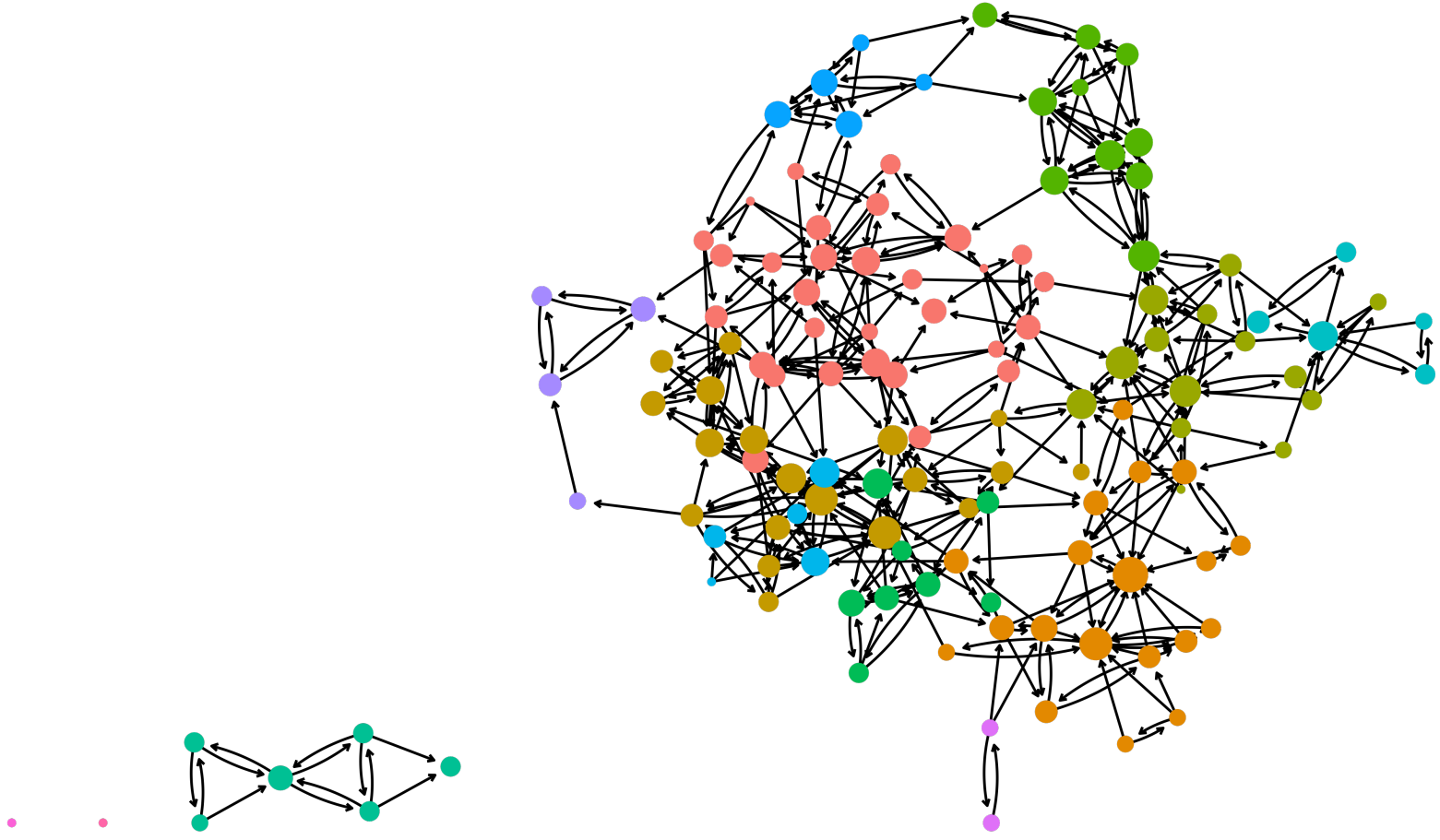
Each node represents one adolescent



Source: Teenage Friends and Lifestyle Study, s129 dataset, 1995

Key idea: suppose node behavior varies with friend group

i.e. nodal features (treatments, controls, outcomes) are homophilous



Nodes colored by estimated friend group

Models of community membership in networks

Homophily via stochastic blockmodels (SBMs)

d "blocks" or communities

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

Z is latent (i.e. unobserved)

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

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

Nodal outcomes Y_i, treatments T_i, and controls C_i related to latent Z

Stochastic blockmodels are good for intuition

...but they do not encode a rich notion of homophily. SBMs can be extended:

- Differing node popularity
- Membership in multiple groups
- Partial memberships in groups
- Membership in multiple groups with group-specific popularity
- Etc, etc

All of these extensions can be captured in a more general model

Sub-gamma random dot product graphs

Definition
Let $A \in \mathbb{R}^{n \times n}$ be a random symmetric matrix, such as the adjacency matrix of an undirected graph. Let $\mathbb{E}[A \mid X] = XX^T$ be the expectation of A conditional on $X \in \mathbb{R}^{n \times d}$, which has independent and identically distributed rows $X_{1\cdot}, \ldots, X_{n\cdot}$. Conditional on X, the elements of the upper triangle of $A - \mathbb{E}[A \mid X]$ are (ν_n, b_n)-subgamma random variables.

Intuition: the latent positions X are very rich measures of community membership and node identity that live in $\mathbb{R}^d$

Sub-gamma random dot product graphs

Remark
We consider symmetric positive semi-definite $\mathbb{E}[A \mid X]$ to simplify notation, but everything works for generic $\mathbb{E}[A \mid X]$.

Remark
X is only identifiable up to orthogonal transformation, since $\mathbb{E}[A \mid X] = XX^T = (XQ)(XQ)^T$ for any $d \times d$ orthogonal matrix Q.

Estimation via the adjacency spectral embedding

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.

These are the "principle components" of the network

The analyst must specify $\widehat{d}$

Uniform consistency of the adjacency spectral embedding

Lemma (Levin et al. (2022))
Under the sub-gamma random dot product model and some additional omitted conditions, if $\widehat{d} = d$*, there is some* $d \times d$ *orthogonal matrix* Q *such that*

$$\max_{i \in [n]} \left\| \widehat{X}_{i\cdot} - X_{i\cdot} Q \right\| = o_p(1).$$

Causal inference on networks

Notation

Network $A \in \mathbb{R}^{n \times n}$

Nodal covariates $W_{1\cdot}, \ldots, W_{n\cdot} \in \mathbb{R}^p$

Latent positions $X_{1\cdot}, \ldots, X_{n\cdot} \in \mathbb{R}^d$

Nodal outcomes $Y_1, \ldots, Y_n \in \mathbb{R}$

Partition $W_i = (T_i, C_{i\cdot})$

Treatment $T_i \in \{0, 1\}$

Controls $C_{i\cdot} \in \mathbb{R}^{p-1}$

The big picture

Homophily is a big deal when you're trying to do causal inference on a network.

Christakis and Fowler (2007) claimed obesity was socially contagious, to much fanfare. Statisticians have spent the subsequent 15 years curb-stomping the paper.

Homophily can act as a powerful confounder (Shalizi and Thomas, 2011).

Community membership can be a confounder

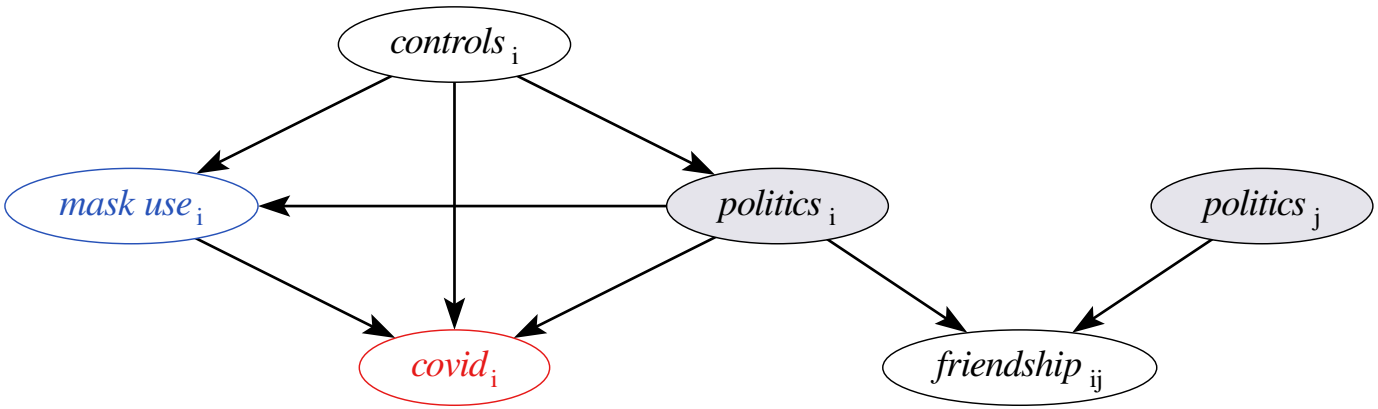


Figure 3: Directed acyclic graph (DAG) for node i. Portions of the DAG corresponding to nodes $j \neq i$ are omitted. X_i and X_j are not observed.

Community membership can be a mediator

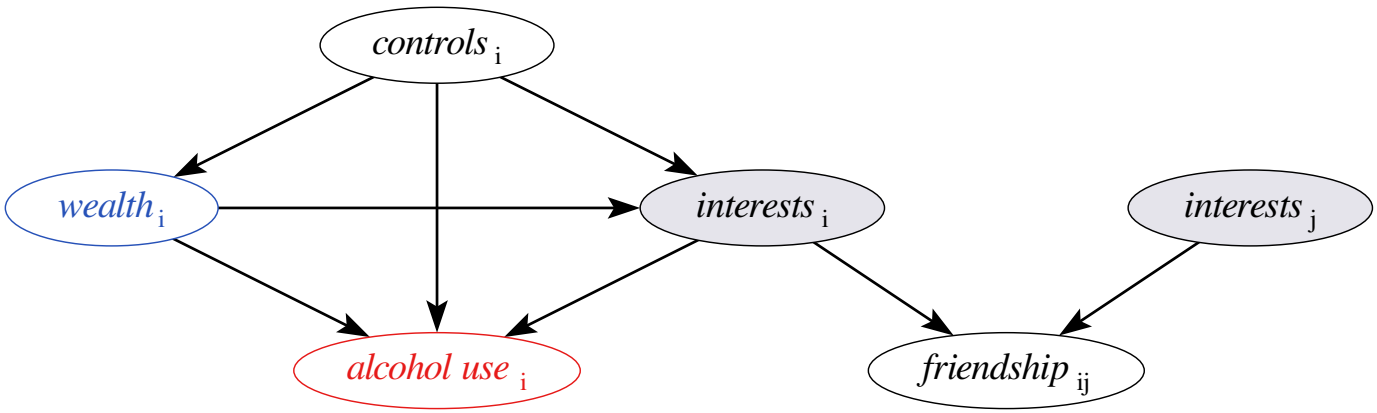


Figure 4: Directed acyclic graph (DAG) for node i. Portions of the DAG corresponding to nodes $j \neq i$ are omitted. X_i and X_j are not observed.

Causation and homophily

Confounding homophily is a big deal (Shalizi and Thomas, 2011), and dealing with confounding homophily is a very active area of research.

Mediating homophily is somehow flying under the radar.

Causal assumptions

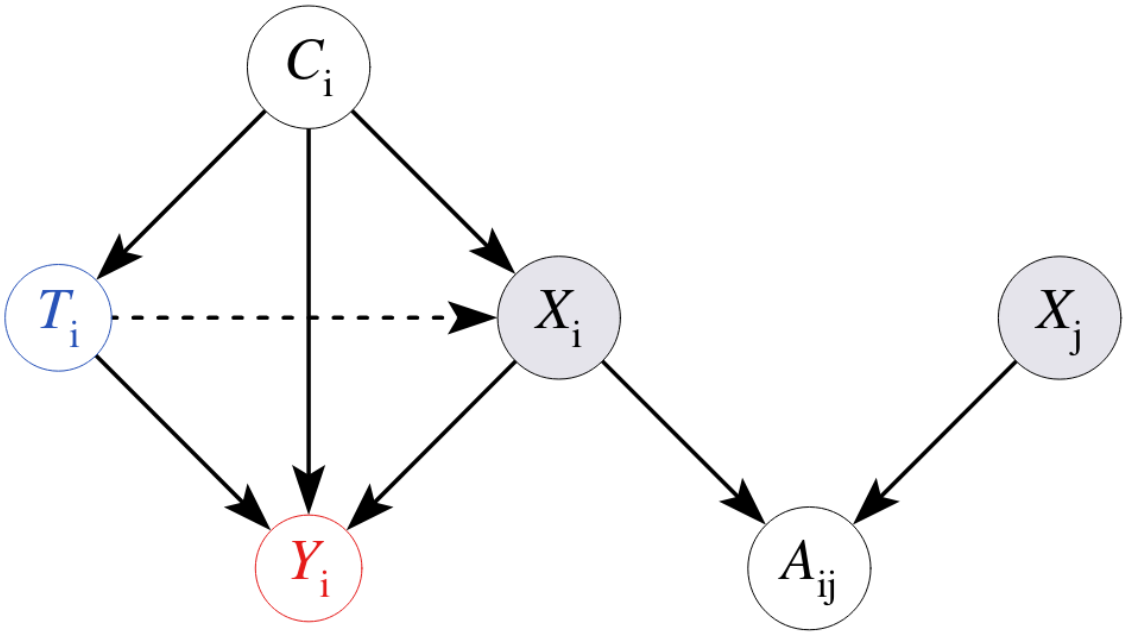


Figure 5: Directed acyclic graph (DAG) for node i. Portions of the DAG corresponding to nodes $j \neq i$ are omitted. X_i and X_j are not observed.

Causal estimands

- Average treatment effect: how much the outcome Y would change on average if the treatment T were changed from $T = t$ to $T = t^*$

$$\Psi_{\text{ate}} = \mathbb{E}[Y_t - Y_{t^*}]$$

- Controlled direct effect: how much the outcome Y would change on average if the mediator X were fixed at level x uniformly in the population, but the treatment were changed from $T = t$ to $T = t^*$

$$\Psi_{\text{cde}} = \mathbb{E}[Y_{tx} - Y_{t^* x}]$$

Causal estimands

Warning: slight change in notation. Now we consider a single node, and let Y_{tX} denote a counterfactual.

- Natural direct effect: how much the outcome Y would change if the exposure T were set at level $T = t^*$ versus $T = t$ but for each individual the mediator X were kept at the level it would have taken, for that individual, if T had been set to t^*

$$\Psi_{\text{nde}} = \mathbb{E}\left[Y_{t X_{t^*}} - Y_{t^* X_{t^*}}\right]$$

- Captures the effect of the exposure on the outcome that would remain if we were to disable the pathway from the exposure to the mediator

Causal estimands

- Natural indirect effect: how much the outcome Y would change on average if the exposure were fixed at level $T = t^*$ but the mediator X were changed from the level it would take if $T = t$ to the level it would take if $T = t^*$

$$\Psi_{\text{nie}} = \mathbb{E}\left[Y_{t X_t} - Y_{t X_{t^*}}\right]$$

- Captures the effect of the exposure on the outcome that operates by changing the mediator

$$\Psi_{\text{ate}} = \Psi_{\text{nde}} + \Psi_{\text{nie}}$$

Regression incorporating network principal components

Regression with rank-$\widehat{d}$ truncated singular value decomposition of A

$$\mathbb{E}[A \mid X] = XX^T = USU^T$$
$$\mathbb{E}[Y_i \mid W_{i\cdot}, X_{i\cdot}] = W_{i\cdot}\beta_{\mathbf{W}} + X_{i\cdot}\beta_{\mathbf{X}}$$

Semi-parametric causal identification

If the mediation DAG holds (non-parametric assumption!) and additionally

$$\underbrace{\mathbb{E}[Y_i \mid T_i, C_{i\cdot}, X_{i\cdot}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\beta_t}_{\mathbb{R}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\beta_c}_{\mathbb{R}^{p'}} + \underbrace{X_{i\cdot}}_{\mathbb{R}^{1\times d}} \underbrace{\beta_x}_{\mathbb{R}^{d}},$$

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

Then:

$$\Psi_{\mathsf{cde}}(t, t^*, x) = \Psi_{\mathsf{nde}}(t, t^*) = (t - t^*)\, \beta_t$$

$$\Psi_{\mathsf{nie}}(t, t^*) = (t - t^*)\, \theta_t\, \beta_x + (t - t^*)\, \mu_C\, \Theta_{tc}\, \beta_x.$$

Causal estimators

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

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

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

Our estimator: plug ASE into ordinary least squares

Let $\widehat{D} = \begin{bmatrix} W & \widehat{X} \end{bmatrix} \in \mathbb{R}^{n \times (p+d)}$. We estimate β_{w} and β_{x} via ordinary least squares as follows

$$\begin{bmatrix} \widehat{\beta}_{\mathsf{w}} \\ \widehat{\beta}_{\mathsf{x}} \end{bmatrix} = \left(\widehat{D}^T \widehat{D}\right)^{-1} \widehat{D}^T Y.$$

Similarly, we estimate Θ via ordinary least squares as

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

Main statistical result

Theorem (Regression coefficients are asymptotically normal)

Under some mild assumptions

$$\sqrt{n}\,\widehat{\Sigma}_{\beta}^{-1/2}\begin{pmatrix}\widehat{\beta}_W - \beta_W \\ Q\,\widehat{\beta}_X - \beta_X\end{pmatrix} \to \mathcal{N}(0, I_d), \text{and}$$

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

where $\widehat{\Sigma}_{\text{vec}(\Theta)}^{-1/2}$ *and* $\widehat{\Sigma}_{\beta}^{-1/2}$ *are the robust covariance estimators based on* $(\widehat{D}, Y, \widehat{\beta})$ *and* $(W, \widehat{X}, \widehat{\Theta})$*, respectively.*

Main causal result

Theorem (Causal estimators are asymptotically normal)

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

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

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

Application to Glasgow data

TODO

Thank you! Questions?

Follow-up work we're interested in

- Better identifiability via varimax rotation
- Extension to GLMs
- Accommodating network interference

Contact me if you'd like to work on any of these!

@alexpghayes

alex.hayes@wisc.edu

https://www.alexpghayes.com

Appendix

Choosing the rank of the network

- Use cross-validated eigenvalues by Chen et al. (2021)
- Check sensitivity of results to choice of $\widehat{d}$

TODO: rank misspecification figure goes here

A more natural parameterization for the regressions

TODO: include result from Appendix of paper

Interference and contagion

TODO: include result from Appendix of paper

Applications outside of network data

TODO: include result from Appendix of paper

Overcontrol bias

TODO: include result from Appendix of paper

Identifying assumptions

We require that natural direct and indirect effects are identified, as follows from consistency, positivity, and sequential ignorability (Imai et al., 2010).

- Consistency:

  if $T = t$, then $X_t = X$ with probability 1, and

  if $T = t$ and $X = x$, then $Y_{tx} = Y$ with probability 1

- Positivity:

$$\mathbb{P}(x \mid T, C) > 0 \text{ for each } x \in \text{supp}(X)$$
$$\mathbb{P}(t \mid C) > 0 \text{ for each } t \in \text{supp}(T)$$

Identifying assumptions

- Sequential ignorability:

$$\{Y_{t^* x}, X_t\} \perp\!\!\!\perp T \mid C$$
$$\{Y_{t^* x}\} \perp\!\!\!\perp X \mid T = t, C$$

This is a criminally strong assumption, in all honesty. Requires the mediator X to be unconfounded with the outcome Y.

Interventions on a network

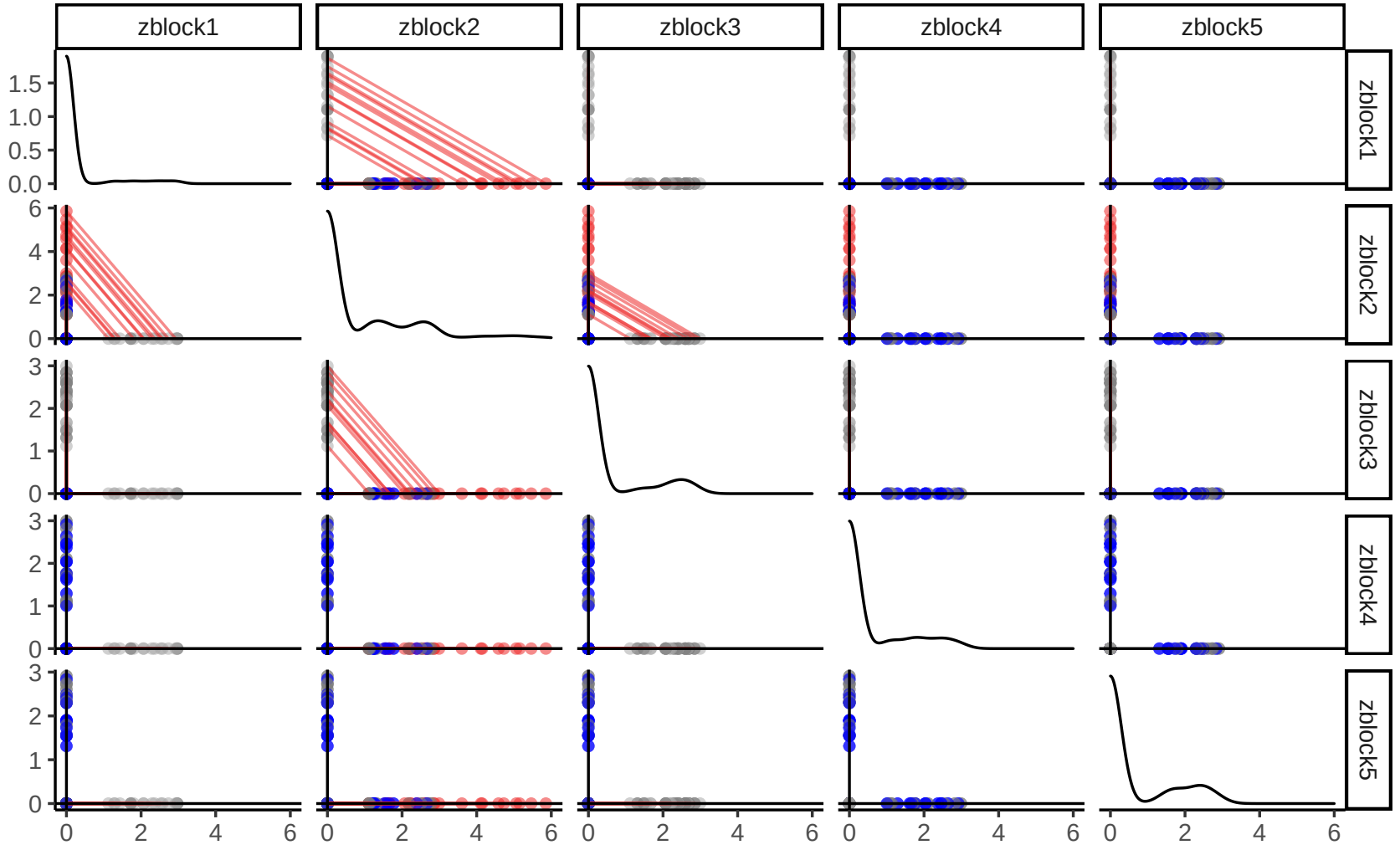


Figure 6: Canonical intervention when C is highly informative.

Interventions on a network

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

In Figure 6, C are latent parameters for a DC-SBM and $\theta_0 = \vec{0}, \theta_t = \vec{0}, \Theta_c = I_k$ and

$$\Theta_{tc} = \begin{bmatrix} -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

References

Chen, F., S. Roch, K. Rohe, and S. Yu (2021, August). Estimating Graph Dimension with Cross-validated Eigenvalues. arXiv:2108.03336 [cs, math, stat].

Christakis, N. A. and J. H. Fowler (2007, July). The Spread of Obesity in a Large Social Network over 32 Years. New England Journal of Medicine 357(4), 370–379.

Imai, K., L. Keele, and T. Yamamoto (2010, February). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. Statistical Science 25(1).

Levin, K., A. Lodhia, and E. Levina (2022). Recovering shared structure from multiple networks with unknown edge distributions. Journal of Machine Learning Research 23, 1–48.

Shalizi, C. R. and A. C. Thomas (2011, May). Homophily and Contagion Are Generically Confounded in Observational

Social Network Studies. Sociological Methods & Research 40(2), 211–239.

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