

# **Estimating indirect effects induced by homophily via spectral network regression**

---

Alex Hayes and Keith Levin

2022-07-07

University of Wisconsin-Madison

# Setting



Network  $A \in \{0, 1\}^{n \times n}$

Nodal covariates  $Z_1, \dots, Z_n \in \mathbb{R}^p$

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

## Regression incorporating network principle components

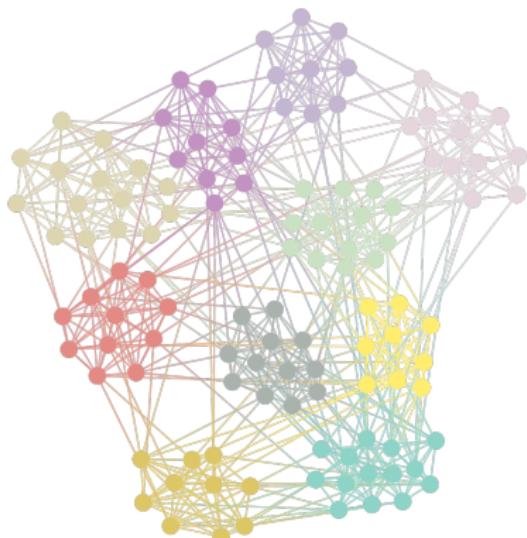
Regression with rank- $k$  truncated eigendecomposition of  $A$

$$\mathbb{E}[A \mid U, S] = \mathbf{X}\mathbf{X}^T$$

$$\mathbb{E}[Y_i \mid \mathbf{Z}_i, U_i, S] = \beta_0 + \mathbf{Z}_i\beta_z + \mathbf{X}_i\beta_x$$

$\beta_x$  has a nice interpretation under stochastic blockmodels

# Network model: stochastic blockmodel (SBM)



## Degree-corrected SBM

$k$  “blocks” or communities

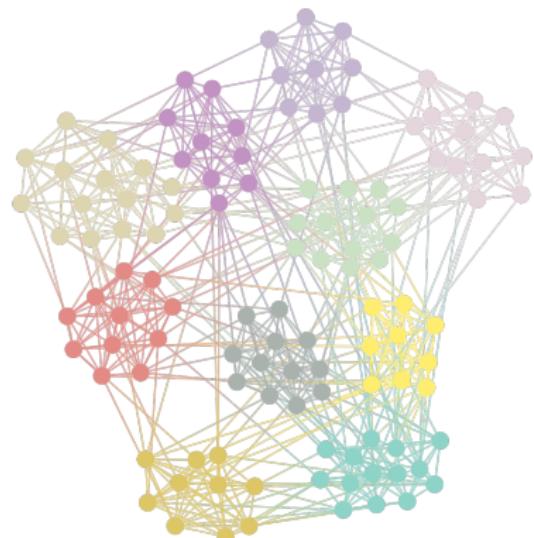
$B \in [0, 1]^{k \times k}$  inter-block edge  
probabilities

$w(i) \in \{1, \dots, k\}$  node  $i$ 's block

$\gamma_i \in [0, 1]$  node  $i$ 's popularity

$$\mathbb{P}[A_{ij} = 1 \mid w, \gamma] = \gamma_i \cdot B_{w(i), w(j)} \cdot \gamma_j$$

# Network model: broad generalization of the SBM



Random dot product graphs

$$A_{ij} \mid \mathbf{X} \sim \text{Bernoulli}\left(\mathbf{x}_i^T \mathbf{x}_j\right)$$

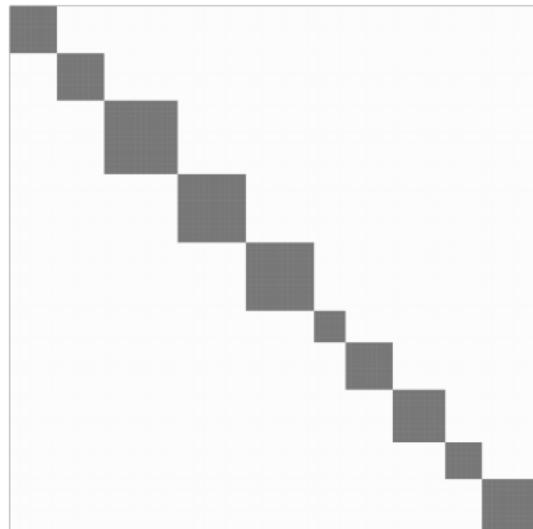
$$\mathbf{X}_i \sim F$$

$F$  a  $k$ -dimensional inner-product distribution

$$(A, \mathbf{X}) \sim \text{RDPG}(F, n)$$

# Random dot product graphs have low rank structure

Population adjacency  $\mathbb{E}[A | \mathbf{X}]$



Sample adjacency  $A$



Stochastic block model with rank  $k = 10$

# Adjacency spectral embedding

## Definition (ASE)

The adjacency spectral embedding  $\widehat{\mathbf{X}} \in \mathbb{R}^{n \times k}$  of a network  $A$  is defined as

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

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

“Principal components” of a network

## Adjacency spectral embedding

**Lemma (Lyzinski et al. (2015), Lemma 5)**

Suppose that  $(A, \mathbf{X}) \sim \text{RDPG}(F, n)$ . Then, letting  $\widehat{\mathbf{X}}_i \in \mathbb{R}^d$  denote the  $i$ -th row of  $\widehat{\mathbf{X}}$ , there exists a universal constant  $C$  and a sequence of orthogonal matrices  $Q_n \in \mathbb{R}^{k \times k}$  such that eventually,

$$\max_{i \in [n]} \|Q_n \widehat{\mathbf{X}}_i - \mathbf{X}_i\| \leq \frac{C \log n}{\sqrt{n}}.$$

This occurs even if  $A$  is observed with sub-gamma noise (Levin et al., 2022).

## Previous approaches to network regression

Regression with rank- $k$  truncated eigendecomposition of  $A$

$$A = USU^T$$

$$\mathbb{E}[Y_i | \mathbf{Z}_i, A] = \beta_0 + \mathbf{Z}_i \boldsymbol{\beta}_z + \mathbf{X}_i \boldsymbol{\beta}_x$$

Why low rank network decompositions?

$\boldsymbol{\beta}_x$  roughly “effect of being in block  $w(i)$  with popularity  $\gamma_i$ ”

$\boldsymbol{\beta}_x$  only identifiable up to rotation

# Regression incorporating network principle components

This has been done in [Le and Li \(2021\)](#)!

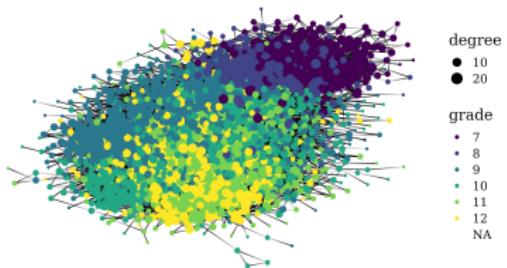
$$\mathbb{E}[Y_i | \mathbf{Z}_i, A] = \mathbf{Z}_i \boldsymbol{\beta}_z + \xi + \alpha$$

Let  $S_k$  be truncated eigenspace of  $\mathbb{E}[A | \mathbf{X}]$  and define  
 $\mathcal{R} = \text{col}(\mathbf{Z}) \cap S_k$ . Require  $\xi \in \mathcal{R}$ ,  $\mathbf{Z}_i \boldsymbol{\beta}_z \perp \mathcal{R}$  and  $\alpha \perp \mathcal{R}$

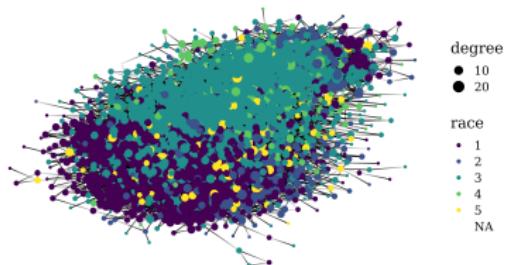
$$\mathbb{E}[Y_i | \mathbf{Z}_i, A] = \underbrace{\mathbf{Z}_i \boldsymbol{\beta}_z}_{\text{covariate effect}} + \underbrace{\mathbf{Z}_i \theta}_{\text{shared effect}} + \underbrace{\alpha}_{\text{network effect}}$$

# AddHealth data from Le and Li (2021)

- Data on 2,152 high school students (nodes)
- 7,986 self-reported friendships (edges)
- Outcome: measure of mental health
- Covariates: race, sex, grade in school



**Figure 1:** Grade based homophily in a high school social network.



**Figure 2:** Race based homophily in a high school social network.

## Results when applied to AddHealth data

- $k$  estimated to be 9
- $\dim(\mathcal{R})$  estimated to be zero, no network-outcome confounding
- $\beta$ : significant sex and grade effects
- $\beta$ : weak effect of race, especially when compared to OLS estimates

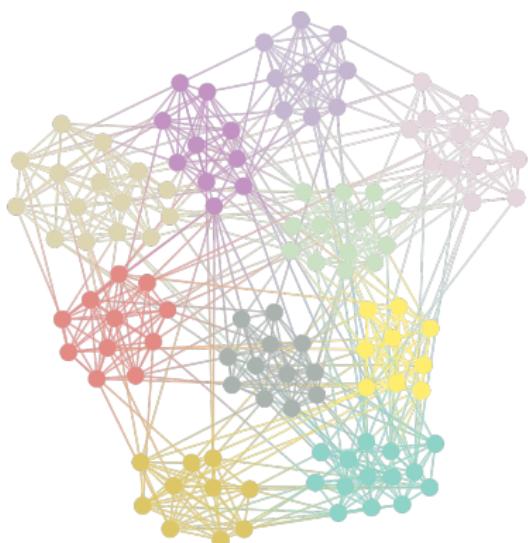
## Results when applied to AddHealth data: a mystery

- $k$  estimated to be 9
- $\dim(\mathcal{R})$  estimated to be zero, no network-outcome confounding
- $\beta$ : significant sex and grade effects
- $\beta$ : weak effect of race, especially when compared to OLS estimates

Why does controlling for latent position in network make the race effect go away?

My hypothesis: race causes group membership, which in turn causes mental health outcomes, i.e. network position mediates rather than confounds race  $\leftrightarrow$  mental health relationship

# Causal network regression



Network  $A \in \{0, 1\}^{n \times n}$

Nodal covariates  $\mathbf{Z}_1, \dots, \mathbf{Z}_n \in \mathbb{R}^p$

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

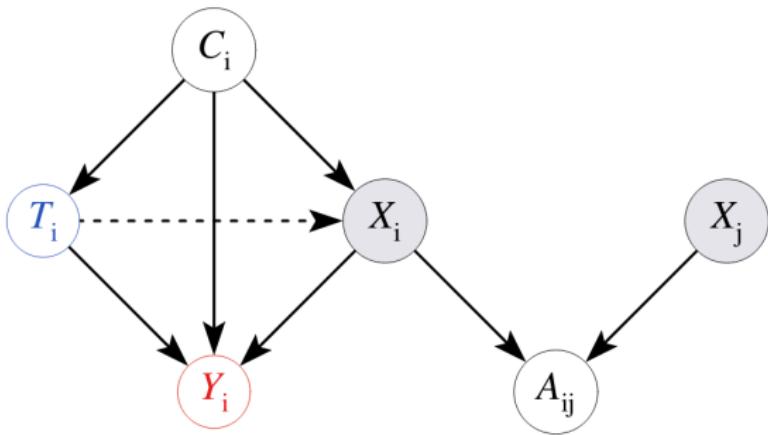
Partition  $\mathbf{Z}_i = (T_i, \mathbf{C}_i)$

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

Controls  $\mathbf{C}_i \in \mathbb{R}^p$

No interference or contagion!

# Mediation in network-linked data for low-rank network models



**Figure 3:** Directed acyclic graph (DAG) relating for node  $i$  in a reduced rank network regression model. Portions of the DAG corresponding to nodes  $j \neq i$  are omitted.  $X_i$  and  $X_j$  are not observed.

## Causal estimands for mediated effects

There are a couple causal effects of interest

$$\Psi_{\text{ate}}(t, t^*) = \mathbb{E}[Y_t - Y_{t^*}] = \Psi_{\text{nde}}(t, t^*) + \Psi_{\text{nie}}(t, t^*),$$

$$\Psi_{\text{cde}}(t, t^*, \mathbf{x}) = \mathbb{E}[Y_{t \mathbf{x}} - Y_{t^* \mathbf{x}}],$$

$$\Psi_{\text{nde}}(t, t^*) = \mathbb{E}[Y_t \mathbf{x}_{t^*} - Y_{t^*} \mathbf{x}_{t^*}], \text{ and}$$

$$\Psi_{\text{nie}}(t, t^*) = \mathbb{E}[Y_t \mathbf{x}_t - Y_t \mathbf{x}_{t^*}].$$

Primarily interested in  $\Psi_{\text{nde}}(t, t^*)$  and  $\Psi_{\text{nie}}(t, t^*)$ , natural direct and natural indirect effects.

# Non-parametric causal identification

## Proposition (non-parametric identification, Imai et al. (2010))

Under our DAG (repeated below),  $\Psi_{nde}(t, t^*)$  and  $\Psi_{nie}(t, t^*)$  are identified, as

$$\mathbb{E}[Y_t \mathbf{x}_{t^*}] = \iint \mathbb{E}[y \mid T = t, \mathbf{c}, \mathbf{x}] p(\mathbf{x} \mid \mathbf{c}, T = t^*) p(\mathbf{c}) d\mathbf{x} d\mathbf{c}.$$

i.e. under assumptions in DAG our counterfactual estimand is strictly a function of observable functionals

# Semi-parametric regression assumptions

Further, if

$$\underbrace{\mathbb{E}[Y | T, \mathbf{C}, \mathbf{X}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{t}_{\{0,1\}} \underbrace{\beta_t}_{\mathbb{R}} + \underbrace{\mathbf{c}}_{\mathbb{R}^{1 \times p}} \underbrace{\beta_c}_{\mathbb{R}^p} + \underbrace{\mathbf{x}}_{\mathbb{R}^{1 \times k}} \underbrace{\beta_x}_{\mathbb{R}^k}$$
$$\underbrace{\mathbb{E}[\mathbf{X} | T, \mathbf{C}]}_{\mathbb{R}^{1 \times k}} = \underbrace{\boldsymbol{\theta}_0}_{\mathbb{R}^{1 \times k}} + \underbrace{t}_{\{0,1\}} \underbrace{\boldsymbol{\theta}_t}_{\mathbb{R}^{1 \times k}} + \underbrace{\mathbf{c}}_{\mathbb{R}^{1 \times p}} \underbrace{\boldsymbol{\Theta}_c}_{\mathbb{R}^{p \times k}}$$

Then:

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

$$\Psi_{\text{nie}}(t, t^*) = (t - t^*) \boldsymbol{\theta}_t \beta_x.$$

## What this means for network regressions

Suppose the latent positions  $\mathbf{X}$  are mediators. Then:

- $\beta_t$  is the natural direct effect of  $T$  on  $Y$ , not the average treatment effect
- If we compute a separate regression, we can compute the natural indirect effect of  $T$  on  $Y$  (i.e. race causes group memberships causes mental health outcomes)
- These two effects add up to the average treatment effect
- Ordinary least squares ignoring the network structure can be used to estimate the average treatment effect

## A causal re-interpretation of Le and Li (2021)'s results

Recall results (new causal interpretations in cyan)

- $k$  estimated to be 9
- $\dim(\mathcal{R})$  estimated to be zero, no network outcome confounding race, grade and sex do not perfectly cause membership in any friend group
- $\beta$ : significant sex and grade direct effects
- $\beta$ : weak direct effect of race estimated by network regression, strong total effect of race estimated by ordinary least squares  
⇒ large indirect effect of race

## Constructing purpose-built causal estimators

Recall that we assume

$$\mathbb{E}[Y \mid T, \mathbf{C}, \mathbf{X}] = \beta_0 + t \beta_t + \mathbf{c} \beta_c + \mathbf{x} \beta_x$$

$$\mathbb{E}[\mathbf{X} \mid T, \mathbf{C}] = \boldsymbol{\theta}_0 + t \boldsymbol{\theta}_t + \mathbf{c} \boldsymbol{\Theta}_c$$

and we want to estimate

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

$$\Psi_{\text{nie}}(t, t^*) = (t - t^*) \boldsymbol{\theta}_t \beta_x.$$

Standard to fit two regressions and multiply coefficients to estimate indirect effect ([VanderWeele and Vansteelandt, 2014](#))

## Our estimator: plug ASE into ordinary least squares

Use least squares (with robust standard errors) to estimate the regression coefficients, then plug into causal estimators

$$\mathbf{Z} = \begin{bmatrix} 1 & \mathbf{T} & \mathbf{C} \end{bmatrix} \in \mathbb{R}^{n \times (1+1+p)}.$$

For estimates  $\hat{\mathbf{X}}$  of  $\mathbf{X}$ , possibly equal to  $\mathbf{X}$  itself, we estimate  $\theta_0$ ,  $\theta_t$ , and  $\Theta_c$

$$\begin{bmatrix} \hat{\theta}_0(\hat{\mathbf{X}}) \\ \hat{\theta}_t(\hat{\mathbf{X}}) \\ \hat{\Theta}_c(\hat{\mathbf{X}}) \end{bmatrix} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{X}}.$$

## Our idea: plug ASE into ordinary least squares

$$\mathbf{z}(\hat{\mathbf{x}}) = \begin{bmatrix} 1 & \boldsymbol{\tau} & \mathbf{C} & \hat{\mathbf{x}} \end{bmatrix} \in \mathbb{R}^{n \times (1+1+p+k)}$$

Then

$$\begin{bmatrix} \hat{\beta}_0(\hat{\mathbf{x}}) \\ \hat{\beta}_t(\hat{\mathbf{x}}) \\ \hat{\beta}_c(\hat{\mathbf{x}}) \\ \hat{\beta}_x(\hat{\mathbf{x}}) \end{bmatrix} = \left[ \mathbf{z}(\hat{\mathbf{x}})^T \mathbf{z}(\hat{\mathbf{x}}) \right]^{-1} \mathbf{z}(\hat{\mathbf{x}})^T \mathbf{y}.$$

## Our idea: plug ASE into ordinary least squares

Plug estimates into standard product-of-coefficients estimator

$$\begin{aligned}\hat{\Psi}_{\text{cde}}(\hat{\mathbf{X}}) &= \hat{\Psi}_{\text{nde}}(\hat{\mathbf{X}}) = (t - t^*) \hat{\beta}_t(\hat{\mathbf{X}}) \\ \hat{\Psi}_{\text{nie}}(\hat{\mathbf{X}}) &= (t - t^*) \hat{\theta}_t(\hat{\mathbf{X}}) \hat{\beta}_x(\hat{\mathbf{X}})\end{aligned}$$

# Theoretical results

## Theorem (informal)

Asymptotically, regression coefficients using  $\hat{\mathbf{X}}$  and  $\mathbf{X}$  converge to the same distribution under generic low-rank models for  $A$  with i.i.d. sub-gamma noise

## Corollary (informal)

Regression coefficients based on  $\hat{\mathbf{X}}$  are asymptotically normal and converge at  $\sqrt{n}$ -rates.

## Corollary (informal)

$\hat{\Psi}_{cde}(\hat{\mathbf{X}})$  and  $\hat{\Psi}_{nie}(\hat{\mathbf{X}})$  are asymptotically normal and converge at  $\sqrt{n}$ -rates.

## Rotational unidentifiability of mediator coefficients

There exists some unknown orthogonal rotation  $Q$  such that

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_0(\widehat{\mathbf{X}}) Q^T - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\theta}}_t(\widehat{\mathbf{X}}) Q^T - \boldsymbol{\theta}_t \\ \widehat{\boldsymbol{\Theta}}_c(\widehat{\mathbf{X}}) Q^T - \boldsymbol{\Theta}_c \end{pmatrix} \rightarrow \text{Normal}(0, \Sigma_{\theta})$$

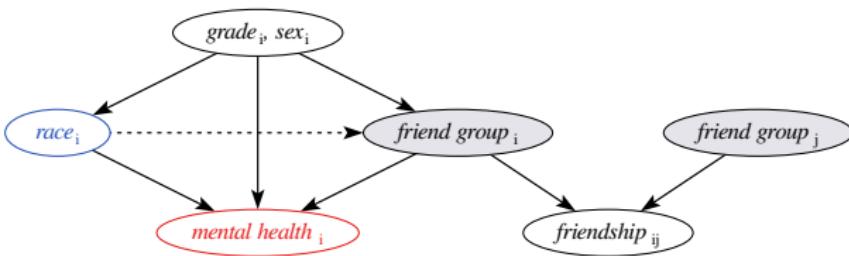
## Rotational unidentifiability of outcome coefficients

There exists some unknown orthogonal rotation  $Q$  (same as in last slide) such that

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_0(\hat{\mathbf{X}}) - \beta_0 \\ \hat{\beta}_t(\hat{\mathbf{X}}) - \beta_t \\ \hat{\beta}_c(\hat{\mathbf{X}}) - \beta_c \\ Q \hat{\beta}_x(\hat{\mathbf{X}}) - \beta_x \end{pmatrix} \rightarrow \text{Normal}(0, \Sigma_{\beta})$$

Rotational unidentifiability of  $\beta_x$  and  $\theta_t$  cancel each other out in causal estimator for  $\Psi_{\text{nie}}(t, t^*)$ !

## Re-analyzing the AddHealth data with our estimators

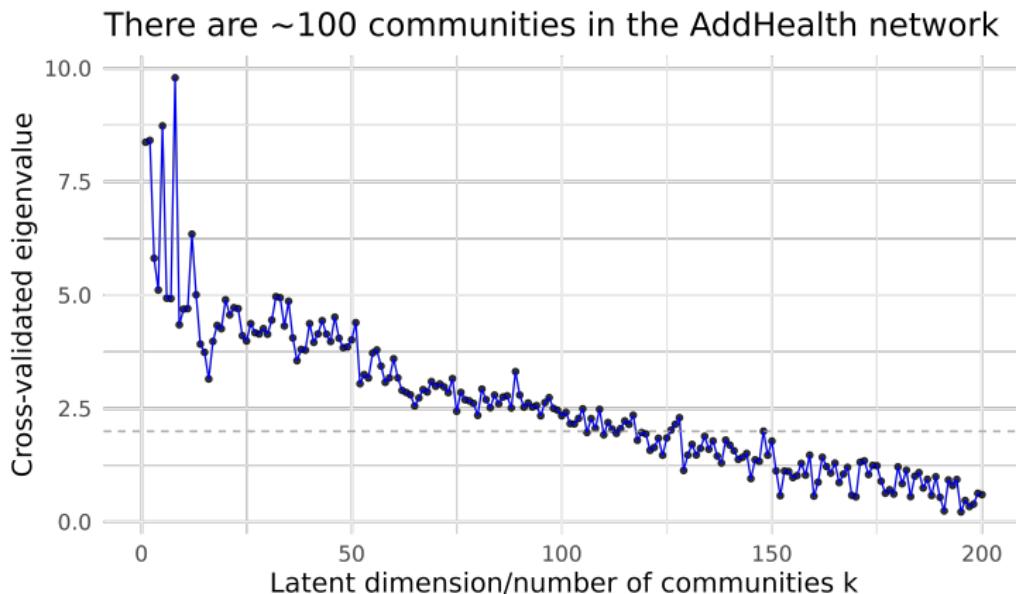


**Figure 4:** Assumptions about the causal structure of the AddHealth data

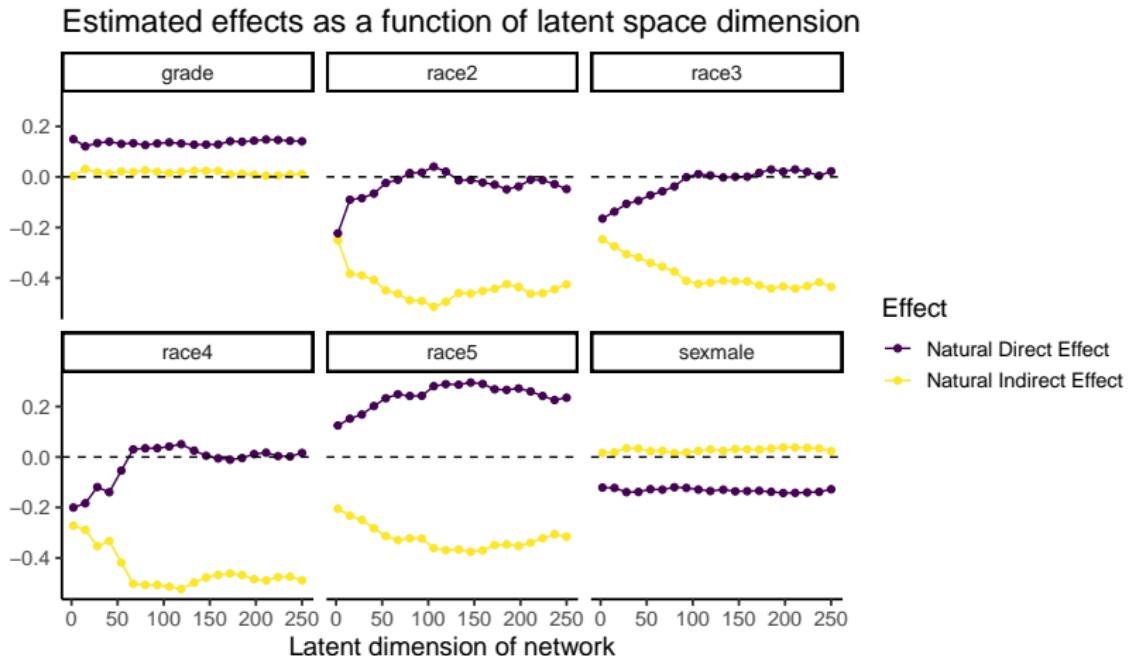
Estimation: need to choose dimension  $k$  of spectral embedding

## Choosing the rank of the network

- Use cross-validated eigenvalues by [Chen et al. \(2021\)](#)
- Check sensitivity of results to choice of  $k$



# Mediated causal effects in the AddHealth data

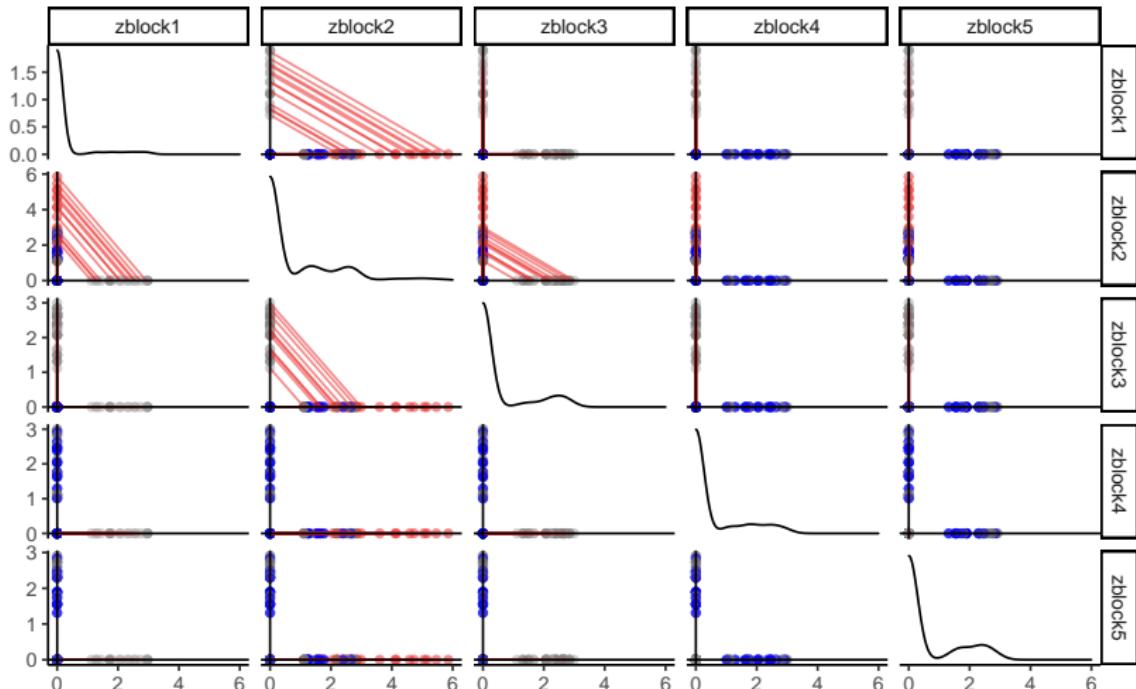


**Figure 5:** Point estimates of natural direct and indirect effects in the AddHealth dataset as a function of varying embedding dimension  $k$ .

**Thank you! Question?**

---

# Interventions on a network



**Figure 6:** Canonical intervention when  $\mathbf{C}$  is highly informative.

## Interventions on a network

$$\mathbb{E}[\mathbf{X} \mid T, \mathbf{C}] = \underbrace{\theta_0}_{\mathbb{R}^{1 \times k}} + \underbrace{t}_{\{0,1\}} \underbrace{\theta_t}_{\mathbb{R}^{1 \times k}} + \underbrace{\mathbf{c}}_{\mathbb{R}^{1 \times p}} \underbrace{\Theta_c}_{\mathbb{R}^{p \times k}}, + \underbrace{t}_{\{0,1\}} \underbrace{\mathbf{c}}_{\mathbb{R}^{1 \times p}} \underbrace{\Theta_{tc}}_{\mathbb{R}^{p \times k}}$$

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

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

# Overcontrol bias from causal misspecification

When  $\mathbf{X}$  is a mediator

$$(t - t^*) \beta_t = \Psi_{nde}(t, t^*) = \Psi_{ate}(t, t^*) - \Psi_{nie}(t, t^*)$$

When  $\mathbf{X}$  is a confounder

$$(t - t^*) \beta_t = \Psi_{ate}(t, t^*)$$

## Abstract

---

The last several years have seen a renewed and concerted effort to incorporate network data into standard tools for regression analysis, and to make network-linked data legible to practicing scientists. Thus far, this literature has primarily developed tools to infer associative relationships between nodal covariates and network structure. In contrast, we augment a statistical model for network regression with counterfactual assumptions and show how causal effects on a network can be partitioned into a direct effect that is uninfluenced by the network, and an indirect effect that is induced by homophily. The method is a conceptually straightforward integration of random dot product models for networks into the well-known product-of-coefficients mediation estimator.

## **References**

---

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

Imai, K., L. Keele, and T. Yamamoto (2010, February). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. [Statistical Science 25](https://doi.org/10.1214/09-SS057)(1).

Le, C. M. and T. Li (2021). Linear regression and its inference on noisy network-linked data. [arXiv:2007.00803 \[stat\]](https://arxiv.org/abs/2007.00803).

Levin, K., A. Lodhia, and E. Levina (2022). Recovering shared structure from multiple networks with unknown edge distributions. [Journal of Machine Learning Research 23](https://doi.org/10.4236/jmlr.v23.i1.13501), 1–48.

Lyzinski, V., D. Sussman, M. Tang, A. Athreya, and C. Priebe (2015, January). Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding. [arXiv:1310.0532 \[stat\]](https://arxiv.org/abs/1310.0532).

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