# Estimating network-mediated causal effects via spectral embeddings

Alex Hayes

2023-03-24 @ IFDS Ideas Forum

Department of Statistics, University of Wisconsin-Madison

Slides available at `https://tinyurl.com/ifds-alex`

# This is joint work!

Mark Fredrickson
University of Michigan

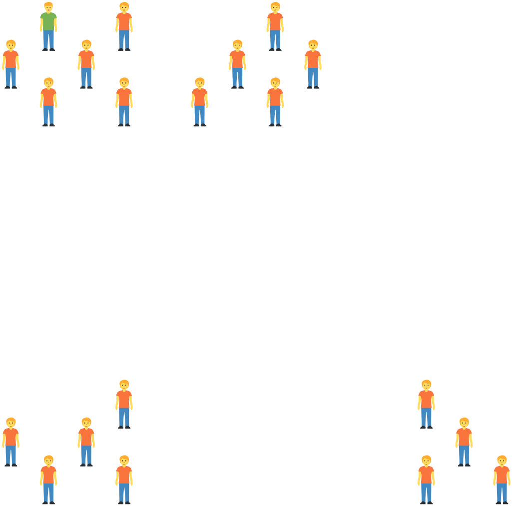
Keith Levin
UW-Madison

# Two questions for the audience

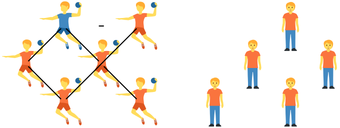
1. How familiar are you with directed acyclic graphs?
2. How familiar are you with stochastic block models?

# A short story about arriving in Madison & the Great Dane

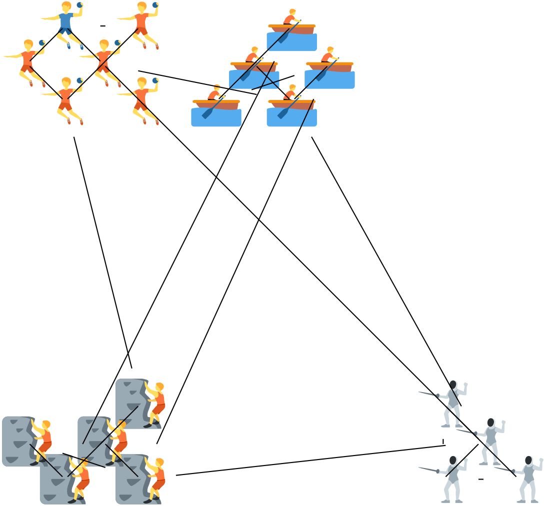
# I didn’t know anyone when I first arrived here



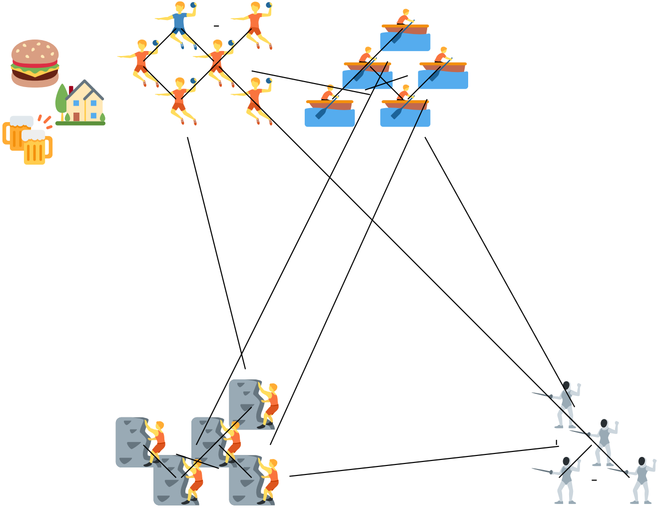
# But! I like frisbee, so I joined a frisbee team!



# The Madison social network



# Madison frisbee players go the Great Dane fairly often!

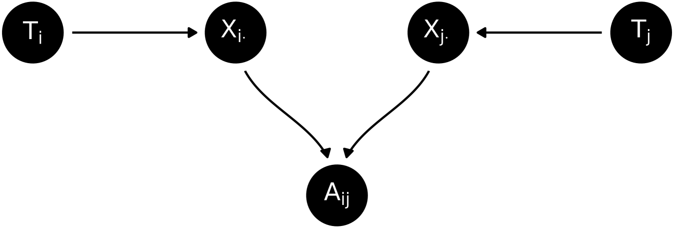


## Idea: social groups mediate the causal effect of individual interests (e.g. frisbee) on visits to the Great Dane

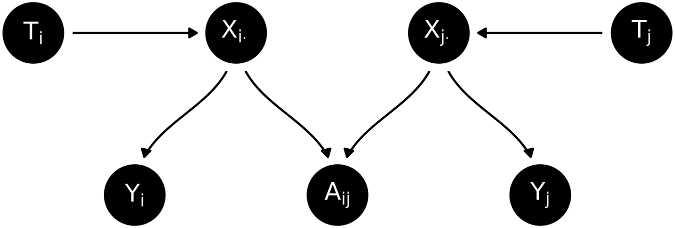
# I like frisbee ($T_i$) so I join a frisbee team ($X_{i\cdot}$)


Assume there are only two people, person *i* and person *j*

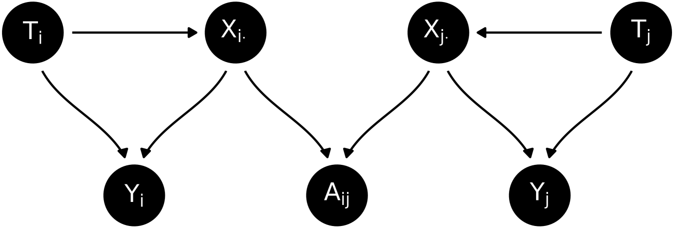
# I’m on a frisbee team, so I form friendships ($A_{ij}$) with other people on my team



# I’m on a frisbee team, so I go to the Great Dane ($Y_i$)

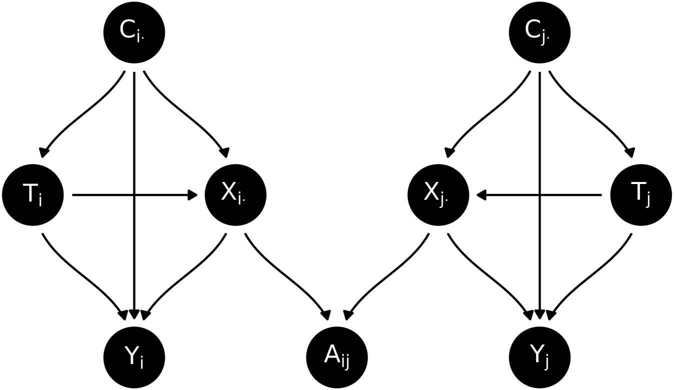


# I like frisbee, and this might directly cause me to go to the Great Dane without my team




i.e. if I want to watch frisbee at the Great Dane

# My individual choices might all be confounded




This is a good place to ask questions

# Formalizing what I think happened with frisbee

First, a recap. I think:

1. enjoying frisbee caused me to visit the Great Dane more frequently than I would have otherwise, and
2. the mechanism proceeded in two stages: first I joined a frisbee team because I liked frisbee, and I went to the Great Dane with the team
3. I don’t think liking frisbee caused me to go to the Great Dane independently of my frisbee team

Can disambiguate these causal effects using mediation analysis

## Average treatment effect (counterfactual quantity)

- Average treatment effect: how much the outcome $Y$ (Great Dane visits) would change on average if the treatment $T$ were changed from $T = t$ (liking frisbee) to $T = t^*$ (not liking frisbee)

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

Note: $Y_i(t)$ is the counterfactual value of $Y_i$ when $T_i$ is set to $t$

I claim $\Psi_{\mathrm{ate}}$ is positive in my example

# Natural indirect effect (counterfactual quantity)

- Natural indirect effect: how much the outcome $Y$ (Great Dane visits) would change on average if the exposure were fixed at level $T = t^*$ (not liking frisbee) but the mediator $X$ (friend group) were changed from the level it would take if $T = t$ (liking frisbee) to the level it would take if $T = t^*$ (not liking frisbee)

$$\Psi_{\text{nie}} = \mathbb{E}[Y_i(t, X_{i\cdot}(t)) - Y_i(t, X_{i\cdot}(t^*))]$$

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

I claim $\Psi_{\text{nie}}$ is positive in my example

# Natural direct effect (counterfactual quantity)

- Natural direct effect: how much the outcome $Y$ (Great Dane visits) would change if the exposure $T$ were set at level $T = t^*$ (liking frisbee) versus $T = t$ (liking frisbee) but for each individual the mediator $X$ (friend group) were kept at the level it would have taken, for that individual, if $T$ had been set to $t^*$ (not liking frisbee)

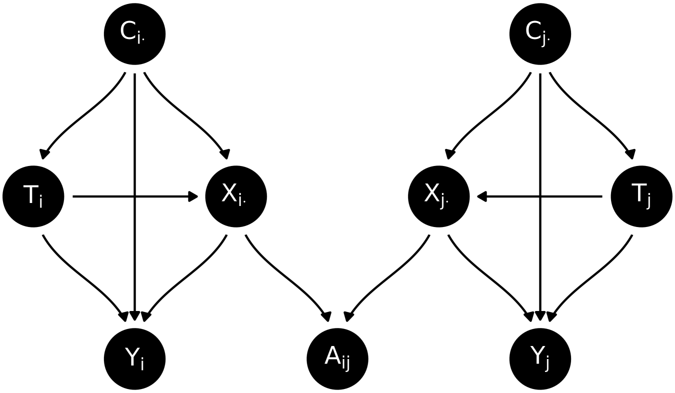
$$\Psi_{\mathrm{nde}} = \mathbb{E}[Y_i(t, X_{i\cdot}(t^*)) - Y_i(t^*, X_{i\cdot}(t^*))]$$

- 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

I claim $\Psi_{\mathrm{nde}}$ is zero in my example.

Note $\Psi_{\mathrm{ate}} = \Psi_{\mathrm{nde}} + \Psi_{\mathrm{nie}}$.

**If we knew the friend groups $X$, we could use standard tools for causal mediation analysis. But we don't know $X$! Instead, we observe the friendship network $A$**



# Semi-parametric network & network regression models

# Intuition: stochastic blockmodels

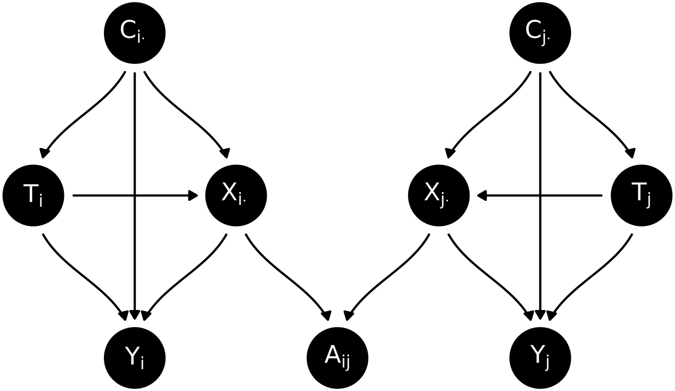
$d$ "blocks" or communities
$X_{i\cdot} \in \{0, 1\}^d$ one-hot indicator of node $i$'s block

$X$ is latent (i.e. unobserved)

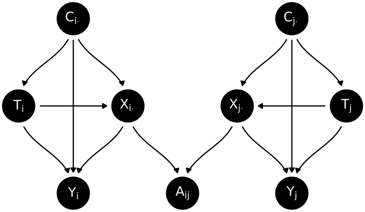
$B \in [0, 1]^{d \times d}$ inter-block edge probabilities

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



# A regression model for friend group membership

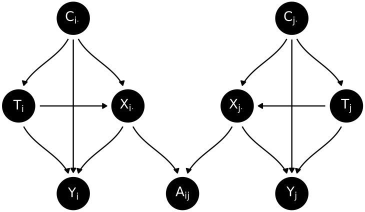



Idea: interventions $T_i$ can cause community membership $X_{i\cdot}$

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

Ex: I like frisbee so I joined an ultimate frisbee team (MUFA)

# A regression model for outcomes




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

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

Ex: I'm on a frisbee team, and the frisbee team goes to the Great Dane together after each game

# Semi-parametric causal identification

Recall the regression models:

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

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

Then:

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

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

# Estimation

# Regression estimators

**Challenge**: regression models depend on $X$, but we never see $X$. Luckily we can estimate it!

**Definition (ASE)**
Given a network $A$, the $\widehat{d}$-dimensional <u>adjacency spectral embedding</u> 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** $\widehat{d}$

# Uniform consistency of the adjacency spectral embedding

Well-known that $\widehat{X}$ is a good estimate of $X$

**Lemma**
*Under a suitably well-behaved network model, if $\widehat{d}$ is correctly specified or consistently estimated, there is some $d \times d$ orthogonal matrix $Q$ such that*

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

# $\widehat{X}$ 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 $\beta_{\mathsf{w}}$ and $\beta_{\mathsf{x}}$ via ordinary least squares as follows

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

Similarly, we estimate $\Theta$ via ordinary least squares as

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

# Causal estimators

To estimate $\Psi_{\mathrm{nde}}$ and $\Psi_{\mathrm{nie}}$ 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 some unknown 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

Slides available at https://tinyurl.com/ifds-alex

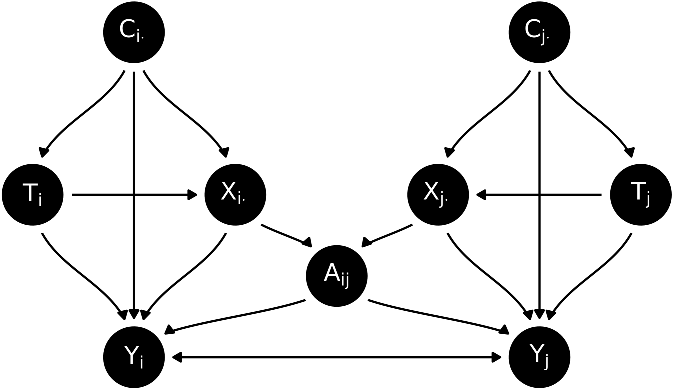
R package netmediate

**Stay in touch**

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

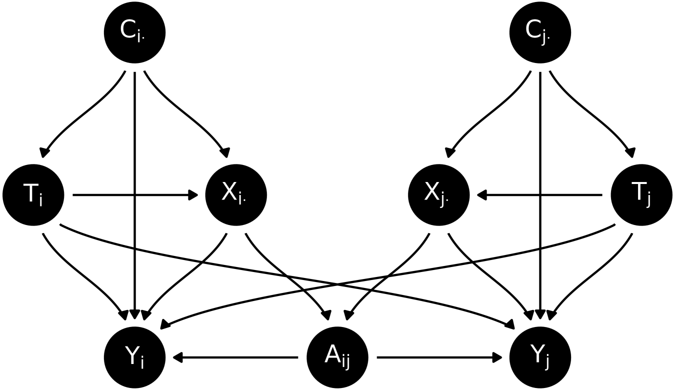
**I'll be on the post-doc market in Fall 2023**

# Disambiguation: contagion is not allowed




Contagion ($Y_j \rightarrow Y_i$) is not allowed

# Disambiguation: interference is not allowed




Interference ($T_j \rightarrow Y_i$) 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 $\beta_{\mathsf{x}} = \beta'_{\mathsf{x}} + X^T Y \delta_{\mathsf{y}}$).

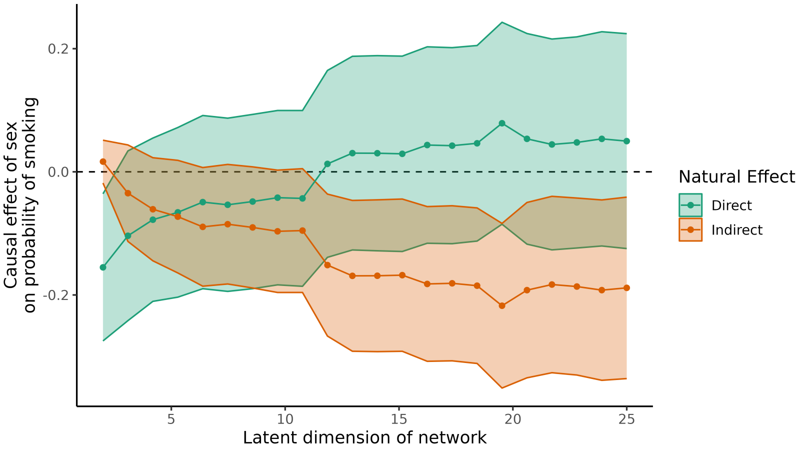
# Semi-parametric network model

Let $A \in \mathbb{R}^{n \times n}$ be a random symmetric matrix, such as the adjacency matrix of an undirected graph. Let $P = \mathbb{E}[A \mid X] = XX^T$ be the expectation of $A$ conditional on $X \in \mathbb{R}^{n \times d}$, which has independent and identically distributed rows $X_{1\cdot}, \ldots, X_{n\cdot}$. That is, $P$ has $\text{rank}(P) = d$ 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 $(\nu_n, b_n)$-sub-gamma random variables.

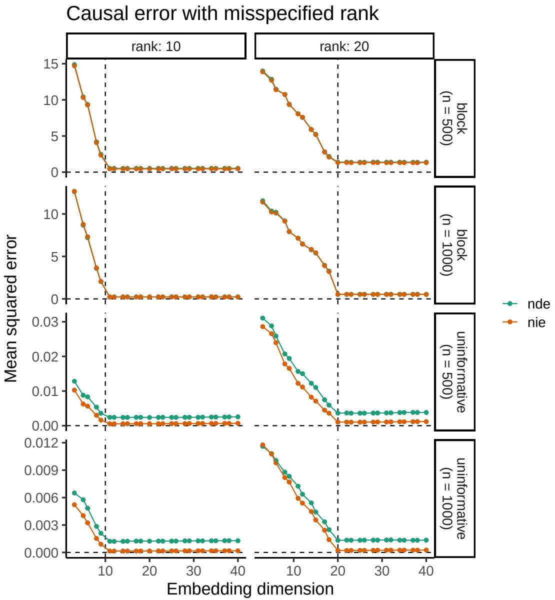
# Semi-parametric network model: identification of $X$

$P = XX^T = (XQ)(XQ)^T$ 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:

   $$\mathbb{P}(x \,|\, T_i, C_{i\cdot}) > 0 \text{ for each } x \in \mathrm{supp}(X_{i\cdot})$$
   $$\mathbb{P}(t \,|\, C_{i\cdot}) > 0 \text{ for each } t \in \mathrm{supp}(T_i)$$

## 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.