Estimating network-mediated causal effects via spectral embeddings

Alex Hayes

2022-10-14

University of Wisconsin-Madison

Motivating example

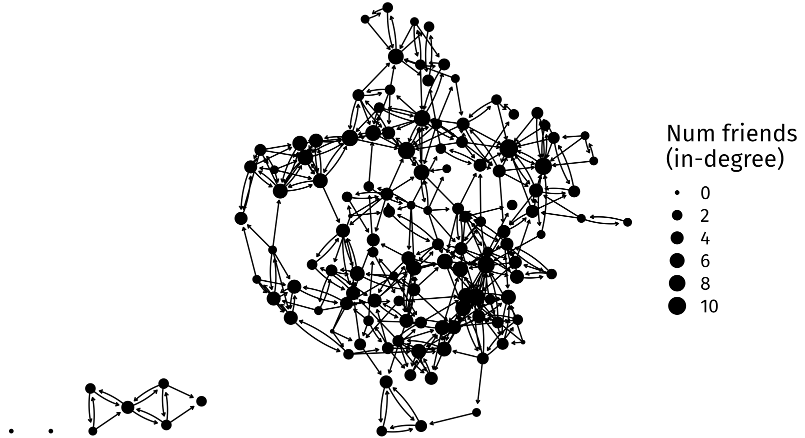
Do age and access to spending money cause adolescents to drink 🍺?

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

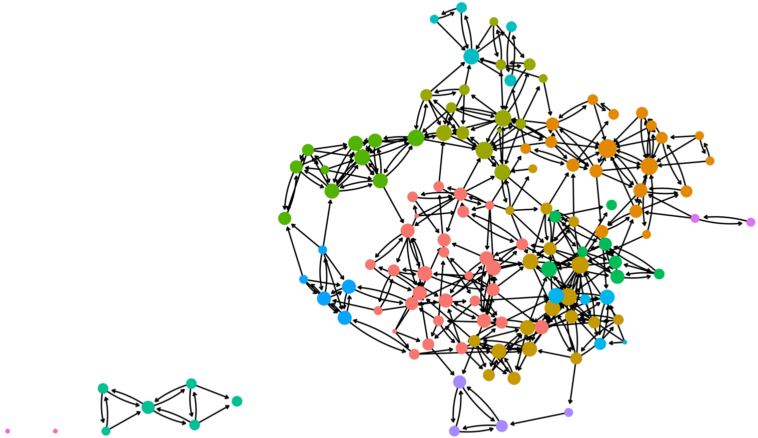
Each node represents one adolescent



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

Key idea: suppose node behavior varies with friend group

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



Nodes colored by estimated friend group

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

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

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 $\mathbb{E}[A \,|\, X] = XX^T$ be the expectation of A conditional on $X \in \mathbb{R}^{n \times d}$, which has independent and identically distributed rows $X_{1\cdot}, \dots, X_{n\cdot}$. Conditional on X, the elements of the upper triangle of $A - \mathbb{E}[A \,|\, X]$ are (ν_n, b_n)-subgamma random variables.

Intuition: the <u>latent positions</u> 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 $\mathbb{E}[A \mid X]$.

This implies possible applications to spatial networks, text data, psychometric surveys, imaging data, and omics panels.

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.

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

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

Behavior that varies over a network

Notation

Network $A \in \mathbb{R}^{n \times n}$
Latent positions $X_{1\cdot}, \ldots, X_{n\cdot} \in \mathbb{R}^d$
Nodal covariates $W_{1\cdot}, \ldots, W_{n\cdot} \in \mathbb{R}^p$
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 $C_{i\cdot} \in \mathbb{R}^{p'}$

A regression model for latent position

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

$$\underbrace{\mathbb{E}[X_{i\cdot} \mid T_i, C_{i\cdot}]}_{\mathbb{R}^{1\times d}} = \underbrace{\theta_0}_{\mathbb{R}^{1\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\theta_{\mathsf{t}}}_{\mathbb{R}^{1\times d}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\Theta_{\mathsf{c}}}_{\mathbb{R}^{p'\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\Theta_{\mathsf{tc}}}_{\mathbb{R}^{p'\times d}} .$$

Example: I like frisbee so I joined an ultimate frisbee team (MUFA) and am likely to form connections to other frisbee players

A regression model for outcomes

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

$$\underbrace{\mathbb{E}[Y_i \mid T_i, C_{i\cdot}, X_{i\cdot}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\beta_{\mathsf{t}}}_{\mathbb{R}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\beta_{\mathsf{c}}}_{\mathbb{R}^{p'}} + \underbrace{X_{i\cdot}}_{\mathbb{R}^{1\times d}} \underbrace{\beta_{\mathsf{x}}}_{\mathbb{R}^{d}}$$

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

Some hints at where the causal stuff is going

The network mediates the relationship between treatments T_i and outcomes Y_i

- I like frisbee, and this might cause me to go the Great Dane more or less often, independent of my friends. This is a direct effect
- I like frisbee, which causes me to be on a frisbee team, which in term causes me to go the Great Dane with my frisbee team. This is an indirect effect
- The total effect of liking frisbee is some combination of the direct and indirect effect

Causal estimands

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

- 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 $T = t^*$

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

Causal estimands

- 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^*

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

- 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

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

Causal identification in the frisbee example

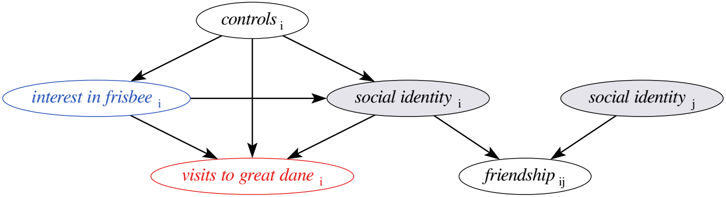


Figure 1: Directed acyclic graph (DAG) for node i. Portions of the DAG corresponding to nodes $j \neq i$ are omitted. Social identities are latent.

Causal identification more general

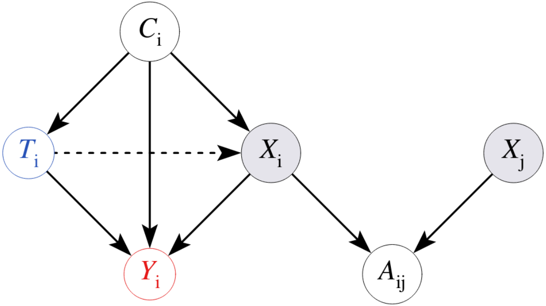


Figure 2: Directed acyclic graph (DAG) for node i. Portions of the DAG corresponding to nodes $j \neq i$ are omitted. X_i and X_j are not observed.

Semi-parametric causal identification

Recall the regression models:

$$\underbrace{\mathbb{E}[Y_i \mid T_i, C_{i\cdot}, X_{i\cdot}]}_{\mathbb{R}} = \underbrace{\beta_0}_{\mathbb{R}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\beta_\mathsf{t}}_{\mathbb{R}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\beta_\mathsf{c}}_{\mathbb{R}^{p'}} + \underbrace{X_{i\cdot}}_{\mathbb{R}^{1\times d}} \underbrace{\beta_\mathsf{x}}_{\mathbb{R}^{d}},$$

$$\underbrace{\mathbb{E}[X_{i\cdot} \mid T_i, C_{i\cdot}]}_{\mathbb{R}^{1\times d}} = \underbrace{\theta_0}_{\mathbb{R}^{1\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{\theta_\mathsf{t}}_{\mathbb{R}^{1\times d}} + \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\Theta_\mathsf{c}}_{\mathbb{R}^{p'\times d}} + \underbrace{T_i}_{\{0,1\}} \underbrace{C_{i\cdot}}_{\mathbb{R}^{1\times p'}} \underbrace{\Theta_\mathsf{tc}}_{\mathbb{R}^{p'\times d}}.$$

Then:

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

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

Challenge: regression models depend on X, but we only have an estimate $\widehat{X}$.

Turns out this is fine. 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

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

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).

Main statistical result

Theorem (Regression coefficients are asymptotically normal)

Under some mild assumptions

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

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

where $\widehat{\Sigma}_{\mathrm{vec}(\Theta)}^{-1/2}$ *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)

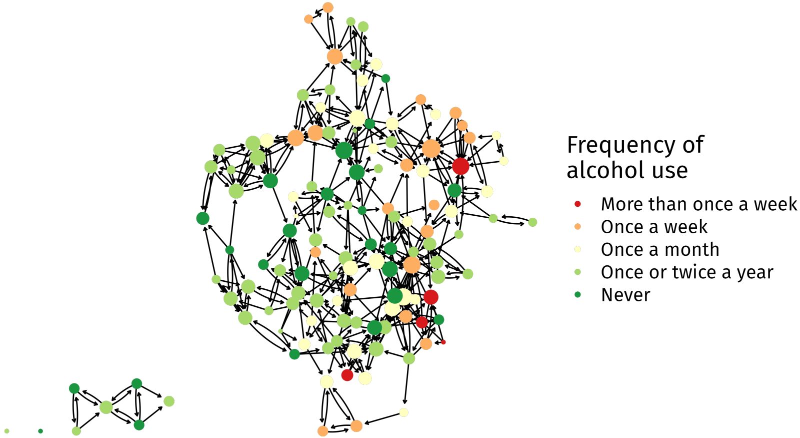
Under the same statistical assumptions as before, plus mediating homophily,

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

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

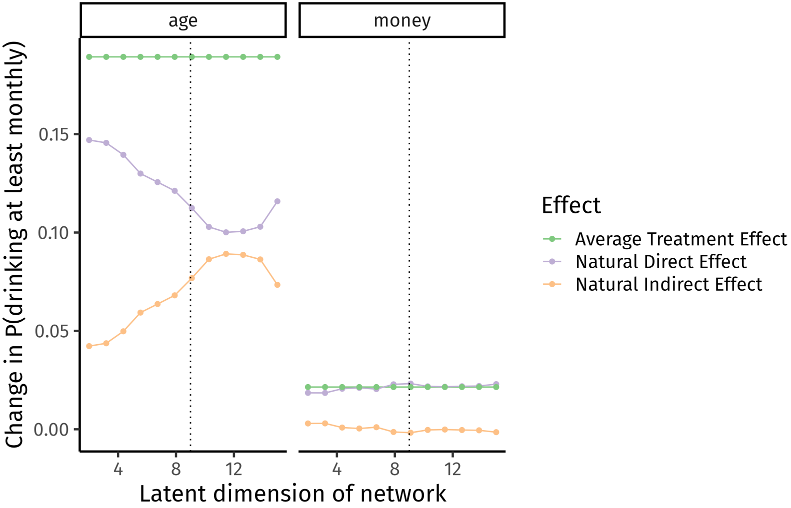
Alcohol use in the adolescent social network

Each node represents one adolescent



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

Estimated effects as a function of latent space dimension



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!

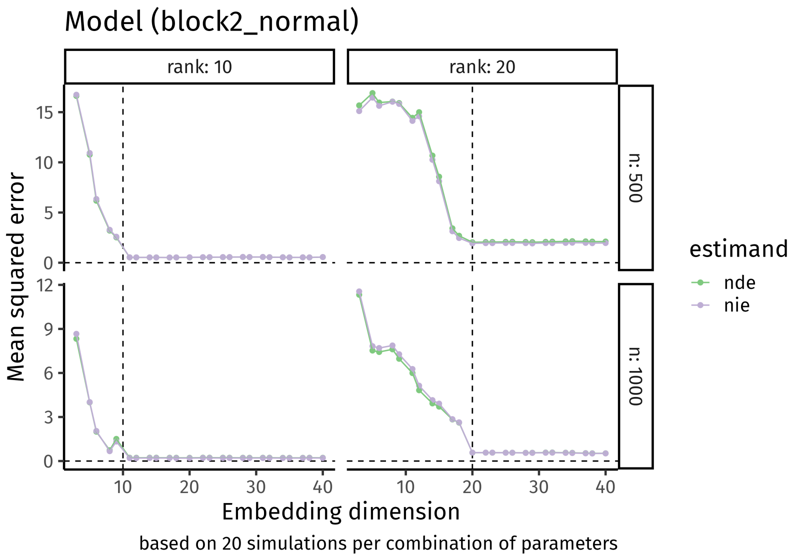
@alexpghayes

alex.hayes@wisc.edu

https://www.alexpghayes.com

Appendix

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



A more natural parameterization for the regressions

X is not the most intuitive way to parameterize a blockmodel. Suppose that $Z \in \mathbb{R}^{n \times d}$ and $B \in \mathbb{R}^{d \times d}$ are arbitrary full-rank matrices such that $\mathbb{E}[A \mid Z, B] = ZBZ^T$. If

$$Z = W\Theta' + \xi', \qquad \text{and} \qquad Y = W\beta_{\mathsf{w}} + Z\beta_{\mathsf{z}} + \varepsilon'$$

then there exist $\Theta, \beta_{\mathsf{x}}, \xi, \varepsilon$ such that

$$X = W\Theta + \xi, \qquad \text{and} \qquad Y = W\beta_{\mathsf{w}} + X\beta_{\mathsf{x}} + \varepsilon.$$

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

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:

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

- Positivity:

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

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

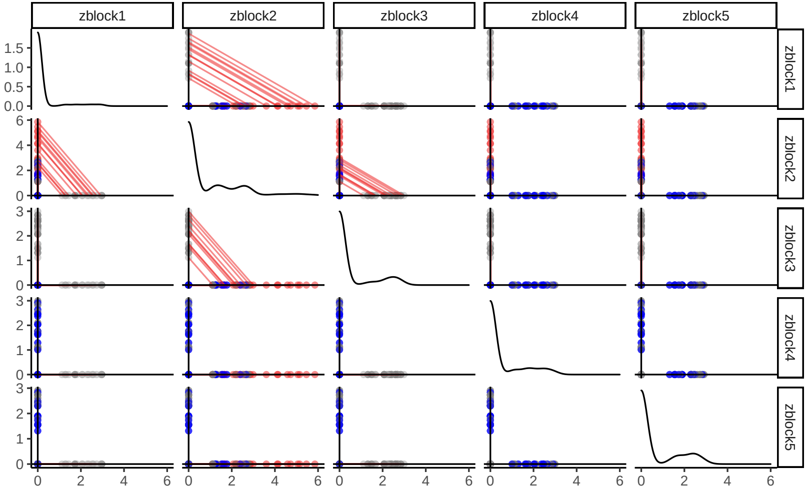


Figure 3: Canonical intervention when C is highly informative.

Interventions on a network

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

In Figure 3, C are latent parameters for a DC-SBM and $\theta_0 = \vec{0}, \theta_t = \vec{0}, \Theta_c = I_k$ and

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

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

References

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.

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