Asymptotic identifiability of peer effects in the linear-in-means model

Alex Hayes | PhD Defense

2024-04-04

Department of Statistics, University of Wisconsin-Madison

Social scientists study social networks

Nodes (people) $i \in \{1, \ldots, n\}$

Social scientists study social networks

Nodes	(people)	i	$\in \{1, \ldots, n\}$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0, 1\}$
Degree	(num friends)	d_i	$\in \{0, 1, 2, \ldots\}$

Social scientists study social networks

Nodes	(people)	i	$\in \{1, \ldots, n\}$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0, 1\}$
Degree	(num friends)	d_i	$\in \{0, 1, 2, \ldots\}$
Outcome	(sick?)	Y_i	$\in \mathbb{R}$
Treatment	(vaccinated?)	T_i	$\in \mathbb{R}$

Peer effects quantify social influence

Contagion: if my friends get sick, I am more likely to get sick

Direct effect: if I get vaccinated, I am less likely to get sick

Interference: if my friends get vaccinated, I am less likely to get sick

Peer effects quantify social influence

Contagion: if my friends get sick, I am more likely to get sick

Direct effect: if I get vaccinated, I am less likely to get sick

Interference: if my friends get vaccinated, I am less likely to get sick

* Can be defined counterfactually, but we define peer effects without counterfactuals in this talk.

Social scientists study peer effects using the linear-in-means model

Outcome	(sick?)	Y_i	$\in \{0,1\}$	Base rate	α	$\in \mathbb{R}$
Treatment	(vaccinated?)	T_i	$\in \{0,1\}$	Contagion	β	$\in (-1,1)$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0,1\}$	Direct effect	γ	$\in \mathbb{R}$
Node degree	(num friends)	d_i	$\in \{0,1,2,\ldots\}$	Interference	δ	$\in \mathbb{R}$

$$Y_i = \alpha$$

Social scientists study peer effects using the linear-in-means model

Outcome	(sick?)	Y_i	$\in \{0,1\}$	Base rate	α	$\in \mathbb{R}$
Treatment	(vaccinated?)	T_i	$\in \{0,1\}$	Contagion	β	$\in (-1,1)$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0,1\}$	Direct effect	γ	$\in \mathbb{R}$
Node degree	(num friends)	d_i	$\in \{0,1,2,\ldots\}$	Interference	δ	$\in \mathbb{R}$

$$Y_i = \alpha + \beta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} Y_j}_{\substack{\text{fraction} \\ \text{sick} \\ \text{friends}}}$$

Social scientists study peer effects using the linear-in-means model

Outcome	(sick?)	Y_i	$\in \{0, 1\}$
Treatment	(vaccinated?)	T_i	$\in \{0, 1\}$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0, 1\}$
Node degree	(num friends)	d_i	$\in \{0, 1, 2, \ldots\}$

Base rate	α	$\in \mathbb{R}$
Contagion	β	$\in (-1, 1)$
Direct effect	γ	$\in \mathbb{R}$
Interference	δ	$\in \mathbb{R}$

$$Y_i = \alpha + \beta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} Y_j}_{\substack{\text{fraction} \\ \text{sick} \\ \text{friends}}} + \gamma T_i$$

Social scientists study peer effects using the linear-in-means model

Outcome	(sick?)	Y_i	$\in \{0,1\}$	Base rate	α	$\in \mathbb{R}$
Treatment	(vaccinated?)	T_i	$\in \{0,1\}$	Contagion	β	$\in (-1,1)$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0,1\}$	Direct effect	γ	$\in \mathbb{R}$
Node degree	(num friends)	d_i	$\in \{0,1,2,\ldots\}$	Interference	δ	$\in \mathbb{R}$

$$Y_i = \alpha + \beta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} Y_j}_{\substack{\text{fraction} \\ \text{sick} \\ \text{friends}}} + \gamma T_i + \delta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}}$$

Social scientists study peer effects using the linear-in-means model

Outcome	(sick?)	Y_i	$\in \{0,1\}$	Base rate	α	$\in \mathbb{R}$
Treatment	(vaccinated?)	T_i	$\in \{0,1\}$	Contagion	β	$\in (-1,1)$
Edge $i \sim j$	(friends?)	A_{ij}	$\in \{0,1\}$	Direct effect	γ	$\in \mathbb{R}$
Node degree	(num friends)	d_i	$\in \{0,1,2,\ldots\}$	Interference	δ	$\in \mathbb{R}$

$$Y_i = \alpha + \beta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} Y_j}_{\substack{\text{fraction} \\ \text{sick} \\ \text{friends}}} + \gamma T_i + \delta \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} + \underbrace{\varepsilon_i}_{\text{error}}$$

Social scientists study peer effects using the linear-in-means model

$[GY]_i =$ "fraction of i's friends who are sick"

$[GT]_i =$ "fraction of i's friends who are vaccinated"

$D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n \times n}$ and $G = D^{-1}A \in \mathbb{R}^{n \times n}$

G computes averages of a value among friends

Social scientists study peer effects using the linear-in-means model

$[GY]_i$ = "fraction of i's friends who are sick"
$[GT]_i$ = "fraction of i's friends who are vaccinated"

$D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n \times n}$ and $G = D^{-1}A \in \mathbb{R}^{n \times n}$
G computes averages of a value among friends

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & [GY]_1 & T_1 & [GT]_1 \\ 1 & [GY]_2 & T_2 & [GT]_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & [GY]_n & T_n & [GT]_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Social scientists study peer effects using the linear-in-means model

$[GY]_i$ = "fraction of i's friends who are sick"
$[GT]_i$ = "fraction of i's friends who are vaccinated"

$D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{R}^{n \times n}$ and $G = D^{-1}A \in \mathbb{R}^{n \times n}$
G computes averages of a value among friends

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & [GY]_1 & T_1 & [GT]_1 \\ 1 & [GY]_2 & T_2 & [GT]_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & [GY]_n & T_n & [GT]_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\mathbb{E}[\varepsilon \mid T, GT] = 0$ but $\text{Cov}(GY, \varepsilon) \neq 0$ because Y is on both sides of the equation

Identification in the linear-in-means model is subtle

Proposition (Bramoullé et al. 2009)
Suppose there are no isolated nodes in the network. Then $(\alpha, \beta, \gamma, \delta)$ *are identified if and only if* $1_n, \mathbb{E}[GY \mid T, GT], T$ *and* GT *are linearly independent.*

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & [GY]_1 & T_1 & [GT]_1 \\ 1 & [GY]_2 & T_2 & [GT]_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & [GY]_n & T_n & [GT]_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Very similar to the usual requirement in linear models that the columns of the design matrix are linearly independent

Identification depends heavily on network structure

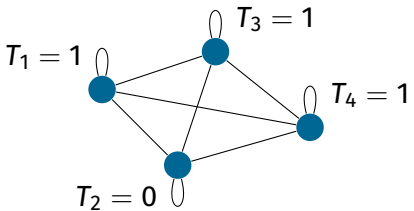
Identification of Endogenous Social Effects: The Reflection Problem

CHARLES F. MANSKI
University of Wisconsin-Madison

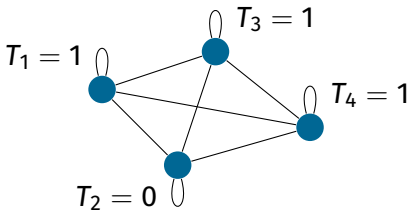
First version received December 1991; *final version accepted December* 1992 (*Eds.*)

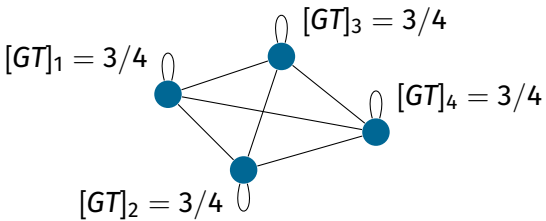
This paper examines the reflection problem that arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group. It is found that inference is not possible unless the researcher has prior information specifying the composition of reference groups. If this information is available, the prospects for inference depend critically on the population relationship between the variables defining reference groups and those directly affecting outcomes. Inference is difficult to impossible if these variables are functionally dependent or are statistically independent. The prospects are better if the variables defining reference groups and those directly affecting outcomes are moderately related in the population.

Too much structure in the network can cause collinearity



Too much structure in the network can cause collinearity





Too much structure in the network can cause collinearity

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & GY_1 & 1 & 3/4 \\ 1 & GY_2 & 0 & 3/4 \\ 1 & GY_3 & 1 & 3/4 \\ 1 & GY_4 & 1 & 3/4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

Can't distinguish base rate α from interference δ

Typical networks don't exhibit problematic structure

Identification of peer effects through social networks

Yann Bramoullé, Habiba Djebbari, Bernard Fortin*

CIRPÉE, Université Laval, Canada
Department of Economics, Université Laval, Canada

ARTICLE INFO

Article history:
Received 5 December 2007
Received in revised form
3 November 2008
Accepted 23 December 2008
Available online 10 February 2009

JEL classification:
D85
L14
Z13
C3

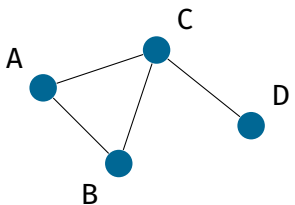
Keywords:
Social network
Peer effects
Identification
Add Health dataset

ABSTRACT

We provide new results regarding the identification of peer effects. We consider an extended version of the linear-in-means model where interactions are structured through a social network. We assume that correlated unobservables are either absent, or treated as network fixed effects. We provide easy-to-check necessary and sufficient conditions for identification. We show that endogenous and exogenous effects are generally identified under network interaction, although identification may fail for some particular structures. We use data from the Add Health survey to provide an empirical application of our results on the consumption of recreational services (*e.g.*, participation in artistic, sports and social activities) by secondary school students. Monte Carlo simulations calibrated on this application provide an analysis of the effects of some crucial characteristics of a network (*i.e.*, density, intransitivity) on the estimates of peer effects. Our approach generalizes a number of previous results due to Manski [Manski, C., 1993. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60 (3), 531–542], Moffitt [Moffitt, R., 2001. Policy interventions low-level equilibria, and social interactions. In: Durlauf, Steven, Young, Peyton (Eds.), Social Dynamics. MIT Press] and Lee [Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140 (2), 333–374].

Peer effects are identified if there is 1+ open triangle in the network

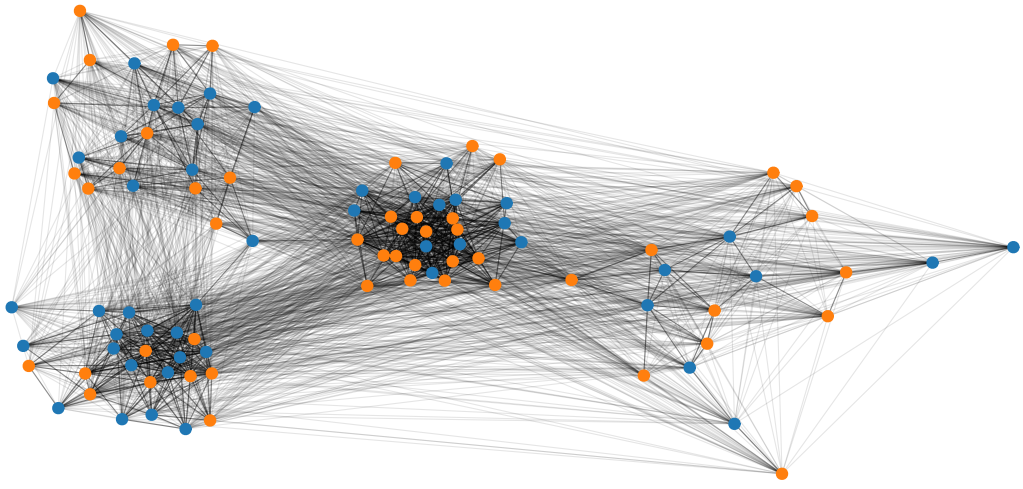
Bramoullé et al. (2009): “intransitivity” identifies peer effects



Closed: $A \leftrightarrow B \leftrightarrow C \leftrightarrow A$

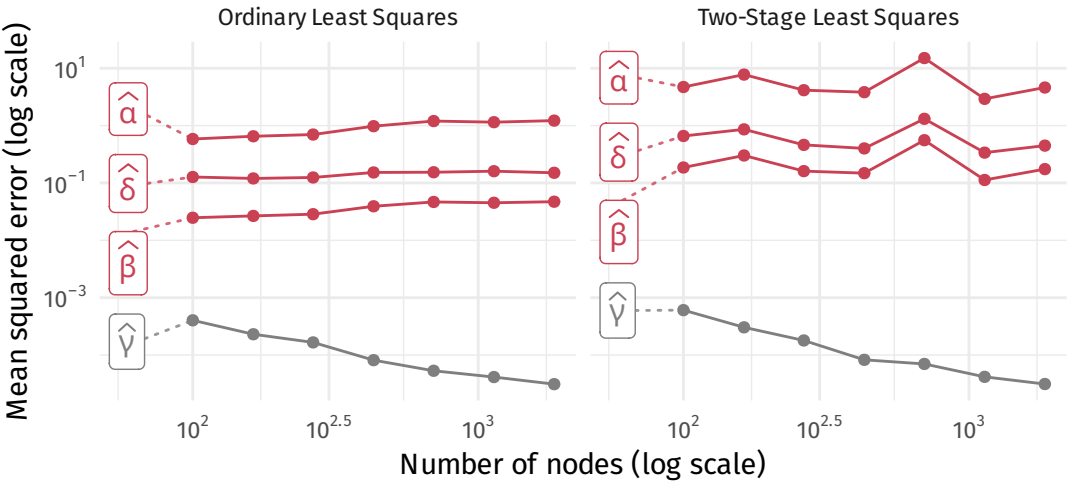
Open: $B \leftrightarrow C \leftrightarrow D \nleftrightarrow B$

We ran an experiment on a model with many open triangles



Treatments are assigned by coin flip and 45% of triangles are open

In the experiment, we couldn't recover the regression coefficients



Based on 100 Monte Carlo replicates

The interference column converges to a constant in large samples

$$\underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}}$$

The interference column converges to a constant in large samples

$$\underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} = \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{average of } d_i \\ \text{i.i.d. coin flips}}}$$

The interference column converges to a constant in large samples

When the network grows ($n \to \infty$),

$$\lim_{n\to\infty} \underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} = \lim_{n\to\infty} \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{average of } d_i \\ \text{i.i.d. coin flips}}}$$

The interference column converges to a constant in large samples

When the network grows ($n \to \infty$), if everyone makes more friends ($d_i \to \infty$)

$$\lim_{n\to\infty} \underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} = \lim_{n\to\infty} \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{average of } d_i \\ \text{i.i.d. coin flips}}} = \frac{1}{2}$$

The interference column converges to a constant in large samples

When the network grows ($n \to \infty$), if everyone makes more friends ($d_i \to \infty$)

$$\lim_{n\to\infty} \underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} = \lim_{n\to\infty} \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\substack{\text{average of } d_i \\ \text{i.i.d. coin flips}}} = \frac{1}{2}$$

For every single node $i = 1, \ldots, n$

Base rates and interence are indistinguishable in large samples

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \underbrace{\begin{matrix} 1_n & GY & T & GT \end{matrix} \begin{bmatrix} 1 & GY_1 & 1 & 1/2 \\ 1 & GY_2 & 0 & 1/2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & GY_n & 1 & 1/2 \end{bmatrix}}_{\text{as } n \to \infty} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Can't distinguish between base rate α and interference δ

Base rates, interference and contagion are indistinguishable in large samples

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \underbrace{\begin{bmatrix} 1_n & GY & T & GT \\ 1 & \eta & 1 & 1/2 \\ 1 & \eta & 0 & 1/2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \eta & 1 & 1/2 \end{bmatrix}}_{\text{as } n \to \infty} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Can't distinguish between base rate α, interference δ and contagion β

Outcomes are generated by diffusing the treatment over the network

$$Y = \alpha 1_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

Outcomes are generated by diffusing the treatment over the network

$$Y = \alpha \mathbf{1}_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

$$Y - \beta GY = \alpha \mathbf{1}_n + \gamma T + \delta GT + \varepsilon$$

Outcomes are generated by diffusing the treatment over the network

$$
\begin{aligned}
Y &= \alpha \mathbf{1}_n + \beta G Y + \gamma T + \delta G T + \varepsilon \\
Y - \beta G Y &= \alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon \\
Y &= (I - \beta G)^{-1} (\alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon)
\end{aligned}
$$

Outcomes are generated by diffusing the treatment over the network

$$
\begin{aligned}
Y &= \alpha \mathbf{1}_n + \beta G Y + \gamma T + \delta G T + \varepsilon \\
Y - \beta G Y &= \alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon \\
Y &= (I - \beta G)^{-1}(\alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon) \\
&\overset{*}{=} \sum_{k=0}^{\infty} \beta^k G^k (\alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon)
\end{aligned}
$$

* using the fact that $(I - \beta G)^{-1} = \sum_{k=0}^{\infty} \beta^k G^k$ when $|\beta| < 1$

Outcomes are generated by diffusing the treatment over the network

$$
\begin{aligned}
Y &= \alpha \mathbf{1}_n + \beta G Y + \gamma T + \delta G T + \varepsilon \\
Y - \beta G Y &= \alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon \\
Y &= (I - \beta G)^{-1}(\alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon) \\
&\overset{*}{=} \sum_{k=0}^{\infty} \beta^k G^k (\alpha \mathbf{1}_n + \gamma T + \delta G T + \varepsilon) \\
&= \frac{\alpha}{1-\beta} \mathbf{1}_n + \gamma T + \underbrace{(\gamma\beta + \delta) \sum_{k=0}^{\infty} \beta^k G^{k+1} T}_{\substack{\text{repeated} \\ \text{averages} \\ \text{of } T}} + \underbrace{\sum_{k=0}^{\infty} \beta^k G^k \varepsilon}_{\substack{\text{repeated} \\ \text{averages} \\ \text{of } \varepsilon}}
\end{aligned}
$$

* using the fact that $(I - \beta G)^{-1} = \sum_{k=0}^{\infty} \beta^k G^k$ when $|\beta| < 1$

The contagion column converges to a constant in large samples

$$GY = \frac{\alpha}{1-\beta}1_n + \underbrace{\gamma GT}_{\substack{\text{neighborhood} \\ \text{average} \to \gamma/2}} + \underbrace{(\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k G^{k+2}T}_{\substack{\text{repeated} \\ \text{averages} \\ \text{of } T^*}} + \underbrace{\sum_{k=0}^{\infty}\beta^k G^{k+1}\varepsilon}_{\substack{\text{repeated} \\ \text{averages} \\ \text{of } \varepsilon \to 0}}$$

Each term converges to a constant, call the sum of these constants η. Then

$$GY \to \eta$$

$$\mathbb{E}[GY \mid T, GT] \to \eta$$

* If $c \in R$ and $GT = c1_n$, then $G^2T = G(GT) = G(c1_n) = c1_n$. i.e., $G^2T = GT$

Base rates, interference and contagion are indistinguishable in large samples

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \underbrace{\begin{matrix} 1_n & GY & T & GT \end{matrix}}_{\text{as } n \to \infty} \begin{bmatrix} 1 & \eta & 1 & 1/2 \\ 1 & \eta & 0 & 1/2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \eta & 1 & 1/2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Can't distinguish between base rate α, interference δ and contagion β

* Recall that identifiability depends on $\mathbb{E}[GY \mid T, GT]$ but last slide also implies $\mathbb{E}[GY \mid T, GT] \to \eta$

We call these indistinguishable parameters asymptotically unidentified

Definition
We say that $(\alpha, \beta, \gamma, \delta)$ are asymptotically identified when the design matrix W_n converges to a limit object W in the sense that

$$\max_{ij} \left| \begin{bmatrix} 1_n & GY & T & GT \end{bmatrix}_{ij} - W_{ij} \right| = o(1)$$

and the columns of W are linearly independent. If the columns of W are linearly dependent, we say that $(\alpha, \beta, \gamma, \delta)$ are asymptotically unidentified.

Peer effects are asymptotically unidentified under very general circumstances

Assumption

1. *$T_1, T_2, \ldots, T_n$ are independent with shared mean $\zeta \in \mathbb{R}$, and T is independent of A.*
2. *$\{T_i - \zeta : i \in [n]\}$ are independent (ν, b)-subgamma random variables.*
3. *$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent subgamma random variables with parameters not depending on n.*
4. *The minimum degree grows strictly faster than $\log n$, such that*

$$\lim_{n \to \infty} \frac{\min_{i \in [n]} d_i}{\log n} = \infty$$

The interference and contagion columns converge uniformly to constants

Lemma
Under the previous assumptions,

$$\max_{i \in [n]} \left\| [GT]_i - \zeta \right\| = o(1) \ \ \textit{almost surely}$$

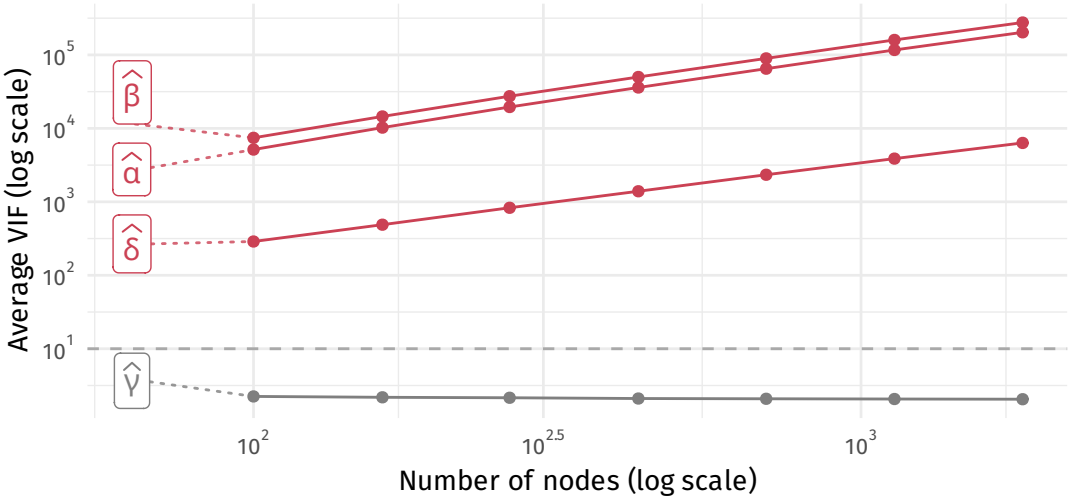
and there exists $\eta = \eta(\zeta, \alpha, \beta, \gamma, \delta) \in \mathbb{R}$ *such that*

$$\max_{i \in [n]} \left\| [GY]_i - \eta \right\| = o(1) \ \ \textit{almost surely.}$$

> *Read: "The fraction of i's friends who are vaccinated* $[GT]_i$ *and fraction of i's friends who are sick* $[GY]_i$ *converge to constants* ζ *and* η*, respectively, for all nodes* $i = 1, \ldots, n$*."*

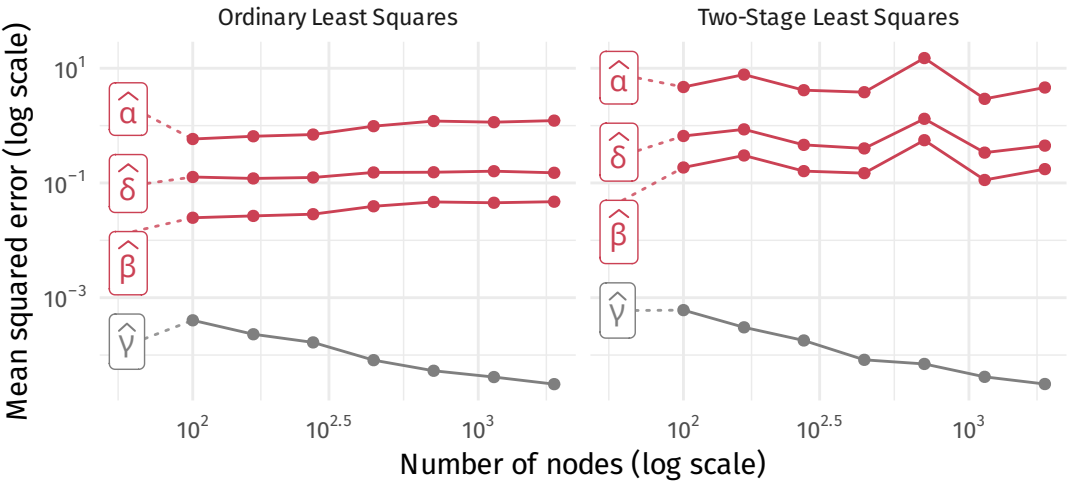
Theorem
α, β *and* δ *are asymptotically unidentified.*

Asymptotically unidentified coefficients become collinear in finite samples



Based on 100 Monte Carlo replicates

Asymptotically unidentified coefficients cannot be estimated



Based on 100 Monte Carlo replicates

Our theory does not apply to sparse networks or fixed covariates

Isolated nodes: If all the connected components that are not singletons satisfy previous assumptions, can recover α but β and δ are still aliased.

Sparse networks: we don't know exactly when the issue kicks in and if sparse networks are in trouble or not

Non-random treatment *T*: theorem doesn't apply

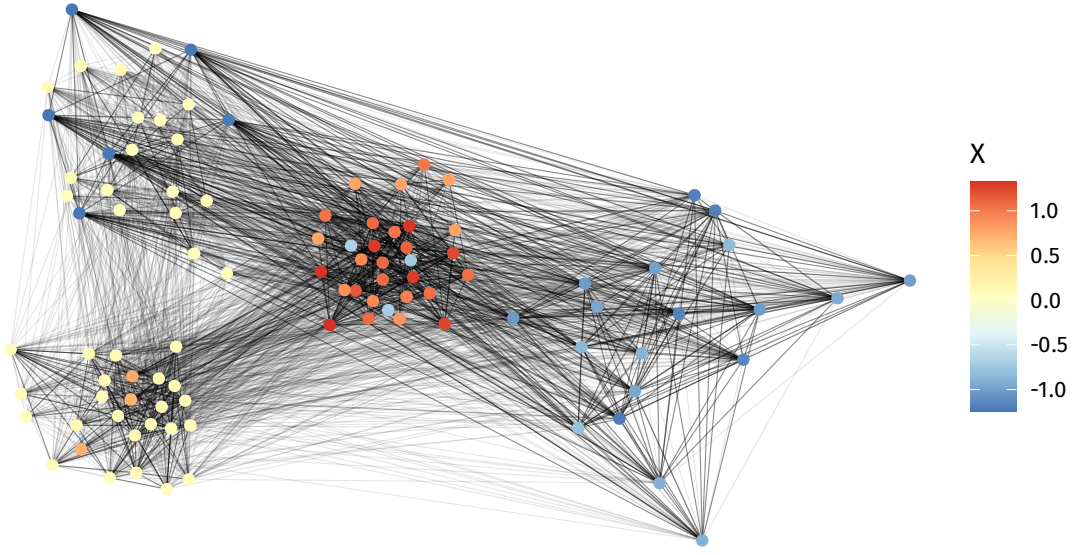
Our theory does apply to weighted and directed networks

Weighted networks: If $A \in \mathbb{R}^{n \times n}$ is a weighted network with (ν, b)-subgamma edges A_{ij}, we require that

$$\max_{i \in [n]} \frac{1}{d_i^2} \sum_{j=1}^{n} A_{ij}^2 = o\left(\frac{1}{\nu \log^2 n}\right) \quad \text{and} \quad \max_{j \in [n]} \frac{A_{ij}}{d_i} = o\left(\frac{1}{b \log n}\right).$$

Directed networks: extension possible, but slightly more involved

What happens when treatment depends on position in network?



Dependence between treatment and network might recover identifiability

$$\underbrace{[GT]_i}_{\substack{\text{fraction} \\ \text{vaccinated} \\ \text{friends}}} = \underbrace{\frac{1}{d_i} \sum_{j\,:\,A_{ij}=1} T_j}_{\text{average of dependent treatments}}$$

GT might not coverge, or might converge to non-constant value

We investigate identification when treatments depend on network structure

Degree-corrected stochastic blockmodels

$$\pi = [\pi_1, \pi_2, ..., \pi_d]$$

$$Z_i \overset{\text{iid}}{\sim} \text{Categorical}(\pi)$$

$$\theta_i \overset{\text{iid}}{\sim} F_\theta$$

$$B = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1d} \\ B_{21} & B_{22} & \dots & B_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{d1} & B_{d2} & \dots & B_{dd} \end{bmatrix} \in [0,1]^{d \times d}$$

$$\mathbb{P}(A_{ij} = 1 \,|\, Z, \theta) = \theta_i Z_i B Z_j^T \theta_j$$

We investigate identification when treatments depend on network structure

$$Z_i = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } Z_j = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{aligned} &\mathbb{P}(A_{ij} = 1 \,|\, Z, \theta) \\ &= \theta_i \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} B_{11} & B_{1\leftrightarrow 2} & B_{1\leftrightarrow 3} & B_{1\leftrightarrow 4} \\ B_{1\leftrightarrow 2} & B_{22} & B_{2\leftrightarrow 3} & B_{2\leftrightarrow 4} \\ B_{1\leftrightarrow 3} & B_{2\leftrightarrow 3} & B_{33} & B_{3\leftrightarrow 4} \\ B_{1\leftrightarrow 4} & B_{2\leftrightarrow 4} & B_{3\leftrightarrow 4} & B_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \theta_j \\ &= \theta_i \cdot B_{2\leftrightarrow 3} \cdot \theta_j \end{aligned}$$

Consider the X_i in place of T_i where

$$X_i = \theta_i Z_i$$

Identification is possible when treatments depend on network structure

Theorem

Suppose that A is sampled from a degree-corrected stochastic blockmodel. Define $X_i = \theta_i Z_i$*. Let*

$$Y = \alpha 1_n + \beta GY + X\gamma + GX\delta + \varepsilon$$

for $\alpha, \beta \in \mathbb{R}$ *and* $\gamma, \delta \in \mathbb{R}^d$*. Suppose that X has* $k \geq 2d$ *distinct rows. Then, under some conditions,*

$$W_n = \begin{bmatrix} 1_n & GY & X & GX \end{bmatrix}$$

converges uniformly to a limit object with rank $2d$ *out of* $2d + 2$*. If any two entries of* $(\alpha, \beta, \delta_1, ..., \delta_d)$ *are set to zero in the data generating process, the limit object of* W_n *is a matrix with full rank.*

* We aren't recommending this model, merely demonstrating that identification is feasible via dependence between the network and nodal covariates

We performed a simulation study to confirm the theoretical results

Sample network from degree-corrected stochastic blockmodel with n nodes

$$\pi = [1/4, 1/4, 1/4, 1/4]$$

$$\theta_i \overset{\text{iid}}{\sim} \text{Uniform}[1, 2]$$

$$B = \begin{bmatrix} 0.5 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.5 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.5 \end{bmatrix}$$

$$n \in \{100, 163, 264, 430, 698, 1135, 1845\}$$

Scale B such that $\mathbb{E}[d_i] = 2n^{0.7}$

The simulations considered asymptotically unidentified and identified models

- RCT model: Treatment random and independent of network. $T_i \overset{\text{iid}}{\sim} \text{Bern}(0.5)$

$$Y = \alpha 1_n + \beta GY + T\gamma + GT\delta + \varepsilon,$$

with $\alpha = 3, \beta = 0.2, \gamma = 4, \delta = 2$ and $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$.

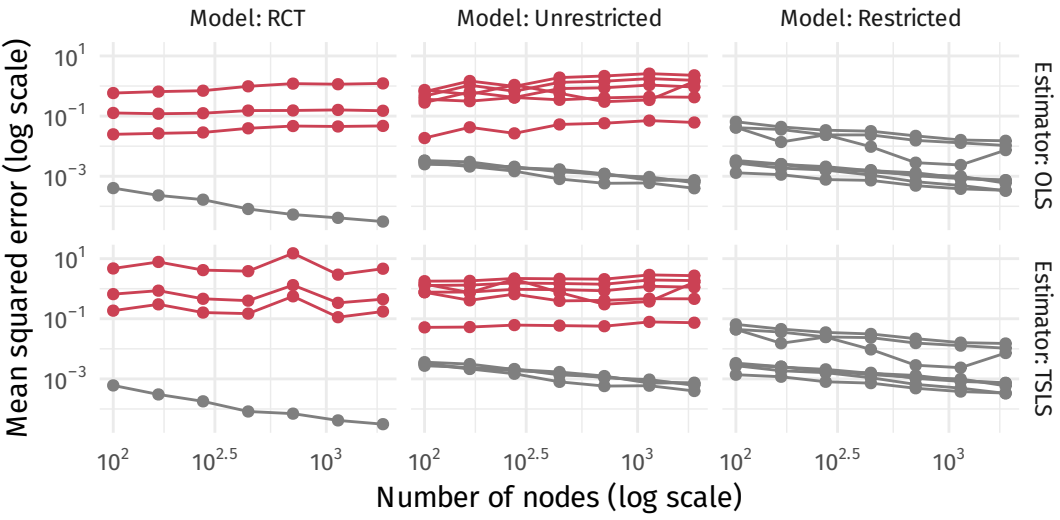
- Unrestricted model: Treatment random and dependent on network. Define $X_i = \theta_i Z_i \in \mathbb{R}^4$

$$Y = \alpha 1_n + \beta GY + X\gamma + GX\delta + \varepsilon,$$

where $\alpha = 3, \beta = 0.2$ and $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.1$. Since $X_i \in \mathbb{R}^4$, $\gamma, \delta \in \mathbb{R}^4$ and we fix $\delta = (2, 2, 2, 2)$ and $\gamma = (1.5, 2.5, 3.5, 4.5)$.

- Restricted model: The *Unrestricted* model, but $\delta = (0, 0, 2, 2)$.

Asymptotically unidentified coefficients cannot be estimated



Based on 100 Monte Carlo replicates

We have important takeaways for social scientists

- Linear-in-means models can be asymptotically unidentified
- Asymptotically unidentified peer effects cannot be estimated
- Treatments independent of network lead to identification failure
- Treatments dependent on network must considered on a case-by-case basis

Our work raises important methodological questions

- Are there realistic models of network-treatment dependence?
- What happens in longitudinal models?
- What happens in sparse networks?
- Are peer effects hopeless?
- What alternatives are there to linear-in-means models?

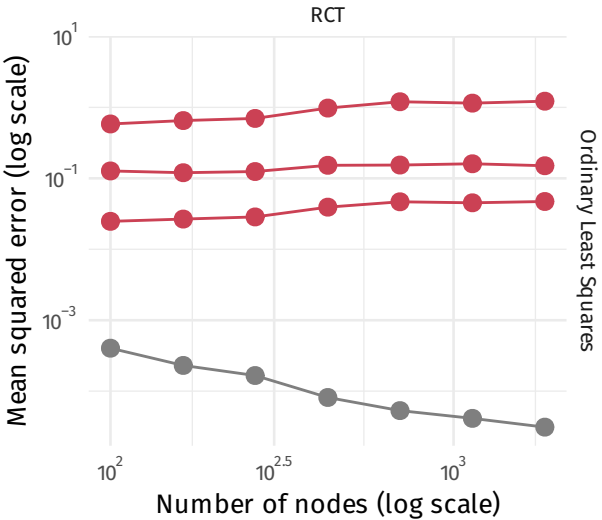
Thank you! Questions?

@alexpghayes

alex.hayes@wisc.edu

alexpghayes.com

github.com/alexpghayes



Estimators

- OLS: $lm(y \sim Gy + T + GT)$
- TSLS: $ivreg(y \sim Gy + T + GT \mid T + GT + G^2T)$

Proposition (Bramoullé et al. 2009)
Suppose $\gamma\beta + \delta \neq 0$ *(i.e., the peer effects are not zero). If* I, G *and* G^2 *are linearly independent, in the sense that* $aI + bG + cG^2 = 0$ *requires* $a = b = c = 0$*, then* α, β, γ *and* δ *are identified.*

Definition (Random Dot Product Graph, Young and Scheinerman 2007)
Let F be a distribution on $\mathbb{R}^d$ such that $0 \leq x^T y$ for all $x, y \in \operatorname{supp} F$ and the convex cone of $\operatorname{supp} F$ is d-dimensional. Draw $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} F$, and collect these in the rows of $X \in \mathbb{R}^{n \times d}$ for ease of notation. Conditional on these n vectors, which we call *latent positions*, generate edges by drawing the edges $\{A_{ij} : 1 \leq i < j \leq n\}$ as independent (ν, b)-subgamma random variables with $\mathbb{E}[A_{ij} \mid X] = \rho X_i^T X_j$, where $\rho \in [0, 1]$. Then we say that A is distributed according to an n-vertex random dot product graph with latent position distribution F, (ν, b)-subgamma edges and sparsity factor ρ. We write $(A, X) \sim \operatorname{RDPG}(F, n)$, with the subgamma and sparsity parameters made clear from the context.

Theorem

Suppose that (A, X) are sampled from a random dot product model where X is rank d with probability 1. Let ε be a vector of mean zero, i.i.d. $(\nu_\varepsilon, b_\varepsilon)$-subgamma random variables, with $(\nu_\varepsilon, b_\varepsilon)$ not depending on n, and let

$$Y = \alpha 1_n + \beta GY + X\gamma + GX\delta + \varepsilon$$

for $\alpha, \beta \in \mathbb{R}$ and $\gamma, \delta \in \mathbb{R}^d$. Suppose that X has $k \geq 2d$ distinct rows. Then, under some conditions,

$$W_n = \begin{bmatrix} 1_n & GY & X & GX \end{bmatrix}$$

converges uniformly to a limit object with rank $2d$ out of $2d + 2$. If any two entries of $(\alpha, \beta, \delta_1, ..., \delta_d)$ are set to zero in the data generating process, the limit object of W_n is a matrix with full rank, and $\alpha, \beta, \gamma, \delta$ are thus asymptotically identifiable.

Proposition

Let $\mu = \mathbb{E}[X] \in \mathbb{R}^d$ and suppose that $Y_1, Y_2, \ldots, Y_d, Z_1, Z_2, \ldots, Z_d \in \mathbb{R}^d$ are rows of $X \in \mathbb{R}^{n \times d}$ such that $Y_1, Y_2, \ldots, Y_d$ are linearly independent and $Z_1, Z_2, \ldots, Z_d$ are linearly independent.

$$H_Y = \operatorname{diag}\left(Y_1^T\mu, Y_2^T\mu, \ldots, Y_d^T\mu\right) \quad \textit{and} \quad H_Z = \operatorname{diag}\left(Z_1^T\mu, Z_2^T\mu, \ldots, Z_d^T\mu\right).$$

Provided that $Z^{-1}H_Z^{-1}Z - Y^{-1}H_Y^{-1}Y \in \mathbb{R}^{d \times d}$ is invertible, then the matrix

$$M = \begin{bmatrix} X & H^{-1}X \end{bmatrix} \in \mathbb{R}^{n \times 2d}$$

has rank $2d$.

Morally: need degree heterogeneity so that X and $D^{-1}X$ are linearly independent

Definition (Boucheron et al. 2013)
Let Z be a mean-zero random variable with cumulant generating function $\psi_Z(t) = \log \mathbb{E}\left[e^{tZ}\right]$. Z is *subgamma* with parameters $\nu \geq 0$ and $b \geq 0$ if

$$\psi_Z(t) \leq \frac{t^2\nu}{2(1-bt)} \quad \text{and} \quad \psi_{-Z}(t) \leq \frac{t^2\nu}{2(1-bt)} \quad \text{for all} \quad t < 1/b.$$

We then write that Z is (ν, b)-subgamma.

Intuition: tails decay at Gamma rates or faster

Examples: Bernoulli, Poisson, Exponential, Gamma, Gaussian, sub-Gaussian, squared sub-Gaussians, bounded distributions, etc

Technical conditions for partial identification result

- $\rho = \omega\left(\frac{\log^2 n}{\sqrt{n}}\right)$ and $\frac{\nu + b^2}{\rho} = \Theta(1)$
- $\min_{i \in [n]} \left| X_i^T \, \mathbb{E}[X_1] \right| = \omega\left(\frac{\log^2 n}{\sqrt{n}\rho}\right)$ almost surely.
- $\max_{i \in [n]} \|X_i\| = o(\sqrt{n})$ almost surely.
- $\mathbb{E}\|X_1\|^2 < \infty$.

References

Boucheron, S., G. Lugosi, and P. Massart (2013, February). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009, May). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Young, S. J. and E. R. Scheinerman (2007). Random Dot Product Graph Models for Social Networks. In A. Bonato and F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph*, Volume 4863, pp. 138–149. Berlin, Heidelberg: Springer Berlin Heidelberg.