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

Alex Hayes | PhD Defense

2024-04-04 @ 12:30 pm

Service Memorial Institute 133, Medical Sciences Center, UW-Madison

Department of Statistics, University of Wisconsin-Madison

What are peer effects?

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

* Not a causal talk. But causally inspired.

The canonical linear model for peer effects

$$\underbrace{Y_i}_{\text{sick?}} = \underbrace{\alpha}_{\substack{\text{base}\\\text{rate}}} + \underbrace{\beta}_{\substack{\text{contagion}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} Y_j}_{\substack{\text{portion}\\\text{sick}\\\text{peers}}} + \underbrace{\gamma}_{\substack{\text{direct}\\\text{effect}}} \underbrace{T_i}_{\text{vaccinated?}} + \underbrace{\delta}_{\substack{\text{interference}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j}_{\substack{\text{portion}\\\text{vaccinated}\\\text{peers}}} + \underbrace{\varepsilon_i}_{\text{error}}$$

Outcome	Y_i	$\in \{0, 1\}$
Treatment	T_i	$\in \{0, 1\}$
Adjacency matrix	A	$\in \{0, 1\}^{n \times n}$
Edge $i \sim j$	A_{ij}	$\in \{0, 1\}$
Node degree	d_i	$\in \mathbb{Z}^+$

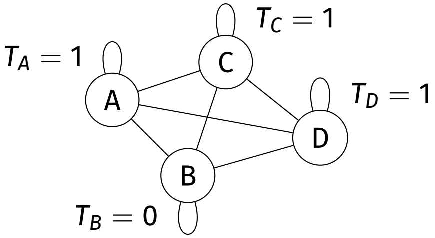
Identification of Endogenous Social Effects: The Reflection Problem

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.

Manski's reflection problem: highly structured networks break the model

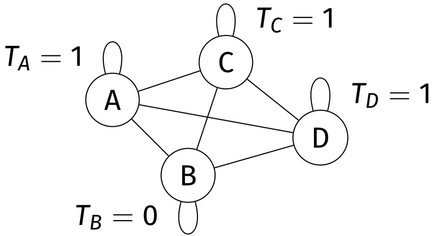


Average value of T amongst peers is the same for all nodes!

$$GT_A = 3/4$$
$$GT_B = 3/4$$
$$GT_C = 3/4$$
$$GT_D = 3/4$$

Problem: cannot distinguish base rate α from interference effect δ

Manski's reflection problem: highly structured networks break the model



$$
\begin{bmatrix} Y_A \\ Y_B \\ Y_C \\ Y_D \end{bmatrix} = \underbrace{\begin{array}{cccc} 1_n & GY & T & GT \\ \end{array} \begin{bmatrix} 1 & GY_A & 1 & 3/4 \\ 1 & GY_B & 0 & 3/4 \\ 1 & GY_C & 1 & 3/4 \\ 1 & GY_D & 1 & 3/4 \end{bmatrix}}_{W} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_A \\ \varepsilon_B \\ \varepsilon_C \\ \varepsilon_D \end{bmatrix}
$$

Problem: Design matrix W becomes collinear!

Identification

Define the degree matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$, where $d_i = \sum_j A_{ij}$. Let $G = D^{-1}A$ be the row-normalized adjacency matrix. Then

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

Definition

We say that $(\alpha, \beta, \gamma, \delta)$ are *identified* when the columns of the design matrix

$$W_n = \begin{bmatrix} 1_n & GY & T & GT \end{bmatrix}.$$

are linearly independent. Otherwise, we say that $(\alpha, \beta, \gamma, \delta)$ are *unidentified*.

We assume that $\mathbb{E}\left[\varepsilon | W_n\right] = 0$.

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

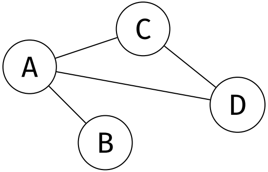
JEL classification:
D85
L14
Z13
C3

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].

Bramoullé: intransivity (i.e, open triangles) fixes the problem

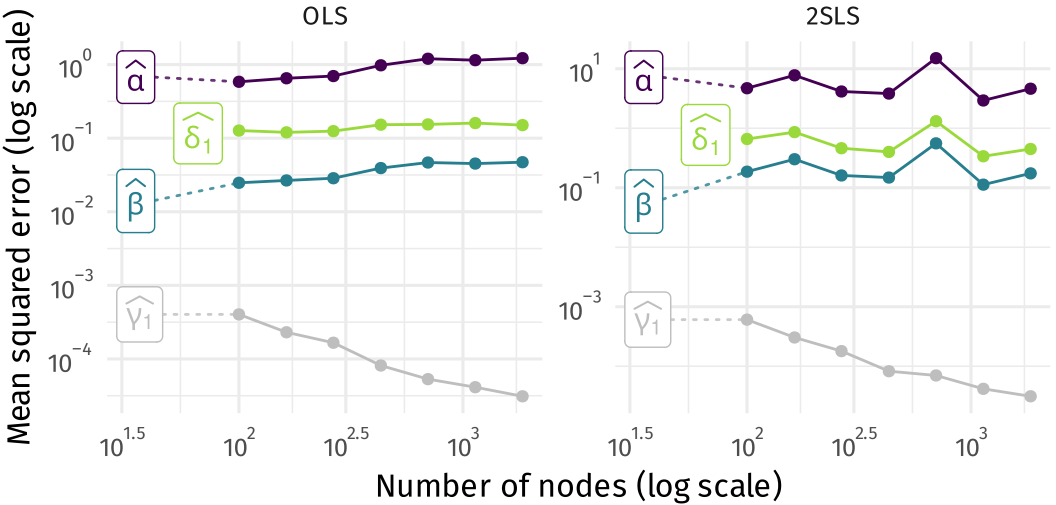


$A \leftrightarrow B \leftrightarrow D$ is an intransitive triangle. If B were friends with D it would “close” the triangle

Since 2009: as long as there is intransitive, $(\alpha, \beta, \gamma, \delta)$ are identified, practitioners are good to use the linear-in-means model

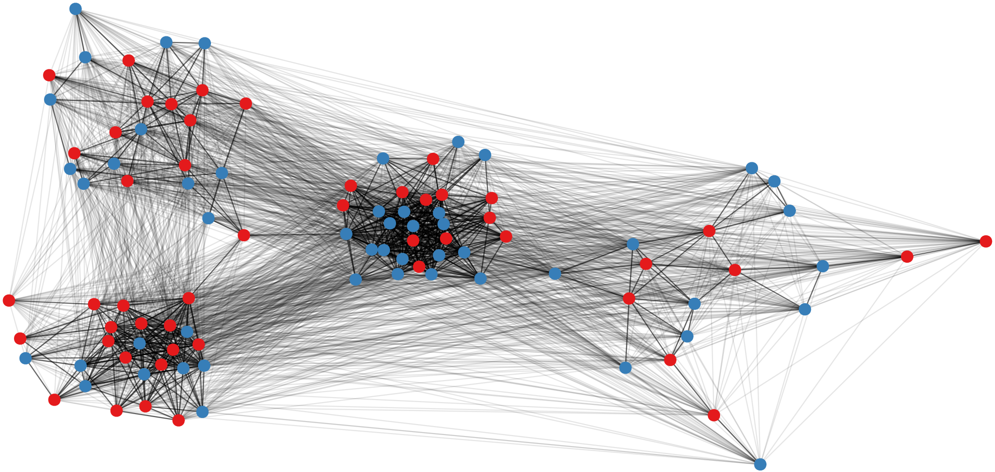
Reflection problem solved!

Simulations on an intransitive network



Randomized experiment on a stochastic blockmodel

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



Let $G_{ij} = A_{ij}/d_i$ be the row-normalization of A. We express the previous model in matrix-vector notation as

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

If $|\beta| < 1$, then $I - \beta G$ is invertible and there is a unique solution

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

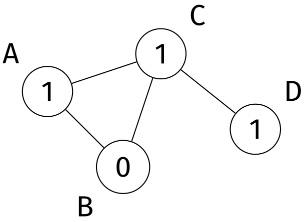


Figure 1: *T*

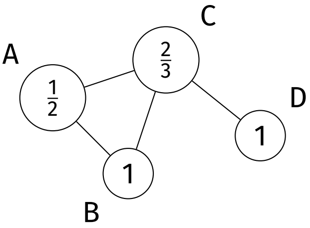


Figure 2: *GT*

Proposition (Finite sample identification, Bramoullé et al. (2009))

Let ε be mean zero, i.i.d. noise and let

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

Suppose that $|\beta| < 1$ and $\gamma\beta + \delta \neq 0$. 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.

$\gamma\beta + \delta \neq 0$ means that there is either some interference effect, or some direct effect and some contagion effect, and if there are both, they don't cancel each other out.

Proposition

If $\gamma\beta + \delta \neq 0$ *and G has three or more distinct eigenvalues, then* α, β, γ *and* δ *are identified.*

Promising! If *G* is from a stochastic blockmodel, identification likely.

Disappearing asymptotic identification: a simple case (Li and Wager, 2022)

$$\underbrace{Y_i}_{\text{sick?}} = \underbrace{\alpha}_{\substack{\text{base}\\\text{rate}}} + \underbrace{\gamma}_{\substack{\text{direct}\\\text{effect}}} \underbrace{T_i}_{\text{vaccinated?}} + \underbrace{\delta}_{\substack{\text{interference}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j}_{\substack{\text{portion}\\\text{vaccinated}\\\text{peers}}} + \underbrace{\varepsilon_i}_{\text{error}}$$

Problem: GT term converges to a constant

$$\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j \to \pi$$

GY can also be a problem

Suppose no node is isolated (otherwise the intercept differs for isolated and connected nodes in reduced form) then the reduced form of Y is given by

$$Y = \frac{\alpha}{1-\beta}1_n + \gamma T + (\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k G^{k+1} T + \sum_{k=0}^{\infty}\beta^k G^k \varepsilon$$

then

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

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

If neighborhood averages GT are converging, repeated neighborhood averages converge as well[1], because $G1_n = 1_n$ (neighborhood average of constants is constant)

[1] Possibly under suitable conditions

Roughly, if

1. neighborhoods d_i are all getting larger in n, and
2. there is no CLT/LLN breaking strong dependence between network A and nodal covariate T

Then

$$\lim_{n \to \infty} GT \to C \qquad C \in \mathbb{R}$$
$$\lim_{n \to \infty} GY \to C' \qquad C' \in \mathbb{R}$$

Then contagion and interference effects are the same for everyone, and indistinguishable from base rates

The direct effect γ is still identified.

Theorem (Identification failure when T independent of A)

Let ε be mean zero, i.i.d. noise and let

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

Suppose $|\beta| < 1$. If T is sub-gamma and independent of the network A, and $\min_{i \in [n]} d_i = \omega(\log n)$, then there exists constants $C, C' \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \sup_{i\in[n]} |GT - 1_n C| = o_p(1), \qquad \text{and}$$

$$\lim_{n\to\infty} \sup_{i\in[n]} \left|GY - 1_n C'\right| = o_p(1).$$

* This covers fixed networks and conditioning on the network

Theorem (Identification failure in random dot product graphs)

Let ε be mean zero, i.i.d. noise and let

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

Suppose $|\beta| < 1$. Let $(A, X) \sim RDPG(F, n)$ and let $T = X \in \mathbb{R}^{n \times d}$. Then there exists constants $C, C' \in \mathbb{R}^d$ such that

$$\lim_{n \to \infty} \sup_{i \in [n]} |GT - 1_n C| = o_p(1), \qquad \textit{and}$$

$$\lim_{n \to \infty} \sup_{i \in [n]} \left| GY - 1_n C' \right| = o_p(1).$$

Why this matters

1. Shalizi and Thomas (2011); McFowland and Shalizi (2021): homophily and contagion are non-parametrically confounded. Must make parametric assumptions to estimate contagion effects.
2. Bramoullé et al. (2020): most parametric contagion identification theory considers fixed n setting but not the $n \to \infty$ limit. Many results won't hold under stochastic blockmodels.

Open questions

- Do *GT* and *GY* ever not converge to constants?
- Are longitudinal contagion models also affected?
- What happens in sparse networks asymptotically?

Thank you! Questions?

Stay in touch

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

References

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

Bramoullé, Y., H. Djebbari, and B. Fortin (2020). Peer Effects in Networks: A Survey. *Annual Review of Economics 12*(1), 603–629.

Li, S. and S. Wager (2022, March). Random Graph Asymptotics for Treatment Effect Estimation under Network Interference. *arXiv:2007.13302 [math, stat]*.

McFowland, E. and C. R. Shalizi (2021, July). Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association 0*(0), 1–12.

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.