# Identifiability of homophily and contagion in social networks

Alex Hayes and Ajinkya Kokandakar

Networks Reading Group

2022-02-23

# What we're distinguishing between

Contagion: peer outcomes influence my outcome

- you get covid and give me covid

c.f. Interference: peer *treatments* influence my outcome

- you don't get vaccinated and give me covid

Homophily: similar people are connected in networks

- we all get vaccinated because we trust the CDC, so none of us get covid

# Timeline

- Christakis and Fowler 2007: flashy claim
- Skepticism by (many) statisticians
- Shalizi and Thomas 2011: negative non-parametric identification result
- McFowland and Shalizi 2021: parametric trick to avoid prior negative identification result

## The Spread of Obesity in a Large Social Network Over 32 Years

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

ABSTRACT

BACKGROUND

The prevalence of obesity has increased substantially over the past 30 years. We performed a quantitative analysis of the nature and extent of the person-to-person spread of obesity as a possible factor contributing to the obesity epidemic.

METHODS

We evaluated a densely interconnected social network of 12,067 people assessed repeatedly from 1971 to 2003 as part of the Framingham Heart Study. The body-mass index was available for all subjects. We used longitudinal statistical models to examine whether weight gain in one person was associated with weight gain in his or her friends, siblings, spouse, and neighbors.

RESULTS

Discernible clusters of obese persons were present in the network at all time points, and the clusters extended to three degrees of separation. These clusters did not appear to be solely attributable to the selective formation of social ties among obese persons. A person's chances of becoming obese increased by 57% (95% confidence interval [CI], 6 to 123) if he or she had a friend who became obese in a given interval. Among pairs of adult siblings, if one sibling became obese, the chance that the other would become obese increased by 40% (95% CI, 21 to 60). If one spouse became obese, the likelihood that the other spouse would become obese increased by 37% (95% CI, 7 to 73). These effects were not seen among neighbors in the immediate geographic location. Persons of the same sex had relatively greater influence on each other as compared with those of the opposite sex. The spread of smoking cessation did not account for the spread of obesity in the network.

CONCLUSIONS

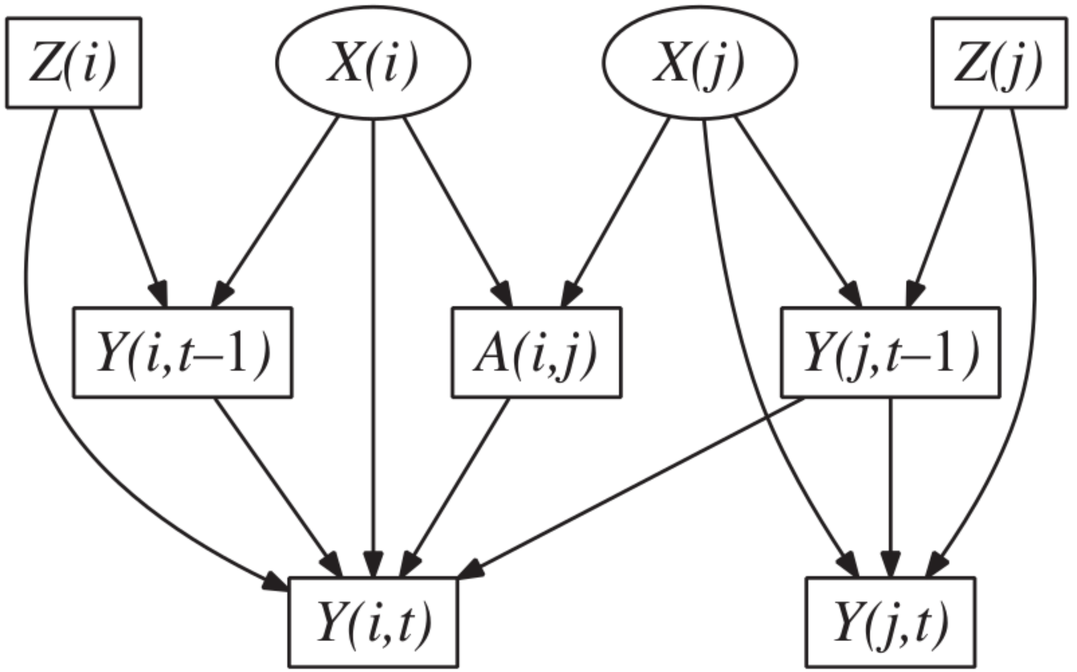
Network phenomena appear to be relevant to the biologic and behavioral trait of obesity, and obesity appears to spread through social ties. These findings have implications for clinical and public health interventions.

From the Department of Health Care Policy, Harvard Medical School, Boston (N.A.C.); the Department of Medicine, Mt. Auburn Hospital, Cambridge, MA (N.A.C.); the Department of Sociology, Harvard University, Boston (N.A.C.); and the Department of Political Science, University of California, San Diego, San Diego (J.H.F.). Address reprint requests to Dr. Christakis at the Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave., Boston, MA 02115, or at christakis@hcp.med.harvard.edu.

N Engl J Med 2007;357:370-9.

# Shalizi and Thomas 2011

- X(i), X(j) encode latent homophily
  - i.e. community membership in blockmodel
- Z(i), Z(j) node covariates
- A(i,j) edges in graph
- Y(i, t) node i outcome at time t




$$\psi = \mathbb{E}[Y(i,t) \mid Y(j,t-1)]$$

Key result: Peer effect is not np-identified under this DAG by a standard confounding/d-separation argument

# Shalizi and Thomas 2011

## Other stuff in there

1. Intuition about homophily and contagion
2. Extended discussion of how Christakis and Fowler's (2007) approach to identification fails **
3. Claim: this is a big and general problem

[**] Egami, Naoki, and Eric J. Tchetgen Tchetgen. “Identification and Estimation of Causal Peer Effects Using Double Negative Controls for Unmeasured Network Confounding.” ArXiv:2109.01933 [Stat], September 4, 2021. http://arxiv.org/abs/2109.01933.

# McFowland and Shalizi 2021

**Parametric contagion model:**

$$Y_{i,t+1} = \alpha_0 + \alpha_1 Y_{i,t} + \beta \frac{\sum_j \left(Y_{j,t} A_{ij}\right)}{\sum_j A_{ij}} + \gamma_1^T C_i + \gamma_2^T X_i + \epsilon_{i,t+1}$$

**Parametric network model:**

1. Stochastic block model

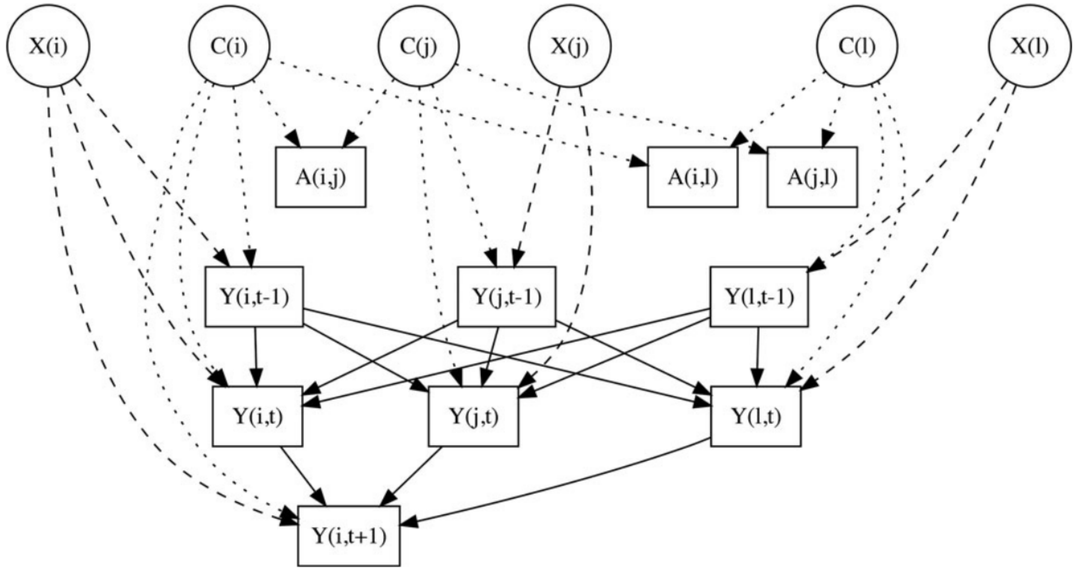
$$\mathbb{P}[A_{ij} \mid C_i, C_j] = C_i^T C_j$$

2. Continuous latent space model

$$\mathbb{P}[A_{ij} \mid C_i, C_j] = \text{logit}(C_i^T C_j)$$

Network assumed to be static over time

Key result: Now the network A is so informative about the latent positions C that we can use estimates of C in OLS to control for the problematic confounding of Shalizi and Thomas 2011



# Discuss!

Realism of assumptions:

- static network
- covariates and latent positions independent
- parametric form of contagion from neighbors

Is identifiability a real problem or a mathematical problem here?