Introduction to Network Interference

Cornell STSCI / INFO / ILRST 3900
Fall 2025
causal3900.github.io

Nov 18 2025

Introduction

- Assistant Professor in Operations Research and Information Engineering
- Research broadly in areas of high dimensional statistics, sequential decision making, and causal inference on networks

Learning goals for today

At the end of class, you will be able to

- Define network interference
- Understand the challenge of network interference
- Use an exposure mapping to express structured interference

Logistics

- Quiz 5 today
- Problem Set 5 peer reviews due today
- Discussion sessions this week will be project help sessions
- Project Check-ins due Nov 25
- PSET 6 due Nov 25; Quiz 6 Dec 2

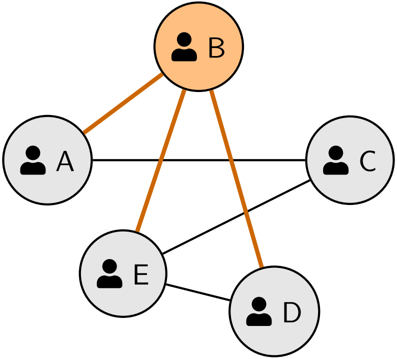
What is Network Interference

- Up to this point we assumed as part of the consistency assumption that the potential outcome of unit i only depends on its own treatment as denoted by Y_i^a
- Under interference the potential outcome of unit i can depend on the treatments of others as well
- Requires a change in notation to indicate the additional dependence, e.g. $Y_i^{\mathbf{a}}$ where $\mathbf{a} = (a_1, a_2, \ldots a_n)$
- In full generality, unit i could have 2^n different potential outcomes, thus it is impossible to do estimation under full generality $\rightarrow$ need some simplifying assumption

Example of Network Interference

Consider LinkedIn experimenting with a new algorithm for ranking posts on an individual's news feed:

- Suppose individual B is treated and all others are control
- B may have a positive increase in engagement with the app
- B may generate activity that is forwarded to their friends feeds
- A, E, D may increase their engagement due to the increased content in their news feed



Example of Network Interference

Interference could show up in many domains

- Public Health
- Economic & Public Policies
- Social Networks
- Online Platforms / Marketplaces
- Advertising / Marketing

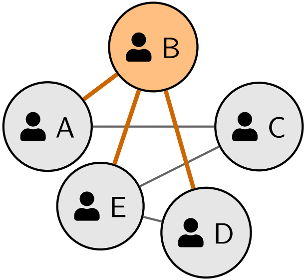
Exposure Mapping under Interference

- Let $\mathbf{a} = (a_1, \ldots, a_n)$ denote vector of treatment assignments.
- Define an **exposure mapping** $e_i = f_i(\mathbf{a})$, which maps from the treatment vector to an exposure level
- Assume that potential outcome $Y_i^{\mathbf{a}}$ depends on $\mathbf{a}$ only through a_i and e_i, written as

$$Y_i^{\mathbf{a}} = Y_i^{(a_i, e_i)} = Y_i^{(a_i, f_i(\mathbf{a}))}$$

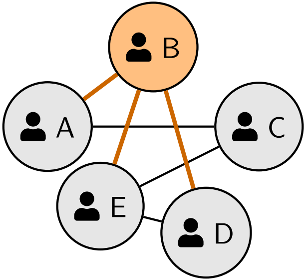
- A_i and E_i denotes the realized treatment and exposure level of unit i,

$$Y_i = Y_i^{(A_i, E_i)}$$



Example of Exposure mappings

- Exposure mapping expresses assumptions on the structure of the interference
- Potential outcome of unit B only depends on a_B and e_B, $Y_B^{\mathbf{a}} = Y_B^{(a_B, e_B)}$
- Ex 1: Neighborhood interference given network is expressed as $e_B = (a_A, a_E, a_D)$
- Ex 2: Anonymous interference where outcome depends only on number of treated neighbors is expressed as $e_B = a_A + a_E + a_D$



Defining Causal Estimands

Need to revisit definition of causal effects under interference:

- Previously causal effect was defined as $E[Y^1] - E[Y^0]$
- Under interference, potential outcomes depend also on exposure level, which needs to be specified
- What causal effects would we potentially care to estimate under interference?
- Depends on what pairs of potential outcomes we want to contrast

Defining Causal Estimands

Consider the setting of neighborhood interference:

- Global Average Treatment Effect

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{(1,\mathbf{1})} - Y_i^{(0,\mathbf{0})}\right)$$

- Direct Average Treatment Effect

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{(1,\mathbf{0})} - Y_i^{(0,\mathbf{0})}\right)$$

- Indirect Average Treatment Effect

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{(0,\mathbf{1})} - Y_i^{(0,\mathbf{0})}\right)$$

Challenges of Estimation under Interference

Return to our LinkedIn example:

- Suppose we conduct a randomized control trial where each unit is independently at random placed in treatment or control
- We want to estimate the Global Average Treatment Effect
- How could the difference in means estimator be misleading?

$$\frac{\sum_{i:A_i=1} Y_i}{\sum_i A_i} - \frac{\sum_{i:A_i=0} Y_i}{\sum_i (1 - A_i)}$$

- Do you expect the difference in means to over-estimate or under-estimate the Global Average Treatment Effect?
- In fact the sign of the bias could be positive or negative depending on if the newly proposed algorithm is better or worse than the status quo

Challenges of Interference

Sometimes the direct / indirect effects could have opposite signs!

- Random subset of unemployed individuals given access to new job training and assistance program
- Outcome - whether an individual gets a job within 90 days
- Positive direct effect on treated individuals
- Negative indirect effect on individuals who did not receive the treatment but are competing for the same set of jobs
- Negative indirect effects could significantly dampen the benefits due to the positive direct effects
- As running such programs are costly, estimating the magnitude of the effect is important

Basic Solutions

When exchangeability holds we talked about two basic methods for estimation:

- Use inverse treatment probability weighted estimator with estimated propensity scores
- Use standardization and parametric g-formula with an outcome model

The earliest solutions for interference consider these two approaches but modified to consider the exposure mappings.

We'll discuss this more next time!

Learning goals for today

At the end of class, you will be able to

- Define network interference
- Understand the challenge of network interference
- Use an exposure mapping to express structured interference