

Observational studies on networks: entangled treatments and calculating propensity scores

Alexander Volfovsky — Statistics Department, Harvard University, Cambridge, MA

Causal inference: potential outcomes

- ▶ Neyman, 1923, Rubin, 1974.
- ▶ There are n units that are *potentially* assigned to treatments (Z_1, \dots, Z_n) .
- ▶ The potential outcome of unit i under assignment vector (Z_1, \dots, Z_n) is given by $Y_i(Z_1, \dots, Z_n)$.
- ▶ Stable unit treatment value assumption: no interference and only one type of treatment means we can write the potential outcome of individual i as a function of their potential treatments Z_i .
- ▶ Under binary treatment: $(Y_i(0), Y_i(1))$ are the potential outcomes of unit i .
- ▶ Under the hood: assumptions about temporality, etc.

Estimands and objects of interest

- ▶ Frequently interested in the average treatment effect (ATE):

$$ATE = \frac{1}{n} \sum_{i=1}^n Y_i(1) - Y_i(0)$$

- ▶ Curse of causal inference... can't observe $Y_i(0)$ and $Y_i(1)$
- ▶ Unbiased estimate of ATE:

$$\widehat{ATE} = \frac{1}{n_t} \sum_{i:Z_i=1} Y_i^{obs} - \frac{1}{n_c} \sum_{i:Z_i=0} Y_i^{obs}$$

- ▶ ... analysis is straightforward in a randomized experiment setting

Classical observational studies

- ▶ For n individuals observe:
 - ▶ outcomes Y_i
 - ▶ self-selected treatment Z_i
 - ▶ pre-treatment covariates X_i
- ▶ Analysis more complicated: Want to make sure the X_i are balanced between the different self-selected treatment levels.
- ▶ Solution: (propensity score) matching:
 - ▶ Fit your favorite model: $P(Z_i = 1|X_i)$
 - ▶ Generate fitted values: $P(\widehat{Z}_i = 1|X_i)$
 - ▶ Match units with $Z_i = 1$ to units with $Z_i = 0$, minimizing $d(P(\widehat{Z}_i = 1|X_i), P(\widehat{Z}_j = 1|X_j))$
 - ▶ Compute your favorite estimate of your favorite estimand based on the matched sample!
- ▶ We take advantage of being able to model the events $\{Z_i = 1\}$ reasonably easily. This can break with networks!

Causal inference with networks

- ▶ There are n units that are connected in some network G^- .
- ▶ Treatment is a function of a change of the network G^- to a network G^+ . For example $Z_i = f_i(G^-, G^+) = d_i(G^+) - d_i(G^-)$
- ▶ No interference but notation still requires us to write $Y_i(Z_1, \dots, Z_n)$ (see toy example where $Y_i(Z_i)$ is not sufficient).
- ▶ Still in an observational framework so need to understand how to perform matching/weighting.
- ▶ Many estimands of interest: $\tau_m = E(Y_i(m+1)) - E(Y_i(m))$.

What goes wrong with observational studies on networks?

- ▶ Classical methods assume that no interference means we can write $Y_i(Z_i)$ and will in turn model the following propensity:

$$e(k, X_i) = P(Z_i = k|X_i, G^-)$$

- ▶ What's the problem here? These $e(k, X_i)$ are actually estimated conditional on the post treatment network G^+ !
- ▶ We need to marginalize over the post treatment network:

$$P(Z_i = k|X, G^-) = \int_{f_i(G^-, G^+) = k} p(G^+|G^-, X) d\mu(G^+)$$

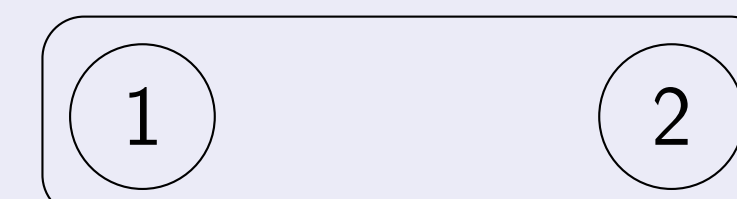
- ▶ This accounts for the uncertainty in the treatment due to the network evolving from G^- to G^+ .

Entangled treatments

- ▶ Most work concentrates on questions of interference of outcomes.
- ▶ Our problem takes a step back from that — what if the treatments are entangled?
- ▶ Treatment: number of new friends in an online game.
- ▶ Treatment: measure of popularity of a website due to new links.
- ▶ Treatment: number of new professional connections.

Entangled treatments: toy example

pre-treatment network: There are two individuals in disconnected G^- :



post-treatment network



No one is treated, that is $Y_1(0, 0)$ and $Y_2(0, 0)$ are observed.

Both are treated, that is $Y_1(1, 1)$ and $Y_2(1, 1)$ are observed.

The treatment is the edge — can't observe one person with an edge and one without.

Theory

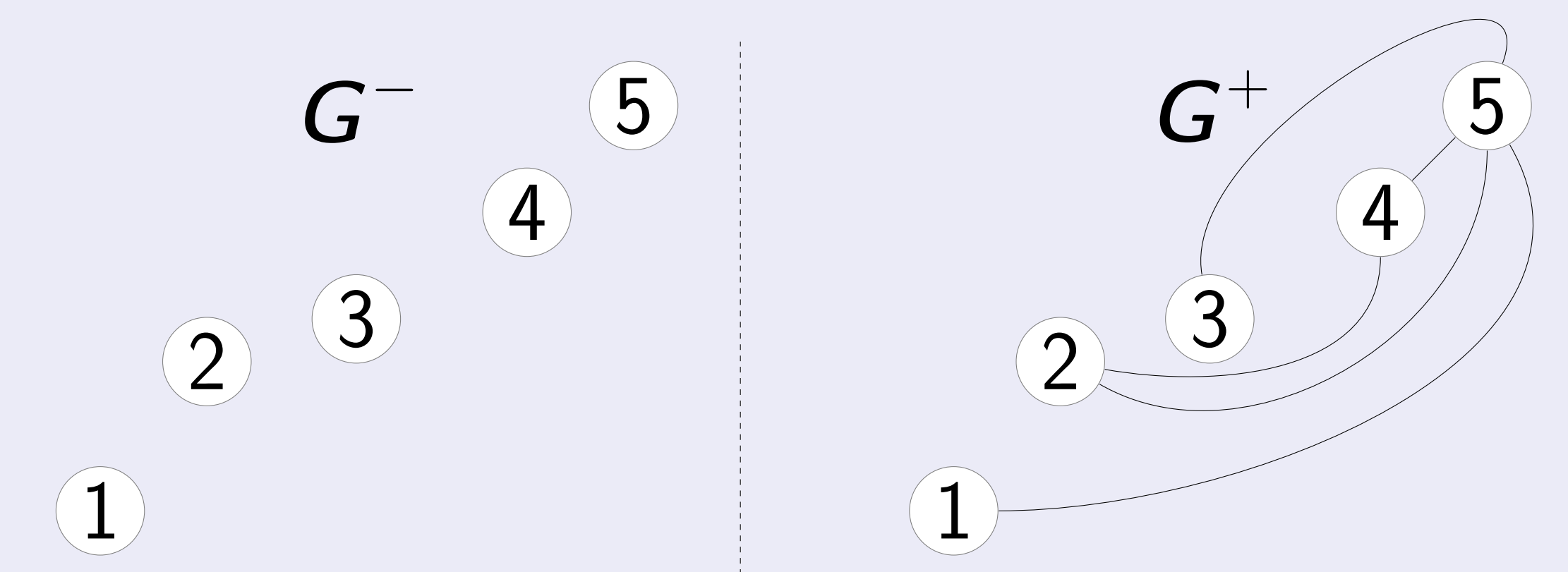
- ▶ Randomized experiment requires the change of one state to another (hence "birth place" is not a treatment usually).
- ▶ For us, G^- changes to G^+ and the individual treatment assignment is given by a contrast function $f_i(\cdot, \cdot)$.
- ▶ The classical setting fits here nicely with G^- being the concentration of acetylsalicylic acid in the blood before the intervention, G^+ the concentration after the intervention and f the individual contrast between those.
- ▶ New strong ignorability for graphs — must be JOINT:

$$P(G^+|G^-, X, Y) = P(G^+|G^-, X)$$

- ▶ Theorem 1: Suppose treatment assignment is strongly graph ignorable and $b(X)$ is a "balancing score" then

$$E[Y(1)|Z = 1, b(X)] - E[Y(0)|Z = 0, b(X)] = E[Y(1) - Y(0)|b(X)]$$

Numerical example



The network G^- is empty and G^+ has independent edges, each of which has probability:

$$P(g_{ij}^+ = 1|G^-, X) \propto \exp(X_i X_j + 1).$$

unit	X_i	Z_i	Y_i^{obs}
1	-5	1	0
2	-1	2	0
3	0	1	1
4	3	2	1
5	10	4	0

Numerical results

Ignoring network information:
 $P(Z_i = k|X_i) \propto \text{Pois}(\exp(\alpha_\beta X_i))$

Incorporating network information:
Follow methodology below

unit (i)	propensity score for $Z_i = \dots$						unit (i)	propensity for $Z_i = \dots$				
	0	1	2	3	4	5		0	1	2	3	4
1	0.37	0.37	0.18	0.06	0.02	0.00	1	0.00	0.27	0.73	0.00	0.00
2	0.24	0.34	0.25	0.12	0.04	0.01	2	0.00	0.24	0.67	0.09	0.00
3	0.21	0.33	0.26	0.13	0.05	0.02	3	0.01	0.06	0.23	0.42	0.28
4	0.13	0.26	0.27	0.19	0.10	0.04	4	0.00	0.24	0.68	0.09	0.00
5	0.02	0.08	0.15	0.20	0.20	0.15	5	0.00	0.27	0.73	0.00	0.00

Set of units that have similar propensities to make one connection or two connections: $\mathcal{S} = \{1, 2, 3, 4\}$

Set of units that have similar propensities to make one connection or two connections: $\mathcal{S} = \{1, 2, 4, 5\}$

Methodology

- ▶ Compute treatment assignment Z_i for each unit i .
- ▶ Model the network G^+ as $G^+ = G^- + \mathcal{E}$, where \mathcal{E} indicates a random graph with a probability distribution $p(\cdot|X, \theta)$.
- ▶ Use the observed new connections in $G^+ - G^-$ to obtain estimates $\hat{\theta}$ of parameter θ , e.g., through maximum-likelihood.
- ▶ Sample J networks $G_{(j)}^+|\hat{\theta}$ conditional on the observed G^- , where J is the total number of samples.
- ▶ Use the samples $\{G_{(j)}^+\}$, $j = 1, \dots, J$ to compute estimates $\hat{e}(k, X_i)$ of the propensity score $e(k, X_i)$:
- ▶ Group/cluster according to estimated propensity scores.
- ▶ Estimate estimand within each group/cluster and combine information across clusters.