Causal Inference with Interference

Yunran Chen

https://github.com/YunranChen/Interference_in_CI

1. Introduction

In observation studies, our target population may have a network structure. Interference may occur if the potential outcome of one unit depends on both individual treatment and neighborhood treatment. The present of interference will break down the traditional assumptions and framework. In this project, we review the paper by Forastiere *et al*. (2016, [1]), which provides a extended framework and solution for the existence of interference. The basic idea for the framework and assumption extension is to take neighbors into consideration, which means to extend potential outcomes, assignment mechanism and average does-response function as a function of both individual and neighborhood treatment. Accordingly, propensity score is generalized to a joint propensity score then decomposed to individual and neighborhood propensity score, from which a estimating procedure is proposed. Here we focus only on causal inference on main effect (individual effect) when interference presents. We reproduced the simulations by Forastiere *et al*. (2016, [1]) based on Facebook data obtained from Stanford Network Analysis Project (SNAP)(Leskovec and Mcaule, 2012 [3]).

2. Problem formulation

2.1. Notations

Let $W_i \in \{0,1\}$, $Y_i \in \mathcal{Y}$ and $\mathbf{X}_i \in \mathcal{X}$ denote the treatment, outcome and covariates of individual i respectively. The covariates can be decomposed to $\mathbf{X}_i^{ind} \in \mathcal{X}^{ind}$ and $\mathbf{X}_i^{neigh} \in \mathcal{X}^{neigh}$, which is only correlated to characteristics of an individual and its neighbors respectively. Denote the undirected network as $G = (\mathcal{N}, \mathcal{E})$. For each node i, consider partitions $(i, \mathcal{N}_i, \mathcal{N}_{-i})$, denoting individual i, i's neighbors $\mathcal{N}_i$, individuals outside i's neighbors $\mathcal{N}_{-i}$ respectively. Therefore, the $\mathbf{W}$ and $\mathbf{Y}$ for the whole population $\mathcal{N}$ can be written as $(W_i, \mathbf{W}_{\mathcal{N}_i}, \mathbf{W}_{\mathcal{N}_{-i}})$ and $(Y_i, \mathbf{Y}_{\mathcal{N}_i}, \mathbf{Y}_{\mathcal{N}_{-i}})$.

2.2. Potential outcomes

Classical stable unit treatment value assumption (SUTVA)(Rubin, 1980 [4], 1986 [5]) assumes consistency and no interference. When the interference presents, the first assumption of SUTVA (consistency) still holds. However, the second assumption of SUTVA is extended to allow neighborhood interference. We called the extended assumption as Stable Unit Treatment Neighborhood Value Assumptions(SUTNVA). Assumption 2.1 indicates there is no other version for the treatment. Assumption 2.2 rules out the effects outside the neighborhoods, but allows effects from neighbors through a specific function $g(\cdot)$. This restriction introduced by the function $g(\cdot)$ can help to reduce the dimension of space of potential outcomes from 2^{N_i} to $|\mathcal{G}_i|$. Let $G_i = g_i(\mathbf{W}_{\mathcal{N}_i})$ denotes the neighborhood treatment. Based on SUTNVA assumption, we can define the potential outcomes as a function of both individual effect W_i and neighborhood effect G_i.

Assumption 2.1 (No Multiple Versions of Treatment(Consistency)) $Y_i = Y_i(\mathbf{W})$

Assumption 2.2 (Neighborhood Interference) *Given a function* $g_i : \{0,1\}^{N_i} \to \mathcal{G}_i$, $\forall i \in \mathcal{N}$, $\forall \mathbf{W}_{\mathcal{N}_{-i}}$, $\mathbf{W}'_{\mathcal{N}_{-i}}$ *and* $\forall \mathbf{W}_{\mathcal{N}_i}, \mathbf{W}'_{\mathcal{N}_i}$,$g_i(\mathbf{W}_{\mathcal{N}_i}) = g_i(\mathbf{W}'_{\mathcal{N}_i})$, *the following equality holds:*

$$Y_i(W_i, \mathbf{W}_{\mathcal{N}_i}, \mathbf{W}_{\mathcal{N}_{-i}}) = Y_i(W_i, \mathbf{W}'_{\mathcal{N}_i}, \mathbf{W}'_{\mathcal{N}_{-i}})$$

Definition 2.1 (Potential outcomes) *For a subset of nodes* $V_g = \{i : g \in \mathcal{G}_i\}$ *where* $G_i = g_i(\mathbf{W}_{\mathcal{N}_i})$

$$Y_i(w, g) = Y_i(W_i = w, G_i = g)$$

2.3. Assignment mechanism

The assignment mechanism is extended to include neighborhood treatment. Since neighborhood treatment G depends on both network structure and treatment of neighbors, G is restricted within a range resulting from $\mathcal{G}$ and $\mathbf{W}$.

Definition 2.2 (Assignment mechanism)

$$P(\mathbf{W}, \mathbf{G}|\mathbf{X}, \{\mathbf{Y}(w, g), w = 0, 1; g \in \mathcal{G}\}) = \begin{cases} P(\mathbf{W}|\mathbf{X}, \{\mathbf{Y}(w, g), w = 0, 1; g \in \mathcal{G}\}) & \textit{if} \quad \mathbf{G} = g(\mathbf{W}) \\ 0 & \textit{otherwise} \end{cases}$$

The unconfoundedness assumption is extended to assume the potential outcomes is conditionally independent to both individual effects and neighborhoods effects.

Assumption 2.3 (Unconfoundness of Individual and Neighborhood Treatment)

$$Y_i(w, g) \perp\!\!\!\perp W_i, G_i|\mathbf{X}_i \qquad \forall w \in \{0, 1\}, g \in \mathcal{G}_i, \forall i$$

2.4. Causal estimands

The main effect $\tau(g)$ and spillover effect $\delta(g; w)$ are defined with holding the neighborhood treatment and the individual treatment as a constant respectively. The overall main effect and overall spillover effect are obtained by marginalizing the neighborhood treatment g. The total effect is defined by setting the controlled unit as untreated for both the individual itself and its neighbors. In addition, the total effect can be written as a sum of over main effect and overall spillover effect at $w = 0$.

Definition 2.3 (Main effect) $\tau(g) = E[Y_i(W_i = 1, G_i = g) - Y_i(W_i = 0, G_i = g)|i \in V_g]$

Definition 2.4 (overall main effect) $\tau = \sum_{g \in \mathcal{G}} \tau(g) P(G_i = g)$

Definition 2.5 (Spillover effect) $\delta(g; w) = E[Y_i(W_i = w, G_i = g) - Y_i(W_i = w, G_i = 0)|i \in V_g]$

Definition 2.6 (overall spillover effect) $\Delta(w) = \sum_{g \in \mathcal{G}} \delta(g; w) P(G_i = g)$

Definition 2.7 (Total effect)

$$TE = \sum_{g \in \mathcal{G}} E[Y_i(W_i = 1, G_i = g) - Y_i(W_i = 0, G_i = 0)|i \in V_g] P(G_i = g) = \tau + \Delta(0)$$

Theorem 2.1 (Identification of Average dose-response function (ADRF)) *If assumption 2.1,2.2,2.3 hold, ADRF is identifiable,* $\forall w \in \{0, 1\}, g \in \mathcal{G}$

$$\mu(w, g) = E[Y_i(w, g)|i \in V_g] = \sum_{x \in \mathcal{X}} E[Y_i|W_i = w, G_i = g, \mathbf{X}_i = \mathbf{x}, i \in V_g] P(\mathbf{X}_i = \mathbf{x}|i \in V_g)$$

3. Generalized Propensity Score Based Estimator

3.1. Definition of GPS

Define the joint propensity score as joint probability of individual and neighborhood treatment conditional on all the covariates. Apply the chain rule to decompose the joint propensity score into individual propensity score and neighborhood propensity score.

Definition 3.1 (Joint propensity score) $\psi(w; g; x) = P(W_i = w, G_i = g|\mathbf{X}_i = \mathbf{x})$

Definition 3.2 (Neighborhood propensity score) $\lambda(g; w; x^g) = P(G_i = g|W_i = w, \mathbf{X}_i^g = \mathbf{x}^g)$

Definition 3.3 (Individual propensity score) $\phi(w; x^w) = P(W_i = w|\mathbf{X}_i^w = \mathbf{x}^w)$

$$\begin{aligned}\psi(w; g; x) &= P(W_i = w, G_i = g|\mathbf{X}_i = \mathbf{x}) \\ &= P(G_i = g|W_i = w, \mathbf{X}_i^g = \mathbf{x}^g)P(W_i = w|\mathbf{X}_i^w = \mathbf{x}^w) \\ &= \lambda(g; w; x^g)\phi(w; x^w)\end{aligned}$$

3.2. Properties of GPS

Based on the extended SUTNVA assumption and unconfoundedness of individual and neighborhood treatment assumption, the generalized proposity score defined previously has the following properties.

Proposition 3.1 (Balancing) $P(W_i, G_i|\mathbf{X}_i, \psi(w; g; \mathbf{X}_i)) = P(W_i, G_i|\psi(w; g; \mathbf{X}_i))$

Proposition 3.2 (Conditional Unconfoundedness (joint))

$$Y_i(w, g) \perp\!\!\!\perp W_i, G_i|\psi(w; g; \mathbf{X}_i) \qquad \forall w \in \{0, 1\}, g \in \mathcal{G}_i$$

Proposition 3.3 (Conditional Unconfoundedness)

$$Y_i(w, g) \perp\!\!\!\perp W_i, G_i|\lambda(g; w; \mathbf{X}_i^g), \phi(1; \mathbf{X}_i^w) \qquad \forall w \in \{0, 1\}, g \in \mathcal{G}_i$$

Proposition 3.1 ensures that if we balance the joint propensity score, we can balance the covariates. Proposition 3.2 ensures the conditional independence of potential outcomes and the treatment given the joint propensity score. Further, the unconfoundedness holds given the individual and neighborhood propensity score. Based on these properties, we can estimate ADRF using individual and neighborhood propensity score.

$$E[E[Y_i|W_i = w, G_i = g, \lambda(g; w; \mathbf{X}_i^g), \phi(1; \mathbf{X}_i^w)]|W_i = w, G_i = g] \tag{1}$$

3.3. Estimating procedure

The GPS-based estimator is conditioned on two propensity scores compared to the ordinary PS-based estimator. The basic idea is to balance one of the propensity score using stratification, then within each stratification, using regression, weighting or matching to estimate ADRF based on the other propensity score. Since the neighborhood PS has multiple levels, we first sub-classify based on individual PS, then within each stratification, using regression conditional on the neighborhood propensity score, individual treatment and neighborhood treatment. At last, we estimate ADRF based on weighted sum of estimation on ADRF within each stratification. The estimating procedure is as follows:

1. Subclassification on $\phi(1; \mathbf{X}_i^w)$
   (a) Estimate $\phi(1; \mathbf{X}_i^w)$: logistic regression $W_i \sim \mathbf{X}_i^w$
   (b) Predict $\phi(1; \mathbf{X}_i^w)$
   (c) Subclassify J subclasses B_j by $\phi(1; \mathbf{X}_i^w)$ where $\mathbf{X}_i^w \perp\!\!\!\perp W_i|i \in B_j$
2. Within B_j, estimate $\mu_j(w, g) = E[Y_i(w, g)|i \in B_j^g]$ where $B_j^g = V_g \cap B_j$
   (a) Estimate $\lambda(g; w; \mathbf{X}_i^g)$: $G_i \sim W_i + \mathbf{X}_i^g$
   (b) Estimate outcome model $Y_i \sim \lambda(g; w; \mathbf{X}_i^g) + W_i + G_i$
   (c) Predict $Y_i(w, g)$
   (d) Estimate $\hat{\mu}_j(w, g) = \sum_{i \in B_j^g} \hat{Y}_i(w, g)/|B_j^g|$
3. Estimate $\hat{\mu}(w, g) = \sum_{j=1}^{J} \hat{\mu}_j(w, g)\pi_j^g$ where $\pi_j^g = |B_j^g|/v_g$

4. Simulation

We estimate the main effect based on a known network and unknown assignment mechanisms. We consider a real friendship network from Facebook, which is obtained from Stanford Network Analysis Project (SNAP)(McAuley and Leskovec, 2012 [3]). This network contains 4039 nodes and 88234 edges. The average clustering coefficient is 0.6055. We consider only two individual covariates $gender_i$, indicating individual's gender, and age_i, a discrete variable indicating individual's age. Consider the node degree and the mean of neighbors' ages and genders as neighborhood covariates. Denote the covariates as $\mathbf{X}_i^{ind} = (gender_i, age_i)$ and $\mathbf{X}_i^{neigh} = (\sum_{k\in\mathcal{N}_i} gender_k/N_i, \sum_{k\in\mathcal{N}_i} age_k/N_i, N_i)$. We consider the function g as the proportion of treated neighbors. We consider two scenarios of dependence between individual treatment W_i and neighborhood treatment G_i.

Scenario 1: $W_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}$. Generate W_i depending on individual gender and age.

Scenario 2: $W_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}, \mathbf{X}_i^{neigh}$. Generate W_i depending on individual gender and age, and on neighbors' gender and age.

We adopt the similar data generating process as the paper, but adjusted the parameters to make sure the balance of covariate distribution and conditional independence. Then we compared the GPS-based estimator to other estimators.

4.1. Assignment mechanism

Scenario 1 We generate individual treatment according to the following propensity score

$$logit(P(W_i = 1)) = -1.5 + 1.2gender_i + 0.3age_i \tag{2}$$

Then we obtain the neighborhood treatment G_i as the proportion of treated neighbors. Here we provide one simulation to show a specific structure induced by the assignment mechanism.

Variables	$\bar{X}_T$	$\bar{X}_C$	Std	$\bar{X}_{G_i\geq 0.5}$	$\bar{X}_{G_i<0.5}$	Std
Gender	1.448	1.178	0.382	1.366	1.283	0.018
Age	3.999	0.563	0.705	3.089	1.188	7.140
Neighbors' Gender	1.337	1.289	0.180	1.340	1.180	0.058
Neighbors' Age	3.123	2.387	0.265	3.154	0.845	9.592
Degree	46.913	37.166	0.130	47.953	12.092	0.480
$G_i(orW_i)$	0.660	0.626	0.144	0.683	0.567	0.165

Table 1. Covariate balance across individual and dichotomized neighborhood treatment arms

Table 1 shows the balance of covariate distributions. For the individual treatment, the individual covariates and the neighborhood covariates are imbalanced, which is due to the propensity score generating process and the homophily of the network respectively. For the dichotomized neighborhood treatment, the imbalance of covariates is due to the homophily. Table 2 shows zero correlation between neighborhood treatment and individual treatment given $\mathbf{X}^{ind}$, which ensures our simulations satisfying $W_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}$.

Variables	Estimate	SE	P-value
Gender	0.055	0.142	0.700
Age	0.003	0.025	0.898
Neighbors' Gender	2.941	0.347	<2e-16
Neighbors' Age	0.702	0.065	<2e-16
Degree	0.082	0.008	<2e-16
W_i	-0.241	0.153	0.117

Table 2. Coefficients of logistic regression

Scenario 2 We generate individual treatment according to the following propensity score

$$logit(P(W_i = 1)) = -5 + 0.5gender_i + 0.1age_i + 2.5neighbors.gender_i + 0.4neighbors.age_i \tag{3}$$

Similarly, for scenario 2, we present a simulation show the structure of the assignment mechanism. Table 3 shows the individual covariates and neighborhood covariates are imbalanced, which is due to the propensity score generating process and homophily. For the dichotomized neighborhood treatment, the imbalance of covariates is due to the propensity score and homophily. Table 4 shows zero correlation between neighborhood treatment and individual treatment given all covariates $\mathbf{X}^w$, but correlation exists if only given individual covariates $\mathbf{X}^{ind}$, which ensures our simulations satisfying $W_i \perp\!\!\!\perp G_i | \mathbf{X}_i^{ind}, \mathbf{X}_i^{neigh}$.

Variables	$\bar{X}_T$	$\bar{X}_C$	Std	$\bar{X}_{G_i \geq 0.5}$	$\bar{X}_{G_i < 0.5}$	Std
Gender	1.422	1.274	0.203	1.402	1.260	0.026
Age	3.948	1.416	0.463	3.669	1.040	10.133
Neighbors' Gender	1.363	1.265	0.378	1.363	1.226	0.053
Neighbors' Age	3.565	1.965	0.616	3.689	1.047	8.187
Degree	53.833	30.154	0.336	53.630	21.180	0.461
G_i(or W_i)	0.665	0.490	0.544	0.676	0.336	0.466

Table 3. Covariate balance across individual and dichotomized neighborhood treatment arms

	$\mathbf{X}^w$			$\mathbf{X}^{ind}$		
Variables	Estimate	SE	P-value	Estimate	SE	P-value
Gender	0.485	0.113	1.95e-05	0.621	0.093	2e-11
Age	0.065	0.022	3.59e-03	0.144	0.018	9.42e-16
Neighbors' Gender	4.323	0.320	<2e-16	-	-	-
Neighbors' Age	1.165	0.061	<2e-16	-	-	-
Degree	-0.001	0.002	0.545	0.025	0.002	<2e-16
W_i	0.064	0.409	<2e-16	0.904	0.101	<2e-16

Table 4. Coefficients of logistic regression

4.2. Outcome models

Here we consider the following outcome models for scenario 1 and 2.

$$Y_i(w, g) \sim N(\mu(w, g, \mathbf{X}_i^{ind}), 1) \tag{4}$$

$$\mu(w, g, \mathbf{X}_i^{ind}) = 15 - 7\mathbf{I}(\phi(1, \mathbf{X}^{ind}) \geq 0.85) - 15w + 3wI(\phi(1, \mathbf{X}^{ind}) \geq 0.85) + \delta g \tag{5}$$

where we consider $\delta \in (-5, -8, -10)$ corresponding to a low, medium and high level of interference. According to the model, we have the true main effect $\tau = -15 + 3\mathbf{I}(\phi(1, \mathbf{X}_i^{ind}) \geq 0.85)$. Notice that the propensity score and the treatment assignment are different for two scenarios.

4.3. Comparison of different estimator

We then compare the GPS-based estimator proposed by with other estimators (Imbens and Rubin, 2015 [2]). Specifically, we run 500 replications of the two scenarios and calculate the following estimator within each replication. The unadjusted estimator is a simple difference in means between treated and untreated units. Regression $\mathbf{X}^{ind}$ and Regression $\mathbf{X}^w$ are ordinary ordinary least squares estimators resulting from $Y \sim W + \mathbf{X}^{ind}$ and $Y \sim W + \mathbf{X}^w$, where $\mathbf{X}^w = (\mathbf{X}^{ind}, \mathbf{X}^{neigh})$. Subclass $\mathbf{X}^{ind}$ and Subclass $\mathbf{X}^w$ are estimators based on subclassification on the individual propensity score estimated by $\hat{\phi}(1, \mathbf{X}^{ind})$

	Unadjusted		Regression$\mathbf{X}^{ind}$		Subclass$\mathbf{X}^{ind}$		Regression$\mathbf{X}^{w}$		Subclass$\mathbf{X}^{w}$		Subclass-GPS	
δ	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
low	-2.260	2.611	-0.572	0.575	-0.283	0.459	-0.573	0.576	-0.334	0.342	-0.039	0.076
med	-2.368	2.369	-0.573	0.578	-0.299	0.487	-0.576	0.580	-0.351	0.363	-0.036	0.072
high	-2.441	2.443	-0.576	0.582	-0.308	0.505	-0.579	0.584	-0.368	0.381	-0.039	0.075
low	-2.990	2.991	-2.175	2.178	-2.126	2.146	-0.270	0.282	-0.181	0.191	-0.001	0.058
med	-3.542	3.544	-2.583	2.585	-2.555	2.580	-0.281	0.295	-0.228	0.240	0.001	0.059
high	-3.915	3.916	-2.859	2.862	-2.846	2.873	-0.292	0.307	-0.262	0.275	-0.000	0.058

Table 5. Estimation of main effect τ

and $\hat{\phi}(1, \mathbf{X}^w)$ respectively. Subclass-GPS is the proposed estimator in section 3.3, which is based on subclassification on the individual propensity score $\hat{\phi}(1, \mathbf{X}^w)$ and model-based (regression) adjustment for the neighborhood propensity score.

Table 5 shows the comparison between different estimators under two different scenarios. The first five estimators are biased and the bias is proportional to the degree of interference δ. This is because the first five estimators ignore the interference. In scenario 1, individual treatment and neighborhood treatment is independent conditional on individual covariates. In this case, the subclassification-based estimators should be unbiasd since they are well adjusted by only individual covariates. The relative large bias may due to the homophily. However, the regression-based estimators are biased due to the misspecified outcome model. The unadjusted estimator is biased due to the confounders $\mathbf{X}^{ind}$. In scenario 2, the individual treatment and neighborhood treatment are not independent if only given the individual covariates. Therefore, the first three estimators are biased with a large deviation due to interference. By adjusting on $\mathbf{X}^{neigh}$, the regression and subclassification based estimators can decrease the bias brought from interference to some degree but cannot eliminate the bias caused by interference. And due to the misspecification, the regression-based estimators have larger bias than the subclassification estimators do. In constrast, the proposed method can eliminate the bias caused by interference through adjusting the individual and neighborhood propensity scores.

5. Acknowledgement

This project is joint work with Dawei Geng. Dawei conducted simulations for scenario 1 and I conducted for scenario 2. Except for the function for proposed estimator (Forastiere *et al*., 2016 [1]), we coded all the simulations on our own. We would like to show our gratitude Dr. Laura Forastiere for providing us the code for Subclass-GPS estimating procedure.

References

[1] L. Forastiere, E. M. Airoldi, and F. Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *arXiv preprint arXiv:1609.06245*, 2016.

[2] G. W. Imbens and D. B. Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[3] J. Leskovec and J. J. Mcauley. Learning to discover social circles in ego networks. pages 539–547, 2012.

[4] D. B. Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

[5] D. B. Rubin. Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81(396):961–962, 1986.