Estimating causal effects of interventions altering social connectivity patterns under network interference

Shinpei Nakamura-Sakai [1] Laura Forastiere [2]

[1]Statistics and Data Science [2]School of Public Health
Yale University

Introduction

Causal inference under network interference is an emerging topic as network data is ubiquitous across multiple disciplines. We say that the treatment effect 'spills over' to other units when one's potential outcomes are affected by the treatment status of other units. Such a phenomenon is often due to social or physical interactions and depends on the population's social structure. An intervention that alters social connectivity would alter the interference mechanism and consequently, the spillover effect. Current methods under interference estimate causal effects conditional on a known and fixed social connectivity graph. [1] [3]

On the other hand, epidemic models have been used to predict the effect of hypothetical interventions altering social connectivity parameters to control the spread of infectious diseases. However, a formal and general definition of the causal effects of such interventions altering the social structure is lacking. We consider a stochastic network formation and propose causal estimands to represent the effects of modifying the network formation mechanism. These causal estimands are defined under interventions shifting the degree distribution or, in general, the network formation mechanism. We develop estimators for the counterfactual estimands under hypothetical interventions altering the network structure, and we investigate these estimators' finite sample bias and large-sample properties.

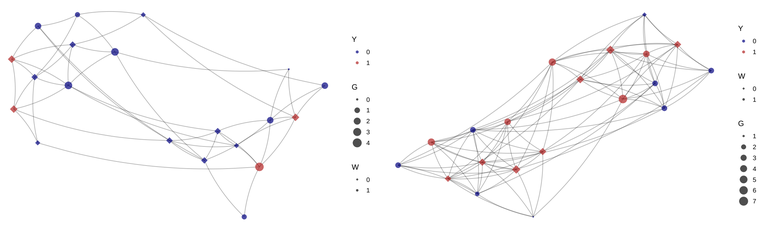
Goals

- Develop causal estimands under the potential outcome framework to define the causal effects of changes in social connectivity under interference
- Develop semi-parametric unbiased estimators for these quantities and investigate their finite sample and large sample properties.
- Provide a realistic simulation study to get an insight into how this framework is useful.

Motivation

An example of the application of this framework is to evaluate the efficacy of a stay-at-home policy as a means of controlling the spread of infection during an epidemiological outbreak and vaccine coverage. Spillover effects of vaccination depend on the social connectivity, which will be affected by state-at-home policies. To assess the effectiveness of the policy, we will compare a specific parameter that captures the propensity for social connectivity under different scenarios. We will visualize the effects of varying this parameter.

In Figure 1, the color of each node represents its infection status with a specific disease. The simulation was designed to reflect the influence of the individual treatment status and the number of unvaccinated neighbors on the infection status. We can observe that a sparser social connectivity pattern results in a lower infection rate.



(a) (b)

Figure 1. (a) Sparser Network, (b) Denser Network

Notation, Estimands, and Estimators

Consider undirected network $(\mathcal{N}, \mathbf{A})$ where $\mathcal{N}$ is a set of nodes or units $\{1, ..., N\}$ and $\mathbf{A} \in \mathbb{R}^{N\times N}$ an adjacency matrix where $A_{ij} \in \{0, 1\}$ indicates whether the unit i and j are connected. Assume that $\mathbf{A}$ is generated by a network formation model $\mathbb{P}_\Theta(\mathbf{A} = \mathbf{a}|\mathbf{X})$ that maps the covariate matrix $\mathbf{X} \in \mathbb{R}^{N\times p}$ to the adjacency matrix $\mathbf{A}$ with parameter Θ. Additionally, denote the treatment assignment $\mathbf{W} \in \{0, 1\}^N$ with assignment mechanism $\mathbb{P}(\mathbf{W} = \mathbf{w}|\mathbf{X})$.

Using exposure mapping function $g : \{0, 1\}^{N-1} \times \{0, 1\}^{\binom{N}{2}} \to \mathbb{R}^l$ where $l \in \mathbb{Z}^+$ we assume the following general form of interference:

$$Y_i(W_i, g(\mathbf{W}_{-i}, \mathbf{A})) = Y_i(W_i, g(\mathbf{W}_{\mathcal{S}_i}, \mathbf{A}))$$

where $\mathcal{S}_i = \nu_i(\mathbf{A})$ is the interference set as a function of $\mathbf{A}$. For example, $\nu_i(\mathbf{A}) = \{j : A_{ij} = 1\}$ indicates $\mathbf{S}_i$ is the set of the first-order neighbours.

$g(\cdot)$ is also a function of $\mathbf{A}$ because we can create weights of $\mathbf{W}_{\mathbf{S}_i}$ from the adjacency matrix. Denote $G_i = g(\mathbf{W}_{\mathcal{S}_i}, \mathbf{A})$. We can denote the potential outcome by $Y_i(w, g)$, where $g \in \mathbb{R}^l$.
Now we define the causal estimand of interest. First, define the hypothetical distribution of the exposure as follows:

$$\begin{aligned}\pi_{HP}(g; \Theta|\mathbf{X}) &= \mathbb{P}_{A,W}(G_i = g; \Theta|\mathbf{X}) \\ &= \sum_{s_i}\sum_{\mathbf{w}} \mathbb{1}(g(\mathbf{w}_{s_i}, \mathbf{a}) = g, \nu_i(\mathbf{a}) = s_i)\mathbb{P}_A(\mathbf{A} = \mathbf{a}; \Theta|\mathbf{X})\mathbb{P}_W(\mathbf{W}_{-i} = \mathbf{w}|\mathbf{X})\end{aligned}$$

where $\mathbb{P}_A(\mathbf{A} = \mathbf{a}; \Theta|\mathbf{X})$ is a network generation model and Θ is a parameter of this model (e.g., the probability of creating a link), which will affect the degree distribution, and in turn G_i.
Then the marginal potential outcome of interest is

$$\overline{Y}(w, \Theta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{g} Y_i(w, g)\pi_{HP}(g; \Theta|\mathbf{X})$$

and the causal estimands of interest are:

- **Direct effect**: effect of receiving the treatment in the same network, $DE(\Theta) = \overline{Y}(1, \Theta) - \overline{Y}(0, \Theta)$
- **Network intervention effect**: effects on a unit's outcome of modifying the network structure, $NIE(w, \Theta, \Theta') = \overline{Y}(w, \Theta) - \overline{Y}(w, \Theta')$

Then the IPW estimators for this estimand for the true network parameter Θ^* would be

- Estimator for the hypothetical network

$$\hat{Y}^{HP}(w, \Theta^{HP}; \Theta^*) = \frac{1}{N}\sum_{i=1}^{N}\frac{\pi_{HP}(g; \Theta^{HP}|\mathbf{X})\mathbb{1}(W_i = w)Y_i^{obs}}{\mathbb{P}_{\Theta^*}(G_i = g(\mathbf{W}_{S_i}^{obs}, \mathbf{A}^{obs})|\mathbf{X})\mathbb{P}(W_i = w|\mathbf{X})}$$

Using the latter estimator we could study the DE and NIE for a hypothetical network

- Estimated direct effect

$$\widehat{DE}(\Theta) = \hat{Y}^{HP}(1, \Theta; \Theta^*) - \hat{Y}^{HP}(0, \Theta; \Theta^*)$$

- Estimated network intervention effect

$$\widehat{NIE}(w, \Theta, \Theta') = \hat{Y}^{HP}(w, \Theta; \Theta^*) - \hat{Y}^{HP}(w, \Theta'; \Theta^*)$$

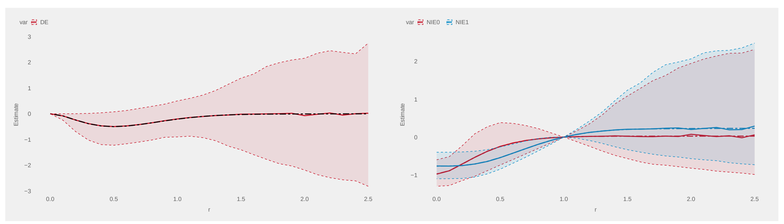
Simulation framework

To get an idea of the performance of our estimators we performed the next simulation for 10 units:

- The potential outcome function is $Y(w, g) = \mathbb{1}(\sigma(-7w + 4g + (1 - w)g - 5) > 0.5))$ where sigma is the logistic function where $Y(w, g)$ represents the status of getting infected with a specific disease and w represents the treatment status of the unit i, and g is the number of untreated units.
- $\mathbf{A}$ is generated from SBM with $\Theta^* = \begin{bmatrix}\theta_{11} & \theta_{12}\\ \theta_{12} & \theta_{22}\end{bmatrix} = \begin{bmatrix}0.6 & 0.4\\ 0.4 & 0.7\end{bmatrix}$
- W_i follows the Bernoulli distribution with rate $p^* = 0.5$.

Simulation result

In this section, we examine the changes in causal estimands of interest for a hypothetical network structure, denoted by $\theta_{ij}^{HP} = \min(r\theta_{ij}^*, 1) \quad \forall i, j \in 1, 2$. Here, r represents the percentage of network connectivity change, where $r = 0.9$ implies a 10% reduction in social connectivity pattern relative to Θ^*. Figure 2 illustrates the direct effect and network intervention effect for each value of r, along with the respective 5% and 95% quantiles on a $5,000$ Monte Carlo simulation. The two black dashed lines represent the true value of the estimand for the given r value. We selected a grid of r values such that Θ^{HP} starts with $\theta_{ij}^{HP} = 0 \quad \forall i, j \in 1, 2$ and ends with $\theta_{ij}^{HP} = 1 \quad \forall i, j \in 1, 2$.



(a) (b)

Figure 2. Direct effect and Network Intervention effect (a) Direct Effect, (b) Network Intervention Effect

We can see that the solid black line for DE and solid lines for the NIE overlaps with the corresponding Monte Carlo estimates, indicating that the estimator is unbiased, as expected. For the direct effect, we observe that as we increase the level of social connection, the interference sets also become larger, leading to greater uncertainty in the treatment effect.
On the right-hand side, we can see how the network intervention effect changes as we vary the hypothetical social connectivity pattern. As the hypothetical parameter moves farther from the true value, we observe an increase in uncertainty.

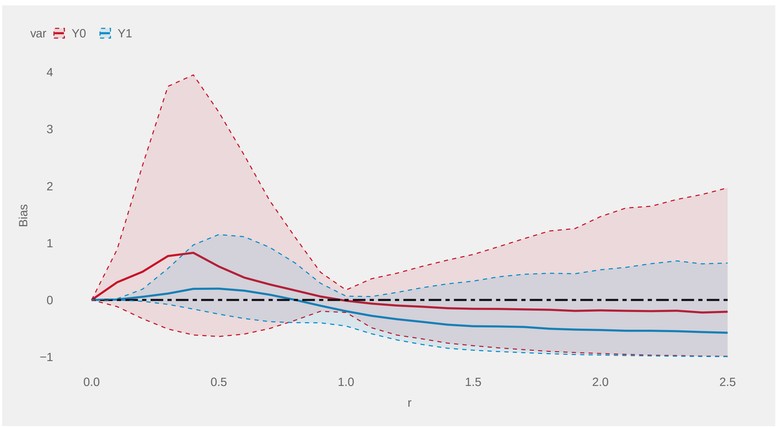
Network interference in observational studies

In observational studies, causal inference under interference is challenging as the traditional unconfoundedness assumption must be relaxed to account for treatment and covariates of neighboring units to ensure the validity of the inference and accurate identification of causal effects. [2] Furthermore, in addition to the propensity score estimation, we need to make an inference on the network formation model.

Namely, we need to estimate $\hat{\Theta}^*$ itself and calculate $\widehat{DE}(\hat{\Theta}^*)$ and $\widehat{NIE}(w, \Theta^{HP}, \hat{\Theta}^*)$. Then, we will examine the bias of our estimators by comparing $|\hat{Y}^{HP}(w, \Theta^{HP}; \hat{\Theta}^*) - \overline{Y}(w, \Theta^{HP})|$. To achieve this, we estimate Θ^* and p^* using Maximum Likelihood Estimation (MLE) and substitute them into our proposed estimator. This allows us to analyze the resulting bias.

Simulation Results for unknown Θ^* and p^*

Simulation study when we estimate both Θ^* and p^* using Maximum-likelihood-estimation (MLE). In figure 3, $|\hat{Y}^{HP}(w, \Theta^{HP}; \hat{\Theta}^*) - \overline{Y}(w, \Theta^{HP})|$ for each hypothetical hyperparameter $\theta_{ij}^{HP} = \min(r\theta_{ij}^*, 1) \quad \forall i, j \in 1, 2$



(a)

Figure 3. Bias when we estimate Θ^* and p^* using MLE

We can observe that the bias increases when the hypothetical parameter is further from the truth.

Discussion/Future work

In this work, we have introduced a framework aimed at estimating the impact of modifying the social connectivity pattern. This framework holds significant value for policymakers as it assists them in determining the extent to which they should restrict social connectivity to maintain a low infection rate.

In future research, we will evaluate the performance of our estimator by examining its asymptotic behavior and convergence rate. Moreover, we will focus on developing an estimator for the variance, considering that Monte Carlo variance is not feasible with real-world data.

References

[1] Peter M. Aronow and Cyrus Samii. "Estimating average causal effects under general interference, with application to a social network experiment". In: *The Annals of Applied Statistics* 11.4 (2017), pp. 1912–1947. ISSN: 19326157.

[2] Laura Forastiere, Edoardo M. Airoldi, and Fabrizia Mealli. "Identification and Estimation of Treatment and Interference Effects in Observational Studies on Networks". In: *Journal of the American Statistical Association* 116.534 (2021), pp. 901–918.

[3] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. "On causal inference in the presence of interference". en. In: *Stat Methods Med Res* 21.1 (Nov. 2010), pp. 55–75.

s.nakamura.sakai@yale.edu