Conformal Causal Inference for Network Intervention: Evidence from Field Experiments

Peilin Yang *

Preliminary Draft

1 Summary of Research

- Network interventions (Valente [2012]) are widely used in social sciences and public health, but it is difficult to estimate the "pure" treatment effect due to the biases introduced by peer effect and spillover effect.
- To improve prediction, we modified conformal inference to accommodate cluster randomized trials with social networks.
- Theoretically, we proved that our conformal algorithm achieves marginal and conditional coverage guarantees under the assumption of data exchangeability.
- In a numerical simulation, we compared our conformal algorithm with off-the- shelf OLS, BART and Causal Forest methods (Fig. 1). We demonstrated superior efficiency of our method, which has shorter intervals and higher coverage rates.
- In a field experiment, our conformal method overcame the misspecification problem due to peer influence in physician social networks (Fig. 2). Our method improved the prediction accuracy of physician burnout and medical errors, thus generalizing the treatment effect to other global health settings.
- We offer the first field-experimental application and validation of the conformal causal inference methodology (Lei and Candès [2021]).

2 Introduction

Estimation bias is widespread in field experiments and A/B testing in the fields of health policy and marketing due to complicated internal nexus. Cluster design of experiments

*yang.peilinc@gmail.com. Preliminary Draft. The data are supported by Lambert Zixin Li. Presented on 2024 Stanford Data Science Conference. The project was supported by the European Research Council (grant number: 882499) and Stanford GSB Business, Government, and Society Research Fund (PI: Sarah A. Soule). The RCT was pre-registered at the AEA RCT Registry: AEARCTR-0010137.

with the social networks has gained increased focus in recent years. Holtz et al. [2024] provided experimental evidence that cluster randomization can counter interference bias in social networks. However, they did not offer a solution that either solves or debiases the estimation. To analyze the estimator under social networks, a spectrum of papers based on modeling approaches exist, such as those by Leung [2020], Goldsmith-Pinkham and Imbens [2013], Bramoullé et al. [2009]. These methods typically assume a linear-in-means setting for the data-generating process. However, in practice, linear settings suffer from significant misspecification problems, such as non-linear effects or omitted variable bias. To improve the efficiency of estimation or reduce the bias in networks with spillover or peer effects, many papers, including those by Basse and Airoldi [2018], Viviano [2020], and Viviano et al. [2023], discuss how to design experiments. In this paper, we approach the problem from another perspective by using a model-free method—conformal inference—to enhance the generalizability and predictive power of the model. Tibshirani et al. [2019] and Lei and Candès [2021] have demonstrated superior predictive abilities. This paper applies conformal prediction to cluster design within social networks.

3 Setting

Assume there are J clusters $j = 1, \ldots, J$ and I_j individuals in each group $i = 1, \ldots, I_j$. The corresponding total population is $N_J = \sum_{j=1}^{J} I_j$.W_{ij} is local network statistics generated from adjacent matrix A of size $N_J \times N_J$; X_{ij} denotes the covariates. We follow the potential outcome framework Rubin [1974] and Neyman [1937]. Assume the population is generated i.i.d. from distribution P^{C_j} control by parameters C_j denoting cluster characteristics which includes cluster treatment T_j is sample from an independent Bernoulli distribution P^T[1] for all i, j.

$$Z_{ij} = (Y_{ij}(1), Y_{ij}(0), X_{ij}, W_{ij}; C_{ij}) \overset{i.i.d.}{\sim} P^{C_j}$$

The individual treatment effect is defined as $\tau_{ij} := Y_{ij}(1) - Y_{ij}(0)$. To allow conformal prediction feasible in this distribution, we need build the exchangeability of Z_{ij} as the premise of conformal prediction. Now, we will discuss some canonical structures to see if the exchangeability still holds in these cases.

Network Characteristics There are many possible structures for adjacent matrix A. One natural question is in what case the exchangeability still holds. Assume the connections are generated by Graphon model. Suppose $\{(X_{ij}, X_{i'j'}, \xi_{ij}, \xi_{i'j'}, C_{ij}, C_{i'j'})\}_{ij,i'j'}$ are generated in an i.i.d. way where $\xi_{ij} \sim U[0,1]$. The adjacent matrix is generated by $A_{ij,i'j'} = A_{i'j',ij} = \mathbf{1}\left(\eta_{ij} \leq \rho_{\mathrm{n}} d\left(\xi_{ij}, \xi_{i'j'}\right)\right)$ where η_{ij} is another independent uniform distribution. ρ_{n} controls the sparsity of the network.

Now let us incorporate two local network characteristics W_{ij} into the model.

1. Average covariate $\tilde{X}_{ij}^k$

[1] This is easy to realize in the cluster field experiment and debias the estimation see Holtz et al. [2024].

Let $\alpha^k_{iji'j'}$ be a binary random variable equal to 1 if the shortest path from node ij to $i'j'$ is of length k.

$$\tilde{D}^k_{ij} = \sum_{i'j' \neq ij} \alpha^k_{iji'j'}, \quad \tilde{X}^k_{ij} = \frac{1}{\tilde{D}^k_{ij}} \sum_{i'j' \neq ij} \alpha^k_{iji'j'} X_{i'j'}$$

2. Average outcome $\tilde{Y}_{ij}$

   Let Φ be a weight matrix contain weighting elements $\phi_{iji'j'}(k)$ depending only on the shortest path k.

$$\tilde{Y}_{ij} = \frac{1}{\sum_{i'j' \neq ij} \phi_{iji'j'}(k)} \sum_{i'j' \neq ij} \phi_{iji'j'}(k) Y_{i'j'}$$

We currently assume Y_{ij} is generated by the same function form f for any i and j.

According to the connection intensity of between inter and between groups, there are three cases of social network[2]:

Case 1 f is independent of cluster characteristics C_j except variable T_j. Apparently, Y_{ij} are exchangeable because these outcomes are total symmetric.

$$Y_{ij} = f\left(X_{ij}, \xi_{ij}, \tilde{Y}_{ij}, \tilde{D}^1_{ij}, \ldots, \tilde{D}^{N_J-1}_{ij}, \tilde{X}^1_{ij}, \ldots, \hat{X}^{N_J-1}_{ij}; T_{ij}\right)$$

Case 2 f depends on cluster characteristics C_{ij}, the social network is sparse. Individuals only have connections with peers in the same cluster. It implies the data are at least with "within" cluster exchangeability. [3]

$$Y_{ij} = f\left(X_{ij}, \xi_{ij}, \tilde{Y}_{ij}, \tilde{D}^1_{ij}, \ldots, \tilde{D}^{I_j-1}_{ij}, \tilde{X}^1_{ij}, \ldots, \hat{X}^{I_j-1}_{ij}; C_{ij}, T_{ij}\right)$$

Case 3 f depends on cluster characteristics and the individuals have connections across the cluster.

$$Y_{ij} = f\left(X_{ij}, \xi_{ij}, \tilde{Y}_{ij}, \tilde{D}^1_{ij}, \ldots, \tilde{D}^{N_J-1}_{ij}, \tilde{X}^1_{ij}, \ldots, \hat{X}^{N_J-1}_{ij}; C_{ij}, T_{ij}\right)$$

3.1 Exchangeability

In this subsection, we will build the exchangeability. Firstly, the adjacent matrix A is generate by paired random variables, so the premise is the exchangeability of pairs.

Assumption 1. Let $Z_{iji'j'} = (Z_{ij}, Z_{i'j'})$. The $Z_{iji'j'}$ is jointly exchangeable for any permutation σ on index set of individuals from each cluster $\{11, \ldots, I_1 1, 12, \ldots I_2 2, \ldots, 1J, \ldots, I_J J\}$. i.e.

$$Z_{\sigma(ij)\sigma(i'j')} \overset{d}{=} Z_{iji'j'}$$

And the adjacent matrix is generated by $Z_{iji'j'}$: $A = A(Z_{iji'j'})$

[2]In these cases, I slightly abuse the notation T_ij and C_ij, actually T_ij is one element of C_ij. For the ease of understanding, we disentangle it from C_ij

[3]Our experiment data are closed to this case.

Assumption 2. For any permutation σ, define the permutation of adjacent matrix $A^{(\sigma)}$ as $A_{\sigma(ij)\sigma(i'j')}$. And $\sigma(ij)$ denotes the position of permutation mapping. The network generating variables W_{ij} is defined by function ξ

$$W_{\sigma(ij)} = \xi\left(A^{(\sigma)}, \left(X_{\sigma(ij)}\right)_{ij}, \left(C_{\sigma(ij)}\right)_{ij}\right)$$

Lemma 1. *(Lunde et al. [2023]) Let X be a random variable taking values in $\mathcal{X}$. $Y = H(X)$ for some mappings H. Further suppose that for some collection of functions $\mathcal{F}$ such that*

$$f(X) \stackrel{d}{=} X \quad \forall f \in \mathcal{F}$$

Furthermore, let $\mathcal{G}$ be a collection of functions on $Y \in \mathcal{Y}$ and suppose that for any $g \in \mathcal{G}$, there exists a corresponding $f \in \mathcal{F}$ such that,

$$g(H(X)) = H(f(X))$$

Then $g(Y) \stackrel{d}{=} Y \ \forall g \in \mathcal{G}$.

Proposition 2. $\left(Y_{ij}, X_{ij}, W_{ij}, C_{ij}\right)_{ij}$ *are exchangeable.*

Proof. By Assumption 1, for permutation mapping $f \in \mathcal{F}$, $f\left(Z_{iji'j'}\right) \stackrel{d}{=} Z_{iji'j'}$. By Assumption 2, for any permutation $g \in \mathcal{G}$, define mapping H:

$$H\left(\left\{Z_{\sigma(ij)\sigma(i'j')}\right\}\right) = \left(\left\{Y_{\sigma(ij)}\right\}_{ij}, \left\{X_{\sigma(ij)}\right\}_{ij}, \xi\left(A^{(\sigma)}, \left(X_{\sigma(ij)}\right)_{ij}, \left(C_{\sigma(ij)}\right)_{ij}\right), \left\{C_{\sigma(ij)}\right\}_{ij}\right)$$

we can always find the corresponding g such that

$$g\left(H\left(\left\{Z_{iji'j'}\right\}\right)\right) = g\left(\left\{Y_{ij}\right\}_{ij}, \left\{X_{ij}\right\}_{ij}, \xi\left(A, \left(X_{ij}\right)_{i,j}, \left(C_{ij}\right)_{ij}\right), \left\{C_{ij}\right\}_{j}\right) = H\left(\left\{Z_{\sigma(ij)\sigma(i'j')}\right\}\right)$$

The exchangeability follows from Lemma 1. □

Assume Y_{ij} is generated by the same function f as case 1 to 3 ;

$$\begin{cases} Y_{11} = f\left(X_{11}, \xi_{11}, \tilde{Y}_{11}, \tilde{D}^{1}_{11}, \ldots, \tilde{D}^{N_J-1}_{11}, \tilde{X}^{1}_{11}, \ldots, \hat{X}^{N_J-1}_{11}; C_{11}, T_{11}\right) \\ \ldots \\ Y_{ij} = f\left(X_{ij}, \xi_{ij}, \tilde{Y}_{ij}, \tilde{D}^{1}_{ij}, \ldots, \tilde{D}^{N_J-1}_{ij}, \tilde{X}^{1}_{ij}, \ldots, \hat{X}^{N_J-1}_{ij}; C_{ij}, T_{ij}\right) \\ \ldots \\ Y_{I_JJ} = f\left(X_{I_JJ}, \xi_{I_JJ}, \tilde{Y}_{I_JJ}, \tilde{D}^{1}_{I_JJ}, \ldots, \tilde{D}^{N_J-1}_{I_JJ}, \tilde{X}^{1}_{I_JJ}, \ldots, \hat{X}^{N_J-1}_{I_JJ}; C_{I_JJ}, T_{I_JJ}\right) \end{cases} \tag{1}$$

In this paper, we assume there exist one single solution because it cannot be identified if there are several solutions.

Assumption 3. (Identification Condition) Equations (1) to solve Y_{ij} have one unique solution.

Proposition 3. *The data generating process mentioned in equations (1) satisfy exchangeability condition on Assumption 1-3 .*

Proof. Define variable

$$U_{ij} = \left(X_{ij}, \xi_{ij}, \widetilde{D}_{ij}^{1}, \ldots, \widetilde{D}_{ij}^{N_J-1}, \tilde{X}_{ij}^{1}, \ldots, \tilde{X}_{ij}^{N_J-1}, \{\eta_{iji'j'}\}_{i'j'}, C_{ij}\right)$$

where ξ_{ij} is the "position" of unit i from cluster j, and $\{\eta_{iji'j'}\}_{i'j'}$ is the random variable to all of other nodes except itself. The shortest paths are invariant to permutation of adjacency matrix. $(X_{ij}, X_{i'j'}, \xi_{ij}, \xi_{i'j'}, \eta_{iji'j'})_{ij,i'j'}$ are jointly exchangeable. By Lemma 1, the $(U_1, U_2, \ldots, U_{N_J})$ are exchangeable. Now, construct the equations

$$y_{ij} = f\left(\frac{1}{\sum_{i'j' \neq ij} \phi_{iji'j'}(k)} \sum_{j \neq i} \phi_{iji'j'}(k) y_{i'j'}; U_{ij}\right) \tag{2}$$

Given set $\{U_{ij}\}$ the equations deliver one unique solution of $\{y_{ij}\}$ by assumption. Define a new mapping H from equation (2)

$$(y_{ij}, U_{ij})_{ij} = H\left(\{U_{ij}\}_{ij}\right)$$

by Lemma 1, the exchangeability $(y_{ij}, U_{ij})_{ij}$ holds. □

4 Conformal Causal Inference for Network Intervention

The exchangeability of data $(Y_{ij}, X_{ij}, W_{ij}, C_{ij})$ provide the foundation for the conformal inference. Let $Z_{ij} = (X_{ij}, W_{ij}, C_{ij})$, and $Q(\alpha; V_{1:n+1})$ denote the α-level quantile of the empirical distribution of V such that $\mathrm{P}\{V_{n+1} \leq Q(\alpha; V_{1:n+1})\} \geq \alpha$. And $V_{ij} = V(Z_{ij}, y; \mathcal{Z})$ is the conformity score where $\mathcal{Z}$ is the dataset.

Given the exchangeability of $(Y_{ij}, X_{ij}, W_{ij}, C_{ij})$, we have the following theorem.

Theorem 4. *(Vovk et al. [2005]; Lei et al. [2018]). Define the conformal band generated by the first N_J data points. The conformal band at $Z_{1,J+1}$,*

$$\hat{C}(Z_{1,J+1}) = \{y : V(Z_{1,J+1}, y) \leq Q(\alpha; V_{1:N_J} \cup V(Z_{1,J+1}, y))\}$$

Then the $\hat{C}(Z_{1,J+1})$ satisfies $\mathbb{P}\left\{Y_{1,J+1} \in \hat{C}(Z_{1,J+1})\right\} \geq 1 - \alpha$.

In this case, we can get a similar algorithm as canonical conformal inference.

Algorithm 1 Split Conformal Algorithm for Cluster Experiment

Input: Set level α, data $(Y_{ij}, X_{ij}, W_{ij}, C_{ij}) \in \mathcal{Z}$, testing point (x, w, c), $T_{ij} \in C_{ij}$ denotes the treatment status of unit i of cluster j.

Procedures

Step 1. Split $\mathcal{Z}$ into a training fold $\mathcal{Z}_{tr} := (Y_{ij}, X_{ij}, W_{ij}, C_{ij})_{ij \in \mathcal{I}_{tr}}$ and a calibration fold $\mathcal{Z}_{ca} := (Y_{ij}, X_{ij}, W_{ij}, C_{ij})_{ij \in \mathcal{I}_{ca}}$

Step 2. For each $ij \in \mathcal{I}_{ca}$, compute the score $V_{ij} = |Y_{ij} - \hat{f}(X_{ij}, W_{ij}, C_{ij})|$

Step 3. Construct the counterfactual interval

For $T_{ij} = 1$,

1. Compute η as the $(1-\alpha)$-th quantile of the distribution $\sum_{ij \in \mathcal{I}_{ca}} \frac{1}{|\mathcal{I}_{ca}|} \delta_{V_{ij}} + \frac{1}{|\mathcal{I}_{ca}|} \delta_{\infty}$ for all $T_{ij} = 1$
2. Construct counterfactual interval $\hat{C}_{ij} = [\hat{Y}_{ij}(1) - Y_{ij}(0) - \eta, \hat{Y}_{ij}(1) - Y_{ij}(0) + \eta]$

For $T_{ij} = 0$,

1. Compute η as the $(1-\alpha)$-th quantile of the distribution $\sum_{ij \in \mathcal{I}_{ca}} \frac{1}{|\mathcal{I}_{ca}|} \delta_{V_{ij}} + \frac{1}{|\mathcal{I}_{ca}|} \delta_{\infty}$ for all $T_{ij} = 0$
2. Construct counterfactual interval $\hat{C}_{ij} = [Y_{ij}(1) - \hat{Y}_{ij}(0) - \eta, Y_{ij}(1) - \hat{Y}_{ij}(0) + \eta]$

Step 5. Given $\hat{C}_{ij} = [\hat{C}_{ij}^L, \hat{C}_{ij}^U]$, use conformal inference on set of lower bound $\hat{C}_{ij}^L$ and upper bound $\hat{C}_{ij}^U$ to deliver conformal inference interval $\hat{\mathcal{C}}_{ITE}$.

Output: $\hat{\mathcal{C}}_{ITE}$

4.1 Local Conformal Inference

To further improve the prediction efficiency of conformal inference, we exploit the strength of local conformal inference.

Algorithm 2 Split Local Conformal Algorithm for Cluster Experiment

Input: Set level α, data $(Y_{ij}, X_{ij}, W_{ij}, C_{ij}) \in \mathcal{Z}$, testing point (x, w, c), $T_{ij} \in C_{ij}$ denotes the treatment status of unit i of cluster j.

Procedures

Step 1. Split $\mathcal{Z}$ into a training fold $\mathcal{Z}_{tr} := (Y_{ij}, X_{ij}, W_{ij}, C_{ij})_{ij \in \mathcal{I}_{tr}}$ and a calibration fold $\mathcal{Z}_{ca} := (Y_{ij}, X_{ij}, W_{ij}, C_{ij})_{ij \in \mathcal{I}_{ca}}$

Step 2. For each $ij \in \mathcal{I}_{ca}$, compute the score $V_{ij} = |Y_{ij} - \hat{f}(X_{ij}, W_{ij}, C_{ij})|$

Step 3. Construct the counterfactual interval for testing point (x, w, c), local weighting function $H(c, C_{ij})$ for $ij \in \mathcal{I}_{ca}$, and the sum of weighting function $S_{i,J+1} = \sum_{j=1}^{J} \sum_{i=1}^{I_j} H(c, C_{i,j}) + 1$.

For $T_{ij} = 1$,

1. Compute η as the $(1-\alpha)$-th quantile of the distribution $\sum_{ij \in \mathcal{I}_{ca}} \frac{H(c,C_{ij})}{S_{i,J+1}} \delta_{V_{ij}} + \frac{1}{S_{i,J+1}} \delta_{V_{ij}} \delta_{\infty}$ for all $T_{ij} = 1$

2. Construct counterfactual interval $\hat{C}_{ij} = [\hat{Y}_{ij}(1) - Y_{ij}(0) - \eta, \hat{Y}_{ij}(1) - Y_{ij}(0) + \eta]$

For $T_{ij} = 0$,

1. Compute η as the $(1-\alpha)$-th quantile of the distribution $\sum_{ij \in \mathcal{I}_{ca}} \frac{H(c,C_{ij})}{S_{i,J+1}} \delta_{V_{ij}} + \frac{1}{S_{i,J+1}} \delta_{V_{ij}} \delta_{\infty}$ for all $T_{ij} = 0$

2. Construct counterfactual interval $\hat{C}_{ij} = [Y_{ij}(1) - \hat{Y}_{ij}(0) - \eta, Y_{ij}(1) - \hat{Y}_{ij}(0) + \eta]$

Step 5. Given $\hat{C}_{ij} = [\hat{C}_{ij}^L, \hat{C}_{ij}^U]$, use conformal inference on set of lower bound $\hat{C}_{ij}^L$ and upper bound $\hat{C}_{ij}^U$ to deliver conformal inference interval $\hat{\mathcal{C}}_{ITE}$.

Output: $\hat{\mathcal{C}}_{ITE}$

Assumption 4. Constant L is given. Assume in the neighborhood region of c_0,

$$P\left(C_{ij} \in \{c : d\left(c_0, c\right) \leq \epsilon\}\right) \geq \frac{\epsilon^{\beta}}{L} \quad \forall \epsilon \leq k$$

The constant k is selected such that $k \ln k \to 0$ when $N_J \to \infty$; and conditional conformity score is Lipschitz continuous with respect to cluster characteristics $\left|P_{V|C}(v) - P_{V|C'}(v)\right| \leq Ld\left(C, C'\right)$.

Under some extra assumptions, we can prove the asymptotic efficiency of the algorithm.

Theorem 5. *Under Assumption 4, the asymptomatic feature of the algorithm B*

$$\lim_{N_J \to \infty} P\left(Y_{i.J+1} \in \hat{C}_L\left(X_{i,J+1}, W_{i,J+1}, C_{i,J+1}\right) \mid C_{i,J+1}\right) = 1 - \alpha$$

Proof. Define the sum of weight $S_{i,J+1} = \sum_{j=1}^{J} \sum_{i=1}^{I_j} H\left(C_{i,j}, C_{i,J+1}\right) + 1$ then the empirical distribution of conformity score is

$$\mathcal{F}_{i,J+1} = \left(\sum_{j=1}^{J} \sum_{i=1}^{I_j} \frac{H\left(C_{i,j}, C_{i,J+1}\right)}{S_{i,J+1}} \delta_{V_{i,j}} + \frac{1}{S_{i,J+1}} \delta_{\infty}\right)$$

Given level $\alpha, V_{i,J+1} \leq Q\left(\alpha; \mathcal{F}_{i,J+1}\right)$, the decomposition of the probability $\mathcal{F}_{i,J+1}\left(V_{i,J+1}\right)$ for point $V_{i,j}$

$$\begin{aligned}\sum_{j=1}^{J}\sum_{i=1}^{I_j}\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}1\left(V_{i,j}<V_{i,J+1}\right) &= \sum_{j=1}^{J}\sum_{i=1}^{I_j}\frac{H\left(C_{i,j},C_{i,J+1}\right)P_{V_{i,j}|C_{i,j}}\left(V_{i,J+1}\right)}{S_{i,J+1}} \\ &+ \sum_{j=1}^{J}\sum_{i=1}^{I_j}\frac{H\left(C_{i,j},C_{i,J+1}\right)\left(1\left(V_{i,j}<V_{i,J+1}\right)-P_{V_{i,j}|C_{i,j}}\left(V_{i,J+1}\right)\right)}{S_{i,J+1}} < \alpha\end{aligned}$$

when $N_J \to \infty$, these two terms (i) For the first term $R_1 = \sum_{j=1}^{J}\sum_{i=1}^{I_j}\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}P_{V_{i,j}|C_{i,j}}\left(V_{i,J+1}\right)$, $A_n = [(n-1)k, k]$

$$S_{i,J+1} = \sum_{j=1}^{J}\sum_{i=1}^{I_j}\exp\left(-\frac{d\left(C_{i,j},C_{i,J+1}\right)}{k}\right) \geq \frac{1}{e}\sum_{j=1}^{J}\sum_{i=1}^{I_j}1\left(d\left(C_{i,j},C_{i,J+1}\right)\in A_1\right)$$

By Chernoff bound and assumption 4 , let $X = \frac{1}{e}\sum_{j=1}^{J}\sum_{i=1}^{I_j}1\left(d\left(C_{i,j},C_{i,J+1}\right)\in A_1\right)$ then $\mathbb{E}X \geq N^J\frac{k^\beta}{eL}$

$$P\left(X \leq \frac{1}{2}\frac{N^Jk^\beta}{eL}\right) \leq P\left(X \leq \frac{1}{2}\mathbb{E}X\right) \leq \exp\left(-\frac{\mathbb{E}X}{8}\right) \leq \exp\left(-\frac{N^Jk^\beta}{8eL}\right)$$
$$P\left(S_{i,J+1} \leq \frac{1}{2}\frac{N^Jk^\beta}{eL}\right) \leq \exp\left(-\frac{N^Jk^\beta}{8eL}\right)$$

Now define the "distance" $\Delta\left(C_{i,j},C_{i,J+1}\right)$

$$\begin{aligned}\Delta\left(C_{i,j},C_{i,J+1}\right) &= \sum_{j=1}^{J}\sum_{i=1}^{I_j}H\left(C_{i,j},C_{i,J+1}\right)\max_v\left|P_{V_{i,J+1}|C_{i,J+1}}(v)-P_{V_{i,j}|C_{i,j}}(v)\right| \\ &\leq \sum_{j=1}^{J}\sum_{i=1}^{I_j}\exp\left(-\frac{d\left(C_{i,j},C_{i,J+1}\right)}{k}\right)Ld\left(C_{i,j},C_{i,J+1}\right) \\ &\leq \min_{n_0}\left\{\sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{n_0}\exp\left(-\frac{d\left(C_{i,j},C_{i,J+1}\right)}{k}\right)eLn_0k + \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=n_0}^{\infty}\exp(-(n-1))eLnk\right\} \\ &\leq eLn_0kS_{i,J+1} + N_JeLn_0k\exp\left(-n_0\right)\end{aligned}$$

when $n_0 > \left\lceil\ln\left(\frac{1}{k}\right)\right\rceil$, there exist constant C_G

$$\frac{\Delta\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}} \leq -C_Gk\ln k$$

Then R_1 is bounded by

$$\left|R_1 - \sum_{j=1}^{J}\sum_{i=1}^{I_j}\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}P_{V_{i,J+1}|C_{i,J+1}}\left(V_{i,J+1}\right)\right| \leq \frac{\Delta\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}} \leq -C_Gk\ln k$$

(ii) For the second term $R_2 = \sum_{j=1}^{J}\sum_{i=1}^{I_j} R_{2ij} = \sum_{j=1}^{J}\sum_{i=1}^{I_j} \frac{H\left(C_{i,j},C_{i,J+1}\right)\left(1\left(V_{i,j}<V_{i,J+1}\right)-P_{V_{i,j}\mid C_{i,j}}\left(V_{i,J+1}\right)\right)}{S_{i,J+1}}$
Condition on dataset $\mathcal{D}$, R_2 is a centered random variable, $R_{2ij} \in \left[-\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}, \frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}\right]$.
By Hoeffding's Lemma,

$$\mathbb{E}_{V_{i,j}}\left[\exp\left(s\frac{H\left(C_{i,j},C_{i,J+1}\right)\left(1\left(V_{i,j}<V_{i,J+1}\right)-P_{V_{i,j}\mid C_{i,j}}\left(V_{i,J+1}\right)\right)}{S_{i,J+1}}\right)\middle|\,\mathcal{D}\right] \le \exp\left(\frac{1}{2}s^2\left(\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}\right)^2\right)$$

$$\begin{aligned}\mathbb{E}\left[\exp\left(s\sum_{j=1}^{J}\sum_{i=1}^{I_j}R_{2ij}\right)\mid\mathcal{D}\right] &= \prod_{j=1}^{J}\prod_{i=1}^{I_j}\exp\left(sR_{2ij}\right)\\ &\le \exp\left(\frac{s^2}{2}\sum_{j=1}^{J}\sum_{i=1}^{I_j}\left(\frac{H\left(C_{i,j},C_{i,J+1}\right)}{S_{i,J+1}}\right)^2\right) \le \exp\left(\frac{s^2}{2}\frac{S_{i,J+1}}{S_{i,J+1}^2}\right)\end{aligned}$$

So $R_2 \sim \operatorname{sub}G\left(\frac{1}{S_{i,J+1}}\right)$, then $P\left(R_2>t\right)\le\exp\left(-\frac{t^2}{2}S_{i,J+1}\right)$. Let $t=\left(\frac{\ln n}{S_{i,J+1}}\right)^{\frac{1}{2}}$, $P\left(R_2>\left(\frac{\ln n}{S_{i,J+1}}\right)^{\frac{1}{2}}\right)\le \exp\left(-\frac{1}{2}\ln n\right)$
Now set $G=\left\{R_2\le\left(\frac{\ln n}{S_{i,J+1}}\right)^{\frac{1}{2}}, S_{i,J+1}>\frac{1}{2}\frac{N^J k^\beta}{eL}\right\}$, then

$$R_1\in\left[P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)-C_G k\ln k, P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)+C_G k\ln k\right]$$

The upper bound

$$\begin{aligned}P\left(R_1+R_2<\alpha\mid C_{i,J+1}\right) &\le P\left(P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)+R_2<\alpha+C_G k\ln k\mid C_{i,J+1}\right)\\ &\le P\left(P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)<\alpha+C_G k\ln k\mid C_{i,J+1}\right)\end{aligned}$$

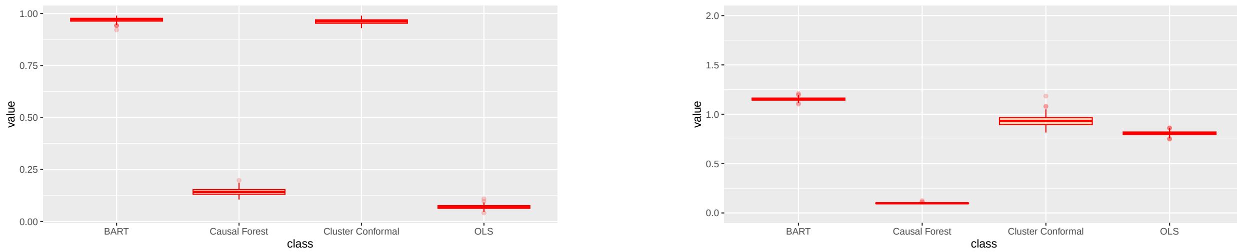
If $V\mid C$ is continuous, $P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)\sim U[0,1]$, $P\left(R_1+R_2<\alpha\mid C_{i,J+1}\right)\le\alpha+C_G k\ln k\to\alpha$
Similarly, the lower bound

$$\begin{aligned}&P\left(R_1+R_2<\alpha\mid C_{i,J+1}\right)\ge P\left(P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)+R_2<\alpha-C_G k\ln k\mid C_{i,J+1}\right)\\ &\ge P\left(P_{V_{i,J+1}\mid C_{i,J+1}}\left(V_{i,J+1}\right)<\alpha-C_G k\ln k-\left(\frac{\ln n}{S_{i,J+1}}\right)^{\frac{1}{2}}\middle|\,C_{i,J+1}\right)P\left(R_1+R_2<\alpha\mid C_{i,J+1}\right)\\ &\ge\alpha-C_G k\ln k-\left(\frac{\ln n}{S_{i,J+1}}\right)^{\frac{1}{2}}\to\alpha\end{aligned}$$

□

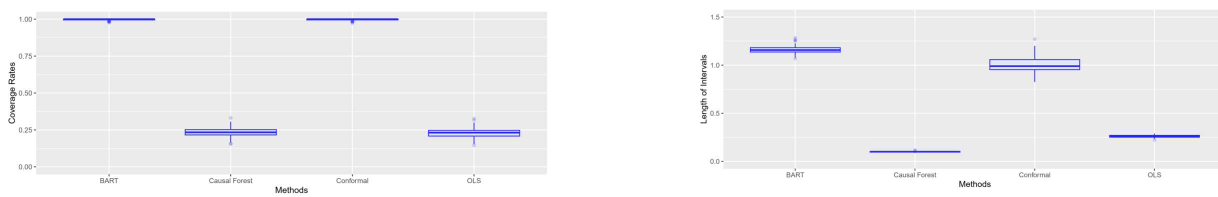
5 Application

We conducted a field experiment ($N = 880$ physicians) to validate our method. In the experiment, we use a psychological interventions aimed at improving communication between



(a) Coverage Rate (b) Length of Interval

Figure 1: Comparison Results of Simulation Experiments



(a) Coverage Rate (b) Length of Interval

Figure 2: Comparison Results of Experiment

nurses and doctors, thereby strengthening social networks, and consequently reducing medical errors and burnout." Our method overcame the misspecification problem due to network effects. Therefore, it generalized the treatment effect on burnout and medical errors to other global health settings.

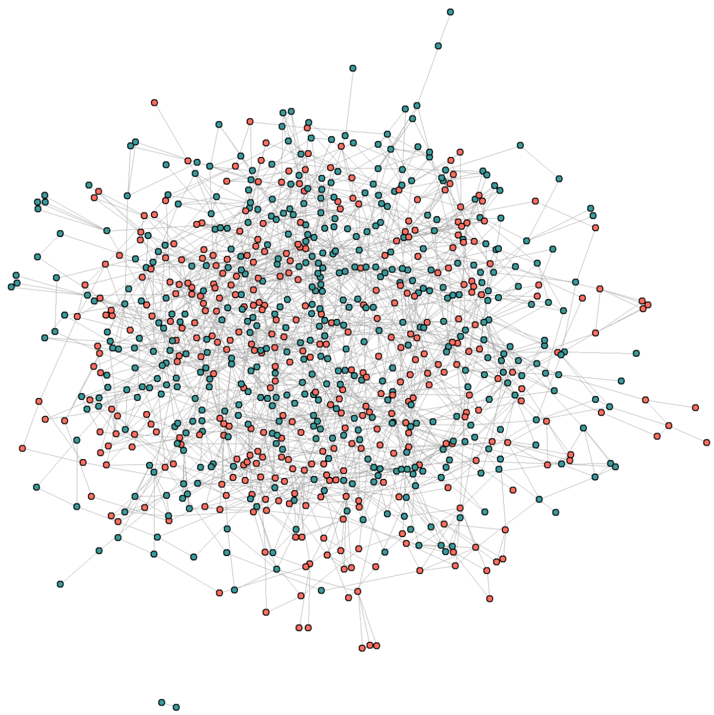


Figure 3: A Social Network from a Healthcare Field Experiment

References

Guillaume W Basse and Edoardo M Airoldi. Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4):849–858, 2018.

Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55, 2009.

Paul Goldsmith-Pinkham and Guido W Imbens. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.

David Holtz, Felipe Lobel, Ruben Lobel, Inessa Liskovich, and Sinan Aral. Reducing interference bias in online marketplace experiments using cluster randomization: Evidence from a pricing meta-experiment on airbnb. *Management Science*, 2024.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Lihua Lei and Emmanuel J Candès. Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5):911–938, 2021.

Michael P Leung. Treatment and spillover effects under network interference. *Review of Economics and Statistics*, 102(2):368–380, 2020.

Robert Lunde, Elizaveta Levina, and Ji Zhu. Conformal prediction for network-assisted regression. *arXiv preprint arXiv:2302.10095*, 2023.

Jerzy Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 236(767):333–380, 1937.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

Thomas W Valente. Network interventions. *science*, 337(6090):49–53, 2012.

Davide Viviano. Experimental design under network interference. *arXiv preprint arXiv:2003.08421*, 2020.

Davide Viviano, Lihua Lei, Guido Imbens, Brian Karrer, Okke Schrijvers, and Liang Shi. Causal clustering: design of cluster experiments under network interference. *arXiv preprint arXiv:2310.14983*, 2023.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.