Contamination-Aware Experimentation on Networks

Mine Su Erturk, Eray Turkel

September 21, 2020

1 Introduction

We study the problem of a decision maker (DM) conducting experiments on a network environment. An 'experiment' in the context of our model is treating a node in the network. Each node responds to the treatment differently, depending on their (observable) characteristics.

The DM's goal is learning the optimal treatment allocation over the network by sequentially conducting local experiments on different nodes. Because of the network setting, treating one node creates spillovers on neighboring nodes.

The distinguishing feature of our model is that every experiment creates negative externalities on previous experiments, which we model as a cost for contaminating the treatment regime decided by a previously run experiment. Moreover, we analyze the DM's problem as a sequential decision making problem inspired by the literature on online learning, instead of using a randomization inference framework, as is common in the network experimentation literature.

1.1 Model

There are N units connected on an undirected, unweighted graph. Denote each unit on the graph by $i \in \{1, \dots, N\}$. The units have observable characteristics, $\{x_1, \dots, x_N\}$ with $x_i \in \mathbb{R}^k$. The DM's objective depends on the outcomes of the units, denoted $\{Y_1, \dots, Y_N\}$ with $Y \in \mathbb{R}$, which respond to a treatment regime that will be determined. The decision maker can run experiments sequentially, at time periods $t \in \{1, \dots, T\}$. Let w_{it} denote the treatment assignment for node i during the period of experimentation t, with $w_{it} \in \{0, 1\}$, and denote the vector of treatment assignments at time t by $w_t = [w_{1t}, \dots, w_{Nt}] \in \{0, 1\}^N$. We assume that the experimentation is local, in the sense that every period, only one node can be treated. Thus for all time periods t, $\sum_{i \leqslant N} w_{it} = 1$.

Each node's outcome depends on their characteristics, the treatment they receive and the exposure to the treatment through their neighbors. Define the exposure for node i at period t as e_{it}, which is (for simplicity) assumed to be:

$$e_{it} = \frac{\sum_{j \leqslant N} \mathbf{1}(j \in nhbd(i)) w_{jt}}{\sum_{j \leqslant N} \mathbf{1}(j \in nhbd(i))}$$

Y is assumed to be a linear function of the unit characteristics and the exposure to the treatment:

$$Y_{it}(w_{it}, e_{it}) = \beta^T x_i + \Gamma^T x_i e_{it} + \epsilon_i,$$

where $\beta, \Gamma \in \mathbb{R}^k$, and for all nodes i, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$.

The DM's goal is to choose a treatment assignment policy over the network based on the value of Γ. Denote the final treatment assignment for node i by w_i, dropping the time subscript. We denote a policy by $\pi(\Gamma)$, where $\pi(\Gamma) = [w_1, \dots, w_N] \in \{0, 1\}^N$. We assume that there is an upper limit on the number of nodes that can be treated in the final deployment, which we call D, i.e., $\sum_{i \leqslant N} w_i \leqslant D$.

At every period t, the DM's experimentation creates a cost, because the experiment interacts with a previous experimenter's treatment allocation, contaminating their results. We assume that the cost is proportional to the number of units exposed to the treatment, and once a unit is exposed, future exposures of that unit do not increase the cost any further. Define the total cost of experimentation at period t by:

$$c_t = c \sum_{i \leqslant N} \mathbf{1}\left(e_{it} > 0 \text{ and } e_{is} = 0, \forall s < t\right).$$

Throughout, we will let $c = 1$ for simplicity. Hence, the total cost incurred through experimentation by the DM can be written as $C = \sum_{t \leqslant T} c_t$. We assume that the DM has to stay below a 'contamination budget' of at most $\bar{C}$ during experimentation. Then, as a function of the cost budget $\bar{C}$, we define the value of a policy as follows:

$$V(\pi, \bar{C}) = \sum_{i \leqslant N} \Gamma^T x_i e_i(\pi),$$

where $e_i(\pi)$ denotes the exposure function determined at the last period for final deployment, induced by the treatment allocation under policy π, learned through experimentation with the cost budget $\bar{C}$. The DM does not observe the vector Γ but has an imperfect estimate of it learned through sequential experimentation, which we will call $\hat{\Gamma}$. The DM maximizes the empirical value function using its estimate:

$$Max_{\hat{\pi}} : \sum_{i \leqslant N} \hat{\Gamma}^T x_i e_i(\hat{\pi}) \quad \text{s.t. } 1^T \hat{\pi} \leqslant D. \tag{1}$$

Let $X_{K \times N}$ denote the matrix of observed covariates for all the nodes on the graph, and let X_s denote the analogous matrix only for nodes that were exposed to the treatment during the experimentation period. Let $\hat{\Gamma}_{1 \times K}$ be the vector of estimated coefficients.

Define the vector of exposure values at time t as $\bar{e}_t = [\mathbf{1}(e_{1t} > 0), \ldots, \mathbf{1}(e_{Nt} > 0)]_{1 \times N}$ and let c_t denote the cost vector at time t. We can recursively write: $c_0 = 0_{1 \times N}, c_1 = \bar{e}_1$, and generally, $c_n = (c_{n-1} + \bar{e}_n) - (Diag(c_{n-1})\bar{e}_n)$. $Diag(c_n)$ denotes the diagonal matrix with the entries of c_n on its diagonal entries. This recursive formulation captures the notion that the cost of second and future exposures to the experiment are zero.

We can also define the final exposure vector e in terms of the adjacency matrix of the network, $A_{N \times N}$ (with self-edges), and the final treatment allocation vector $\pi \in \{0,1\}^N$. Representing e as a vector, we have: $Diag(A\mathbf{1}_N)^{-1} A\pi = e,$, where $\mathbf{1}_N$ is an $(N \times 1)$ vector of 1's. Therefore, we can rewrite the maximization problem (1) as a convenient linear program:

$$Max_{\hat{\pi}} : \left(\hat{\Gamma}^T X Diag(A\mathbf{1}_N)^{-1} A\right) \hat{\pi}, \text{ subject to: } 1^T \hat{\pi} \leqslant D, \forall i, \hat{\pi}_i \in \{0,1\}$$

1.2 Regret Analysis

Let $z(\hat{\Gamma})$ denote the optimal objective function value of the optimization problem (1) given that the DM has estimated the coefficient vector as $\hat{\Gamma}$. Note that an oracle with full knowledge of the coefficient vector Γ would be able to achieve $z(\Gamma)$. Thus, we can define the regret as follows:

$$R = z(\Gamma) - z(\hat{\Gamma}).$$

To analyze the regret of the empirical value maximizing policy $\hat{\pi}$, let us introduce some notation. Let $\sigma_j(A)$ denote the j^{th} eigenvalue of the matrix A, with σ_1 being the largest. Suppose that the covariate matrices X, X_s are standardized, therefore the matrix XX^T is the symmetric correlation matrix.

Theorem 1 *The regret is bounded above by K with the following probability:*

$$\mathbb{P}(R \leqslant K) \geqslant 1 - \exp\left(-\frac{1}{2}(\tau - 1)^2\right) \tag{2}$$

where

$$K = \delta \frac{\max\{\sqrt{n + D^2}, ||\Gamma|| \sigma_1(XX^T)\}}{\sqrt{n+1}} \frac{\sqrt{n + D^2}}{\sqrt{n+1} - \delta} \tag{3}$$

and $\delta = \tau \sigma_\epsilon \sum_{j=1}^{k} \sigma_j((X_s X_s^T)^{-1})$

The proof is derived from the perturbation analysis of the LP in (1). Assuming $n >> ||\Gamma||$, the term K in theorem 1 becomes $\delta \frac{n+D^2}{n+1-\delta\sqrt{n+1}}$, which converges to 1 as $n \to \infty$.

1.3 Simulations and Proposed Algorithm

We propose an algorithm based on information-directed sampling and the literature on linear bandits, which adaptively chooses experimental subjects that will maximize information gain, or reduce the variance of the coefficient estimates. Since our setting has an additional component which is the contamination cost due to externalities on other experiments, we use a knapsack-based heuristic $c - \mu$ rule, which chooses the most informative nodes to experiment on, taking into account the cost of contaminating other nodes which is determined by the network topology.

References

[1] James Renegar. Some perturbation theory for linear programming. Technical report, Cornell University Operations Research and Industrial Engineering, 1993.