Published in final edited form as:

Stat (Int Stat Inst). 2019 ; 8(1): . doi:10.1002/sta4.219.

Modelling and estimation for optimal treatment decision with interference

Lin Su, Wenbin Lu, and Rui Song

Department of Statistics, North Carolina State University, Raleigh, North Carolina

Abstract

In many network-based intervention studies, treatment applied on an individual or his or her own characteristics may also affect the outcome of other connected people. We call this interference along network. Approaches for deriving the optimal individualized treatment regimen remain unknown after introducing the effect of interference. In this paper, we propose a novel network-based regression model that is able to account for interaction between outcomes and treatments in a network. Both Q-learning and A-learning methods are derived. We show that the optimal treatment regimen under our model is independent from interference, which makes its application in practice more feasible and appealing. The asymptotic properties of the proposed estimators are established. The performance of the proposed model and methods is illustrated by extensive simulation studies and an application to a mobile game network data.

Keywords

A-learning; interference; network; optimal treatment regimen; Q-learning

1 | INTRODUCTION

Interference, which refers to treatment of one individual that has influence on connected nodes, is observed in many network-based intervention studies. For example, the coverage of vaccination in a neighbourhood can affect the infection rate for a nonvaccinated individual (Halloran & Struchiner, 1995). Getting a private tutor may affect the grades of other students in the same study group. Encouraging users to vote on social network can improve the voting rate among his or her friends (Bond et al., 2012). More examples can be found in Sobel (2006), Hong and Raudenbush (2006), Rosenbaum (2007), and so on.

The existence of interference introduces challenges to traditional statistical analyses. In order to handle the interaction between treatments of different individuals, new methodologies for experiment design have been developed. Related work include Basse and

Correspondence: Wenbin Lu, Department of Statistics, North Carolina State University, Raleigh, NC 27695. lu@stat.ncsu.edu.

DATA AVAILABILITY STATEMENT

Simulated data and the associated Python codes for implementing the proposed Q-learning and A-learning estimation methods are provided in the Supporting Information.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Airoldi (2015) and Eckles, Karrer, and Ugander (2017). In the area of causal inference for social network data, a common assumption is called partial interference, which basically states that interference exists within partitioned small groups but not between different groups. Under this framework, different types of causal inference have been studied by many authors, for example, Hudgens and Halloran (2008), Tchetgen and VanderWeele (2012), Aronow and Samii (2013), Liu and Hudgens (2014), Liu, Hudgens, and Becker-Dreps (2016), and Sussman and Airoldi (2017).

In this paper, we consider the derivation of the optimal individualized treatment regimen in the presence of interference. As far as we know, no such methods have been studied previously. For the related work mentioned above, multiple covariates of the individuals are usually collected along with their network structure, which provide us the opportunity to learn the optimal treatment decision. In the vaccination example, our goal is to get the lowest overall infection rate in a neighbourhood. In such case, because of the existence of interference, it is not necessary that everyone gets the vaccine to achieve the optimal overall outcome. For instance, there can be nonignorable side effect for young children to get the vaccine, but as long as other family members are vaccinated, they are well protected. Therefore, on the basis of covariates such as age, health status, vaccine history, and some others, we hope to have an optimal individualized decision rule with network information taken into account. Another example is a recent psychological study (Harrison et al., 2017) on resilience-based intervention for children affected by parental HIV/AIDS in order to improve their school outcome. There are three types of treatments, that is, child, caregiver, and community. All the participating children are 6 to 17 years old and spend most of their time at school. Therefore, one's psychological status can be heavily affected by his or her friends. In this situation, a careful choice of intervention can benefit not only individual but his or her friends indirectly as well.

Deriving the optimal individualized treatment regimen with the consideration of interference is interesting and challenging, for which the existing models for deriving the optimal treatment regimen cannot be applied directly. We propose a new network-based regression model in this paper, which includes the interference effects when modelling the response. Although the response for one individual depends on the covariates and treatments of others, we can show that the optimal treatment rule, if considering the overall response for the population, will only depend on the individual's covariates; that is, there are no interference effects involved in the decision rule. This is appealing because it makes the proposed model feasible in practice.

The optimal treatment rule based on our proposed model is derived using both Q-learning and A-learning approaches. Q-learning (Watkins, 1989; Watkins & Dayan, 1992) optimizes the corresponding value function derived from a parametric model of responses given observed covariates and treatment assignments, and it results in an optimal decision rule. A-learning (Murphy, 2003; Robins, 2004), in contrast, is a semiparametric method, which derives from a model that directly describes the difference between treatments, with the baseline remaining unspecified. For illustration purposes, we focus on a single decision time point throughout this paper. Computational details for parameter estimation and their asymptotic variance are provided for both Q-learning and A-learning.

The rest of this paper is organized as follows. In Section 2, we propose the network-based regression model that explicitly characterizes the interference effects. The parameter and variance estimators, as well as the corresponding optimal treatment regimen estimator by Q-learning approach, are derived. In Section 3, all estimators and their computational details are provided in the scope of A-learning. Extensive simulation studies and an application to a mobile game network data are conducted in Sections 4 and 5, respectively, in order to illustrate the proposed methods. A conclusion follows in Section 6.

2 | Q-LEARNING

2.1 | Model formulation

The data consist of n observations. When subject i and subject j are friends (connected), we denote it as $j \sim i$ and the interference set of subject i is defined as

$\mathcal{J}_i = \{j : j \sim i \text{ and } j = 1, 2, \dots, n_i\}$, where n_i is the number of friends of subject i . We consider two possible treatment assignments, denoted by a random variable $A_i \in \{0, 1\}$, whereas a_i represents the observed treatment assignment for subject i . Let S_i be the vector of treatment assignments for subjects in the interference set of subject i ; that is, $S_i = (A_{i_1}, A_{i_2}, \dots, A_{i_{n_i}})$, and s_i is the corresponding observed treatment assignments $(a_{i_1}, a_{i_2}, \dots, a_{i_{n_i}})$. Let $\mathbf{x}_i$ denote a p -

dimensional vector of covariates for subject i . The covariate set of the interference set of subject i is defined as $\chi_i = \{\mathbf{x}_{i_j} : i_j \in \mathcal{J}_i\}$. Moreover, let, $Y_i^*(a_i, s_i)$ and $Y_i(a_i, s_i)$ and $Y_i(a_i, s_i)$ denote the potential and observed outcomes, respectively. Note that the outcome for subject i also depends on the treatments assigned to his or her friends because of the existence of interference. As is customary, the stable unit treatment value assumption (Rubin, 1978) and the no unmeasured confounder assumption are necessary but need to be extended under the existence of interference. In other words, we assume that (a)

$Y_i(a_i, s_i) = I(A_i = a_i, S_i = s_i)Y_i^*(a_i, s_i)$, and (b) $Y_i^*(a_i, s_i)$ is independent of (A_i, S_i) conditional on $\mathbf{x}_i$ and χ_i . Let $g(\mathbf{x}_i, \chi_i)$ denote the optimal treatment rule for subject i . For now, we assume that it is a function of not only $\mathbf{x}_i$ but also χ_i because we want to consider the influence

between friends. It is infeasible to implement if the optimal treatment decision of an individual relies on all the covariates of his or her friends. We will show later that under our proposed model, the optimal treatment regimen for subject i is a function of $\mathbf{x}_i$ only. Let $\mathcal{G}_i = \{g(\mathbf{x}_i, \chi_{i_j}) : i_j \in \mathcal{J}_i\}$, which is the optimal treatment decision set of the interference set of i . Without loss of generality, we assume that larger value of outcome is better, so the optimal treatment rule is given by maximizing

$$\frac{1}{n} \sum_{i=1}^n E\{Y_i^*(g(\mathbf{x}_i, \chi_i), G_i) | \mathbf{x}_i, \chi_i\}, \quad (1)$$

which is equivalent to maximizing $\frac{1}{n} \sum_{i=1}^n E\{Y_i(a_i, s_i) | \mathbf{x}_i, \chi_i, a_i = g(\mathbf{x}_i, \chi_i), s_i = \mathcal{G}_i\}$ under assumptions (a) and (b).

In order to characterize the influence of interference, we propose the following network-based regression model for $Y_i(a_i, s_i)$:

$$Y_i(a_i, s_i) = \alpha + \boldsymbol{\beta}' \mathbf{x}_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \boldsymbol{\beta}' \mathbf{x}_j + \eta a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j + a_i \boldsymbol{\theta}' \mathbf{x}_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \boldsymbol{\theta}' \mathbf{x}_j + \epsilon_i, \quad (2)$$

where ϵ_i is a random error term with mean 0 and $\mathbf{W} = (W_{ij}) \in \{0, 1\}^{n \times n}$ is the adjacency matrix with $W_{ij} = 1$ if $i \sim j$ and $W_{ij} = 0$ otherwise. Note that parameters γ_1 , γ_2 , and γ_3 quantify the dependence of network, characterizing the effects of covariates, treatments, and covariates–treatment interactions of connected nodes, respectively. Based on our model, the response of an individual is related to not only his or her own covariates or treatment assignment but also his or her friends' covariates and treatment assignments. Define $\Theta = (\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\theta}, \gamma_1, \gamma_2, \gamma_3)^T$.

A main motivation for us to consider model (2) for incorporating network interference effects of treatments is that it will lead to a simple optimal treatment rule. Note that $W_{ii} = 0$ for all i . With some simple algebra, we can show that Equation 1 is equivalent to

$$\frac{1}{n} \sum_{i=1}^n \left(\alpha + \boldsymbol{\beta}' \mathbf{x}_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \boldsymbol{\beta}' \mathbf{x}_j \right) + \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}_i, \chi_i) \left[\left(1 + \gamma_2 \sum_{j \neq i}^n W_{ij} \right) \eta + \left(1 + \gamma_3 \sum_{j \neq i}^n W_{ij} \right) \boldsymbol{\theta}' \mathbf{x}_i \right].$$

The optimal treatment rule for subject i is thus given by

$$g^{\text{opt}}(\mathbf{x}_i, \chi_i) = g^{\text{opt}}(\mathbf{x}_i) = I \left[\left(1 + \gamma_2 \sum_{j \neq i}^n W_{ij} \right) \eta + \left(1 + \gamma_3 \sum_{j \neq i}^n W_{ij} \right) \boldsymbol{\theta}' \mathbf{x}_i > 0 \right], \quad (3)$$

where $I(\cdot)$ is the indicator function. From Equation 3, it is obvious to see that the optimal decision rule of subject i only depends on his or her own covariates, although we model the response in the presence of interference. The only information of network needed in the decision rule is $\sum_{j \neq i}^n W_{ij}$, the number of friends, which is easy to collect in practice.

2.2 | Model fitting

The parameter estimations are obtained by minimizing the quadratic loss between the observed responses and their means, that is,

$$\min_{\Theta} \frac{1}{n} \sum_{i=1}^n \left[Y_i - \left(\alpha + \beta' x_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \beta' x_j + \eta a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j + a_i \theta' x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \theta' x_j \right) \right]^2.$$

In order to find the solutions, we first fix γ_1 , γ_2 , and γ_3 at some initial values, and we get the ordinary least square (OLS) estimates for α , β , η , and θ . In our implementation, we set the initial values of γ_1 , γ_2 , and γ_3 as 0 because they are usually very small in practical applications. Then, we update γ_1 , γ_2 , and γ_3 by solving OLS after fixing α , β , η , and θ at the previous iteration, and we alternate between these two OLS problems till convergence. From our numerical experience, the proposed iterative algorithm usually converges within 50 iterations for a reasonably large sample size. Let $\hat{\Theta}$ denote the resulting estimators at convergence. Let Θ_0 denote the true values of the parameters. The asymptotic distribution of the estimators $\hat{\Theta}$ is given in Theorem 1.

Theorem 1. Assume the following conditions:

C1 The covariates $\mathbf{x}$ are bounded.

C2 The adjacency matrix $\mathbf{W}$ satisfies $\sum_{j \neq i} W_{ij} \leq K_0$ for all i , for some constant K_0 .

C3 The parameters Θ belong to a compact space that contains Θ_0 as an inner point.

$$\text{C4 } A(\Theta_0) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left[- \frac{\partial \phi(Y_i; \Theta)}{\partial \Theta} \right]_{\Theta = \Theta_0} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i(\Theta_0) \{d_i(\Theta_0)\}^T \text{ is}$$

finite and nondegenerate.

$$\text{C5 } B(\Theta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(Y_i; \Theta_0) \{\phi(Y_i; \Theta_0)\}^T \text{ is positive definite.}$$

We have, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\Theta} - \Theta_0) \xrightarrow{d} N\left(\mathbf{0}, A^{-1}(\Theta_0) B(\Theta_0) \{A^{-1}(\Theta_0)\}^T\right),$$

where the definitions of $d_i(\Theta)$ and $\phi(Y_i; \Theta)$ are given in Appendix A.

Remark 1. Conditions C1–C3 are standard regularity conditions. In particular, Condition C2 assumes that the network has a sparse structure so that node degrees are bounded by a constant even when the sample size is large. Such a boundedness condition can facilitate the derivation of large sample properties of the estimator and is widely assumed in causal inference for network-based data. Conditions C4 and C5 are needed for establishing the asymptotic normal distribution of the estimator.

3 | A-LEARNING

3.1 | Model formulation

From the decision rule (3) of the proposed Q-learning model, we notice that the optimal treatment assignment is independent of the baseline. Therefore, we propose another more robust semiparametric A-learning model:

$$Y_i(a_i, s_i) = \mu(x_i, \chi_i) + \eta a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j + a_i \theta' x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \theta' x_j + \epsilon_i,$$

where $\mu(x_i, \chi_i)$ is an unspecified baseline function. Note that $\mu(x_i, \chi_i)$ can be a function of either x_i itself or both x_i and χ_i . By adopting the same technique of switching the double summations, the optimal decision rule in the above model can be shown to be the same as Equation 3. Let $\Omega = (\eta, \theta, \gamma_2, \gamma_3)^T$, which is the vector of the parameters involved in the decision rule. Inspired by the doubly robust estimating equation proposed by Robins, Rotnitzky, and Zhao (1994), we propose the following estimating equation for Ω :

$$U_{n,\Omega} \equiv \sum_{i=1}^n \begin{pmatrix} (a_i - \pi_i) + \gamma_2 \sum_{j \neq i}^n W_{ij} (a_j - \pi_j) \\ (a_i - \pi_i) x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} (a_j - \pi_j) x_j \\ \sum_{j \neq i}^n W_{ij} \eta (a_j - \pi_j) \\ \sum_{j \neq i}^n W_{ij} (a_j - \pi_j) \theta' x_j \end{pmatrix} \quad (4)$$

$$\times \left\{ Y_i - \mu(x_i, \chi_i; \delta) - \eta a_i - \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j - a_i \theta' x_i - \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \theta' x_j \right\} = 0,$$

where π_i is the posited propensity score model for subject i , which can be a function of x_i and χ_i , and $\mu(x_i, \delta)$ is a posited parametric model for the baseline mean. When at least one of π_i and $\mu(x_i, \delta)$ is correctly specified, Equation 4 is an unbiased estimating equation for Ω .

This can be shown by using the iterated conditional expectation, first conditioning on x_i, a_i ,

χ_i , and s_i and then conditioning on $\mathbf{x}_i$ and χ_i . In addition, the estimating equation for $\boldsymbol{\delta}$ is given by

$$U_{n,\boldsymbol{\delta}} \equiv \sum_{i=1}^n \left(\frac{1}{x_i} \right) \left\{ Y_i - \mu(\mathbf{x}_i, \chi_i; \boldsymbol{\delta}) - \eta a_i - \gamma_2 \sum_{j \neq i} W_{ij} \eta a_j - a_i \boldsymbol{\theta}' \mathbf{x}_i - \gamma_3 \sum_{j \neq i} W_{ij} a_j \boldsymbol{\theta}' \mathbf{x}_j \right\} = 0. \quad (5)$$

In the following discussion, for simplicity, we posit a standard linear model for the baseline mean function, that is, $\mu(\mathbf{x}_i, \chi_i; \boldsymbol{\delta}) = \alpha + \boldsymbol{\beta}' \mathbf{x}_i$, where $\boldsymbol{\delta} = (\alpha, \boldsymbol{\beta}^T)^T$. In addition, we posit a standard logistic regression model for the propensity score $\pi_i = \pi(\mathbf{x}_i, \Phi)$. However, the derivation based on other parametric models is similar.

3.2 | Model fitting

The parameter Φ in the propensity score model can be first estimated using the generalized estimating equation

$$U_{n,\Phi} \equiv \sum_{i=1}^n \frac{\partial \pi(\mathbf{x}_i, \Phi)}{\partial \Phi} \{a_i - \pi(\mathbf{x}_i, \Phi)\} = 0.$$

Let $\widehat{\Phi}$ denote the resulting solution. Write $\widehat{\pi}_i = \pi(\mathbf{x}_i, \widehat{\Phi})$. To improve the stability of solving the estimating equations (4) and (5), we consider to the following least square minimization problem at step k :

$$\min_i \sum_{i=1}^n \left\{ Y_i - \alpha - \boldsymbol{\beta}' \mathbf{x}_i - f(\mathbf{x}_i; \Omega^{(k-1)}) - \eta(a_i - \widehat{\pi}_i) - \gamma_2 \sum_{j \neq i} W_{ij} \eta(a_j - \widehat{\pi}_j) - \boldsymbol{\theta}' \mathbf{x}_i(a_i - \widehat{\pi}_i) - \gamma_3 \sum_{j \neq i} W_{ij} (a_j - \widehat{\pi}_j) \boldsymbol{\theta}' \mathbf{x}_j \right\}^2,$$

where $f(\mathbf{x}_i; \Omega) = \eta \widehat{\pi}_i + \gamma_2 \sum_{j \neq i} W_{ij} \eta \widehat{\pi}_j + \boldsymbol{\theta}' \mathbf{x}_i \widehat{\pi}_i + \gamma_3 \sum_{j \neq i} W_{ij} \widehat{\pi}_j \boldsymbol{\theta}' \mathbf{x}_j$ and $\Omega^{(k-1)}$ is the estimator of the parameters Ω at the $(k-1)$ th step.

To compute the updated least square estimates at the k th step, we consider a similar alternate least square estimation as studied in Q-learning. Specifically, we first fix γ_2 and γ_3 at their estimates $\gamma_2^{(k-1)}$ and $\gamma_3^{(k-1)}$ in the $(k-1)$ th step and update all other parameters α , $\boldsymbol{\beta}$, η , and $\boldsymbol{\theta}$ by OLS. Next, we update the parameters in $f(\mathbf{x}_i; \Omega)$ using the estimates from the previous step, and then we solve γ_2 and γ_3 after fixing the other parameters. Finally, $f(\mathbf{x}_i; \Omega)$ is updated again by the current estimates of γ_2 and γ_3 . From an initial estimator $\Omega^{(0)}$, which can be obtained using the Q-learning method, we iterate the above steps till convergence. In our implementation, the convergence criterion is set as the absolute differences between estimates of all the parameters in two consecutive iterations are no larger than 10^{-3} . Based on our numerical studies, this algorithm usually converges within 50 iterations for a reasonably large sample size. Let $\widehat{\Omega}$ and $\widehat{\boldsymbol{\delta}}$ denote the estimates of Ω and $\boldsymbol{\delta}$, respectively.

Define $u_{\Phi}(\Phi) = \lim_{n \rightarrow \infty} n^{-1} U_{n, \Phi}(\Phi)$ and $u_{\delta}(\delta, \Omega) = \lim_{n \rightarrow \infty} n^{-1} U_{n, \delta}(\delta, \Omega)$. Let Ω_0 denote the true value of Ω . In addition, assume that the equations $u_{\Phi}(\Phi) = 0$ and $u_{\delta}(\delta, \Omega_0) = 0$ have unique solutions Φ^* and δ^* respectively. Following White (1982) and Li and Duan (1989) for misspecified models, under certain regularity conditions, it can be shown that $\hat{\Phi}$ converges in probability to Φ^* and has an asymptotic linear representation. In the following theorem, we establish the asymptotic distribution of $\hat{\Omega}$ and $\hat{\delta}$.

Theorem 2. Assume that either the baseline mean or the propensity score model is correctly specified. In addition, assume Conditions C1–C3 and

C6 $\sqrt{n}(\hat{\Phi} - \Phi^*) = n^{-1/2} \sum_{i=1}^n \xi_i + o_p(1)$, where ξ_i are independent mean-zero random vectors with $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \xi_i \xi_i^T < \infty$.

C7 $A(\Omega_0, \delta^*, \Phi^*)$ is finite and nondegenerate, and $B(\Omega_0, \delta^*, \Phi^*)$ is positive definite.

We have, as $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \hat{\Omega} - \Omega \\ \hat{\delta} - \delta^* \end{pmatrix} \xrightarrow{d} N \left(\mathbf{0}, A^{-1}(\Omega_0, \delta^*, \Phi^*) B(\Omega_0, \delta^*, \Phi^*) \{ A^{-1}(\Omega_0, \delta^*, \Phi^*) \}^T \right).$$

The definitions of $A(\Omega_0, \delta^*, \Phi^*)$ and $B(\Omega_0, \delta^*, \Phi^*)$ are given in Appendix B.

4 | SIMULATION STUDIES

In this section, several simulation studies are conducted to evaluate the empirical performance of the proposed methods. In all settings, the underlying network is generated by the stochastic block model (Holland, Laskey, & Leinhardt, 1983), in which the probability that node i and node j are connected is modelled as $P_{C_i C_j}$, where C_i and C_j are the

communities to which node i and j belong, respectively, and P is a matrix whose element is the probability that two communities are connected. We set the number of communities as 5.

To illustrate the performance of the proposed Q-learning method, the total number of subjects is chosen as $n = 1,000$ or $2,000$. When $n = 1,000$, the numbers of subjects in each community are (250, 250, 200, 200, 100), and when $n = 2,000$, the community sizes are (500, 500, 400, 400, 200). We also try two different densities of the network. The dense network is set as $P_{11} = P_{33} = 0.05$, $P_{22} = P_{44} = P_{55} = 0.1$, and $P_{C_i C_j}$ for $C_i \neq C_j$. For the

sparse network, the connecting probabilities are half of those in the dense one. The responses are generated from model (2). Two covariates are considered, where $x_{11} \sim \text{Bernoulli}(0.5)$ and $x_{12} \sim \text{Uniform}(-1, 1)$. Treatment A_j is randomly assigned to 0/1 with equal probability; that is, $\pi_j = 0.5$. In our simulations, treatment assignment indicators are assumed independent between individuals for simplicity. More general settings can be considered as well. The parameters are set as $\alpha = 0.5$, $\beta = (-1, 1)^T$, $\eta = 1$, $\theta = (-0.5, 0.5)^T$, $\gamma_1 = 0.05$, and $\gamma_3 = 0.2$. Variance of ϵ_j is 1, and γ_2 can be 0.1 or 0. Simulation results based on 1,000 replicates are shown in Table 1. The average percentage of making correct decisions (PCD) over 1,000 simulations based on the estimated optimal treatment regimen

(denoted by PCD_{NET}) is reported. We also report the PCD based on the estimated optimal treatment regimen, ignoring the network structure, that is, assuming $\gamma = (\gamma_1 = \gamma_2 = \gamma_3)^T = \mathbf{0}$ when fitting the model (denoted by $PCD_{\gamma=0}$). We can see that PCD_{NET} is always much higher than $PCD_{\gamma=0}$ as expected. Moreover, we report the average value functions of model (2) based on the estimated optimal treatment regimen considering the network structure, the estimated optimal treatment regimen ignoring the network structure, that is, $\gamma = \mathbf{0}$, and the true optimal treatment regimen, denoted by Q_{NET}^{opt} , $Q_{\gamma=0}^{opt}$, and Q^{opt} . It can be seen that Q_{NET}^{opt} is always larger than $Q_{\gamma=0}^{opt}$ and slightly smaller but close to the true optimal value Q^{opt} . In addition, we show the performance of the parameter estimation. Here, standard errors of the estimates are computed using the plug-in method based on the asymptotic variances established in Theorems 1 and 2. Details can be found in the Appendix. It can be seen that, for all settings, all the parameter estimators are almost unbiased. The means of the estimated standard errors are close to the standard deviations of the estimators, and the empirical coverage probabilities of the 95% Wald-type confidence interval are close to the nominal level.

For the proposed A-learning model, we fix the total number of subjects as $n = 2,000$ with community sizes (500, 500, 400, 400, 200). The network density is the same as the sparse network described above; that is, $P_{11} = P_{33} = 0.025$, $P_{22} = P_{44} = P_{55} = 0.05$, and $P_{C_i C_j}$ for c_i

C_j . Here, γ_2 is set at 0.1. All other parameters are the same as in the Q-learning model. Covariates are also generated in the same way. In order to show the double robustness property of the proposed A-learning model, we consider both the correctly and incorrectly specified models for the baseline $\mu(x_i, \chi_i)$ and the propensity score π_i . In particular, we always fit a standard linear model $h(x_i, \chi_i) = \alpha + \beta'x_i$ for the baseline, which is correctly specified when $\mu(x_i, \chi_i) = \alpha + \beta'x_i$, but misspecified when

$\mu(x_i, \chi_i) = \alpha + \beta'x_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \beta'x_j$ as in the Q-learning model. For the propensity score, A_i is generated by a logistic model; that is, $\text{logit}(\pi_i) = \psi + \phi'x_i$, where $\psi = 0.5$ and $\phi = (-1, 1)^T$. In considered settings, the propensity score model only depends on x_i but not χ_i for simplicity in implementation. As discussed before, the proposed A-learning method can handle the more general case where the propensity score model depends on both x_i and χ_i . However, more numerical studies need to be conducted to examine the performance of the proposed method under such a general setting. The same logistic model is fitted when it is correctly specified, whereas a random assignment $\pi_i = p$ is fitted when it is incorrectly specified. Simulation results are shown in Table 2. We can see that as long as either $h(x_i, \chi_i)$ or π_i is correctly specified, PCD_{NET} of the estimated optimal treatment regimen taking into account the network interference is much higher than $PCD_{\gamma=0}$, which ignores the network interference. Similarly, the average optimal value Q_{NET}^{opt} based on the estimated optimal treatment regimen with the network interference is higher than $Q_{\gamma=0}^{opt}$ that ignores the network effect, and it is close to the true optimal value. When both the baseline and the propensity score are incorrect, the PCD and optimal value of the network-based optimal treatment regimen are just slightly worse than those ignoring the network effect. In addition,

the estimators for the parameters involved in the decision rule are almost unbiased when at least one of the baseline and the propensity score is correctly specified. The means of their estimated standard errors are close to the standard deviations of the estimators, and the coverage probabilities of the 95% Wald-type confidence interval are also close to the nominal level.

5 | APPLICATION TO MOBILE GAME DATA

In this section, we use the covariates and network structure collected in one of the largest online social network platforms to illustrate our proposed methods (the platform that provided us data has requested anonymity). This data set consists of $n = 961$ users and their friendship network as shown in Figure 1. The numbers of friends are from 0 to 154, and the median is 6. Several covariates of these users are also available. We include age and activity level in the following analysis, where activity level indicates how active a user is on the network. All covariates are centred and scaled first. This data set was collected to study the propagation of a particular mobile game. The game sends invitations to players' friends through the network asking them to join the game.

For illustration purpose, we assume there are two kinds of invitations: One type of invitation notifies you that your friends are playing this game, whereas the other type also mentions the names of these friends. The invitation is considered as treatment in terms of promoting the popularity of this game, denoted by 0 or 1 for either type. The propensity score is assumed to follow a logistic model, and treatment for each user is generated with parameter $\Phi = (1, -0.1, 0.1)^T$. Let T_i denote the time that the i th user starts to play the game since it was launched. We first generate the response $Y_i = -\log(T_i)$ based on our proposed Q-learning model. In order to advertise the game, we hope that more users start playing it in a short time. It indicates that the larger value of the response is better. The parameters that we use to generate the responses are chosen as $\alpha = 1$, $\beta = (1, -1)^T$, $\eta = 1$, $\theta = (0.5, -1)^T$, $\gamma_1 = 0.1$, $\gamma_2 = 0.1$, and $\gamma_3 = 0.2$, and the error term follows a standard normal distribution. For the Q-learning model, we assume that the responses come from the proposed model (2). The results are shown in Table 3. First, the proposed estimators are nearly unbiased. Second, by considering the interference between friends, the estimated optimal treatment regimen achieves higher PCD and value than that ignoring the network structure in decision making. In addition, we plot the assignment based on the true optimal treatment regimen (left panel) and the assignment based on the estimated optimal treatment regimen obtained by Q-learning (right panel) in Figure 1. It can be seen that the estimated optimal treatment regimen and the true optimal treatment regimen give very consistent assignment (the PCD is 0.986), showing the good performance of the proposed estimation method.

Next, we evaluate the performance of the proposed A-learning method based on the mobile game data. As in simulations, we always fit a simple linear model for the baseline; that is, $h(x_i, \chi_i) = \alpha + \beta'x_i$. We consider cases when the baseline model or the propensity score or both is correctly specified. Results in each situation are presented in Table 4. As expected, as long as at least one of the baseline function and the propensity score is correctly specified, the proposed estimator is nearly unbiased, and the estimated optimal treatment regimen with interference achieves higher PCD and value than that ignoring the network structure.

6 | CONCLUSIONS

In this paper, we propose a novel network-based regression model for deriving the optimal treatment regimen in the presence of interference. One advantage of our proposed model is that the true optimal treatment regimen only depends on one's own covariates and number of connected nodes in the network. Such property makes its implementation in reality become practical. In future work, we can extend the model to include interaction between others' covariates/treatment assignment and one's own covariates/treatment assignment, which makes the model more flexible but results in more complexity in the optimal decision rule. In addition, our current study is focused on a single decision time point. We may extend the proposed methods to incorporate multiple decision time points based on backward induction. Moreover, the proposed model assumed a linear interaction between covariates and treatment, which can also be relaxed. These are interesting topics that need further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

APPENDIX A:: PROOF OF THEOREM 1

The parameter estimations for the Q-learning model are obtained by minimizing the loss function $L(Y; \Theta)$, which is defined as

$$L(Y; \Theta) = \sum_{i=1}^n \left[Y_i - \left(\alpha + \beta'x_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \beta'x_j + \eta a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j + a_i \theta'x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \theta'x_j \right) \right]^2,$$

where $\Theta = (\alpha, \beta, \eta, \theta, \gamma_1, \gamma_2, \gamma_3)^T$. Therefore, the minimizer $\hat{\Theta}$ can be obtained by solving the estimating equation $\sum_{i=1}^n \phi(Y_i; \hat{\Theta}) = 0$, where $\phi(Y_i; \Theta) = d_i(\Theta) \{Y_i - m_i(\Theta)\}$ with

$$m_i(\Theta) = \alpha + \beta'x_i + \gamma_1 \sum_{j \neq i}^n W_{ij} \beta'x_j + \eta a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} \eta a_j + a_i \theta'x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j \theta'x_j,$$

and

$$d_i(\Theta) = \begin{pmatrix} 1 \\ x_i + \gamma_1 \sum_{j \neq i}^n W_{ij} x_j \\ a_i + \gamma_2 \sum_{j \neq i}^n W_{ij} a_j \\ a_i x_i + \gamma_3 \sum_{j \neq i}^n W_{ij} a_j x_j \\ \sum_{j \neq i}^n W_{ij} \beta' x_j \\ \sum_{j \neq i}^n W_{ij} \eta a_j \\ \sum_{j \neq i}^n W_{ij} a_j \theta' x_j \end{pmatrix}.$$

By the Taylor expansion, we have

$$\begin{aligned} \sqrt{n}(\hat{\Theta} - \Theta_0) &= \left[-\frac{1}{n} \sum_{i=1}^n \frac{\partial \phi(Y_i; \Theta)}{\partial \Theta^T} \bigg|_{\Theta = \Theta_0} \right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(Y_i; \Theta_0) \right] + \sqrt{n} R_n^* \\ &= \{A(\Theta_0)\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(Y_i; \Theta_0) + \sqrt{n} R_n^* + o_p(1). \end{aligned}$$

Following Stefanski and Boos (2002), under Conditions C1–C3, we can show that $\sqrt{n} R_n^* \xrightarrow{P} 0$.

Moreover, by conditioning on the network $\mathbf{W}$ and all the treatment assignments a_s and covariates $\mathbf{x}_s$, the terms $d_i(\Theta)$ s are constants. Under Conditions C4 and C5, by the multiplier central limit theorem (Kosorok, 2008, p. 184, Lemma 10.5), it can be shown that

$$n^{-1/2} \sum_{i=1}^n \phi(Y_i; \Theta_0) \xrightarrow{d} N(\mathbf{0}, B(\Theta_0)), \text{ where } B(\Theta_0) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \phi(Y_i; \Theta_0) \{\phi(Y_i; \Theta_0)\}^T.$$

Then the asymptotic distribution of $\hat{\theta}$ given in Theorem 1 is proved.

The asymptotic variance of the estimates can be estimated by

$$\widehat{\text{cov}}(\hat{\Theta}) = n^{-1} A_n^{-1}(\hat{\Theta}) B_n(\hat{\Theta}) \{A_n^{-1}(\hat{\Theta})\}^T, \text{ where}$$

$$\begin{aligned} B_n(\Theta) &= \frac{1}{n} \sum_{i=1}^n \phi(Y_i; \Theta) \{\phi(Y_i; \Theta)\}^T = \frac{1}{n} \sum_{i=1}^n d_i(\Theta) \{d_i(\Theta)\}^T \{Y_i - m_i(\Theta)\}^2, \\ A_n(\Theta) &= \frac{1}{n} \sum_{i=1}^n d_i(\Theta) \{d_i(\Theta)\}^T. \end{aligned}$$

Equations (4) and (5) can be written as $U_n(\Omega, \delta, \hat{\Phi}) = \sum_{i=1}^n g_i(\Omega, \delta, \hat{\Phi}) \epsilon_i(\Omega, \delta) = 0$, where

$$g_i(\Omega, \delta, \Phi) = \begin{pmatrix} 1 \\ \mathbf{x}_i \\ (a_i - \pi_i) + \gamma_2 \sum_{j \neq i}^n W_{ij}(a_j - \pi_j) \\ (a_i - \pi_i)\mathbf{x}_i + \gamma_3 \sum_{j \neq i}^n W_{ij}(a_j - \pi_j)\mathbf{x}_j \\ \sum_{j \neq i}^n W_{ij}\eta(a_j - \pi_j) \\ \sum_{j \neq i}^n W_{ij}(a_j - \pi_j)\theta'\mathbf{x}_j \end{pmatrix}, \quad \epsilon_i(\Omega, \delta) = Y_i - \alpha - \beta'\mathbf{x}_i - \eta a_i - \gamma_2 \sum_{j \neq i}^n W_{ij}\eta a_j - a_i\theta'\mathbf{x}_i - \gamma_3 \sum_{j \neq i}^n W_{ij}a_j\theta'\mathbf{x}_j$$

By the Taylor expansion and some empirical process approximation techniques, we have

$$0 = U_n(\widehat{\Omega}, \widehat{\delta}, \widehat{\Phi}) = U_n(\Omega_0, \delta^*, \Phi^*) + \frac{\partial U_n(\Omega, \delta, \Phi^*)}{\partial(\Omega^T, \delta^T)} \bigg|_{\Omega = \Omega_0, \delta = \delta^*} \begin{pmatrix} \widehat{\Omega} - \Omega_0 \\ \widehat{\delta} - \delta^* \end{pmatrix} + \frac{\partial U_n(\Omega_0, \delta^*, \Phi)}{\partial \Phi^T} \bigg|_{\Phi = \Phi^*} (\widehat{\Phi} - \Phi^*) + o_p(\sqrt{n}).$$

$$\text{Let } A(\Omega_0, \delta^*, \Phi^*) = \lim_{n \rightarrow \infty} \left(-n^{-1} \frac{\partial U_n(\Omega, \delta, \Phi^*)}{\partial(\Omega^T, \delta^T)} \bigg|_{\Omega = \Omega_0, \delta = \delta^*} \right) \text{ and } H(\Omega_0, \delta^*, \Phi^*). \text{ Then, we have}$$

$$= \lim_{n \rightarrow \infty} n^{-1} \frac{\partial U_n(\Omega_0, \delta^*, \Phi)}{\partial \Phi^T} \bigg|_{\Phi = \Phi^*}$$

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\widehat{\Omega}-\Omega_0 \\ \widehat{\delta}-\delta^*\end{pmatrix} &= \{A(\Omega_0, \delta^*, \Phi^*)\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{g_i(\Omega_0, \delta^*, \Phi^*)\epsilon_i(\Omega_0, \delta^*) + H(\Omega_0, \delta^*, \Phi^*)\xi_i\} + o_p(1) \\ &\equiv \{A(\Omega_0, \delta^*, \Phi^*)\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_i(\Omega_0, \delta^*, \Phi^*) + o_p(1).\end{aligned}$$

Note that conditioning on the network $\mathbf{W}$ and all the treatment assignments a_s and covariates $\mathbf{x}_s$, $\Psi_i(\Omega_0, \delta^*, \Phi^*)$ s are independent random vectors with mean zero. Let $B(\Omega_0, \delta^*, \Phi^*) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \Psi_i(\Omega_0, \delta^*, \Phi^*) \{\Psi_i(\Omega_0, \delta^*, \Phi^*)\}^\top$, which is assumed to be positive definite by Condition C7. By the multiplier central limit theorem (Kosorok, 2008, p. 184, Lemma 10.5), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_i(\Omega_0, \delta^*, \Phi^*) \xrightarrow{d} N(\mathbf{0}, B(\Omega_0, \delta^*, \Phi^*))$$

The asymptotic distribution of $\widehat{\Omega}$ and $\widehat{\delta}$ established in Theorem 2 then follows. As before, the asymptotic variance of the estimators can be estimated using the simple plug-in method.

REFERENCES

- Aronow PM, & Samii C (2013). Estimating average causal effects under general interference, with application to a social network experiment. arXiv preprint arXiv:1305.61569.
- Basse GW, & Airoldi EM (2015). Optimal design of experiments in the presence of network-correlated outcomes. ArXiv e-prints arXiv:1507.00803v1.
- Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, Settle JE, & Fowler JH (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298. [PubMed: 22972300]
- Eckles D, Karrer B, & Ugander J (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1), 1–23.
- Halloran ME, & Struchiner CJ (1995). Causal inference in infectious diseases. *Epidemiology*, 6(2), 142–151. [PubMed: 7742400]
- Harrison SE, Li X, Zhang J, Chi P, Zhao J, & Zhao G (2017). Improving school outcomes for children affected by parental HIV/AIDS: Evaluation of the childcare intervention at 6-, 12-, and 18-months. *School Psychology International*, 38(3), 264–286.
- Holland PW, Laskey KB, & Leinhardt S (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137.
- Hong G, & Raudenbush SW (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101(475), 901–910.
- Hudgens MG, & Halloran ME (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482), 832–842. [PubMed: 19081744]
- Kosorok MR (2008). Introduction to empirical processes and semiparametric inference. New York: Springer.
- Li K-C, & Duan N (1989). Regression analysis under link violation. *Annals of Statistics*, 17(4), 1009–1052.
- Liu L, & Hudgens MG (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association*, 109(505), 288–301. [PubMed: 24659836]

- Liu L, Hudgens MG, & Becker-Dreps S (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika*, 103(4), 829–842. [PubMed: 29422692]
- Murphy SA (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2), 331–355.
- Robins JM (2004). Optimal structural nested models for optimal sequential decisions In *Proceedings of the Second Seattle Symposium in Biostatistics*. New York: Springer, pp. 189–326.
- Robins JM, Rotnitzky A, & Zhao LP (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427), 846–866.
- Rosenbaum PR (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477), 191–200.
- Rubin DB (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 6(1), 34–58.
- Sobel ME (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476), 1398–1407.
- Stefanski LA, & Boos DD (2002). The calculus of M-estimation. *The American Statistician*, 56(1), 29–38.
- Sussman DL, & Airoidi EM (2017). Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*.
- Tchetgen EJT, & VanderWeele TJ (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1), 55–75. [PubMed: 21068053]
- Watkins CJCH (1989). *Learning from delayed rewards*. (Ph.D. Thesis). Cambridge, UK: King's College.
- Watkins CJ, & Dayan P (1992). Q-learning. *Machine learning*, 8(3–4), 279–292.
- White H (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1), 1–25.

FIGURE 1.

Plots of the assignment based on the true optimal treatment regimen (left panel) and the assignment based on the estimated optimal treatment regimen obtained by Q-learning (right panel). The percentage of making correct decisions is 0.986 comparing the estimated optimal treatment regimen with the true optimal regime

TABLE 1

Q-learning simulation results

Setting	PCD/Q	Est. (SD)	Parameters	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
(a) $n = 1,000$												
Dense	PCD _{NET}	0.962 (0.0009)	Est.	0.500	-1.005	1.002	0.998	-0.493	0.492	0.049	0.101	0.211
$\gamma_2 = 0.1$	PCD _{$\gamma=0$}	0.838 (0.0011)	<i>SD</i>	0.091	0.084	0.074	0.081	0.106	0.101	0.012	0.014	0.055
	$Q_{\text{NET}}^{\text{opt}}$	1.310 (0.0030)	<i>SE</i>	0.093	0.085	0.075	0.084	0.110	0.099	0.012	0.014	0.057
	$Q_{\gamma=0}^{\text{opt}}$	1.192 (0.0034)	CP	0.953	0.946	0.951	0.957	0.947	0.935	0.945	0.951	0.961
	Q^{opt}	1.321 (0.0029)										
	PCD _{NET}	0.939 (0.0012)	Est.	0.489	-1.000	1.000	1.004	-0.506	0.497	0.049	0.001	0.207
$\gamma_2 = 0$	PCD _{$\gamma=0$}	0.535 (0.0012)	<i>SD</i>	0.094	0.083	0.074	0.080	0.104	0.099	0.012	0.013	0.051
	$Q_{\text{NET}}^{\text{opt}}$	0.221 (0.0024)	<i>SE</i>	0.092	0.083	0.074	0.082	0.105	0.096	0.012	0.013	0.052
	$Q_{\gamma=0}^{\text{opt}}$	-0.415 (0.0045)	CP	0.938	0.947	0.950	0.961	0.947	0.936	0.942	0.940	0.952
	Q^{opt}	0.235 (0.0023)										
	PCD _{NET}	0.954 (0.0011)	Est.	0.502	-1.005	1.002	0.998	-0.493	0.492	0.049	0.100	0.211
$\gamma_2 = 0.1$	PCD _{$\gamma=0$}	0.876 (0.0011)	<i>SD</i>	0.084	0.085	0.076	0.083	0.113	0.105	0.017	0.019	0.067
	$Q_{\text{NET}}^{\text{opt}}$	1.014 (0.0021)	<i>SE</i>	0.085	0.087	0.076	0.086	0.117	0.103	0.017	0.019	0.070
	$Q_{\gamma=0}^{\text{opt}}$	0.972 (0.0022)	CP	0.953	0.948	0.950	0.960	0.952	0.932	0.949	0.946	0.969
	Q^{opt}	1.024 (0.0021)										
	PCD _{NET}	0.935 (0.0011)	Est.	0.496	-1.002	1.002	1.000	-0.499	0.494	0.049	0.000	0.210
$\gamma_2 = 0$	PCD _{$\gamma=0$}	0.675 (0.0012)	<i>SD</i>	0.086	0.085	0.076	0.083	0.112	0.104	0.017	0.019	0.065
	$Q_{\text{NET}}^{\text{opt}}$	0.402 (0.0018)	<i>SE</i>	0.085	0.086	0.076	0.085	0.114	0.102	0.017	0.018	0.067
	$Q_{\gamma=0}^{\text{opt}}$	0.158 (0.0026)	CP	0.939	0.943	0.948	0.958	0.952	0.930	0.946	0.938	0.967
	Q^{opt}	0.413 (0.0018)										
	PCD _{NET}	0.979 (0.0005)	Est.	0.497	-1.000	1.003	1.001	-0.500	0.498	0.050	0.100	0.203
$\gamma_2 = 0.1$	PCD _{$\gamma=0$}	0.806 (0.0009)	<i>SD</i>	0.068	0.060	0.051	0.059	0.072	0.064	0.006	0.008	0.032
	$Q_{\text{NET}}^{\text{opt}}$	1.919 (0.0033)	<i>SE</i>	0.070	0.059	0.052	0.058	0.073	0.067	0.006	0.008	0.033
	$Q_{\gamma=0}^{\text{opt}}$	1.622 (0.0043)	CP	0.950	0.945	0.951	0.947	0.949	0.958	0.962	0.964	0.957
	Q_{opt}	1.925 (0.0033)										
	PCD _{NET}	0.966 (0.0006)	Est.	0.480	-0.992	0.998	1.015	-0.528	0.512	0.050	0.002	0.195
$\gamma_2 = 0$	PCD _{$\gamma=0$}	0.421 (0.0010)	<i>SD</i>	0.069	0.059	0.050	0.056	0.068	0.060	0.006	0.006	0.026
	$Q_{\text{NET}}^{\text{opt}}$	-0.031 (0.0024)	<i>SE</i>	0.069	0.055	0.050	0.055	0.066	0.062	0.006	0.006	0.026
(b) $n = 2,000$												

Setting	PCD/Q	Est. (SD)	Parameters	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
Sparse $\gamma_2 = 0.1$	$Q_{\gamma=0}^{\text{opt}}$	-1.580 (0.0068)	CP	0.933	0.931	0.947	0.945	0.928	0.952	0.957	0.949	0.930
	Q^{opt}	-0.024 (0.0024)										
	PCD _{NET}	0.974 (0.0006)	Est.	0.498	-1.000	1.003	1.002	-0.500	0.498	0.050	0.100	0.203
	PCD _{$\gamma=0$}	0.833 (0.0008)	<i>SD</i>	0.063	0.061	0.052	0.059	0.077	0.067	0.008	0.009	0.037
	$Q_{\text{NET}}^{\text{opt}}$	1.311 (0.0020)	<i>SE</i>	0.065	0.060	0.052	0.059	0.077	0.070	0.008	0.010	0.037
	$Q_{\gamma=0}^{\text{opt}}$	1.183 (0.0025)	CP	0.955	0.947	0.947	0.944	0.940	0.948	0.971	0.961	0.962
Sparse $\gamma_2 = 0$	Q^{opt}	1.316 (0.0020)										
	PCD _{NET}	0.957 (0.0008)	Est.	0.489	-0.996	1.002	1.007	-0.512	0.502	0.050	0.001	0.200
	PCD _{$\gamma=0$}	0.530 (0.0009)	<i>SD</i>	0.066	0.061	0.051	0.058	0.075	0.066	0.008	0.009	0.034
	$Q_{\text{NET}}^{\text{opt}}$	0.232 (0.0016)	<i>SE</i>	0.065	0.058	0.052	0.058	0.074	0.068	0.008	0.009	0.034
	$Q_{\gamma=0}^{\text{opt}}$	-0.422 (0.0033)	CP	0.945	0.940	0.949	0.942	0.935	0.946	0.967	0.947	0.947
	Q^{opt}	0.238 (0.0016)										

Note. PCD_{NET}: average percentage of making correct decisions by the network-based decision rule; PCD _{$\gamma=0$} : average percentage of making correct decisions when ignoring network effect; $Q_{\text{NET}}^{\text{opt}}$: average estimated value using the network-based optimal decision; $Q_{\gamma=0}^{\text{opt}}$: average estimated value of the optimal treatment when ignoring the network effect; Q^{opt} : average value using the true optimal regime; Est.: mean of estimators; *SD*: standard deviation of estimators; *SE*: mean of estimated standard errors; CP: empirical coverage probability of 95% confidence intervals; corresponding *SD* is in parentheses.

TABLE 2

A-learning simulation results

Baseline	PS	PCD/Q	Est. (SD)	Parameters	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
Correct	Correct	PCD _{NET}	0.969 (0.0008)	Est.	1.000	-0.500	0.503	0.100	0.203
		PCD _{$\gamma=0$}	0.837 (0.0010)	SD	0.064	0.087	0.078	0.017	0.047
		Q_{NET}^{opt}	1.717 (0.0016)	SE	0.065	0.088	0.080	0.017	0.046
		$Q_{\gamma=0}^{opt}$	1.596 (0.0022)	CP	0.950	0.950	0.957	0.954	0.941
		Q^{opt}	1.725 (0.0016)						
Correct	Incorrect	PCD _{NET}	0.940 (0.0015)	Est.	1.004	-0.508	0.509	0.099	0.197
		PCD _{$\gamma=0$}	0.837 (0.0010)	SD	0.062	0.084	0.073	0.011	0.059
		Q_{NET}^{opt}	1.699 (0.0019)	SE	0.064	0.083	0.075	0.011	0.059
		$Q_{\gamma=0}^{opt}$	1.596 (0.0021)	CP	0.952	0.946	0.955	0.960	0.956
		Q^{opt}	1.725 (0.0016)						
Incorrect	Correct	PCD _{NET}	0.960 (0.0010)	Est.	1.000	-0.498	0.501	0.102	0.204
		PCD _{$\gamma=0$}	0.837 (0.0010)	SD	0.068	0.095	0.084	0.021	0.053
		Q_{NET}^{opt}	1.712 (0.0017)	SE	0.067	0.091	0.083	0.018	0.050
		$Q_{\gamma=0}^{opt}$	1.596 (0.0022)	CP	0.946	0.935	0.951	0.935	0.942
		Q^{opt}	1.725 (0.0016)						
Incorrect	Incorrect	PCD _{NET}	0.817 (0.0004)	Est.	1.002	-0.501	0.499	0.098	-0.063
		PCD _{$\gamma=0$}	0.837 (0.0010)	SD	0.072	0.102	0.092	0.016	0.090
		Q_{NET}^{opt}	1.568 (0.0018)	SE	0.075	0.105	0.093	0.014	0.082
		$Q_{\gamma=0}^{opt}$	1.596 (0.0021)	CP	0.953	0.944	0.955	0.903	0.038
		Q^{opt}	1.725 (0.0016)						

Note. PCD_{NET}: average percentage of making correct decisions by the network-based decision rule; PCD _{$\gamma=0$} : average percentage of making correct decisions when ignoring the network effect; Q_{NET}^{opt} : average estimated value using the network-based optimal decision; $Q_{\gamma=0}^{opt}$: average estimated value of the optimal treatment when ignoring network effect; Q^{opt} : average value using the true optimal regime; Est.: mean of estimators; SD: standard deviation of estimators; SE: mean of estimated standard errors; CP: empirical coverage probability of 95% confidence intervals; corresponding SD is in parentheses.

TABLE 3

Q-learning results for mobile game data

PCD	Est.	Q	Est.	Parameters	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
PCD _{NET}	0.986	Q_{NET}^{opt}	3.011	Est.	0.939	0.974	-1.017	1.029	0.514	-0.951	0.076	0.099	0.232
PCD _{$\gamma=0$}	0.906	$Q_{\gamma=0}^{opt}$	2.924	SE	0.064	0.063	0.070	0.070	0.061	0.081	0.016	0.008	0.030
		Q^{opt}	3.013										

Note. PCD_{NET}: percentage of making correct decisions by the network-based decision rule; PCD _{$\gamma=0$} : percentage of making correct decisions when ignoring network effect; Q_{NET}^{opt} : estimated value using the network-based optimal decision; $Q_{\gamma=0}^{opt}$: estimated value of the optimal treatment when ignoring network effect; Q^{opt} : value using the true optimal regime; Est.: estimators; SE: estimated standard errors.

TABLE 4

A-learning results for mobile game data

PCD	Est.	Q	Est.	Parameters	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
(a) Baseline correct, propensity score correct												
PCD _{NET}	0.991	Q_{NET}^{opt}	3.087	Est.	0.912	0.975	-1.010	1.011	0.474	-0.964	0.106	0.213
PCD _{$\gamma=0$}	0.888	$Q_{\gamma=0}^{opt}$	2.976	SE	0.396	0.109	0.114	0.073	0.154	0.079	0.048	0.051
		Q^{opt}	3.087									
(b) Baseline correct, propensity score incorrect												
PCD _{NET}	0.995	Q_{NET}^{opt}	3.087	Est.	0.896	0.977	-1.002	1.013	0.466	-0.967	0.111	0.230
PCD _{$\gamma=0$}	0.892	$Q_{\gamma=0}^{opt}$	2.980	SE	0.379	0.100	0.077	0.075	0.147	0.088	0.060	0.043
		Q^{opt}	3.087									
(c) Baseline incorrect, propensity score correct												
PCD _{NET}	0.989	Q_{NET}^{opt}	3.086	Est.	0.764	1.057	-1.090	1.009	0.504	-0.973	0.117	0.232
PCD _{$\gamma=0$}	0.893	$Q_{\gamma=0}^{opt}$	2.980	SE	0.543	0.156	0.118	0.078	0.203	0.116	0.068	0.082
		Q^{opt}	3.087									

Note. PCD_{NET}: percentage of making correct decisions by the network-based decision rule; PCD _{$\gamma=0$} : percentage of making correct decisions when ignoring network effect; Q_{NET}^{opt} : estimated value using the network-based optimal decision; $Q_{\gamma=0}^{opt}$: estimated value of the optimal treatment when ignoring network effect; Q^{opt} : value using the true optimal regime; Est.: estimators; SE: estimated standard errors.