Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

Modeling and Estimation for Optimal Treatment Decision with Interference

Lin Su[a], Wenbin Lu[a]*, Rui Song[a]

Received 00 Month 2012; Accepted 00 Month 2012

In many network-based intervention studies, treatment applied on an individual or his/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 regime 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- and A-learning methods are derived. We show that the optimal treatment regime 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 are illustrated by extensive simulation studies and an application to a mobile game network data. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: A-learning; interference; network; optimal treatment regime; Q-learning

1. Introduction

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

The existence of interference introduces challenges to the 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 & Airoldi (2015) and Eckles et al. (2017). In the area of causal inference for social network data, a common assumption is called partial interference, which basically states that interference exists in partitioned small groups but not between different groups. Under this framework, different types of causal inference have been

[a]**Department of Statistics, North Carolina State University, Raleigh, NC 27695**
*__Email: lu@stat.ncsu.edu__

Prepared using staauth.cls [Version: 2012/05/12 v1.00]

studied by many authors, for example, Hudgens & Halloran (2008), Tchetgen & VanderWeele (2012), Aronow & Samii (2013), Liu & Hudgens (2014), Liu et al. (2016) and Sussman & Airoldi (2017).

In this paper, we consider the derivation of the optimal individualized treatment regime 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 neighborhood. 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 non-ignorable side effect for young children to get the vaccine, but as long as other family members are vaccinated, they are well protected. Therefore, based on 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, i.e., child, caregiver and community. All the participated 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/her friends. In this situation, a careful choice of intervention can not only benefit individual but his/her friends indirectly as well.

Deriving the optimal individualized treatment regime with the consideration of interference is interesting and challenging, for which the existing models for deriving the optimal treatment regime cannot be applied directly. We propose a new network-based regression model in this paper, which includes the interference effects when modeling 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, i.e., there is no interference effects involved in the decision rule. This is appealing since 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 results in an optimal decision rule. A-learning (Robins, 2004; Murphy, 2003), in contrast, is a semi-parametric method, which derives from a model that directly describes the difference between treatments, with the baseline remaining unspecified. For illustration purpose, 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- 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 regime 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 Section 4 and Section 5 respectively in order to illustrate the proposed methods. A conclusion is followed 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 $\mathcal{I}_i = \{i_j : i_j \sim i \text{ and } j = 1, 2, \ldots, n_i\}$, where n_i is the number of friends

of subject i. We consider two possible treatment assignments, denoted by a random variable $A_i \in \{0,1\}$, while a_i represents the observed treatment assignment. Let $\boldsymbol{S}_i$ be the vector of treatment assignments for subjects in the interference set of subject i, i.e., $\boldsymbol{S}_i = (A_{i_1}, A_{i_2}, \ldots, A_{i_{n_i}})$, and $\boldsymbol{s}_i$ is the corresponding observed treatment assignments $(a_{i_1}, a_{i_2}, \ldots, a_{i_{n_i}})$. Let $\boldsymbol{x}_i$ denote a p-dimensional vector of covariates for subject i. The covariates set of the interference set of subject i is defined as $\mathcal{X}_i = \{\boldsymbol{x}_{i_j} : i_j \in \mathcal{I}_i\}$. Moreover, let $Y_i^*(a_i, \boldsymbol{s}_i)$ and $Y_i(a_i, \boldsymbol{s}_i)$ denote the potential and observed outcomes, respectively. Note that the outcome for subject i also depends on the treatments assigned to his/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 $Y_i(a_i, \boldsymbol{s}_i) = I(A_i = a_i, \boldsymbol{S}_i = \boldsymbol{s}_i)Y_i^*(a_i, \boldsymbol{s}_i)$, and $Y_i^*(a_i, \boldsymbol{s}_i)$ is independent of $(A_i, \boldsymbol{S}_i)$ conditional on $\boldsymbol{x}_i$ and $\mathcal{X}_i$. Let $g(\boldsymbol{x}_i, \mathcal{X}_i)$ denote the optimal treatment rule for subject i. For now, we assume that it is a function of not only $\boldsymbol{x}_i$ but also $\mathcal{X}_i$ 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/her friends. We will show later that under our proposed model, the optimal treatment regime for subject i is a function of $\boldsymbol{x}_i$ only. Let $\mathcal{G}_i = \{g(\boldsymbol{x}_{i_j}, \mathcal{X}_{i_j}) : i_j \in \mathcal{I}_i\}$, 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

$$\frac{1}{n}\sum_{i=1}^{n} E\left\{Y_i^*\left(g(\boldsymbol{x}_i, \mathcal{X}_i), \mathcal{G}_i\right) | \boldsymbol{x}_i, \mathcal{X}_i\right\}, \tag{1}$$

which is equivalent to maximizing $\frac{1}{n}\sum_{i=1}^{n} E\left\{Y_i(a_i, \boldsymbol{s}_i) | \boldsymbol{x}_i, \mathcal{X}_i, a_i = g(\boldsymbol{x}_i, \mathcal{X}_i), \boldsymbol{s}_i = \mathcal{G}_i\right\}$ under the assumptions mentioned above.

In order to characterize the influence of interference, we propose the following network-based regression model for $Y_i(a_i, \boldsymbol{s}_i)$:

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

where ϵ_i is a random error term with mean 0, $\boldsymbol{W} = (W_{ij}) \in \{0,1\}^{n\times n}$ is the adjacency matrix with $W_{ij} = 1$ if $i \sim j$ and $W_{ij} = 0$ otherwise. $\Theta = (\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\theta}, \gamma_1, \gamma_2, \gamma_3)^T$ are the parameters involved. Note that $\gamma_1, \gamma_2, \gamma_3$ quantify the dependence of network. Based on our model, the response of an individual is not only related to his/her own covariates or treatment assignment, but also his/her friends' covariates and treatment assignments. Note that $W_{ii} = 0$ for $\forall i = 1, 2, \ldots, n$, with some simple algebra, (1) is equivalent to

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

The optimal treatment rule for subject i is thus given by

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

where $I(\cdot)$ is the indicator function. From (3), it is obvious to see that the optimal decision rule of subject i only depends on his/her own characteristics, 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, i.e.,

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

In order to find the solutions, we first fix γ_1, γ_2, and γ_3 at some initial values, and get the ordinary least square (OLS) estimates for α, $\boldsymbol{\beta}$, η and $\boldsymbol{\theta}$. Then, we update γ_1, γ_2, and γ_3 by solving OLS after fixing α, $\boldsymbol{\beta}$, η and $\boldsymbol{\theta}$ at the previous iteration, alternate between these two OLS problems till convergence. After getting the parameter estimates, the estimated optimal treatment $\hat{g}^{opt}(\boldsymbol{x}_i)$ is obtained by plugging in the estimates into (3). The asymptotic distribution of the estimators is given in Theorem 1.

Theorem 1
Let $\Theta = (\alpha, \boldsymbol{\beta}, \eta, \boldsymbol{\theta}, \gamma_1, \gamma_2, \gamma_3)^T$. Suppose $Y_1, Y_2, \ldots, Y_n$ are samples from model (2). Let Θ_0 denote the underlying true values of parameters, which is the unique solution to the equation $\sum_{i=1}^{n}[\boldsymbol{\phi}(Y_i;\Theta_0)] = 0$. Then, we have

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

where the definition of $\boldsymbol{\phi}(Y_i;\Theta_0)$, $A(\Theta)$ and $B(\Theta)$ can be found in Appendix A.

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 semi-parametric A-learning model:

$$Y_i(a_i, \boldsymbol{s}_i) = \mu(\boldsymbol{x}_i, \mathcal{X}_i) + \eta a_i + \gamma_2\sum_{j\neq i}^{n} W_{ij}\eta a_j + a_i\boldsymbol{\theta}'\boldsymbol{x}_i + \gamma_3\sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j + \epsilon_i,$$

where $\mu(\boldsymbol{x}_i, \mathcal{X}_i)$ is an unspecified baseline function. Note that $\mu(\boldsymbol{x}_i, \mathcal{X}_i)$ can be a function of either $\boldsymbol{x}_i$ itself or both $\boldsymbol{x}_i$ and $\mathcal{X}_i$. By the same technique of switching the double summations, the optimal decision rule is the same as (3). Let $\Omega = (\eta, \boldsymbol{\theta}, \gamma_2, \gamma_3)^T$, which is the vector of the parameters involved in the decision rule. Inspired by the doubly robust estimating equation proposed by Robins et al. (1994), we propose the following estimating equation for Ω:

$$\sum_{i=1}^{n}\begin{pmatrix} (a_i - \pi_i) + \gamma_2\sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j) \\ (a_i - \pi_i)\boldsymbol{x}_i + \gamma_3\sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j)\boldsymbol{x}_j \\ \sum_{j\neq i}^{n} W_{ij}\eta(a_j - \pi_j) \\ \sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j)\boldsymbol{\theta}'\boldsymbol{x}_j \end{pmatrix} \times \left[Y_i - \mu(\boldsymbol{x}_i, \mathcal{X}_i) - \eta a_i - \gamma_2\sum_{j\neq i}^{n} W_{ij}\eta a_j - a_i\boldsymbol{\theta}'\boldsymbol{x}_i - \gamma_3\sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j\right] = 0, \quad (4)$$

where π_i is the propensity score for subject i, which can be a function of $\boldsymbol{x}_i$ and $\mathcal{X}_i$. When at least one of π_i and $\mu(\boldsymbol{x}_i, \mathcal{X}_i)$ is correctly specified, (4) is an unbiased estimating equation for Ω. This can be established by using the

iterated conditional expectation, firstly conditioning on $\boldsymbol{x}_i$, a_i, $\mathcal{X}_i$ and $\boldsymbol{s}_i$, and then conditioning on $\boldsymbol{x}_i$ and $\mathcal{X}_i$. In the following discussion, we posit a linear model for the baseline function as an example, i.e $\mu(\boldsymbol{x}_i, \mathcal{X}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_i$, but derivation based on other parametric models is similar. Specifically, the estimating equation for $(\alpha, \boldsymbol{\beta})^T$ is

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ \boldsymbol{x}_i \end{pmatrix} \left[Y_i - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_i - \eta a_i - \gamma_2 \sum_{j \neq i}^{n} W_{ij}\eta a_j - a_i\boldsymbol{\theta}'\boldsymbol{x}_i - \gamma_3 \sum_{j \neq i}^{n} W_{ij} a_j \boldsymbol{\theta}'\boldsymbol{x}_j \right] = 0. \tag{5}$$

3.2. Model fitting

To improve the stability of solving the estimating equations (4) and (5), we consider to solve the following minimization problem instead:

$$\min \sum_{i=1}^{n} \Bigg[Y_i - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_i - f(\boldsymbol{x}_i;\Omega) - \eta(a_i - \pi_i) - \gamma_2 \sum_{j \neq i}^{n} W_{ij}\eta(a_j - \pi_j) - \boldsymbol{\theta}'\boldsymbol{x}_i(a_i - \pi_i) - \gamma_3 \sum_{j \neq i}^{n} W_{ij}(a_j - \pi_j)\boldsymbol{\theta}'\boldsymbol{x}_j \Bigg]^2,$$

where $f(\boldsymbol{x}_i;\Omega) = \eta\pi_i + \gamma_2 \sum_{j \neq i}^{n} W_{ij}\eta\pi_j + \boldsymbol{\theta}'\boldsymbol{x}_i\pi_i + \gamma_3 \sum_{j \neq i}^{n} W_{ij}\pi_j\boldsymbol{\theta}'\boldsymbol{x}_j$. The model for propensity score is firstly fitted and we plug in the predicted $\hat{\pi}_i$ in the estimating equation in order to get the estimation for other parameters. Note that when $f(\boldsymbol{x}_i;\Omega)$ is fixed on the true parameters Ω_0, minimizing the above equation is equivalent to solving the estimating equations (4) and (5).

In order to get the estimation for α, $\boldsymbol{\beta}$ and Ω, we fix $f(\boldsymbol{x}_i;\Omega)$ at some initial value $f(\boldsymbol{x}_i;\Omega^{(0)})$. Then we fix γ_2 and γ_3 at some initial value as well and update all other parameters by OLS in the same way as in the Q-learning. Next, we update the parameters in $f(\boldsymbol{x}_i;\Omega)$ using the estimates from the previous step, and then solve γ_2, γ_3 after fixing other parameters. $f(\boldsymbol{x}_i;\Omega)$ is updated again by the current estimates of γ_2, γ_3. All the steps are repeated till convergence. The convergence criterion requires the absolute difference between two iterations for all the parameters to be within 10^{-3}. Based on our numerical studies, this algorithm usually converges after about 20 iterations. Theorem 2 shows the asymptotic distribution of $\Omega = (\eta, \boldsymbol{\theta}, \gamma_2, \gamma_3)^T$.

Theorem 2
Let $\Omega = (\eta, \boldsymbol{\theta}, \gamma_2, \gamma_3)^T$ denote the parameters in the decision rule, $\boldsymbol{\delta}$ be the parameters in the baseline function, and Φ be the parameters in the propensity score model. Under the assumption that either the baseline function or the propensity score model is correctly specified, (4) is an unbiased estimating equation of the underlying true Ω_0. Then, we have

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

where $\boldsymbol{\delta}^*$ and Φ^* are the corresponding population parameters of the posited baseline and propensity score models. The definition of $A(\Omega_0, \boldsymbol{\delta}^*, \Phi^*)$ and $B(\Omega_0, \boldsymbol{\delta}^*, \Phi^*)$ can be found 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 et al., 1983), in which the

probability that node i and node j are connected is modeled as $\boldsymbol{P}_{C_iC_j}$, where C_i and C_j are the communities to which node i and j belong respectively, and $\boldsymbol{P}$ 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 $n = 1000$ or 2000. When $n = 1000$, the numbers of subjects in each community are $(250, 250, 200, 200, 100)$, and when $n = 2000$, the community sizes are $(500, 500, 400, 400, 200)$. We also try two different densities of the network. The dense network is set as $\boldsymbol{P}_{11} = \boldsymbol{P}_{33} = 0.05$, $\boldsymbol{P}_{22} = \boldsymbol{P}_{44} = \boldsymbol{P}_{55} = 0.1$, and $\boldsymbol{P}_{C_iC_j} = 10^{-4}$ 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 $x_{i1} \sim$ Bernoulli(0.5), and $x_{i2} \sim$ Uniform$(-1, 1)$. Treatment A_i is randomly assigned to 0/1 with equal probability, i.e., $\pi_i = 0.5$. The parameters are set as $\alpha = 0.5$, $\boldsymbol{\beta} = (-1, 1)^T$, $\eta = 1$, $\boldsymbol{\theta} = (-0.5, 0.5)^T$, $\gamma_1 = 0.05$, $\gamma_3 = 0.2$. Variance of ϵ_i is 1, and γ_2 can be 0.1 or 0. Simulation results based on 1000 replicates are shown in Table 1. The average percentage of making correct decisions (PCD) over 1000 simulations based on the estimated optimal treatment regime (denoted by PCD_{NET}) is reported. We also report the PCD based on the estimated optimal treatment regime ignoring the network structure, that is, assuming $\boldsymbol{\gamma} = (\gamma_1 = \gamma_2 = \gamma_3)^T = 0$ when fitting the model (denoted by $\text{PCD}_{\boldsymbol{\gamma}=0}$). We can see that PCD_{NET} is always much higher than $\text{PCD}_{\boldsymbol{\gamma}=0}$ as expected. Moreover, we report the average value functions of model (2) based on the estimated optimal treatment regime considering the network structure, the estimated optimal treatment regime ignoring the network structure, i.e. $\boldsymbol{\gamma} = 0$, and the true optimal treatment regime, denoted by Q^{opt}_{NET}, $Q^{opt}_{\boldsymbol{\gamma}=0}$ and Q^{opt}. It can be seen that Q^{opt}_{NET} is always larger than $Q^{opt}_{\boldsymbol{\gamma}=0}$, and slightly smaller but close to the true optimal value Q^{opt}. In addition, we show the performance of the parameter estimation. 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 $n = 2000$ with community sizes $(500, 500, 400, 400, 200)$. The network density is the same as the sparse network described above, that is, $\boldsymbol{P}_{11} = \boldsymbol{P}_{33} = 0.025$, $\boldsymbol{P}_{22} = \boldsymbol{P}_{44} = \boldsymbol{P}_{55} = 0.05$, and $\boldsymbol{P}_{C_iC_j} = 0.5 \times 10^{-4}$ for $C_i \neq C_j$. 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 $\mu(\boldsymbol{x}_i, \mathcal{X}_i)$ and the propensity score π_i. In particular, we always fit a standard linear model $h(\boldsymbol{x}_i, \mathcal{X}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_i$ for the baseline, which is correctly specified when $\mu(\boldsymbol{x}_i, \mathcal{X}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_i$, but misspecified when $\mu(\boldsymbol{x}_i, \mathcal{X}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_i + \gamma_1 \sum_{j \neq i}^{n} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j$ as in the Q-learning model. For the propensity score, A_i is generated by a logistic model, i.e., $\text{logit}(\pi_i) = \psi + \boldsymbol{\phi}'\boldsymbol{x}_i$, where $\psi = 0.5$ and $\boldsymbol{\phi} = (-1, 1)^T$. The same logistic model is fitted when it is correctly specified, while 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 $h(\boldsymbol{x}_i, \mathcal{X}_i)$ or π_i is correctly specified, PCD_{NET} of the estimated optimal treatment regime taking into account the network interference is much higher than $\text{PCD}_{\boldsymbol{\gamma}=0}$, which ignores the network interference. Similarly, the average optimal value Q^{opt}_{NET} based on the estimated optimal treatment regime with the network interference is higher than $Q^{opt}_{\boldsymbol{\gamma}=0}$ that ignores the network effect, and close to the true optimal value. When both of the baseline and the propensity score are incorrect, the PCD and optimal value of the network-based optimal treatment regime 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 Tencent QQ Game Data

In this section, we use the covariates and network structure collected in the Tencent QQ game data to illustrate our proposed methods. Tencent QQ is a popular instant chatting software in China. This data set consists of $n = 961$ users of Tencent QQ and their friendship network as showed 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 QQ level in the following analysis, where QQ level indicates how active a user is on Tencent QQ. All covariates are centered and scaled firstly. This data set was collected to study the propagation of a particular QQ game. Because of confidentiality, the name of this game is excluded here. The game sends invitations to players' friends through QQ 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, while 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 ith 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. Parameters we use to generate the responses are chosen as $\alpha = 1$, $\boldsymbol{\beta} = (1, -1)^T$, $\eta = 1$, $\boldsymbol{\theta} = (0.5, -1)^T$, $\gamma_1 = 0.1$, $\gamma_2 = 0.1$, $\gamma_3 = 0.2$, 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 regime 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 regime (left panel) and the assignment based on the estimated optimal treatment regime obtained by Q-learning (right panel) in Figure 1. It can be seen that the estimated optimal treatment regime and the true optimal treatment regime 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 QQ game data. As in simulations, we always fit a simple linear model for the baseline, i.e. $h(\boldsymbol{x}_i, \mathcal{X}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_i$. 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 regime 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 regime in the presence of interference. One advantage of our proposed model is that the true optimal treatment regime only depends on one's own covariates and number of connected nodes in the network. Such property makes its implementation in reality become practical. Approaches for parameter and variance estimation under the framework of both Q- and A-learning are studied. 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. In addition, the proposed model assumed a linear interaction between covariates and treatment, which can also be relaxed. These are interesting topics that need further investigation.

A. Proof of Theorem 1

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

$$\begin{aligned}
&L(\boldsymbol{Y};\Theta)\\
&=\sum_{i=1}^{n} L(Y_i;\Theta)\\
&=\sum_{i=1}^{n}\left[Y_i-\left(\alpha+\boldsymbol{\beta}'\boldsymbol{x}_i+\gamma_1\sum_{j\neq i}^{n}W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j+\eta a_i+\gamma_2\sum_{j\neq i}^{n}W_{ij}\eta a_j+a_i\boldsymbol{\theta}'\boldsymbol{x}_i+\gamma_3\sum_{j\neq i}^{n}W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j\right)\right]^2,
\end{aligned}$$

where $\Theta=(\alpha,\boldsymbol{\beta},\eta,\boldsymbol{\theta},\gamma_1,\gamma_2,\gamma_3)^T$. Therefore, the estimating equation is $\sum_{i=1}^{n}\boldsymbol{\phi}(Y_i;\hat{\Theta})=0$, where $\boldsymbol{\phi}=(\phi_1,\phi_2,\ldots,\phi_7)^T$,

$$\begin{aligned}
\phi_1(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\alpha}=Y_i-m_i(\Theta),\\
\phi_2(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\boldsymbol{\beta}}=(Y_i-m_i(\Theta))(\boldsymbol{x}_i+\gamma_1\sum_{j\neq i}^{n}W_{ij}\boldsymbol{x}_j),\\
\phi_3(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\eta}=(Y_i-m_i(\Theta))(a_i+\gamma_2\sum_{j\neq i}^{n}W_{ij}a_j),\\
\phi_4(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\boldsymbol{\theta}}=(Y_i-m_i(\Theta))(a_i\boldsymbol{x}_i+\gamma_3\sum_{j\neq i}^{n}W_{ij}a_j\boldsymbol{x}_j),\\
\phi_5(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\gamma_1}=(Y_i-m_i(\Theta))(\sum_{j\neq i}^{n}W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j),\\
\phi_6(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\gamma_2}=(Y_i-m_i(\Theta))(\sum_{j\neq i}^{n}W_{ij}\eta a_j),\\
\phi_7(Y_i;\Theta)&=-\frac{1}{2}\cdot\frac{\partial L(Y_i;\Theta)}{\partial\gamma_3}=(Y_i-m_i(\Theta))(\sum_{j\neq i}^{n}W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j),
\end{aligned}$$

and $m_i(\Theta)=\alpha+\boldsymbol{\beta}'\boldsymbol{x}_i+\gamma_1\sum_{j\neq i}^{n}W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j+\eta a_i+\gamma_2\sum_{j\neq i}^{n}W_{ij}\eta a_j+a_i\boldsymbol{\theta}'\boldsymbol{x}_i+\gamma_3\sum_{j\neq i}^{n}W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j$.

Let Θ_0 denote the true values of parameters. By Taylor expansion,

$$\sqrt{n}(\hat{\Theta}-\Theta_0)=\left[-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\boldsymbol{\phi}(Y_i;\Theta)}{\partial\Theta^T}\Big|_{\Theta=\Theta_0}\right]^{-1}\cdot\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\phi}(Y_i;\Theta_0)\right]+\sqrt{n}R_n^*.$$

As discussed in Stefanski & Boos (2002), $\sqrt{n}R_n^*\xrightarrow{p}0$. Conditioning on the network $\boldsymbol{W}$ and all covariates $\boldsymbol{x}_i$, and assuming that $\sum_{j\neq i}^{n}W_{ij}$ is bounded for all i, it can be shown that $\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\phi}(Y_i;\Theta_0)\right]\xrightarrow{d}N(0,B(\Theta_0))$ by the multiplier central limit theorem (van der Vaart & Wellner, 1996), where $B(\Theta_0)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\phi}(Y_i;\Theta)\boldsymbol{\phi}^T(Y_i;\Theta)\right]$. Let $A(\Theta_0)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\partial\boldsymbol{\phi}(Y_i;\Theta)}{\partial\Theta}\Big|_{\Theta=\Theta_0}\right]$, then the asymptotic distribution of $\hat{\Theta}$ is proved to follow Theorem 1.

In order to obtain the estimated $\hat{\text{cov}}(\hat{\Theta})$, we can define

$$\boldsymbol{d}_i(\Theta) = \begin{pmatrix} 1 \\ \boldsymbol{x}_i + \gamma_1 \sum_{j\neq i}^{n} W_{ij}\boldsymbol{x}_j \\ a_i + \gamma_2 \sum_{j\neq i}^{n} W_{ij}a_j \\ a_i\boldsymbol{x}_i + \gamma_3 \sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{x}_j \\ \sum_{j\neq i}^{n} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j \\ \sum_{j\neq i}^{n} W_{ij}\eta a_j \\ \sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j \end{pmatrix},$$

then $\boldsymbol{\phi}(Y_i;\Theta) = \boldsymbol{d}_i(\Theta)(Y_i - m_i(\Theta))$. The estimated asymptotic variance of parameters can be obtained by $\hat{\text{cov}}(\hat{\Theta}) = \frac{1}{n}A_n^{-1}(\hat{\Theta})B_n(\hat{\Theta})[A_n^{-1}(\hat{\Theta})]^T$, where

$$\begin{aligned} B_n(\Theta) =& \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\phi}(Y_i;\Theta)\boldsymbol{\phi}^T(Y_i;\Theta)\right] = \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{d}_i(\Theta)\boldsymbol{d}_i^T(\Theta)(Y_i - m_i(\Theta))^2\right], \\ A_n(\Theta) =& \frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\partial\boldsymbol{\phi}(Y_i;\Theta)}{\partial\Theta^T}\right] = \frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\partial\boldsymbol{d}_i(\Theta)}{\partial\Theta^T}(Y_i - m_i(\Theta)) + \boldsymbol{d}_i(\Theta)\boldsymbol{d}_i^T(\Theta)\right], \end{aligned}$$

and

$$\frac{\partial\boldsymbol{d}_i(\Theta)}{\partial\Theta^T} = \begin{pmatrix} 0 & 0_p^T & 0 & 0_p^T & 0 & 0 & 0 \\ 0_p & 0_{p\times p} & 0_p & 0_{p\times p} & \sum_{j\neq i}^{n} W_{ij}\boldsymbol{x}_j & 0_p & 0_p \\ 0 & 0_p^T & 0 & 0_p^T & 0 & \sum_{j\neq i}^{n} W_{ij}a_j & 0 \\ 0_p & 0_{p\times p} & 0_p & 0_{p\times p} & 0_p & 0_p & \sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{x}_j \\ 0 & \sum_{j\neq i}^{n} W_{ij}\boldsymbol{x}_j^T & 0 & 0_p^T & 0 & 0 & 0 \\ 0 & 0_p^T & \sum_{j\neq i}^{n} W_{ij}a_j & 0_p^T & 0 & 0 & 0 \\ 0 & 0_p^T & 0 & \sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{x}_j^T & 0 & 0 & 0 \end{pmatrix},$$

where p is the dimension of covariates, that is, the dimension of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. 0_p is a p-vector of 0 and $0_{p\times p}$ is a $p\times p$ matrix of 0.

B. Proof of Theorem 2

Estimating equations (4) and (5) can be written as $U_n(\Omega,\boldsymbol{\delta},\Phi) = \sum_{i=1}^{n}\boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi) = 0$, where

$$\begin{aligned} \boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi) = & \begin{pmatrix} 1 \\ \boldsymbol{x}_i \\ (a_i - \pi_i) + \gamma_2\sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j) \\ (a_i - \pi_i)\boldsymbol{x}_i + \gamma_3\sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j)\boldsymbol{x}_j \\ \sum_{j\neq i}^{n} W_{ij}\eta(a_j - \pi_j) \\ \sum_{j\neq i}^{n} W_{ij}(a_j - \pi_j)\boldsymbol{\theta}'\boldsymbol{x}_j \end{pmatrix} \times \\ & \left[Y_i - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_i - \eta a_i - \gamma_2\sum_{j\neq i}^{n} W_{ij}\eta a_j - a_i\boldsymbol{\theta}'\boldsymbol{x}_i - \gamma_3\sum_{j\neq i}^{n} W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j\right]. \end{aligned}$$

By Taylor expansion,

$$0 = \frac{1}{n}U_n(\hat{\Omega},\hat{\boldsymbol{\delta}},\hat{\Phi}) = \frac{1}{n}U_n(\Omega_0,\boldsymbol{\delta}^*,\Phi^*) + \frac{1}{n}\frac{\partial U_n(\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Omega^T)}\Big|_{\Omega=\Omega_0,\boldsymbol{\delta}=\boldsymbol{\delta}^*,\Phi=\Phi^*}\begin{pmatrix}\hat{\Omega}-\Omega_0\\ \hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^*\end{pmatrix} + \frac{1}{n}H(\Omega_0,\boldsymbol{\delta}^*,\Phi^*)I_n^{-1}(\Phi^*)\sum_{i=1}^n S(Y_i;\Phi^*) + o_p(1),$$

where $H(\Omega,\boldsymbol{\delta},\Phi) = \frac{\partial U_n(\Omega,\boldsymbol{\delta},\Phi)}{\partial\Phi^T}$, $S(Y_i;\Phi)$ is the score function for the propensity score, and $I_n(\Phi)$ is the corresponding information matrix of Φ. Therefore,

$$\sqrt{n}\begin{pmatrix}\hat{\Omega}-\Omega_0\\ \hat{\boldsymbol{\delta}}-\boldsymbol{\delta}^*\end{pmatrix} = \left[-\frac{1}{n}\frac{\partial U_n(\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Omega^T)}\Big|_{\Omega=\Omega_0,\boldsymbol{\delta}=\boldsymbol{\delta}^*,\Phi=\Phi^*}\right]^{-1}\cdot\sqrt{n}\left[\frac{1}{n}U_n(\Omega_0,\boldsymbol{\delta}^*,\Phi^*) + \frac{1}{n}H(\Omega_0,\boldsymbol{\delta}^*,\Phi^*)I_n^{-1}(\Phi^*)\sum_{i=1}^n S(Y_i;\Phi^*)\right] + \sqrt{n}R_n^*.$$

As discussed in (Stefanski & Boos, 2002), $\sqrt{n}R_n^* \xrightarrow{p} 0$, and by the multiplier central limit theorem (van der Vaart & Wellner, 1996) assuming that $\sum_{j\neq i}^n W_{ij}$ is bounded and the limits exist, we have

$$\sqrt{n}\left[\frac{1}{n}U_n(\Omega_0,\boldsymbol{\delta}^*,\Phi^*) + \frac{1}{n}H(\Omega_0,\boldsymbol{\delta}^*,\Phi^*)I_n^{-1}(\Phi^*)\sum_{i=1}^n S(Y_i;\Phi^*)\right] \xrightarrow{d} N(0,B(\Omega_0,\boldsymbol{\delta}^*,\Phi^*)),$$

where $B(\Omega_0,\boldsymbol{\delta}^*,\Phi^*) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left[\boldsymbol{\psi}^*(Y_i;\Omega_0,\boldsymbol{\delta}^*,\Phi^*)\boldsymbol{\psi}^{*T}(Y_i;\Omega_0,\boldsymbol{\delta}^*,\Phi^*)\right]$, and $\boldsymbol{\psi}^*(Y_i;\Omega,\boldsymbol{\delta},\Phi) = \boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi) + H(\Omega,\boldsymbol{\delta},\Phi)I_n^{-1}(\Phi)S(Y_i;\Phi)$. Define $A(\Omega,\boldsymbol{\delta},\Phi) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n[-\frac{\partial\boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Theta^T)}]$ and assume its existence. Therefore, the asymptotic distribution of $\hat{\Omega}$ is as described in Theorem 2. The covariance can be estimated as $\hat{\text{cov}}(\hat{\boldsymbol{\delta}},\hat{\Omega}) = \frac{1}{n}A_n^{-1}(\hat{\Omega},\hat{\boldsymbol{\delta}},\hat{\Phi})B_n(\hat{\Omega},\hat{\boldsymbol{\delta}},\hat{\Phi})[A_n^{-1}(\hat{\Omega},\hat{\boldsymbol{\delta}},\hat{\Phi})]^T$, where $A_n(\Omega,\boldsymbol{\delta},\Phi) = \frac{1}{n}\sum_{i=1}^n[-\frac{\partial\boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Theta^T)}]$ and $B_n(\Omega,\boldsymbol{\delta},\Phi) = \frac{1}{n}\sum_{i=1}^n[\boldsymbol{\psi}^*(Y_i;\Omega,\boldsymbol{\delta},\Phi)\boldsymbol{\psi}^{*T}(Y_i;\Omega,\boldsymbol{\delta},\Phi)]$. Details are as followings.

If we define $\boldsymbol{d}_i(\Omega,\boldsymbol{\delta},\Phi)$ and $\epsilon_i(\Omega,\boldsymbol{\delta})$ as

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

$$\epsilon_i(\Omega,\boldsymbol{\delta}) = Y_i - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_i - \eta a_i - \gamma_2\sum_{j\neq i}^n W_{ij}\eta a_j - a_i\boldsymbol{\theta}'\boldsymbol{x}_i - \gamma_3\sum_{j\neq i}^n W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j.$$

Then we have $\boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi) = \boldsymbol{d}_i(\Omega,\boldsymbol{\delta},\Phi)\cdot\epsilon_i(\Omega,\boldsymbol{\delta})$. Suppose Φ is a q-dimension vector, then

$$H(\Omega,\boldsymbol{\delta},\Phi) = \sum_{i=1}^n\begin{pmatrix}0_q^T\\ 0_q^T\\ -\frac{\partial\pi_i}{\partial\phi^T}-\gamma_2\sum_{j\neq i}^n W_{ij}\frac{\partial\pi_j}{\partial\phi^T}\\ -\boldsymbol{x}_i\frac{\partial\pi_i}{\partial\phi^T}-\gamma_3\sum_{j\neq i}^n W_{ij}\boldsymbol{x}_j\frac{\partial\pi_j}{\partial\phi^T}\\ -\sum_{i=1}^n W_{ij}\eta\frac{\partial\pi_j}{\partial\phi^T}\\ -\sum_{j\neq i}^n W_{ij}\boldsymbol{\theta}'\boldsymbol{x}_j\frac{\partial\pi_j}{\partial\phi^T}\end{pmatrix}\cdot\epsilon_i(\Omega,\boldsymbol{\delta}).$$

Besides, $\frac{\partial \boldsymbol{\psi}(Y_i;\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Theta^T)} = \frac{\partial \boldsymbol{d}_i(\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Theta^T)}\epsilon_i(\Omega,\boldsymbol{\delta}) + \boldsymbol{d}_i(\Omega,\boldsymbol{\delta},\Phi)\frac{\partial \epsilon_i(\Omega,\boldsymbol{\delta})}{\partial(\boldsymbol{\delta}^T,\Theta^T)}$, $\frac{\partial \epsilon_i(\Omega,\boldsymbol{\delta})}{\partial(\boldsymbol{\delta}^T,\Theta^T)} = -(1, \boldsymbol{x}_i^T, a_i + \gamma_2 \sum_{j\neq i}^n W_{ij}a_j, a_i\boldsymbol{x}_j^T + \gamma_3 \sum_{j\neq i}^n W_{ij}a_j\boldsymbol{x}_j^T, \sum_{j\neq i}^n \eta a_j, \sum_{j\neq i}^n W_{ij}a_j\boldsymbol{\theta}'\boldsymbol{x}_j)$, and

$$\frac{\partial d_i(\Omega,\boldsymbol{\delta},\Phi)}{\partial(\boldsymbol{\delta}^T,\Theta^T)} = \begin{pmatrix} 0 & 0_p^T & 0 & 0_p^T & 0 & 0 \\ 0_p & 0_{p\times p} & 0_p & 0_{p\times p} & 0_p & 0_p \\ 0 & 0_p^T & 0 & 0_p^T & \sum_{j\neq i}^n W_{ij}\Delta_j & 0 \\ 0_p & 0_{p\times p} & 0_p & 0_{p\times p} & 0 & \sum_{j\neq i}^n W_{ij}\Delta_j\boldsymbol{x}_j \\ 0 & 0_p^T & \sum_{j\neq i}^n W_{ij}\Delta_j & 0_p^T & 0 & 0 \\ 0 & 0_p^T & 0 & \sum_{j\neq i}^n W_{ij}\Delta_j\boldsymbol{x}_j^T & 0 & 0 \end{pmatrix},$$

where $\Delta_j = (a_j - \pi_j)$ and p is the dimension of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

References

Aronow, P & Samii, C (2013), 'Estimating average causal effects under general interference, with application to a social network experiment,' *arXiv preprint arXiv:1305.6156*.

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), pp. 295–298.

Eckles, D, Karrer, B & Ugander, J (2017), 'Design and analysis of experiments in networks: Reducing bias from interference,' *Journal of Causal Inference*, **5**(1).

Halloran, ME & Struchiner, CJ (1995), 'Causal inference in infectious diseases,' *Epidemiology*, pp. 142–151.

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*, p. 0143034316689589.

Holland, PW, Laskey, KB & Leinhardt, S (1983), 'Stochastic blockmodels: First steps,' *Social Networks*, **5**(2), pp. 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), pp. 901–910.

Hudgens, MG & Halloran, ME (2008), 'Toward causal inference with interference,' *Journal of the American Statistical Association*, **103**(482), pp. 832–842.

Liu, L, Hudgens, M & Becker-Dreps, S (2016), 'On inverse probability-weighted estimators in the presence of interference,' *Biometrika*, **103**(4), pp. 829–842.

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), pp. 288–301.

Murphy, SA (2003), 'Optimal dynamic treatment regimes,' *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(2), pp. 331–355.

Robins, JM (2004), 'Optimal structural nested models for optimal sequential decisions,' in *Proceedings of the second seattle Symposium in Biostatistics*, 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), pp. 846–866.

Rosenbaum, PR (2007), 'Interference between units in randomized experiments,' *Journal of the American Statistical Association*, **102**(477), pp. 191–200.

Rubin, DB (1978), 'Bayesian inference for causal effects: The role of randomization,' *The Annals of statistics*, pp. 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), pp. 1398–1407.

Stefanski, LA & Boos, DD (2002), 'The calculus of m-estimation,' *The American Statistician*, **56**(1), pp. 29–38.

Sussman, DL & Airoldi, 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), pp. 55–75.

van der Vaart, AW & Wellner, JA (1996), *Multiplier Central Limit Theorems*, Springer New York, New York, NY, pp. 176–189, doi:10.1007/978-1-4757-2545-2_21.

Watkins, CJ & Dayan, P (1992), 'Q-learning,' *Machine learning*, **8**(3-4), pp. 279–292.

Watkins, CJCH (1989), *Learning from delayed rewards*, Ph.D. thesis, King's College, Cambridge.

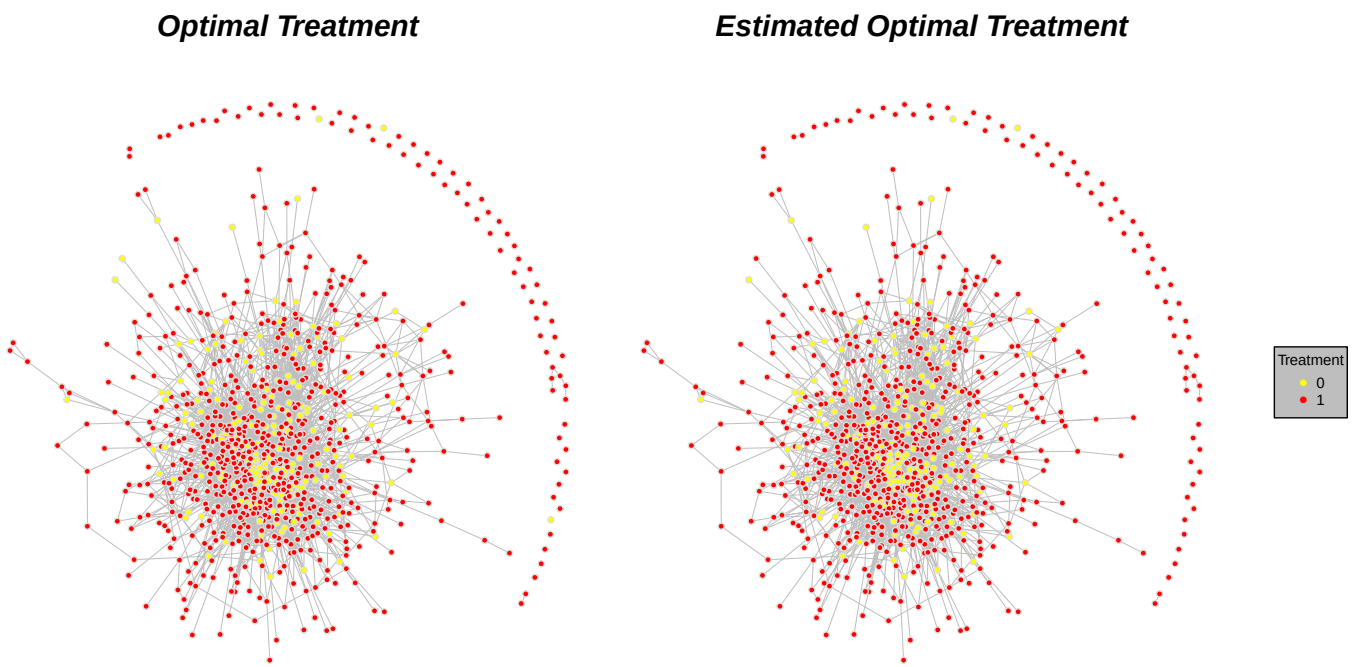

Figure 1. Plots of the assignment based on the true optimal treatment regime (left panel) and the assignment based on the estimated optimal treatment regime obtained by Q-learning (right panel). The PCD is 0.986 comparing the estimated optimal treatment regime with the true optimal regime.

Table 1. Q-learning simulation results.

(a) $n = 1000$

Settings				$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
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
$\gamma_2 = 0.1$	$PCD_{\boldsymbol{\gamma}=0}$	0.838 (0.0011)	SD	0.091	0.084	0.074	0.081	0.106	0.101	0.012	0.014	0.055
	Q^{opt}_{NET}	1.310 (0.0030)	SE	0.093	0.085	0.075	0.084	0.110	0.099	0.012	0.014	0.057
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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)										
dense	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
$\gamma_2 = 0$	$PCD_{\boldsymbol{\gamma}=0}$	0.535 (0.0012)	SD	0.094	0.083	0.074	0.080	0.104	0.099	0.012	0.013	0.051
	Q^{opt}_{NET}	0.221 (0.0024)	SE	0.092	0.083	0.074	0.082	0.105	0.096	0.012	0.013	0.052
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	-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)										
sparse	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
$\gamma_2 = 0.1$	$PCD_{\boldsymbol{\gamma}=0}$	0.876 (0.0011)	SD	0.084	0.085	0.076	0.083	0.113	0.105	0.017	0.019	0.067
	Q^{opt}_{NET}	1.014 (0.0021)	SE	0.085	0.087	0.076	0.086	0.117	0.103	0.017	0.019	0.070
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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)										
sparse	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
$\gamma_2 = 0$	$PCD_{\boldsymbol{\gamma}=0}$	0.675 (0.0012)	SD	0.086	0.085	0.076	0.083	0.112	0.104	0.017	0.019	0.065
	Q^{opt}_{NET}	0.402 (0.0018)	SE	0.085	0.086	0.076	0.085	0.114	0.102	0.017	0.018	0.067
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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)										

(b) n=2000

Settings				$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
dense	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
$\gamma_2 = 0.1$	$PCD_{\boldsymbol{\gamma}=0}$	0.806 (0.0009)	SD	0.068	0.060	0.051	0.059	0.072	0.064	0.006	0.008	0.032
	Q^{opt}_{NET}	1.919 (0.0033)	SE	0.070	0.059	0.052	0.058	0.073	0.067	0.006	0.008	0.033
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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)										
dense	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
$\gamma_2 = 0$	$PCD_{\boldsymbol{\gamma}=0}$	0.421 (0.0010)	SD	0.069	0.059	0.050	0.056	0.068	0.060	0.006	0.006	0.026
	Q^{opt}_{NET}	-0.031 (0.0024)	SE	0.069	0.055	0.050	0.055	0.066	0.062	0.006	0.006	0.026
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	-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)										
sparse	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
$\gamma_2 = 0.1$	$PCD_{\boldsymbol{\gamma}=0}$	0.833 (0.0008)	SD	0.063	0.061	0.052	0.059	0.077	0.067	0.008	0.009	0.037
	Q^{opt}_{NET}	1.311 (0.0020)	SE	0.065	0.060	0.052	0.059	0.077	0.070	0.008	0.010	0.037
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	1.183 (0.0025)	CP	0.955	0.947	0.947	0.944	0.940	0.948	0.971	0.961	0.962
	Q^{opt}	1.316 (0.0020)										
sparse	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
$\gamma_2 = 0$	$PCD_{\boldsymbol{\gamma}=0}$	0.530 (0.0009)	SD	0.066	0.061	0.051	0.058	0.075	0.066	0.008	0.009	0.034
	Q^{opt}_{NET}	0.232 (0.0016)	SE	0.065	0.058	0.052	0.058	0.074	0.068	0.008	0.009	0.034
	$Q^{opt}_{\boldsymbol{\gamma}=0}$	-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)										

PCD_{NET}, average percentage of making correct decisions by the network-based decision rule; $PCD_{\boldsymbol{\gamma}=0}$, average percentage of making correct decisions when ignoring network effect; Q^{opt}_{NET}, average estimated value using the network-based optimal decision; $Q^{opt}_{\boldsymbol{\gamma}=0}$, 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 the parenthesis.

Table 2. A-learning simulation results.

Baseline	PS				$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
Correct	Correct	PCD_{NET}	0.969 (0.0008)	Est.	1.000	-0.500	0.503	0.100	0.203
		$PCD_{\boldsymbol{\gamma}=0}$	0.837 (0.0010)	SD	0.064	0.087	0.078	0.017	0.047
		Q^{opt}_{NET}	1.717 (0.0016)	SE	0.065	0.088	0.080	0.017	0.046
		$Q^{opt}_{\boldsymbol{\gamma}=0}$	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
		$PCD_{\boldsymbol{\gamma}=0}$	0.837 (0.0010)	SD	0.062	0.084	0.073	0.011	0.059
		Q^{opt}_{NET}	1.699 (0.0019)	SE	0.064	0.083	0.075	0.011	0.059
		$Q^{opt}_{\boldsymbol{\gamma}=0}$	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
		$PCD_{\boldsymbol{\gamma}=0}$	0.837 (0.0010)	SD	0.068	0.095	0.084	0.021	0.053
		Q^{opt}_{NET}	1.712 (0.0017)	SE	0.067	0.091	0.083	0.018	0.050
		$Q^{opt}_{\boldsymbol{\gamma}=0}$	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
		$PCD_{\boldsymbol{\gamma}=0}$	0.837 (0.0010)	SD	0.072	0.102	0.092	0.016	0.090
		Q^{opt}_{NET}	1.568 (0.0018)	SE	0.075	0.105	0.093	0.014	0.082
		$Q^{opt}_{\boldsymbol{\gamma}=0}$	1.596 (0.0021)	CP	0.953	0.944	0.955	0.903	0.038
		Q^{opt}	1.725 (0.0016)						

PCD_{NET}, average percentage of making correct decisions by the network-based decision rule; $PCD_{\boldsymbol{\gamma}=0}$, average percentage of making correct decisions when ignoring network effect; Q^{opt}_{NET}, average estimated value using the network-based optimal decision; $Q^{opt}_{\boldsymbol{\gamma}=0}$, 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 the parenthesis.

Table 3. Q-learning results for Tencent QQ game data.

					$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
PCD_{NET}	0.986	Q^{opt}_{NET}	3.011	Est.	0.939	0.974	-1.017	1.029	0.514	-0.951	0.076	0.099	0.232
$PCD_{\boldsymbol{\gamma}=0}$	0.906	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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										

PCD_{NET}, percentage of making correct decisions by the network-based decision rule; $PCD_{\boldsymbol{\gamma}=0}$, percentage of making correct decisions when ignoring network effect; Q^{opt}_{NET}, estimated value using the network-based optimal decision; $Q^{opt}_{\boldsymbol{\gamma}=0}$, 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 Tencent QQ game data.

(a) Baseline correct, propensity score correct.

					$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
PCD_{NET}	0.991	Q^{opt}_{NET}	3.087	Est.	0.912	0.975	-1.010	1.011	0.474	-0.964	0.106	0.213
$PCD_{\boldsymbol{\gamma}=0}$	0.888	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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.

					$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
PCD_{NET}	0.995	Q^{opt}_{NET}	3.087	Est.	0.896	0.977	-1.002	1.013	0.466	-0.967	0.111	0.230
$PCD_{\boldsymbol{\gamma}=0}$	0.892	$Q^{opt}_{\boldsymbol{\gamma}=0}$	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.

					$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\eta}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
PCD_{NET}	0.989	Q^{opt}_{NET}	3.086	Est.	0.764	1.057	-1.090	1.009	0.504	-0.973	0.117	0.232
$PCD_{\boldsymbol{\gamma}=0}$	0.893	$Q^{opt}_{\boldsymbol{\gamma}=0}$	2.980	SE	0.543	0.156	0.118	0.078	0.203	0.116	0.068	0.082
		Q^{opt}	3.087									

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