Testing and Estimation of Social Network Dependence with Time to Event Data

Lin Su, Wenbin Lu, Rui Song and Danyang Huang*

Abstract

Nowadays, events are spread rapidly along social network. We are interested in whether people's responses to an event are affected by their friends' characteristics. For example, how soon will a person starts playing a game given that his/her friends like it? Studying social network dependence is an emerging research area. In this work, we propose a novel latent spatial autocorrelation Cox model to study social network dependence with time-to-event data. The proposed model introduces a latent indicator to characterize whether a person's survival time might be affected by his or her friends' features. We first propose a score-type test for detecting the existence of social network dependence. If it exists, we further develop an EM-type algorithm to estimate the model parameters. The performance of the proposed test and estimators are illustrated by simulation studies and an application to a time-to-event data set about playing a popular mobile game from one of the largest online social network platforms.

Keywords: Cox model; EM algorithm; Social network dependence; Time-to-event data.

*Lin Su is graduate student (E-mail: lsu3@ncsu.edu), Wenbin Lu is Professor (E-mail: lu@stat.ncsu.edu), and Rui Song is Associate Professor (rsong@ncsu.edu), Department of Statistics, North Carolina State University, Raleigh, NC 27695. Danyang Huang is Assistant Professor (E-mail: dy-huang89@126.com), School of Statistics, Remin University, Beijing, China 100872.

1 Introduction

With the development of internet, information propagates quickly along social network. People can easily share information, such as ideas, pictures, articles or videos, to a lot of friends through large social network platforms like Facebook, Twitter and QQ. Social network data have become popular and various models have been proposed for social network data, which include but are not limited to the stochastic block model (Wang and Wong 1987; Nowicki and Snijders 2001), exponential random graph model (Frank and Strauss 1986; Hunter et al. 2008) and latent space model (Hoff et al. 2002; Hoff 2003; Chang et al. 2018).

It has been widely studied that people are likely to be influenced by their friends on social network. Studying social dependence is an important question in social network analysis (Manski 1993; Shalizi and Thomas 2011; Huang et al. 2016; Zhu et al. 2017). There are mainly three types of social network dependence, namely contextual (exogenous), endogenous (contagion) and correlated effects (external causation). Contextual effect means friends have similar responses because they share similar characteristics. Endogenous effect refers to the situation when one's outcome depends on others' outcome. Correlated effect exists when an external effect results in the similarity between friends.

Various methods have been proposed for modeling social network dependence. One popular approach is to introduce a network-based penalty on individual node effects, for example, see Li et al. (2016). In their work, a cohesion penalty similar to the graph Laplacian regularization is posited on individual node effects, which encourages similarity between effects of linked nodes. Similar ideas have also been widely used for building prediction models for studying gene-network data (e.g. Li and Li 2008, 2010; Sun et al. 2014). Another popular approach is to consider spatial autoregression models, with the parameter of spatial autocorrelation that quantifies the interactive dependence between connected nodes in a network (Chen et al. 2013). The maximum likelihood estimation for various spatial autocorrelation models have been studied in the economics literature (e.g. Ord 1975; Anselin 1980; Bramoullé et al. 2009; Lee et al. 2010). Recently, Zhou et al. (2017) proposed several likelihood-based estimation methods for spatial autocorrelation in a linear regression setting based on sampled network data.

In some applications, it is interesting to find whether interaction between friends can affect the propagation of events. For example, when people start playing an online game and send invitations to their friends to join in, it is likely to see that some of their friends will follow and start playing the same game. Statistical analysis for online game data has drawn great attention recently. Chen et al. (2017) proposed a hazards regression for freemium products based on a competing risk approach.

In this work, we are interested in how friendship affects the propagation of an event along network. Our study is motivated by data collected from players of a popular mobile game from one of the largest online social network platforms. Due to the confidentiality, the platform that provided us data has requested anonymity. The network we considered is the users' friendship. Since friends can collaborate to win more experience and tools, the game sends invitations to players' friends asking them to join the game. The times when users joined the game are recorded. Here, the time-to-event of interest is defined as the time from the starting point, when the game was launched, to the endpoint when a user joined to play the game. If a person never started playing the game during the study period, the event time of this person is considered to be censored. In addition, some demographic information of users, such as age and gender, is available.

Our goals here are to detect whether certain type of social network dependence exists with time-to-event data and to quantify this dependence if it exists. In our considered mobile game data application, an important feature for studying social network dependence is that not all users will be influenced by friends. For example, for some users, whether they will start playing the mobile game will not depend on their friends' characteristics. Because there is a cost on information targeting, it is of great interest to identify the subgroup of people that are more likely to be influenced by their friends on a social network.

Towards these goals, we propose a latent Cox model with contextual effect. Our model differs from the existing models in two aspects. First, the existing models are mainly for uncensored data and most of them are based on linear regression models for responses. Here, we incorporate the network dependence term in the conditional hazard function of the event time to model the dependence between event times of connected users. Second, a key difference is that most existing models (e.g. Zhu et al. 2017; Zhou et al. 2017) assume the response of any user in the social network will be affected by his or her friends and the magnitude of such dependence is common; while in our model, a latent binary indicator is introduced, indicating whether a user is susceptible to the influence of his or her friends. Here, introducing the susceptible indicator not only increases the flexibility for practical applications but also provides a way to estimate the probability that a user might be affected by his or her friends' characteristics in the social network. Therefore, it can help to identify a subgroup of users who are more likely to be influenced by their friends.

We first develop a score-type test for detecting the existence of the social network dependence based on the proposed latent cox model with contextual effect. When the social network dependence exists, we further develop an EM-type algorithm to estimate the model parameters and derive the associated inference procedure. The remainder of this paper is organized as follows. In Section 2, we introduce the proposed model. In Section 3, we present the proposed test statistic and estimation method. The asymptotic properties of the proposed test and estimators are also established here. In section 4, simulations are conducted to evaluate the empirical performance of the proposed test and estimators. An application of the proposed methods to the time-to-event data for playing a popular mobile game is given in Section 5, followed by discussions given in Section 6. All the technical derivations are provided in the Appendix.

2 Latent Cox Model with Contextual Effects

Let $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})^T \in \mathbb{R}^{n\times p}$ denote the covariate matrix which contains feature information of n individuals in the network, such as age and gender of each person in the network. Let $\boldsymbol{W} = (W_{ij}) \in \{0,1\}^{n\times n}$ be the adjacency matrix of a network involving n nodes, where $W_{ij} = 1$ means node i and node j are connected and $W_{ij} = 0$ otherwise. For each node i, let T_i denote the time to an event of interest and C_i the associated censoring time. Define $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Our goal is to test and estimate the dependence of event times among friends in the social network. A salient feature of social network dependence is that not all the individuals are susceptible to their friends' influence. To characterize the heterogeneity in susceptibility of individuals, we propose the following latent Cox model with contextual effects for the conditional hazard function for subject i:

$$\lambda_i(t|\boldsymbol{W}, \boldsymbol{X}, \xi_i) = \lambda(t) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho\xi_i \sum_{j\neq i}^{n} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j}, \tag{1}$$

where $\lambda(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a p-dimensional vector of parameters and $\xi_i = 0/1$ denotes the susceptibility indicator of individual i. In particular, when $\xi_i = 0$, the event time of individual i does not depend on his or her friends' characteristics. Moreover, we assume

$$P(\xi_i = 1|\boldsymbol{x}_i) = \frac{e^{\boldsymbol{\gamma}'\boldsymbol{x}_i^*}}{1 + e^{\boldsymbol{\gamma}'\boldsymbol{x}_i^*}}, \tag{2}$$

where $\boldsymbol{x}_i^* = (1, \boldsymbol{x}_i')'$ and γ is a $(p+1)$-dimensional vector of parameters. Note that ρ is identifiable only when $\boldsymbol{\beta}$ does not equal to 0 and $\boldsymbol{W}$ is not a zero matrix. Throughout this paper, we make these assumptions.

Note that in our model, the adjacency matrix $\boldsymbol{W}$ is not normalized by row. The proposed method can be easily extended to incorporate the row-normalized matrix $\boldsymbol{W}$, where the sum of each row of $\boldsymbol{W}$ is 1. However, the interpretation of the model would be different. When $\boldsymbol{W}$ is row-normalized, the network effect from friends is assumed to be an average, which does not depend on the number of friends of a node. On the other hand, when $\boldsymbol{W}$ is not row-normalized, as what we consider here, the number of friends also plays a role to the magnitude of network effect.

The parameter ρ describes the magnitude of the dependence of a susceptible node to its connected nodes, which is similar to the spatial autocorrelation parameter studied in Zhou et al. (2017). When $\rho = 0$, there is no social network dependence between event times of connected nodes. Under such a situation, the parameter γ is not estimable. In the next section, we will first propose a test for the null hypothesis: $H_0 : \rho = 0$, and then develop an EM-type algorithm for estimating the model parameters when $\rho \neq 0$. For convenience, it is assumed that C_i is independent of T_i. For example, in the mobile game application, all the censoring times are equal to the total duration of the study. This assumption is satisfied. However, this assumption can be relaxed as that C_i is independent of T_i given covariates $\boldsymbol{x_i}$ and those $\boldsymbol{x_j}$'s with $W_{ij} = 1$. Our proposed test and estimators are still valid.

3 Testing and Estimation Methods

3.1 Test for $H_0 : \rho = 0$

We propose a score-type tests statistic. Firstly suppose that $\boldsymbol{\xi} \equiv (\xi_1, \ldots, \xi_n)'$ is known. With the same argument as in Cox (1975), the partial likelihood function of the proposed model (1) is

$$L(\boldsymbol{\eta}; \boldsymbol{\xi}) = \prod_{i=1}^{n} \left[\frac{e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho\xi_i \sum_{j\neq i}^{n} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j}}{\sum_{l=1}^{n} e^{\boldsymbol{\beta}'\boldsymbol{x}_l + \rho\xi_l \sum_{j\neq l}^{n} W_{lj}\boldsymbol{\beta}'\boldsymbol{x}_j} I(\tilde{T}_l \geq \tilde{T}_i)} \right]^{\delta_i}, \tag{3}$$

where $\boldsymbol{\eta} = (\rho, \boldsymbol{\beta}')'$. Under H_0, model (1) becomes the standard Cox proportional hazards model. Let $\tilde{\boldsymbol{\beta}}$ denote the maximum partial likelihood estimator under the null. Define $\tilde{\boldsymbol{\eta}} = (0, \tilde{\boldsymbol{\beta}}')'$. Then, the score statistic is given by

$$\begin{aligned} S_1(\tilde{\boldsymbol{\eta}}; \boldsymbol{\xi}) =& \left.\frac{\partial \log(L)}{\partial \rho}\right|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}} = \sum_{i=1}^{n} \delta_i \left\{ \hat{Z}_i - \frac{\sum_{l=1}^{n} e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq \tilde{t}_i)\hat{Z}_l}{\sum_{l=1}^{n} e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq \tilde{t}_i)} \right\} \\ =& \sum_{i=1}^{n} \int_0^{\tau} \left\{ \hat{Z}_i - \frac{\sum_{l=1}^{n} e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)\hat{Z}_l}{\sum_{l=1}^{n} e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)} \right\} d\hat{M}_i(s), \end{aligned} \tag{4}$$

where τ is the total study duration, $\hat{Z}_i = \xi_i \sum_{j\neq i}^{n} W_{ij}\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_j$ and

$$\hat{M}_i(s) = N_i(s) - \int_0^s e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_i} I(\tilde{T}_i \geq u) d\tilde{\Lambda}(u),$$

with $N_i(s) = \delta_i I(\tilde{T}_i \leq s)$ and $\tilde{\Lambda}(u) = \int_0^u \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{j=1}^{n} I(\tilde{T}_j \geq t) e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_j}}$ being the Breslow estimator of the baseline cumulative hazard function under the null.

Since $\boldsymbol{\xi}$ is unknown in practice, we replace ξ_i with its expectation $p_i \equiv P(\xi_i = 1|\boldsymbol{x}_i)$ given in (2). Specifically, define $\hat{Z}_i^* = p_i \sum_{j\neq i}^{n} W_{ij}\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_j$. By replacing $\hat{Z}_i$ with $\hat{Z}_i^*$ in equation

(4), we obtain a new score-type statistic, denoted by $S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma})$. Note that $\boldsymbol{\gamma}$ is not identifiable under the null. Following the similar technique used in Fan et al. (2016) for testing the existence of a subgroup with an enhanced treatment effect, we propose the following supremum score test statistic:

$$T_n = \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\{S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma})\}^2}{\sum_{i=1}^n \left\{\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}},\tilde{\Lambda};\boldsymbol{\gamma})\right\}^2}.$$

Here, $\boldsymbol{\Gamma}$ is the domain of $\boldsymbol{\gamma}$, which is usually $\mathbb{R}^{p+1}$. In practice, the supreme is obtained by a grid search over $\boldsymbol{\Gamma}$. $\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}},\tilde{\Lambda};\boldsymbol{\gamma})$ in the denominator is defined as

$$\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}},\tilde{\Lambda};\boldsymbol{\gamma}) = \int_0^\tau \Bigg[\hat{Z}_i^* - \frac{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)\hat{Z}_l^*}{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)} - \boldsymbol{\mathcal{I}}_{12,n}^*(\tilde{\boldsymbol{\eta}})\boldsymbol{\mathcal{I}}_{22,n}^{-1}(\tilde{\boldsymbol{\eta}})\left\{\boldsymbol{x}_i - \frac{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)\boldsymbol{x}_l}{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)}\right\}\Bigg] d\hat{M}_i(s),$$

where $\boldsymbol{\mathcal{I}}_{22,n}(\tilde{\boldsymbol{\eta}}) = -\frac{\partial^2 \log(L)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}}$, $\boldsymbol{\mathcal{I}}_{12,n}(\tilde{\boldsymbol{\eta}}) = -\frac{\partial^2 \log(L)}{\partial\rho\partial\boldsymbol{\beta}'}|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}}$, and $\boldsymbol{\mathcal{I}}_{12,n}^*(\tilde{\boldsymbol{\eta}})$ is obtained from $\boldsymbol{\mathcal{I}}_{12,n}(\tilde{\boldsymbol{\eta}})$ by replacing ξ_i with p_i.

In the Appendix, we show that under the null,

$$\frac{1}{\sqrt{n}} S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i^*(\tilde{\boldsymbol{\eta}}_0,\tilde{\Lambda}_0;\boldsymbol{\gamma}) + o_p(1),$$

where $\tilde{\boldsymbol{\eta}}_0 = (0,\tilde{\boldsymbol{\beta}}_0')'$ and $\psi_i^*(\tilde{\boldsymbol{\eta}}_0,\tilde{\Lambda}_0;\boldsymbol{\gamma})$ can be obtained from $\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}},\tilde{\Lambda};\boldsymbol{\gamma})$ by replacing $\tilde{\boldsymbol{\beta}}$ with $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\Lambda}$ with $\tilde{\Lambda}_0$. Here, $\tilde{\boldsymbol{\beta}}_0$ and $\tilde{\Lambda}_0$ are the true values of $\boldsymbol{\beta}$ and Λ, respectively, under the null. By applying the martingale central limit theorem, we have $n^{-1/2}S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma})$ converges in distribution to a mean-zero normal variable under the null, with the asymptotic variance being consistently estimated by $n^{-1}\sum_{i=1}^n\{\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}},\tilde{\Lambda};\boldsymbol{\gamma})\}^2$. In particular, we need the following conditions:

C1. The function $\tilde{\Lambda}_0(t)$ is strictly increasing and continuously differentiable, and $\tilde{\Lambda}_0(\tau) < \infty$. The true parameters $\tilde{\boldsymbol{\beta}}_0$ under the null lie in the interior of a compact set.

C2. The covariates vector $\boldsymbol{x}$ is bounded in the sense that $P(|\boldsymbol{x}| < m) = 1$ for some constant $m > 0$. In addition, the vector $\boldsymbol{x}$ is linearly independent.

C3. The adjacency matrix $\boldsymbol{W}$ satisfies $\sum_{j \neq i} W_{ij} \leq K_0$ for all i, for some constant K_0.

Conditions C1 - C3 ensure the boundedness of the asymptotic variance of $n^{-1/2} S_1^*(\tilde{\boldsymbol{\eta}}; \boldsymbol{\gamma})$. In particular, Condition C3 assumes that the number of friends of each node is bounded, which implies sparsity of the network connectivity. Such a condition can facilitate the derivation of the large sample results. Under these conditions, we can establish the asymptotic null distribution of the test statistic T_n in the following theorem.

Theorem 1. *T_n converges in distribution to $\sup_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} G^2(\boldsymbol{\gamma})$ under H_0 as $n \to \infty$, where $\{G(\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \boldsymbol{\Gamma}\}$ is a mean zero Gaussian process with the covariance function*

$$\boldsymbol{\Sigma}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \lim_{n \to \infty} \frac{\sum_{i=1}^n \psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0; \boldsymbol{\gamma}_1) \psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0; \boldsymbol{\gamma}_2)}{\sqrt{\sum_{i=1}^n \{\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda}; \boldsymbol{\gamma_1})\}^2 \sum_{i=1}^n \{\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda}; \boldsymbol{\gamma_2})\}^2}},$$

for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \boldsymbol{\Gamma}$.

The proof of Theorem 1 is given in Appendix A. To obtain the critical value of the asymptotic null distribution of T_n, we adopt a resampling method. Specifically, we consider the perturbed test statistic

$$T_n^* = \sup_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \frac{\left\{ \sum_{i=1}^n \phi_i \hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda}; \boldsymbol{\gamma}) \right\}^2}{\sum_{i=1}^n \left\{ \hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda}; \boldsymbol{\gamma}) \right\}^2},$$

where $\phi_1, \ldots, \phi_n$ are i.i.d. standard normal variables. It can be shown that T_n and T_n^* have the same asymptotic null distribution. Therefore, we can generate a large number of perturbed statistics and use the empirical upper α-quantile of the perturbed statistics to estimate the critical value C_α for an α-level test. The null hypothesis is rejected if $T_n > C_\alpha$.

3.2 Parameter Estimation

Throughout this section, we assume $\rho \neq 0$. Under such an assumption, the parameters in models (1) and (2) are identifiable. We develop an EM-type algorithm to estimate the model parameters, denoted by $\boldsymbol{\Theta} = (\boldsymbol{\beta}', \rho, \boldsymbol{\gamma}')'$ and $\Lambda(t) = \int_0^t \lambda(u)du$. Define $\Lambda_i(t) = e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho\xi_i \sum_{j\neq i}^n W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j}\Lambda(t)$. The complete log likelihood function is

$$l(\boldsymbol{\Theta}, \Lambda) = \sum_{i=1}^n \left[\delta_i\{\log\lambda(\tilde{T}_i) + \boldsymbol{\beta}'\boldsymbol{x}_i + \rho\xi_i \sum_{j\neq i}^n W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j\} - \Lambda_i(\tilde{T}_i) + \xi_i\boldsymbol{\gamma}'\boldsymbol{x}_i^* - \log(1 + e^{\boldsymbol{\gamma}'\boldsymbol{x}_i^*})\right]. \tag{5}$$

Let $\hat{\boldsymbol{\Theta}}^{(k)}$ and $\hat{\Lambda}^{(k)}$ denote the estimators of $\boldsymbol{\Theta}$ and Λ at the kth iteration, respectively, and $\boldsymbol{\Omega}$ denote the observed data, $\{(\tilde{T}_i, \delta_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$ and $\boldsymbol{W}$. At the $(k+1)$th iteration, in the E-step, we calculate the conditional expectation of $l(\boldsymbol{\Theta}, \Lambda)$ given observed data $\boldsymbol{\Omega}$ and current estimators $\hat{\boldsymbol{\Theta}}^{(k)}$ and $\hat{\Lambda}^{(k)}$ of the parameters. Specifically,

$$\begin{aligned} &Q(\boldsymbol{\Theta}, \boldsymbol{\Lambda}|\hat{\boldsymbol{\Theta}}^{(k)}, \hat{\Lambda}^{(k)}) \equiv E\{l(\boldsymbol{\Theta}, \Lambda)|\boldsymbol{\Omega}, \hat{\boldsymbol{\Theta}}^{(k)}, \hat{\Lambda}^{(k)}\} \\ =&\sum_{i=1}^n \left[\delta_i\{\log\lambda(\tilde{T}_i) + \boldsymbol{\beta}'\boldsymbol{x}_i + \rho A_i^{(k)} \sum_{j\neq i}^n W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j\} - B_i^{(k)} + A_i^{(k)}\boldsymbol{\gamma}'\boldsymbol{x}_i^* - \log(1 + e^{\boldsymbol{\gamma}'\boldsymbol{x}_i^*})\right], \end{aligned} \tag{6}$$

where

$$\begin{aligned} A_i^{(k)} &= E(\xi_i|\boldsymbol{\Omega}, \hat{\boldsymbol{\Theta}}^{(k)}, \hat{\Lambda}^{(k)}) \\ &= \frac{e^{\delta_i\hat{\rho}^{(k)}\sum_{j\neq i}^n W_{ij}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_j} e^{-e^{\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_i + \hat{\rho}^{(k)}\sum_{j\neq i}^n W_{ij}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_j}\hat{\Lambda}^{(k)}(\tilde{T}_i)}\hat{p}_i^{(k)}}{e^{\delta_i\hat{\rho}^{(k)}\sum_{j\neq i}^n W_{ij}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_j} e^{-e^{\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_i + \hat{\rho}^{(k)}\sum_{j\neq i}^n W_{ij}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_j}\hat{\Lambda}^{(k)}(\tilde{T}_i)}\hat{p}_i^{(k)} + e^{-e^{\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_i}\hat{\Lambda}^{(k)}(\tilde{T}_i)}(1 - \hat{p}_i^{(k)})}, \end{aligned}$$

$$B_i^{(k)} = E\{\Lambda_i(\tilde{T}_i)|\boldsymbol{\Omega}, \hat{\boldsymbol{\Theta}}^{(k)}, \hat{\Lambda}^{(k)}\} = (1 - A_i^{(k)})e^{\boldsymbol{\beta}'\boldsymbol{x}_i}\Lambda(\tilde{T}_i) + A_i^{(k)}e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho\sum_{j\neq i}^n W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j}\Lambda(\tilde{T}_i),$$

and $\hat{p}_i^{(k)} = \exp(\hat{\boldsymbol{\gamma}}^{(k)\prime}\boldsymbol{x}_i^*)/\{1 + \exp(\hat{\boldsymbol{\gamma}}^{(k)\prime}\boldsymbol{x}_i^*)\}$.

The function $Q(\boldsymbol{\Theta}, \boldsymbol{\Lambda}|\hat{\boldsymbol{\Theta}}^{(k)}, \hat{\Lambda}^{(k)})$ can be written as the summation of $l_1(\boldsymbol{\beta}, \rho, \Lambda)$ and

$l_2(\boldsymbol{\gamma})$, where

$$\begin{aligned} l_1(\boldsymbol{\beta}, \rho, \Lambda) &= \sum_{i=1}^{n} \Big[\delta_i \{ \log \lambda(\tilde{T}_i) + \boldsymbol{\beta}' \boldsymbol{x}_i + \rho A_i^{(k)} \sum_{j \neq i}^{n} W_{ij} \boldsymbol{\beta}' \boldsymbol{x}_j \} \\ &\qquad - e^{\boldsymbol{\beta}' \boldsymbol{x}_i} \Lambda(\tilde{T}_i) \{ (1 - A_i^{(k)}) + A_i^{(k)} e^{\rho \sum_{j \neq i}^{n} W_{ij} \boldsymbol{\beta}' \boldsymbol{x}_j} \} \Big], \\ l_2(\boldsymbol{\gamma}) &= \sum_{i=1}^{n} \Big\{ A_i^{(k)} \boldsymbol{\gamma}' \boldsymbol{x}_i^* - \log(1 + e^{\boldsymbol{\gamma}' \boldsymbol{x}_i^*}) \Big\}. \end{aligned}$$

In the M-step, we maximize the functions $l_1(\boldsymbol{\beta}, \rho, \Lambda)$ and $l_2(\boldsymbol{\gamma})$ separately. Note that $l_2(\boldsymbol{\gamma})$ has a form similar to the log likelihood function for a logistic regression. It can be maximized directly using many existing gradient-based methods. Let $\hat{\boldsymbol{\gamma}}^{(k+1)}$ denote the resulting maximizer. The function $l_1(\boldsymbol{\beta}, \rho, \Lambda)$ involves the nonparametric function Λ. To maximize $l_1(\boldsymbol{\beta}, \rho, \Lambda)$, a log profile likelihood is first constructed. Similar to the arguments in Johansen (1983) and Klein (1992), when $\boldsymbol{\beta}$ and ρ are fixed, the nonparametric estimator that maximizes $l_1(\boldsymbol{\beta}, \rho, \Lambda)$ is given by

$$\hat{\Lambda}^{(k+1)}(t; \boldsymbol{\beta}, \rho) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} I(\tilde{T}_j \geq s) e^{\boldsymbol{\beta}' \boldsymbol{x}_j} \{ (1 - A_j^{(k)}) + A_j^{(k)} e^{\rho \sum_{l \neq j}^{n} W_{jl} \boldsymbol{\beta}' \boldsymbol{x}_l} \}}. \tag{7}$$

Plugging $\hat{\Lambda}^{(k+1)}(t; \boldsymbol{\beta}, \rho)$ into $l_1(\boldsymbol{\beta}, \rho, \Lambda)$, the log profile likelihood function for $\boldsymbol{\beta}$ and ρ, up to some constant, is

$$\begin{aligned} pl_1(\boldsymbol{\beta}, \rho) = \sum_{i=1}^{n} \delta_i \Bigg(& \boldsymbol{\beta}' \boldsymbol{x}_i + \rho A_i^{(k)} \sum_{j \neq i}^{n} W_{ij} \boldsymbol{\beta}' \boldsymbol{x}_j \\ & - \log \left[\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i) e^{\boldsymbol{\beta}' \boldsymbol{x}_j} \left\{ (1 - A_j^{(k)}) + A_j^{(k)} e^{\rho \sum_{l \neq j}^{n} W_{jl} \boldsymbol{\beta}' \boldsymbol{x}_l} \right\} \right] \Bigg). \end{aligned}$$

The log profile likelihood function $pl_1(\boldsymbol{\beta}, \rho)$ is not concave in $\boldsymbol{\beta}$ and ρ. To maximize it, we propose an iterative optimization method. Specifically, given $\boldsymbol{\beta}$, $pl_1(\boldsymbol{\beta}, \rho)$ is a univariate concave function of ρ, so it can be easily maximized with respect to ρ. Let $\hat{\rho}^{(k+1)} = \arg\max_{\rho} pl_1(\hat{\boldsymbol{\beta}}^{(k)}, \rho)$. Updating $\boldsymbol{\beta}$ given ρ is not straightforward. To facilitate the optimization with the computational stability, we fix $\rho = \hat{\rho}^{(k+1)}$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(k)}$ in the terms $\rho A_i^{(k)} \sum_{j\neq i}^n W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j$ and $A_j^{(k)} e^{\rho\sum_{l\neq j}^n W_{jl}\boldsymbol{\beta}'\boldsymbol{x}_l}$ of $pl_1(\boldsymbol{\beta},\rho)$. Then, the log profile likelihood function $pl_1(\boldsymbol{\beta},\rho)$ can be written as, up to some constant,

$$\sum_{i=1}^n \delta_i \left(\boldsymbol{\beta}'\boldsymbol{x}_i - \log\left[\sum_{j=1}^n I(\tilde{T}_j \geq \tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_j}\left\{(1-A_j^{(k)}) + A_j^{(k)} e^{\hat{\rho}^{(k+1)}\sum_{l\neq j}^n W_{jl}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_l}\right\}\right]\right),$$

which is equivalent to fit a Cox model with regression parameters $\boldsymbol{\beta}$ and an offset $\log\{(1-A_j^{(k)}) + A_j^{(k)} e^{\hat{\rho}^{(k+1)}\sum_{l\neq j}^n W_{jl}\hat{\boldsymbol{\beta}}^{(k)\prime}\boldsymbol{x}_l}\}$ for the jth subject, $j = 1,\ldots,n$. Let $\hat{\boldsymbol{\beta}}^{(k+1)}$ denote the maximizer of the above function. Define $\hat{\Lambda}^{(k+1)}(t) = \hat{\Lambda}^{(k+1)}(t;\hat{\boldsymbol{\beta}}^{(k+1)},\hat{\rho}^{(k+1)})$. We iterate the E-step and M-step until convergence. Let $\hat{\boldsymbol{\Theta}}$ and $\hat{\Lambda}$ denote the resulting estimators of $\boldsymbol{\Theta}$ and Λ, respectively, at convergence. Ideally, at each iteration of the EM algorithm, ρ and $\boldsymbol{\beta}$ should be updated iteratively till convergence. However, to make the algorithm faster and stable, we just update ρ and $\boldsymbol{\beta}$ once in each EM iteration. When the algorithm converges, the final estimators approximately maximize the observed likelihood function. In our EM algorithm, we chose the initial estimators of the parameters as follows: $\hat{\rho}^{(0)} = 0$, $\hat{\boldsymbol{\gamma}}^{(0)} = \mathbf{0}$, $\hat{\boldsymbol{\beta}}^{(0)}$ is the maximum partial likelihood estimator and $\hat{\Lambda}^{(0)}$ is the Breslow's estimator under the standard Cox model when $\rho = 0$. In addition, the convergence criteria was set as $||\hat{\boldsymbol{\Theta}}^{(k+1)} - \hat{\boldsymbol{\Theta}}^{(k)}||_\infty < 10^{-6}$. Based on our numerical experience, the proposed EM algorithm usually converges within 50 iterations. It is worth noting that the term $A_i^{(k)}$ at convergence denotes the posterior probability that the ith user might be affected by his or her friends' behavior. We name it the posterior "susceptible" probability.

Let $\boldsymbol{\Theta}_0$ and Λ_0 denote the true values of $\boldsymbol{\Theta}$ and Λ, respectively. To establish the asymptotic properties of the proposed estimators, we need Conditions C2 and C3, and the following conditions:

C4. The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable, and $\Lambda_0(\tau) < \infty$. The true parameters $\boldsymbol{\Theta}_0$ lie in the interior of a compact set.

C5. The information matrix $\boldsymbol{I}(\boldsymbol{\Theta}_0)$ defined in Appendix B is finite and positive definite.

Conditions C2 - C5 are commonly assumed for establishing the asymptotic properties of the nonparametric maximum likelihood estimators for semiparametric survival models (e.g. Zeng and Lin (2006), Lu (2008)).

Theorem 2. *As $n \to \infty$,*

$$\sup_{t\in[0,\tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| \to 0 \ \ and \ \ ||\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0|| \to 0 \ \ a.s.$$

In addition, $\sqrt{n}\{\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0\}$ converges in distribution to a mean-zero multivariate normal variable with variance $\{\boldsymbol{I}(\mathbf{\Theta}_0)\}^{-1}$.

The proof of Theorem 2 and the definition of $\boldsymbol{I}(\mathbf{\Theta}_0)$ are given in Appendix B. Next, we derive a method for estimating the asymptotic variance of $\hat{\mathbf{\Theta}}$. We adopt the techniques developed in Lange (1999) and Hunter and Lange (2004) for estimating the variance of MM estimators. Specifically, define $g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)}) = pl_1(\boldsymbol{\beta}, \rho) + l_2(\boldsymbol{\gamma})$. Let $\nabla^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})$ denote the second derivative of $g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})$ with respect to $\mathbf{\Theta}$. Note that $\nabla^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})$ has explicit expressions, which are provided in Appendix C. Then, the observed information matrix can be approximated by

$$I(\hat{\mathbf{\Theta}}) = -\nabla^2 g(\hat{\mathbf{\Theta}}|\hat{\mathbf{\Theta}}) \left\{ I - \nabla M(\hat{\mathbf{\Theta}}) \right\}, \tag{8}$$

where $M(\boldsymbol{\nu}) = \arg\max_{\mathbf{\Theta}} g(\mathbf{\Theta}|\boldsymbol{\nu})$ and $\nabla M(\boldsymbol{\nu}) = \partial M(\boldsymbol{\nu})/\partial \boldsymbol{\nu}'$. The inverse of $I(\hat{\mathbf{\Theta}})$ is an estimator of the asymptotic covariance matrix of $\hat{\mathbf{\Theta}}$. Here, $\nabla M(\boldsymbol{\nu})$ does not have an explicit expression and we compute it by numerical differentiation. Specifically, write $M(\hat{\mathbf{\Theta}}) = \{M_1(\hat{\mathbf{\Theta}}), \ldots, M_q(\hat{\mathbf{\Theta}})\}'$, where $q = 2(p+1)$. The (i,j)th element of $\nabla M(\hat{\mathbf{\Theta}})$ is computed by $\{M_i(\hat{\mathbf{\Theta}} + de_j) - M_i(\hat{\mathbf{\Theta}})\}/d$, where d is a small positive value and e_j is the basis vector with the jth element as 1 and others as 0. Note that when the EM algorithm converges, $M(\hat{\mathbf{\Theta}}) = \hat{\mathbf{\Theta}}$. To compute $M(\hat{\mathbf{\Theta}} + de_j)$, we fix $\boldsymbol{\nu} = \hat{\mathbf{\Theta}} + de_j$ and compute the maximizer of $g(\mathbf{\Theta}|\boldsymbol{\nu})$ using the proposed EM algorithm. In our implementation, we chose $d = d_0/n$, where d_0 is a small positive constant. We have tried a few values of d_0, ranging from 1 to 10, and found that $d_0 = 5$ gives reasonable variance estimates for all cases.

4 Simulation Studies

In this section, we conduct simulations to evaluate the empirical performance of the proposed test and estimators. The underlying network is generated from the stochastic block model (Holland et al. 1983). Let K be the number of communities in the network. The stochastic block model is defined by

$$P(W_{ij} = 1) = 1 - P(W_{ij} = 0) = \boldsymbol{P}_{C_i C_j}, \tag{9}$$

where $\boldsymbol{P}$ is a $K \times K$ symmetric matrix whose (C_i, C_j)th element $\boldsymbol{P}_{C_i C_j}$ records the probability that communities C_i and C_j are connected. The total number of nodes n is set to be 2000 and K is set to be 5 or 10. For $K = 5$, the numbers of nodes contained in each community are $(500, 500, 400, 400, 200)$ and the corresponding $\boldsymbol{P}$ matrix has elements $\boldsymbol{P}_{11} = \boldsymbol{P}_{33} = 0.05$, $\boldsymbol{P}_{22} = \boldsymbol{P}_{55} = 0.1$, $\boldsymbol{P}_{44} = 0.2$ and $\boldsymbol{P}_{C_i C_j} = 10^{-4}$ for $C_i \neq C_j$. Similarly, when $K = 10$, community sizes are $(100, 100, 100, 100, 200, 200, 200, 200, 400, 400)$ and $\boldsymbol{P}_{11} = \boldsymbol{P}_{55} = 0.05$, $\boldsymbol{P}_{22} = \boldsymbol{P}_{66} = \boldsymbol{P}_{99} = \boldsymbol{P}_{10,10} = 0.1$, $\boldsymbol{P}_{33} = \boldsymbol{P}_{77} = 0.2$, $\boldsymbol{P}_{44} = \boldsymbol{P}_{88} = 0.3$ and $\boldsymbol{P}_{C_i C_j} = 10^{-4}$ if $C_i \neq C_j$. For $K = 1$, we generate the network from a pseudo 5-community stochastic block model with community numbers same as in $K = 5$, and $\boldsymbol{P}_{11} = \boldsymbol{P}_{22} = \boldsymbol{P}_{33} = \boldsymbol{P}_{44} = 0.01$, $\boldsymbol{P}_{55} = 0.3$, $\boldsymbol{P}_{C_i C_j} = 0.015$ for $C_i \neq C_j$. All communities are quite connected without clear separation, while a subset of nodes are connected more closely. Such network structure is similar to the observation in the real data example.

The baseline hazard function is chosen as $\lambda_0(t) = 0.5$. Two covariates are included, where the first covariate is generated from a Bernoulli distribution with the success probability 0.5 and the second is generate from a uniform distribution on (-1, 1). We set $\boldsymbol{\beta} = (1, -1)'$ and $\boldsymbol{\gamma} = (0, 1, -1)'$. The censoring time is generated from a uniform distribution on $(0, c)$, where the constant c is chosen to yield the approximately 15% or 30% censoring rate. We conduct 1000 replicates for each simulation setting.

4.1 Simulation Results for Testing

We consider the following values of ρ: 0, 0.01, 0.015, 0.02 and 0.05, and conduct the proposed test with the alpha-level as 0.05. When computing the p-value of the test statistic, we generate 1000 perturbed test statistics as described in Section 3.1. The empirical type I error and power of the proposed test are reported in Table 1. It can be seen that the proposed test gives proper type I error rates under the null when $\rho = 0$. In addition, the power of the test increases as ρ increases and the censoring rate decreases as expected.

Table 1: Type I error and power of the proposed test for simulated network.

K	ρ	CR	RR	K	ρ	CR	RR	K	ρ	CR	RR
	0.000	0.15	0.050		0.000	0.15	0.063		0.000	0.15	0.050
	0.010	0.15	0.855		0.010	0.15	0.818		0.010	0.15	0.655
10	0.015	0.15	0.994	5	0.015	0.15	0.986	1	0.015	0.15	0.888
	0.020	0.15	1.000		0.020	0.15	0.999		0.020	0.15	0.973
	0.050	0.15	1.000		0.050	0.15	1.000		0.050	0.15	1.000
	0.000	0.30	0.051		0.000	0.30	0.055		0.000	0.30	0.060
	0.010	0.30	0.814		0.010	0.30	0.775		0.010	0.30	0.593
10	0.015	0.30	0.982	5	0.015	0.30	0.976	1	0.015	0.30	0.881
	0.020	0.30	0.999		0.020	0.30	0.997		0.020	0.30	0.969
	0.050	0.30	1.000		0.050	0.30	1.000		0.050	0.30	1.000

CR, censoring rate; RR, rejection rate over 1000 replicates.

In the above simulations, the values of ρ were chosen to be small because $\boldsymbol{W}$ is not row-normalized in our model. Next, we examine the performance of the proposed test when $\boldsymbol{W}$ is or is not row-normalized in our model. In particular, we will illustrate how the power depends on the magnitude of ρ under these two choices of $\boldsymbol{W}$. In this simulation, we use the network from the considered mobile game data to demonstrate the performance of our proposed test on a real-world network without any known structure. Details about this data will be discussed in Section 5. Covariates and responses are generated in the same way as previous, and regression parameters are set the same as before. For unnormalized $\boldsymbol{W}$, the values of ρ are chosen the same as in Table 1, from where we know the power is already close to 1 when ρ approaches 0.02. For normalized $\boldsymbol{W}$, the corresponding ρ value needs to be much larger to yield reasonable power. Here, we increase the values of ρ tenfold for unnormalized $\boldsymbol{W}$. The type I error and power results are reported in Table 2. Based on the results, it can be seen that our test gives proper type I error for both normalized and unnormalized $\boldsymbol{W}$. Moreover, when $\boldsymbol{W}$ is normalized, the proposed test has a reasonable power ($\geq 80\%$) only when ρ increases to 0.5. This is not a surprise since with normalized $\boldsymbol{W}$, ρ measures the dependence on the average network effects of connected nodes.

Table 2: Type I error and power based on the considered mobile game network data.

CR	$\boldsymbol{W}$ Unnormalized		$\boldsymbol{W}$ Normalized	
	ρ	RR	ρ	RR
0.15	0.000	0.046	0.000	0.054
	0.010	0.351	0.100	0.103
	0.015	0.567	0.150	0.161
	0.020	0.764	0.200	0.244
	0.050	0.994	0.500	0.834
0.30	0.000	0.058	0.000	0.057
	0.010	0.290	0.100	0.088
	0.015	0.507	0.150	0.146
	0.020	0.713	0.200	0.214
	0.050	0.993	0.500	0.807

CR, censoring rate; RR, rejection rate over 1000 replicates.

4.2 Simulation Results for Estimation

Here, we set $\rho = 0.05$ or 0.1, and use the simulated network with $K = 1, 5,$ or 10. Other settings are the same as in the previous section. We report the mean and standard deviation of the proposed estimators, mean of the estimated standard errors and empirical coverage probability of Wald-type 95% confidence intervals. Simulation results based on 1000 replicates are given in Table 3. Based on the results, we can see that all the estimators are nearly unbiased, the means of estimated standard errors are close to the standard deviations of the estimators, and the empirical coverage probabilities are close to the nominal level for most parameters except that those for ρ are slightly larger than 95%. For comparisons, we also fit the classical Cox model without including the network information. The corresponding results are reported in Table 4. Since the network dependence parameter ρ is chosen to be small in our simulations, the estimates of the regression parameters obtained under the classical Cox model are close to those obtained under the proposed Cox model with network structure. However, the estimators from the classical Cox model are apparently more biased for most scenarios, especially when ρ is 0.1.

Table 3: Simulation results of parameter estimation for the proposed model.

K	ρ	CR		$\hat{\rho}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
10	0.05	0.15	Est.	0.051	1.000	-1.001	-0.017	1.029	-1.022
			SD	0.006	0.090	0.076	0.273	0.367	0.316
			SE	0.007	0.085	0.074	0.256	0.340	0.304
			CP	0.979	0.923	0.937	0.935	0.921	0.943
	0.10	0.15	Est.	0.100	1.003	-1.003	-0.001	1.005	-0.997
			SD	0.007	0.069	0.068	0.128	0.180	0.169
			SE	0.009	0.070	0.063	0.127	0.178	0.159
			CP	0.980	0.953	0.934	0.947	0.949	0.938
	0.05	0.30	Est.	0.050	1.003	-1.000	-0.006	1.020	-1.024
			SD	0.006	0.098	0.084	0.294	0.393	0.339
			SE	0.007	0.095	0.082	0.280	0.365	0.326
			CP	0.983	0.936	0.954	0.948	0.925	0.952
	0.10	0.30	Est.	0.100	1.004	-1.004	-0.001	1.006	-0.996
			SD	0.008	0.077	0.073	0.132	0.186	0.172
			SE	0.010	0.077	0.069	0.132	0.184	0.165
			CP	0.979	0.943	0.945	0.948	0.946	0.945
5	0.05	0.15	Est.	0.051	1.001	-0.998	-0.003	1.027	-1.036
			SD	0.006	0.091	0.078	0.285	0.381	0.331
			SE	0.007	0.088	0.077	0.271	0.360	0.321
			CP	0.974	0.944	0.950	0.942	0.943	0.948
	0.10	0.15	Est.	0.101	0.998	-1.000	-0.005	1.018	-1.011
			SD	0.008	0.076	0.067	0.133	0.184	0.170
			SE	0.009	0.072	0.066	0.133	0.185	0.166
			CP	0.978	0.932	0.939	0.949	0.946	0.946
	0.05	0.30	Est.	0.051	1.003	-1.000	0.003	1.023	-1.032
			SD	0.006	0.099	0.087	0.303	0.403	0.353
			SE	0.007	0.097	0.085	0.293	0.384	0.341
			CP	0.980	0.943	0.937	0.946	0.947	0.947
	0.10	0.30	Est.	0.101	0.999	-0.999	-0.004	1.016	-1.014
			SD	0.009	0.083	0.075	0.140	0.194	0.179
			SE	0.010	0.079	0.072	0.139	0.193	0.173
			CP	0.977	0.928	0.931	0.950	0.955	0.937
1	0.05	0.15	Est.	0.051	0.999	-0.998	-0.025	1.056	-1.059
			SD	0.007	0.097	0.091	0.377	0.510	0.475
			SE	0.007	0.094	0.083	0.367	0.493	0.439
			CP	0.973	0.944	0.928	0.952	0.947	0.945
	0.10	0.15	Est.	0.101	1.000	-0.999	-0.013	1.016	-1.013
			SD	0.009	0.088	0.079	0.182	0.242	0.221
			SE	0.010	0.083	0.076	0.183	0.245	0.219
			CP	0.971	0.929	0.950	0.958	0.953	0.942
	0.05	0.30	Est.	0.050	1.006	-1.002	0.007	1.024	-1.041
			SD	0.007	0.104	0.094	0.405	0.528	0.486
			SE	0.008	0.101	0.090	0.388	0.517	0.461
			CP	0.970	0.926	0.937	0.946	0.945	0.948
	0.10	0.30	Est.	0.101	1.001	-0.998	-0.010	1.014	-1.017
			SD	0.010	0.098	0.088	0.196	0.257	0.240
			SE	0.011	0.090	0.082	0.195	0.259	0.231
			CP	0.967	0.924	0.927	0.954	0.952	0.938

CR, censoring rate; Est., mean of estimators; SD, standard deviation of estimators; SE, mean of estimated standard errors; CP, empirical coverage probability of 95 % confidence intervals.

Table 4: Simulation results of parameter estimation for the classical Cox model.

K	ρ	CR		β_1	β_2
10	0.05	0.15	Est.	0.995	-0.991
			SD	0.052	0.048
	0.10	0.15	Est.	0.902	-0.897
			SD	0.051	0.047
	0.05	0.30	Est.	1.001	-0.997
			SD	0.057	0.052
	0.10	0.30	Est.	0.899	-0.892
			SD	0.056	0.051
5	0.05	0.15	Est.	1.009	-1.005
			SD	0.052	0.048
	0.10	0.15	Est.	0.924	-0.920
			SD	0.051	0.047
	0.05	0.30	Est.	1.016	-1.012
			SD	0.057	0.052
	0.10	0.30	Est.	0.925	-0.919
			SD	0.056	0.052
1	0.05	0.15	Est.	1.040	-1.037
			SD	0.053	0.048
	0.10	0.15	Est.	0.981	-0.977
			SD	0.054	0.048
	0.05	0.30	Est.	1.048	-1.044
			SD	0.058	0.052
	0.10	0.30	Est.	0.992	-0.987
			SD	0.059	0.052

CR, censoring rate; Est., mean of estimators; SD, standard deviation of estimators.

For each simulated data set, we can calculate the estimated posterior "susceptible" probability for each subject as described in Section 3.2. Figure 1 shows the true ξ_i's (left panel) and the estimated posterior "susceptible" probability (right panel), $\hat{P}(\xi_i = 1|\mathbf{\Omega})$ for one randomly selected simulated data set with $K = 5$, $\rho = 0.1$ and censoring rate 30%. The root mean squared difference between them is 0.35. The values are represented by color intensity varying from white (corresponds to 0) to red (corresponds to 1). The similarity of these two plots demonstrates the ability of the estimated posterior "susceptible" probability for predicting which users are more likely to be influenced by his or her friends' behavior.

5 Application to a Mobile Game Data

This popular mobile game was officially launched in 2013, supported on either Android or IOS with free download. It is a role-playing game, where users can choose from multiple different characters with various skills. The game can be played either solo or in a team in order to complete a task. Users can log in using their account, which is one of the largest online social network platforms. The friends of users on the network will become their friends in the game, and they can compete or collaborate in the game. Users can also share their scores or experience to all their friends. Due to the confidentiality, the platform that provided us data has requested anonymity.

In this data, $n = 961$ users are included. We have the friendship structure of these users, which can be transformed to the adjacency matrix $\boldsymbol{W}$. The friend network is plotted in Figure 3, where nodes refer to users and edges denote the friendship between two users. Note that the friendship has to be mutual, so the network is undirected and $\boldsymbol{W}$ is symmetric. The numbers of friends of users vary from 0 to 154 with median 6. In addition, the data records the time when the user started to play the mobile game since it was launched, which is the event time of our interest. The survival curve for the Kaplan-Meier estimator of the event times is showed in Figure 2. It can be seen that the estimated survival curve has a quick descent in the first 10 days, then a gradual decent from 10 to 40 days, and finally a steep decent beyond 40 days. Besides the event times, three covariates: age, gender and activity level (denoted by a-level), are included in our analysis. Here, the activity level measures how long and how often the user has been using the platform. Higher activity level indicates that the user is more active on the platform.

We first apply the proposed score test in Section 3.1. The value of the test statistic is 5.59 and the associated p-value is 0.028. At level 0.05, we reject the null hypothesis and claim that there is social network dependence among users' event times. Next, we estimate the model parameters using the proposed EM algorithm. Results are reported in Table 5. The estimated value of the exogenous effect parameter ρ is 0.056 with the Wald test p-value 0.02, indicating a significant positive social network dependence among those susceptible users. This agrees with the score test result. Note that the Wald test requires the full model defined by (1) and (2) being correctly specified, while the score test is more robust since it only requires the standard Cox model being correctly specified under the null. In terms of the other covariates, age and gender are not significant but activity level has a very significant effect on the event time. This implies that an individual's activity level will have an effect on his or her event time, and if this individual is susceptible, his or her friends' activity level will also have an effect on this individual's event time. To further investigate this finding, we divide users into two groups by the median of activity level and plot the Kaplan-Meier curve for each group as well in Figure 2. We can observe that the group with smaller activity levels has larger survival probabilities than the group with larger activity levels. This implies that users with larger activity levels tend to play the game earlier than those with smaller activity levels and thus may have more impact on their friends, which agrees with our intuition.

In addition, we calculate the estimated posterior "susceptible" probability for each user and plot these values using different colors in Figure 3, similar to Figure 1 in simulations. The nodes with color closer to red indicate users that are more likely to be influenced by friends. The histogram of the estimated posterior is presented in Figure 4. The right skewness of the "susceptible" probability shows that the probability that people are affected by friends is low for most individuals, but can be high for a group of people. Specifically, the mean of the estimated posterior probability is 0.254 while the median is 0.081.

Table 5: Analysis results based on the proposed model for the mobile game data.

	$\hat{\rho}$	$\hat{\beta}_{\text{age}}$	$\hat{\beta}_{\text{gender}}$	$\hat{\beta}_{\text{a-level}}$
Estimate	0.056	0.056	0.051	0.140
SE	0.024	0.040	0.035	0.038
p-value	0.020	0.162	0.145	0.000

SE, estimated standard error.

Next, we investigate the necessity of introducing the latent susceptible indicators ξ_i's in the proposed model, where we assume there are two types of people, one can be affected by friends with a common coefficient ρ, while the other cannot. If we ignore these two subgroups, and assume that the dependence for the whole population is ρ, then the model is simplified to

$$\lambda_i(t|\boldsymbol{W}, \boldsymbol{X}) = \lambda(t) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho \sum_{j\neq i}^{n} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j}. \tag{10}$$

A score test statistic for ρ can be derived as $T = \frac{\{S_1^0(\tilde{\eta})\}^2}{\sum_{i=1}^n \left\{\hat{\psi}_i^0(\tilde{\boldsymbol{\eta}},\tilde{\Lambda})\right\}^2}$, where $S_1^0(\tilde{\boldsymbol{\eta}})$ and $\hat{\psi}_i^0(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda})$ are in the same format as $S_1^*(\tilde{\boldsymbol{\eta}}; \boldsymbol{\gamma})$ and $\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda}; \boldsymbol{\gamma})$ in Section 3.1 by replacing all ξ_i or p_i by 1. It is straightforward to show that T follows a chi-square distribution with the degree freedom of 1. For the mobile game data, the test statistic value is 3.48, and the p-value is 0.062. Therefore, the network dependence ρ is not significant under model (10) at level 0.05. Such result indicates that it is hard to detect the existence of social network dependence without considering the two subgroups. Moreover, we estimate the parameters based on model (10) and the results are given in Table 6. The parameter estimation is obtained by maximizing the partial likelihood through coordinate descent, i.e., optimize ρ and $\boldsymbol{\beta}$ alternatively till convergence. The standard errors are estimated by inverting the information matrix. The Wald test for ρ is not significance with the p-value of 0.226 based on (10). This also agrees with the score test result.

To check the goodness-of-fit of different models, we further calculate the AIC values as the summation of the negative log observed likelihood function and the number of parameters multiplying by 2 based on the fits for both the proposed model (1) and the simplified model (10). The AIC values are 13168.8 and 13191.0 respectively. Therefore, model (1) fits the data better than model (10). All above analyses support the need of including individual latent susceptible indicators for studying the social network dependence with the considered data application. To complete the discussion, we also fit a classical Cox model ignoring the network structure for the mobile game data. The parameter estimates for age, gender and a-level are -0.019, 0.041, 0.154 with standard errors 0.040, 0.033, 0.036, and the AIC value of the fitted Cox model is 13193.9. The larger AIC value here indicates that the proposed model fits the data better than the classical Cox model ignoring the network structure.

Table 6: Analysis results of simplified model (10) for the mobile game data.

	$\hat{\rho}$	$\hat{\beta}_{\text{age}}$	$\hat{\beta}_{\text{gender}}$	$\hat{\beta}_{\text{a-level}}$
Estimate	0.011	-0.042	0.035	0.138
SE	0.009	0.042	0.032	0.036
p-value	0.226	0.319	0.283	0.000

SE, estimated standard error.

6 Discussion

In this paper, we developed a latent Cox model with contextual effect for studying social network dependence with time to event data. An important feature of the proposed model is to allow that only a subset of users are susceptible to the influence of their friends. Both testing and estimation methods for the proposed model are investigated. In addition, our estimation method naturally provides an estimate of the individual posterior "susceptible" probability, which measures how likely a user might be influenced by his or her friends.

In the proposed model, it is assumed that there is a mixture of susceptible and non-susceptible users, defined by a latent binary variable. For those susceptible users, the social network dependence is characterized by the same parameter ρ. However, we may consider a more general case to have an individual parameter ρ_i for each person, instead of $\rho\xi_i$, and assume ρ_i follows some distribution, for example, a Beta distribution with the mean and variance depending on covariates. In addition, we can include a subject-specific offset in the proposed model to explain the further dependence among event times of connected nodes and add a cohesion penalty on offsets as in Li et al. (2016).

For the considered mobile game data application, we have type I censoring. In our work, we consider a general framework to incorporate random censoring. On the other hand, a Tobit model (Tobin 1958) with the considered network structure can be developed for handling type I censoring. However, to incorporate the latent susceptible indicator, the derivation of associated testing and estimation procedures is not a straightforward extension of the proposed methods, and requires further investigation. In addition, in the mobile game data application, it is possible to have the event of "never start playing the game", which corresponds to a cured subject in survival analysis (Lu and Ying 2004; Lu 2008). In the literature, a mixture cure-rate model can be fitted for survival data with a cure fraction. However, it usually requires a sufficiently long follow-up. For the considered mobile game data application, the study duration is very limited, which may make it difficult to fit a mixture cure-rate model due to the non-identifiability in finite samples. Alternatively, a competing risks approach can be considered as studied in Chen et al. (2017). In such a competing risks approach, the event of "never start playing the game" can be naturally incorporated. However, it needs further investigation under our considered latent Cox model with contextual effects.

Finally, in many network data applications, the scale of data can be incredibly large. Our current method is developed for network-based event-time data with a moderate size. For handling data of a large scale, some computational techniques need to be developed. For example, parallel computing can be used when calculating the likelihood function and the test statistics under different $\boldsymbol{\gamma}$ values. In addition, more efficient algorithms can be used to handle the large matrix algebra in the proposed EM estimation method. This is an interesting topic that warrants a future research.

Acknowledgment

The authors thank an associate editor and referees for their thoughtful and constructive comments that help to improve the work. This work was partly supported by a NIH grant P01 CA142538 and a NSF grant DMS 1555244.

A Proof of Theorem 1

Recall that $\tilde{\boldsymbol{\beta}}$ is the standard maximum partial likelihood under the null. Its asymptotic representation is well known. By Taylor expansion and some empirical process approximation techniques, we have

$$\frac{1}{\sqrt{n}}S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}S_1^*(\tilde{\boldsymbol{\eta}}_0;\boldsymbol{\gamma}) - \boldsymbol{\mathcal{I}}_{12,n}^*(\tilde{\boldsymbol{\eta}}_0)\boldsymbol{\mathcal{I}}_{22,n}^{-1}(\tilde{\boldsymbol{\eta}}_0)\frac{1}{\sqrt{n}}S_2(\tilde{\boldsymbol{\eta}}_0;\boldsymbol{\gamma}) + o_p(1),$$

where $S_2(\tilde{\boldsymbol{\eta}}_0;\boldsymbol{\gamma}) = \partial \log(L)/\partial\beta|_{\boldsymbol{\eta}=\tilde{\boldsymbol{\eta}}_0}$. Therefore,

$$\begin{aligned}\frac{1}{\sqrt{n}}S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \Bigg[& Z_i^* - \frac{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_l} I(\tilde{T}_l \geq s) Z_l^*}{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)} - \\ & \boldsymbol{\mathcal{I}}_{12,n}^*(\tilde{\boldsymbol{\eta}}_0)\boldsymbol{\mathcal{I}}_{22,n}^{-1}(\tilde{\boldsymbol{\eta}}_0)\left\{\boldsymbol{x}_i - \frac{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)\boldsymbol{x}_l}{\sum_{l=1}^n e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_l} I(\tilde{T}_l \geq s)}\right\}\Bigg] dM_i(s) + o_p(1) \\ = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0;\boldsymbol{\gamma}) + o_p(1), &\end{aligned}$$

where $Z_i^* = p_i \sum_{j\neq i}^n W_{ij}\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_j$ and $M_i(t) = N_i(t) - \int_0^t I(\tilde{T}_i \geq s)e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_i}d\tilde{\Lambda}_0(s)$ is a mean-zero martingale process under the null.

Under Conditions C1-C3, it can be shown that $\sigma^2(\boldsymbol{\gamma}) \equiv \lim_{n\to} n^{-1}\sum_{i=1}^n\{\psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0;\boldsymbol{\gamma})\}^2$ is finite and can be consistently estimated by $n^{-1}\sum_{i=1}^n\{\hat{\psi}_i^*(\tilde{\boldsymbol{\eta}}, \tilde{\Lambda};\boldsymbol{\gamma})\}^2$. Therefore, for fixed $\boldsymbol{\gamma}$, by the martingale central limit theorem, we have $n^{-1/2}S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma})$ converges in distribution to a mean-zero normal random variable with variance $\sigma^2(\boldsymbol{\gamma})$ under the null. In addition, write $n^{-1/2}\psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0;\boldsymbol{\gamma}) = \int_0^\tau H_{n,i}(s;\boldsymbol{\gamma})dM_i(s)$. We have

$$\sum_{i=1}^n\int_0^\tau H_{n,i}^2(s;\boldsymbol{\gamma})I\{|H_{n,i}(s;\boldsymbol{\gamma})| \geq \epsilon\}I(\tilde{T}_i \geq s)e^{\tilde{\boldsymbol{\beta}}_0'\boldsymbol{x}_i}d\tilde{\Lambda}_0(s) \to 0$$

for any $\epsilon > 0$ and fixed $\boldsymbol{\gamma}$. By the martingale central limit theorem of Fleming and Harrington (1991), we have $n^{-1/2}S_1^*(\tilde{\boldsymbol{\eta}};\boldsymbol{\gamma})$ converges weakly to a Gaussian process with mean 0 and covariance matrix $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0;\boldsymbol{\gamma}_1)\psi_i^*(\tilde{\boldsymbol{\eta}}_0, \tilde{\Lambda}_0;\boldsymbol{\gamma}_2)$ for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \boldsymbol{\Gamma}$. The results in Theorem 1 then follows.

B Proof of Theorem 2

First, we prove the consistency of the proposed estimators. Define $H_i(\boldsymbol{\beta}) = \sum_{j\neq i} W_{ij}\boldsymbol{\beta}'\boldsymbol{x}_j$. The observed likelihood function is given by

$$
\begin{aligned}
L(\Lambda, \boldsymbol{\Theta}) =& \prod_{i=1}^{n} \left[\{\lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}\}^{\delta_i} e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} p_i + \{\lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}\}^{\delta_i} e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}} (1-p_i) \right] \\
=& \prod_{i=1}^{n} \left\{ p_i \lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})} e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} + (1-p_i) \lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i} e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}} \right\}^{\delta_i} \times \\
& \quad \left\{ p_i e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} + (1-p_i) e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}} \right\}^{1-\delta_i}.
\end{aligned}
$$

Let $l_n(\Lambda, \boldsymbol{\Theta}) = n^{-1}\log\{L(\Lambda, \boldsymbol{\Theta})\}$. Then

$$
l_n(\Lambda, \boldsymbol{\Theta}) = \frac{1}{n}\sum_{i=1}^{n} \left[\delta_i \log\left\{ \lambda(\tilde{T}_i) g_i(\Lambda, \boldsymbol{\Theta}) e^{\boldsymbol{\beta}'\boldsymbol{x}_i} \right\} + \log S(\tilde{T}_i, \Lambda, \boldsymbol{\Theta}) \right],
$$

where

$$
\begin{aligned}
S(\tilde{T}_i, \Lambda, \boldsymbol{\Theta}) =& p_i e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} + (1-p_i) e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}}, \\
g_i(\Lambda, \boldsymbol{\Theta}) =& \frac{p_i e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} e^{\rho H_i(\boldsymbol{\beta})} + (1-p_i) e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}}}{p_i e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i + \rho H_i(\boldsymbol{\beta})}} + (1-p_i) e^{-\Lambda(\tilde{T}_i) e^{\boldsymbol{\beta}'\boldsymbol{x}_i}}}.
\end{aligned}
$$

Write $\hat{\boldsymbol{\Theta}} = \hat{\boldsymbol{\Theta}}_n$ and $\hat{\Lambda} = \hat{\Lambda}_n$, to show their dependence on n. It can be shown that $\sup_n \hat{\Lambda}_n(\tau) < \infty$ based on Condition C3.2. By Helly's theorem (Ash 1972), there exists a convergent subsequence of $(\hat{\Lambda}_n, \hat{\boldsymbol{\Theta}}_n)$, say $(\hat{\Lambda}_{n_k}, \hat{\boldsymbol{\Theta}}_{n_k}) \to (\Lambda^*, \boldsymbol{\Theta}^*)$ a.s. for some $\boldsymbol{\Theta}^*$ and an increasing function Λ^*. Define

$$
\Lambda_n^0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} I(\tilde{T}_j \geq s) e^{\boldsymbol{\beta}_0' x_j} \{(1-A_j) + A_j e^{\rho_0 H_j(\boldsymbol{\beta}_0)}\}},
$$

where $A_j = E(\xi_j | \boldsymbol{\Omega}, \boldsymbol{\Theta}_0, \Lambda_0)$. It is easy to show that $\Lambda_n^0(t) \to \Lambda_0(t)$ uniformly on $[0, \tau]$. In addition, we have

$$
0 \leq l_{n_k}(\hat{\Lambda}_{n_k}, \hat{\boldsymbol{\Theta}}_{n_k}) - l_{n_k}(\Lambda_{n_k}^0, \boldsymbol{\Theta}_0).
$$

The right-hand side of the above equation can be shown to converge to

$$E\left(\int_0^\tau \left[\log\left\{\frac{g(\Lambda^*, \boldsymbol{\Theta}^*)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}}{g(\Lambda_0, \boldsymbol{\Theta}_0)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}}\gamma(t)\right\} - \left\{\frac{g(\Lambda^*, \boldsymbol{\Theta}^*)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}}{g(\Lambda_0, \boldsymbol{\Theta}_0)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}}\gamma(t) - 1\right\}\right] Y(t)g(\Lambda_0, \boldsymbol{\Theta}_0)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}d\Lambda_0(t)\right), \tag{11}$$

where $Y(t) = I(\tilde{T} \geq t)$ and

$$\gamma(t) = \frac{E\left[Y(t)e^{\boldsymbol{\beta}_0' x}\{(1-A) + Ae^{\rho_0 H(\boldsymbol{\beta}_0)}\}\right]}{E\left[Y(t)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}\{(1-A^*) + A^* e^{\rho^* H(\boldsymbol{\beta}^*)}\}\right]}.$$

Here $A^* = E(\xi|\boldsymbol{\Omega}, \boldsymbol{\Theta}^*, \Lambda^*)$. Equation (11) is the negative Kullback-Leibler information, which is less or equal to 0. As a result, the Kullback-Leibler information must equal to zero, that is,

$$\begin{aligned}
&\int_0^\tau \log\{\lambda^*(t)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}g(\Lambda^*, \boldsymbol{\Theta}^*)\}dN(t) - \int_0^\tau Y(t)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}g(\Lambda^*, \boldsymbol{\Theta}^*)d\Lambda^*(t) \\
&= \int_0^\tau \log\{\lambda_0(t)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}g(\Lambda_0, \boldsymbol{\Theta}_0)\}dN(t) - \int_0^\tau Y(t)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}g(\Lambda_0, \boldsymbol{\Theta}_0)d\Lambda_0(t).
\end{aligned} \tag{12}$$

Equation (12) holds for (i) $Y(\tau) = 1$, $N(\tau) = 0$, and (ii) $Y(t) = 1$, $N(t-) = 0$ and $N(t) = 1$ for $\forall t \in (0, \tau]$. Taking the difference of these two cases, we have

$$\lambda^*(t)e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}g(\Lambda^*, \boldsymbol{\Theta}^*) = \lambda_0(t)e^{\boldsymbol{\beta}_0'\boldsymbol{x}}g(\Lambda_0, \boldsymbol{\Theta}_0), \qquad \forall t \in (0, \tau].$$

By integrating both sides from 0 to t, this implies that $\log\{S(t, \Lambda_0, \boldsymbol{\Theta}_0)\} = \log\{S(t, \Lambda^*, \boldsymbol{\Theta}^*)\}$ for all $t \in (0, \tau]$. Thus, we have

$$\begin{aligned}
&(p - p^*)\left\{e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x} + \rho_0 H(\boldsymbol{\beta}_0)}} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}}\right\} \\
&\quad - p^*\left\{e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x} + \rho^* H(\boldsymbol{\beta}^*)}} - e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x} + \rho_0 H(\boldsymbol{\beta}_0)}} + e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}}\right\} \\
&= e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}},
\end{aligned} \tag{13}$$

where $p = \frac{e^{\boldsymbol{\gamma}'\boldsymbol{x}}}{1+e^{\boldsymbol{\gamma}'\boldsymbol{x}}}$ and $p^* = \frac{e^{\boldsymbol{\gamma}^{*\prime}\boldsymbol{x}}}{1+e^{\boldsymbol{\gamma}^{*\prime}\boldsymbol{x}}}$. Suppose that $p \neq 0$ and $p^* \neq 0$. First, when $\rho_0 = 0$, the above equation equation reduces to

$$-p^* \left\{ e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}+\rho^* H(\boldsymbol{\beta}^*)}} - e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} \right\} = e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}}.$$

Because the right-hand side of is independent of $\boldsymbol{\gamma}^*$, it implies

$$e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}+\rho^* H(\boldsymbol{\beta}^*)}} - e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} = 0.$$

This equation holds only when $\rho^* = 0$ or $\boldsymbol{\beta}^* = \mathbf{0}$. However, if $\boldsymbol{\beta}^* = \mathbf{0}$, we have $e^{-\Lambda^*(\tilde{T})} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}} = 0$. Then, $\boldsymbol{\beta}_0$ must be $\mathbf{0}$ based on condition 3.1, which contradicts with our model assumption. Therefore, we have $\rho^* = 0$. Then, $e^{-\Lambda^*(\tilde{T})e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{x}}} - e^{-\Lambda_0(\tilde{T})e^{\boldsymbol{\beta}_0'\boldsymbol{x}}} = 0$. This further implies that $\Lambda^* = \Lambda_0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. Next, when $\rho_0 \neq 0$, because the right-hand side (13) is independent of $\boldsymbol{\gamma}$, we can conclude that $p = p^*$, which implies $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_0$. Furthermore, because the right-hand side of (13) is also independent of $\boldsymbol{\gamma}^*$, we can show that $\Lambda^* = \Lambda_0$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, which further implies that $\rho^* = \rho_0$. Therefore, we have $\boldsymbol{\Theta}^* = \boldsymbol{\Theta}_0$ and $\Lambda^* = \Lambda_0$. Then, by Helly's theorem, we have $\hat{\boldsymbol{\Theta}}_n \to \boldsymbol{\Theta}_0$ and $\hat{\Lambda}_n \to \Lambda_0$ a.s. Since both $\hat{\Lambda}_n$ and Λ_0 are increasing and bounded functions on $[0, \tau]$, the point-wise convergence can be strengthened to the uniform convergence. The consistency results are proved.

The asymptotic normality of $\hat{\boldsymbol{\Theta}}$ can be similarly derived following the nonparametric maximum likelihood estimation theory for censored data (e.g. Zeng and Lin 2007) and its proof is omitted for brevity. Here, we provide the definition of the information matrix $\boldsymbol{I}(\boldsymbol{\Theta}_{\mathbf{0}})$. Recall that the observed log likelihood for subject i is

$$l_i(\Lambda, \boldsymbol{\Theta}) = \delta_i \log \left\{ \lambda(\tilde{T}_i) g_i(\Lambda, \boldsymbol{\Theta}) e^{\boldsymbol{\beta}'\boldsymbol{x}_i} \right\} + \log S(\tilde{T}_i, \Lambda, \boldsymbol{\Theta}).$$

The score for Θ is

$$\frac{\partial l_i}{\partial \boldsymbol{\Theta}} = \int_0^\tau W_i(t, \Lambda, \boldsymbol{\Theta}) dM_i(t, \Lambda, \boldsymbol{\Theta}),$$

where $M_i(t,\Lambda,\boldsymbol{\Theta}) = N_i(t) - \int_0^t Y_i(s) g_i(\Lambda,\boldsymbol{\Theta}) e^{\boldsymbol{\beta}'\boldsymbol{x}_i} d\Lambda(s)$, and $W_i(t,\Lambda,\boldsymbol{\Theta}) = (\boldsymbol{x}_i' + \frac{\frac{\partial g_i(\Lambda,\boldsymbol{\Theta})}{\partial \boldsymbol{\beta}'}}{g_i(\Lambda,\boldsymbol{\Theta})}, \frac{\frac{\partial g_i(\Lambda,\boldsymbol{\Theta})}{\partial \rho}}{g_i(\Lambda,\boldsymbol{\Theta})}, \frac{\frac{\partial g_i(\Lambda,\boldsymbol{\Theta})}{\partial \boldsymbol{\gamma}'}}{g_i(\Lambda,\boldsymbol{\Theta})})'$. The efficient score for $\boldsymbol{\Theta}$ is defined as

$$S_{eff,i} = \int_0^\tau \left[W_i(t,\Lambda_0,\boldsymbol{\Theta}_0) - a_{eff,i}(t) + \{1 - g_i(\Lambda_0,\boldsymbol{\Theta}_0)\} e^{\boldsymbol{\beta}_0'\boldsymbol{x}_i} \int_0^t a_{eff,i}(s) d\Lambda_0(s) \right] dM_i(t,\Lambda_0,\boldsymbol{\Theta}_0),$$

where $a_{eff,i}$ satisfies that

$$\begin{aligned} E\Bigg(& \int_0^\tau \left[W_i(t,\Lambda_0,\boldsymbol{\Theta}_0) - a_{eff,i} + \{1 - g_i(\Lambda_0,\boldsymbol{\Theta}_0)\} e^{\boldsymbol{\beta}_0'\boldsymbol{x}_i} \int_0^t a_{eff,i}(s) d\Lambda_0(s) \right]' dM_i(t,\Lambda_0,\boldsymbol{\Theta}_0) \\ & \times \int_0^\tau \left[a^*(t) - \{1 - g_i(\Lambda_0,\boldsymbol{\Theta}_0)\} e^{\boldsymbol{\beta}_0'\boldsymbol{x}_i} \int_0^t a^*(s) d\Lambda_0(s) \right] dM_i(t,\Lambda_0,\boldsymbol{\Theta}_0) \Bigg) = 0 \end{aligned}$$

for all a^*. Then the information matrix for $\boldsymbol{\Theta}$ is defined as $\boldsymbol{I}(\boldsymbol{\Theta}_0) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n E(S_{eff,i} S_{eff,i}')$.

C Calculation of $\nabla^2 g(\hat{\boldsymbol{\Theta}}|\hat{\boldsymbol{\Theta}})$

Here, we provide the expression of $\nabla^2 g(\hat{\boldsymbol{\Theta}}|\hat{\boldsymbol{\Theta}})$ in (8). Define $D_j = e^{\boldsymbol{\beta}'\boldsymbol{x}_j}(1 - A_j^{(k)})$, $H_j = \sum_{l\neq j}^n W_{jl} \boldsymbol{\beta}'\boldsymbol{x}_l$, $\boldsymbol{G}_j = \sum_{l\neq j}^n W_{jl} \boldsymbol{x}_l$, $E_j = e^{\boldsymbol{\beta}'\boldsymbol{x}_j} A_j^{(k)} e^{\rho H_j}$, $\boldsymbol{M}_j = \boldsymbol{x}_j + \rho \boldsymbol{G}_j$, and $F_i = \sum_{j=1}^n I(\tilde{T}_j \geq$

$\tilde{T}_i)(D_j + E_j)$. Then, we have

$$\frac{\partial g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \rho} = \sum_{i=1}^{n} \delta_i \left\{ A_i^{(k)} H_i - \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i) E_j H_j}{F_i} \right\},$$

$$\frac{\partial g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{x}_i + \rho A_i^{(k)} \boldsymbol{G_i} - \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(D_j \boldsymbol{x}_j + E_j \boldsymbol{M}_j)}{F_i} \right\},$$

$$\frac{\partial g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{x}_i^* \left(A_i^{(k)} - \frac{e^{\boldsymbol{\gamma}' \boldsymbol{x}_i^*}}{1 + e^{\boldsymbol{\gamma}' \boldsymbol{x}_i^*}} \right),$$

$$\frac{\partial^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \rho^2} = -\sum_{i=1}^{n} \delta_i \frac{F_i \sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i) E_j H_j^2 - \{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i) E_j H_j\}^2}{F_i^2},$$

$$\frac{\partial^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = -\sum_{i=1}^{n} \boldsymbol{x}_i^* \boldsymbol{x}_i^{*\prime} \frac{e^{\boldsymbol{\gamma}' \boldsymbol{x}_i^*}}{(1 + e^{\boldsymbol{\gamma}' \boldsymbol{x}_i^*})^2},$$

$$\frac{\partial^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n} \left\{ \delta_i \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(D_j \boldsymbol{x}_j \boldsymbol{x}_j' + E_j \boldsymbol{M}_j \boldsymbol{M}_j')}{F_i} \right.$$
$$\left. + \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(D_j \boldsymbol{x}_j + E_j \boldsymbol{M}_j) \sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(D_j \boldsymbol{x}_j' + E_j \boldsymbol{M}_j')}{F_i^2} \right\},$$

$$\frac{\partial^2 g(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)})}{\partial \rho \partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \delta_i \left\{ A_i^{(k)} \boldsymbol{G}_i - \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(E_j H_j \boldsymbol{M}_j + E_j \boldsymbol{G}_j)}{F_i} \right.$$
$$\left. + \frac{\sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i)(D_j \boldsymbol{x}_j + E_j \boldsymbol{M}_j) \sum_{j=1}^{n} I(\tilde{T}_j \geq \tilde{T}_i) E_j H_j}{F_i^2} \right\}.$$

In addition, $\frac{\partial^2 g}{\partial \rho \partial \boldsymbol{\gamma}}$ and $\frac{\partial^2 g}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'}$ are 0. In the above expressions, setting $\mathbf{\Theta} = \hat{\mathbf{\Theta}}^{(k)} = \hat{\mathbf{\Theta}}$, we obtain the corresponding components of $\nabla^2 g(\hat{\mathbf{\Theta}}|\hat{\mathbf{\Theta}})$.

References

Anselin, L. (1980), "Estimation Methods for Spatial Autoregressive Structures," *Regional Science Dissertation and Monograph Series 8*, Cornell University, Ithaca, NY.

Ash, R. (1972), *Real Analysis and Probability*, San Diego: Academic Press.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009), "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150(1), 41–55.

Chang, X., Huang, D., and Wang, H. (2018), "A Popularity Scaled Latent Space Model for Large-Scale Directed Social Network," *Statistica Sinica*, To appear.

Chen, D., Li, J., and Chong, J. K. (2017), "Hazards regression for freemium products and services: a competing risks approach," *Journal of Statistical Computation and Simulation*, 87(9), 1863–1876.

Chen, X., Chen, Y., and Xiao, P. (2013), "The impact of sampling and network topology on the estimation of social intercorrelations," *Journal of Marketing Research*, 50(1), 95–110.

Cox, D. R. (1975), "Partial Likelihood," *Biometrika*, pp. 269–276.

Fan, A., Song, R., and Lu, W. (2016), "Change-Plane Analysis for Subgroup Detection and Sample Size Calculation," *Journal of the American Statistical Association*, pp. 1–31.

Fleming, T. R., and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: Wiley.

Frank, O., and Strauss, D. (1986), "Markov graphs," *Journal of the American Statistical Association*, 81(395), 832–842.

Hoff, P. (2003), "Random Effects Models for Network Data," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, eds. R. Breiger, K. Carley, and P. Pattison The National Academies Press, pp. 303–312.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, 97(460), 1090–1098.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), "Stochastic Blockmodels: First Steps," *Social Networks*, 5(2), 109–137.

Huang, D., Yin, J., Shi, T., and Wang, H. (2016), "A statistical model for social network labeling," *Journal of Business and Economic Statistics*, 34(3), 368–374.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008), "ergm: A package to fit, simulate and diagnose exponential-family models for networks," *Journal of Statistical Software*, 24(3), nihpa54860.

Hunter, D. R., and Lange, K. (2004), "A Tutorial on MM Algorithms," *The American Statistician*, 58(1), 30–37.

Johansen, S. (1983), "An Extension of Cox's Regression Model," *International Statistical Review/Revue Internationale de Statistique*, pp. 165–174.

Klein, J. P. (1992), "Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm," *Biometrics*, pp. 795–806.

Lange, K. (1999), *Numerical Analysis for Statisticians*, New York: Springer-Verlag.

Lee, L. f., Liu, X., and Lin, X. (2010), "Specification and Estimation of Social Interaction Models with Network Structures," *The Econometrics Journal*, 13(2), 145–176.

Li, C., and Li, H. (2008), "Network-constrained Regularization and Variable Selection for Analysis of Genomic Data," *Bioinformatics*, 24(9), 1175–1182.

Li, C., and Li, H. (2010), "Variable Selection and Regression Analysis for Graph-structured Covariates with An Application to Genomics," *The Annals of Applied Statistics*, 4(3), 1498.

Li, T., Levina, E., and Zhu, J. (2016), "Prediction Models for Network-linked Data," *arXiv preprint*, arXiv:1602.01192.

Lu, W. (2008), "Maximum likelihood estimation in the proportional hazards cure model," *Annals of the Institute of Statistical Mathematics*, 60(3), 545–574.

Lu, W., and Ying, Z. (2004), "On semiparametric transformation cure models," *Biometrika*, 91(2), 331–343.

Manski, C. F. (1993), "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60(3), 531–542.

Nowicki, K., and Snijders, T. A. B. (2001), "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, 96(455), 1077–1087.

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70(349), 120–126.

Shalizi, C. R., and Thomas, A. C. (2011), "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies," *Sociological Methods & Research*, 40(2), 211–239.

Sun, H., Lin, W., Feng, R., and Li, H. (2014), "Network-regularized High-dimensional Cox Regression for Analysis of Genomic Data," *Statistica Sinica*, 24(3), 1433.

Tobin, J. (1958), "Estimation of relationships for limited dependent variables," *Econometrica: Journal of the Econometric Society*, pp. 24–36.

Wang, Y. J., and Wong, G. Y. (1987), "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, 82(397), 8–19.

Zeng, D., and Lin, D. (2006), "Efficient estimation of semiparametric transformation models for counting processes," *Biometrika*, 93(3), 627–640.

Zeng, D., and Lin, D. (2007), "Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4), 507–564.

Zhou, J., Tu, Y., Chen, Y., and Wang, H. (2017), "Estimating Spatial Autocorrelation with Sampled Network Data," *Journal of Business & Economic Statistics*, 35(1), 130–138.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017), "Network vector autoregression," *The Annals of Statistics*, 45(3), 1096–1123.

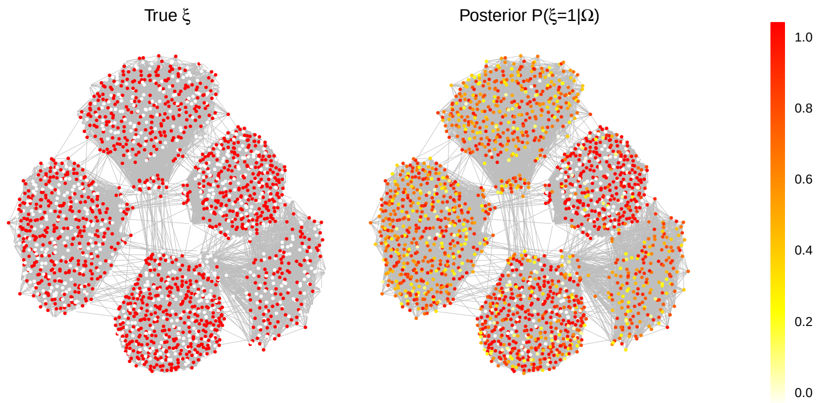

Figure 1: Plots of the true susceptible status and estimated posterior susceptible probabilities based on a simulated data. Note $\mathbf{\Omega}$ denotes the observed data, and ξ is the susceptibility indicator, whose value is represented by color intensity varying from white (corresponds to 0) to red (corresponds to 1).

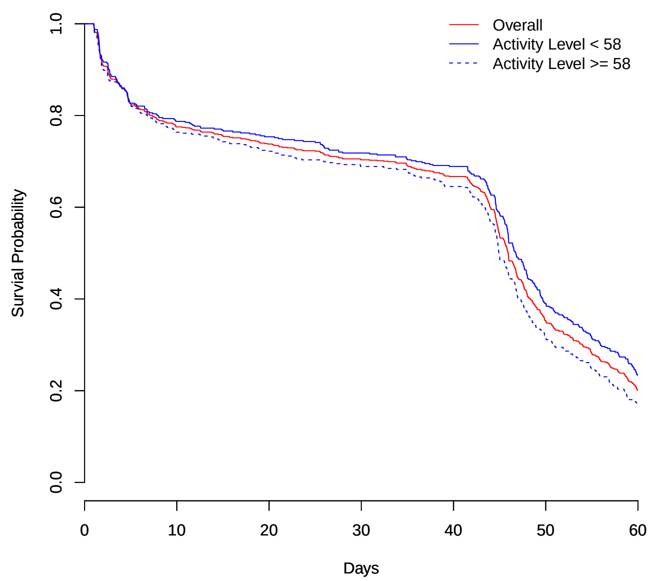

Figure 2: Kaplan-Meier curves for times to becoming the players of the mobile game.

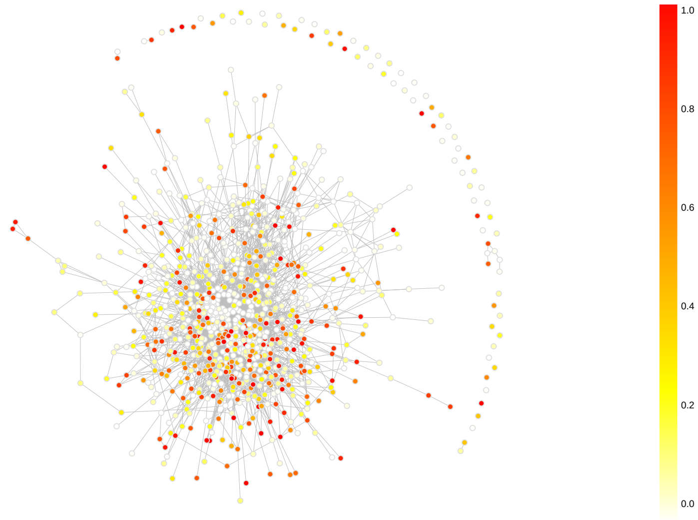

Figure 3: Plot of the network and the estimated posterior susceptible probabilities for the mobile game data. The values of the estimated posterior susceptible probabilities are represented by the color intensity (white corresponds to 0, while red corresponds to 1).

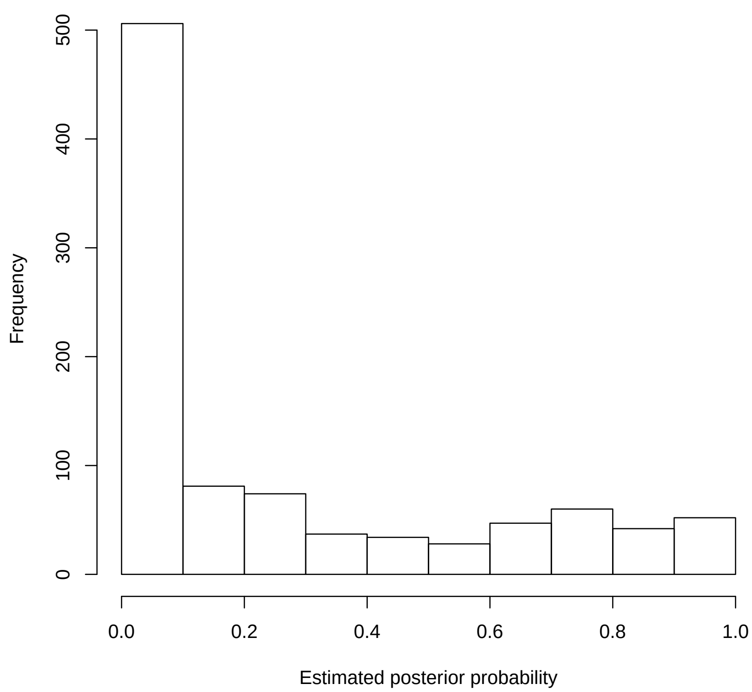

Figure 4: Distribution of the estimated posterior susceptible probabilities for the users in the mobile game data.