Statistica Sinica Preprint No: SS-2018-0222

Title	MODELING AND ESTIMATION OF CONTAGION-BASED SOCIAL NETWORK DEPENDENCE WITH TIME-TO-EVENT DATA
Manuscript ID	SS-2018-0222
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0222
Complete List of Authors	Lin Yu Wenbin Lu and Danyang Huang
Corresponding Author	Wenbin Lu
E-mail	lu@stat.ncsu.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

MODELING AND ESTIMATION OF CONTAGION-BASED SOCIAL NETWORK DEPENDENCE WITH TIME-TO-EVENT DATA

Lin Yu[1], Wenbin Lu[1], Danyang Huang[2]

[1] *North Carolina State University, and*

[2] *Renmin University*

Abstract: Social network data consists of social ties, node characteristics and behaviors over time. It is known that people who are close to each other in a social network are more likely to behave in a similar way. One of the reasons they act similarly is due to the peer influence and social contagion that acts along the network ties. A primary interest of social network data analysis is to identify the contagion-based social correlation. In this work, we model and estimate the contagion-based social network dependence based on time-to-event data. A generalized linear transformation model is proposed for the conditional survival probability at each observed event time, which uses a time-varying covariate to incorporate the network structure and quantify the contagion-based social correlation. We develop the nonparametric maximum likelihood estimation for the proposed model. The consistency and asymptotic normality of the resulting estimators for the regression parameters are established. Simulations are conducted to evaluate the empirical performance of the proposed estimators. We further apply the proposed method to analyze a time-to-event dataset about playing a popular mobile game from one of the largest online social network platforms and find that there is a significant contagion-based social correlation in times to play the game.

Key words and phrases: Contagion-based social correlation, generalized linear transformation model, nonparametric maximum likelihood estimation, social network, time-to-event data.

1. Introduction

Over the past few decades, there are plenty of studies focusing on the network research in a wide range of areas, such as psychology, geography, economics, health care, online networking, treatment recommendation, etc. It is undoubtable that the study of social networks is a quickly widening multidisciplinary area and has caught a lot of attention. Quantitative study on social networks was initially discussed by Moreno (1934), who developed a new technique called 'sociometry' to study the structure of groups and the positions of individuals within groups. Since then, the study of social networks has become quite popular in the sociological and behavioral sciences.

The field of social network analysis is fast growing, and one important research question in social network analysis is to study the social correlation between individual nodes through the node covariates. In general, the social correlation might involve three factors: homophily, social contagion and external influence (Shalizi and Thomas, 2011). The evidence of homophily can be referred as an increased rate of interactions among individuals sharing similar characteristics, and the related literatures can be found in (Lee, 2004; Lee, Liu and Lin, 2010; Zhou, Tu, Chen and Wang, 2015; Li, Levina and Zhu , 2016). In particular, spatial autoregressive models (e.g. Lee, 2004; Zhou, Tu, Chen and Wang, 2015) have been widely used to study contextual effects (one popular type of homophily) in social network dependence.

The second one is social contagion, or social influence, which occurs when one's emotions, opinions or behaviors are triggered by his/her friends' recent actions(Burt, 1987; Aral and Walker, 2011; Iyengar, Van den Bulte and Valente, 2011; Pacheco, 2012). The last one is external influence, where the external factors impact two individuals' behaviors or other measurable responses. However, sometimes these three sources might be confounded with each other (Shalizi and Thomas, 2011). In general, identifying contagion (from confounding effects such as homophily) is not feasible if one has only one snapshot of the network or measurements at a single time point. However, when multiple snapshots or measurements over time are available, it becomes feasible.

Researchers have asserted that it is important to identify the situations where the social influence is the source of correlation. In the literature, Snijders, Van der Bunt and Steglich (2010) introduce stochastic actor-based models for analyzing the dynamics of networks and behavior, which can be used to test hypotheses about many different tendencies, such as reciprocity, homophily, social influence for dynamics of behavior, and estimate the parameters expressing their strengths. In their model, it is assumed that the number of observation moments is usually between 2 and 10, which ensuring the large value of the total number of changes between consecutive observations and in turn making sure that there are enough information for estimating the parameters. However, this assumption is often violated in practice, especially given the continuous observed time-to-event data, which will carry more information than the discrete observation time points.

On the basis of continuously observed longitudinal data, Anagnostopoulos, Kumar and Mahdian (2008) propose a way to model the contagion-based social correlation and develop statistical tests for the existence of such a correlation. Specifically, they model a specified action of users in a social network via a logistic regression with the number of "active" friends included as a covariate. Here, an active friend of a user is one of his friends who had taken certain action in the past. The regression coefficient associated with this covariate measures the magnitude of the social influence. In their method, time is discretized and a logistic regression is built on each discrete time point in an ad hoc fashion.

In this work, we are interested in modeling and estimation of the contagion-based social network dependence with time-to-event data. Our work is motivated by a study of the initial playing times of a popular mobile game from one of the largest online social network platforms (the platform that provided us data has requested anonymity). This study involves 966 users. The users can send messages to their friends asking them to join the game. The endpoint of interest is the time at which a user begins to play the game since it was launched. In addition, some characteristics of these users, such as age, gender, location and active level, and their friend network are recorded. We would like to test whether an individual begins to play this game because his/her friends have started to played it, and estimate how much the influence will be. To do this, we utilize a generalized linear transformation model (Dabrowska and Doksum, 1988; Cheng, Wei and Ying, 1995) for the conditional survival probability at each observed event time and we also use a time-varying covariate for the number of active friends to model the contagion-based social network dependence. We develop an efficient estimation procedure for the model parameters based on the nonparametric maximum likelihood.

The rest of this paper is organized as follows. In Section 2, we introduce the proposed generalized linear transformation model for network-based time-to-event data and its associated data generation procedure, and describe our methodology for parameters estimation. The asymptotic properties of the proposed estimators are studied in Section 3. In Section 4, the numerical performance of the estimators is assessed using simulations. In Section 5, we further illustrate our method with an application to a data set for initial times of playing a mobile game. Section 6 concludes with discussions. All proofs are contained in the Appendix.

2. Methodology

Consider a social network with n individuals and the adjacency matrix W, where $W_{i,j} = 1$ means individual i and j are friends, and $W_{i,j} = 0$ otherwise. By convention, all the diagonal entries of W are assumed to be zero. To model the contagion-based social network dependence for time-to-event data, we consider a new data generating mechanism. Specifically, let $T_{(k)}$ denote the time to the kth event in the network. In our motivating example, this represents the kth smallest time that a user started to play the mobile game since it was launched. During a fixed study period, we totally observe M_n event times, that is, $0 < T_{(1)} < \cdots < T_{(M_n)} \leq \tau$, where τ is the total study duration. Let i_k denote the index of the user who experienced the event at time $T_{(k)}$, $k = 1, \ldots, M_n$. Here, the adjacency matrix W is assumed to be static over the study period. One of the possible future researches would be considering a dynamic social network, where the adjacency matrix is a time-dependent covariate.

In order to incorporate the network structure and quantify the contagion-based social correlation, we introduce a time-varying covariate $a_{j,k}$ for $j = 1, \ldots, n$ and $k = 1, \ldots, M_n$, which is defined as the number of active friends of individual j up to time $T_{(k)}$. Here, active friends of individual j up to time $T_{(k)}$ refer to those friends of individual j who had experienced the event before $T_{(k)}$. Let Z_j denote the p-dimensional baseline covariates of individual j. For simplicity, we define the covariates of individual j up to time $T_{(k)}$ as $X_{j,k} = (Z_j^T, g(a_{j,k}))^T$, where $g(\cdot)$ is a known non-decreasing function with $g(0) = 0$. For example, we can take $g(a) = \log(a + 1)$. Let N_k denote the index set of individuals who are at risk up to time $T_{(k)}$. Note that $i_k \in N_k$. Then, the observed data can be summarized as

$$\{i_k, T_{(k)}, (X_{j,k}, j \in N_k);\ k = 1, \ldots, M_n\}.$$

Different from the classical survival data, the above data representation is not recorded based on individuals but according to sequential event times.

Such a representation can facilitate the modeling of the contagion-based social correlation. In addition, it is assumed that the censoring can only occur at the end of the study, which is generally true for the social network study, for example, as in the considered mobile game application.

2.1 Proposed Model

In classical survival model, individual failure times can be generated independently. However, in social network study, the event time of individual i may depend on the status of his or her friends. Hence, we generate $T_{(1)}, \ldots, T_{(M_n)}$ sequentially based on the following conditional survival model. Specifically, suppose we have generated the first $(k-1)$ event times: $T_{(1)}, \ldots, T_{(k-1)}$, $k \geq 1$. Then, we know $(i_1, ..., i_{k-1})$ and the covariates $(X_{jk}, j \in N_k)$ on the interval $(T_{(k-1)}, T_{(k)}]$ for those individuals who are at risk for the kth event. Note that $T_{(0)} = 0$ and $N_1 = (1, \ldots, n)$. At the baseline, there are no active nodes in the network, that is, $a_{j,1} \equiv 0$ for all j. Therefore, $X_{j,1} = (Z_j^T, 0)^T$. To generate $T_{(k)}$, we introduce a latent event time $T_{j,k}$ at which individual j first plays the game after $T_{(k-1)}$. To be specific, all the $T_{j,k}$ for $j \in N_k$ are not observed, they are only used as latent event time to characterize the kth observed event time $T_{(k)}$. Here,

$T_{j,k}$ is generated from the following conditional survival model

$$\begin{aligned} &P(T_{j,k} > t | T_{j,k} > T_{(k-1)}, X_{j,k}) \\ &\qquad = \exp\left(- \left[G\left\{ \Lambda(t) e^{\theta^T X_{j,k}} \right\} - G\left\{ \Lambda(T_{(k-1)}) e^{\theta^T X_{j,k}} \right\} \right] \right), \quad (2.1) \end{aligned}$$

for $t > T_{k-1}$, where $\theta = (\beta^T, \beta_a)^T$ is the $(p+1)$-dimensional parameters of interest, $\Lambda(t)$ is an unspecified monotone increasing function with $\Lambda(0) = 0$ and $G(\cdot)$ is a specified monotone increasing transformation function, for example, a class of logarithmic transformation

$$G(x) = \begin{cases} \frac{1}{s} \log(1 + sx), & s > 0 \\ x. & s = 0 \end{cases} \quad (2.2)$$

where s is a prespecified parameter. Here, $s = 0$ corresponds to the proportional hazards model while $s = 1$ refers to the proportional odds model. It can be seen that the above model is a generalization of the linear transformation model of Zeng and Lin (2006) for the conditional survival probability. Note that the parameter β_a measures the magnitude of the contagion-based social correlation. In model 2.2, the simultaneous estimation of s and model parameters will be challenging since the information about s may be weak. In practice, people often select s using some information criteria as done in our real data application presented in Section 5.

Then, we define

$$T_{(k)} = \min_{j \in N_k} T_{j,k}, \quad i_k = \arg\min_{j \in N_k} T_{j,k}.$$

In addition, the numbers of active friends are updated by $a_{j,k+1} = a_{j,k} + W_{j,i_k}$ for $j \in N_{k+1} = N_k \backslash \{i_k\}$, which stay the same on the interval $(T_{(k)}, T_{(k+1)}]$. Repeat the above step until all the event times are generated. Based on the proposed data generating mechanism, the observed log-likelihood is given by

$$\ell_n(\theta, \Lambda) = \sum_{k=1}^{M_n} \Bigg(\log \lambda(T_{(k)}) + \theta^T X_{i_k,k} + \log \dot{G} \left\{ \Lambda(T_{(k)}-) e^{\theta^T X_{i_k,k}} \right\} \\ - \sum_{j \in N_k} \left[G \left\{ \Lambda(T_{(k)}) e^{\theta^T X_{j,k}} \right\} - G \left\{ \Lambda(T_{(k-1)}) e^{\theta^T X_{j,k}} \right\} \right] \Bigg), \quad (2.3)$$

where $\lambda(t) = d\Lambda(t)/dt$ and $\dot{G}(u) = dG(u)/du$.

2.2 Nonparametric Maximum Likelihood Estimation

Here, we derive the nonparametric maximum likelihood estimation based on the likelihood function (2.3). The maximum of (2.3) does not exist if $\Lambda(\cdot)$ is restricted to be absolutely continuous. As widely studied in the literature for the nonparametric maximum likelihood estimation, we assume that $\Lambda(\cdot)$

is a non-decreasing step function with jumps only at observed event times $T_{(1)}, \ldots, T_{(M_n)}$. Let $\Lambda\{T_{(k)}\}$ be the jump size at time $T_{(k)}$. Then, we have $\lambda(T_{(k)}) = \Lambda\{T_{(k)}\}$ for $k = 1, \ldots, M_n$.

To simplify the estimation, we consider a reparameterization. Define $\gamma_k = \log \Lambda\{T_{(k)}\}$. We have $\Lambda(T_{(k)}) = \sum_{\ell=1}^{k} e^{\gamma_\ell} = e^{\gamma_k} + \Lambda(T_{(k-1)})$, $k = 1, \ldots, M_n$. Thus, the log-likelihood function can be rewritten as

$$\ell_n(\theta, \gamma) = \sum_{k=1}^{M_n} \left(\gamma_k + \theta^T X_{i_k,k} + \log \dot{G} \left\{ (\sum_{\ell=1}^{k-1} e^{\gamma_\ell}) e^{\theta^T X_{i_k,k}} \right\} \right. \\ \left. - \sum_{j \in N_k} \left[G \left\{ (\sum_{\ell=1}^{k} e^{\gamma_\ell}) e^{\theta^T X_{j,k}} \right\} - G \left\{ (\sum_{\ell=1}^{k-1} e^{\gamma_\ell}) e^{\theta^T X_{j,k}} \right\} \right] \right). \tag{2.4}$$

In the following, we consider the logarithmic transformation function (2.2) and present the estimation of the parameters (θ, γ) in two cases: $s = 0$ and $s > 0$. However, the proposed estimation method can be easily extended to other specified transformation functions.

First, consider the case $s = 0$ with $G(x) = x$ and $\dot{G}(x) \equiv 1$. Then, the observed log-likelihood is reduced to

$$\ell_n(\theta, \gamma) = \sum_{k=1}^{M_n} \left(\gamma_k + \theta^T X_{i_k,k} - e^{\gamma_k} \sum_{j \in N_k} e^{\theta^T X_{j,k}} \right). \tag{2.5}$$

Taking the derivative of (2.5) with respect to γ_k and setting them equal to 0, we can obtain an explicit solution for γ_k as $\hat{\gamma_k}(\theta) = -\log\left(\sum_{j\in N_k} e^{\theta^T X_{j,k}}\right)$, for $k = 1, \ldots, M_n$. Then, plugging $\hat{\gamma_k}(\theta)$ back into model (2.5), we obtain the profile log-likelihood for θ

$$p\ell_n(\theta) = \sum_{k=1}^{M_n} \left\{ \theta^T X_{i_k,k} - \log\left(\sum_{j\in N_k} e^{\theta^T X_{j,k}}\right)\right\}, \tag{2.6}$$

which is similar to the log partial likelihood function for the proportional hazards model. Let $\hat{\theta}_n$ denote the resulting maximizer of θ. The asymptotic variance-covariance matrix of $\hat{\theta}_n$ can be estimated by $\mathrm{I}^{-1}(\hat{\theta}_n)$, where $\mathrm{I}(\hat{\theta}_n)$ is the negative of the second derivative of $p\ell_n(\theta)$ with respect to θ.

Next, we consider the case $s > 0$. The log-likelihood function (2.4) reduces to

$$\begin{aligned}\ell_n(\theta,\gamma) = \sum_{k=1}^{M_n} \Bigg(&\gamma_k + \theta^T X_{i_k,k} - \log\left\{1 + s\left(\sum_{\ell=1}^{k-1} e^{\gamma_\ell}\right) e^{\theta^T X_{i_k,k}}\right\} \\ &- \frac{1}{s}\sum_{j\in N_k}\left[\log\left\{1 + s\left(\sum_{\ell=1}^{k} e^{\gamma_\ell}\right) e^{\theta^T X_{j,k}}\right\} - \log\left\{1 + s\left(\sum_{\ell=1}^{k-1} e^{\gamma_\ell}\right) e^{\theta^T X_{j,k}}\right\}\right]\Bigg).\end{aligned} \tag{2.7}$$

Then, the estimates of θ and γ_k's can be obtained by the following procedure.

Step 1. Choose the initial estimator $\theta^{(0)}$, for example, $\theta^{(0)} = 0$;

Step 2. Given $\theta^{(0)}$, we solve $\gamma_k^{(1)}$ sequentially by maximizing $\ell_n(\theta^{(0)}, \gamma_1, \ldots, \gamma_{M_n})$ using a coordinate descent algorithm;

Step 3. Given $\{\gamma_k^{(1)}, k = 1, \ldots, M_n\}$, we further update $\theta^{(1)}$ by maximizing $\ell_n(\theta, \gamma^{(1)}, \ldots, \gamma_{M_n}^{(1)})$;

Step 4. Iterate *Step* 2 and *Step* 3 until a convergence criterion is met.

It can be shown that the objective function ℓ_n is not convex with respect to θ and γ_k's, for $k = 1, \ldots, M_n$. However, empirically we found that our algorithm usually converges within 20 iterations and works well provided that the starting values are not far from the truth. Let $\hat{\theta}_n$ and $\hat{\Lambda}_n$ denote the resulting estimators of θ and Λ, respectively, at convergence. The asymptotic variance-covariance matrix of $\hat{\theta}_n$ can be obtained by the following numerical differentiation method. For a small value $\delta > 0$, let $\hat{\gamma}_{n,j}^{+}$ and $\hat{\gamma}_{n,j}^{-}$ denote the solutions for γ obtained by maximizing $\ell_n(\theta, \gamma)$ with θ fixed at $\hat{\theta}_n + \delta e_j$ and $\hat{\theta}_n - \delta e_j$, respectively, where e_j is a $(p+1)$-vector with the jth component as 1 and others as 0, $j = 1, \ldots, p+1$. Let $\ell_{n,k}(\theta, \gamma)$ denote the kth summand in $\ell_n(\theta, \gamma)$. Define $S_{k,j}(\hat{\theta}_n) = \{\ell_{n,k}(\hat{\theta}_n + \delta e_j, \hat{\gamma}_{n,j}^{+}) - \ell_{n,k}(\hat{\theta}_n - \delta e_j, \hat{\gamma}_{n,j}^{-})\}/(2\delta)$ and $S_k(\hat{\theta}_n) = \{S_{k,1}(\hat{\theta}_n), \ldots, S_{k,p+1}(\hat{\theta}_n)\}^T$. Then, the observed information matrix is $I(\hat{\theta}_n) = \sum_{k=1}^{M_n} S_k(\hat{\theta}_n) S_k(\hat{\theta}_n)^T$ and the asymptotic variance-covariance matrix of $\hat{\theta}_n$ can be estimated by $\{I(\hat{\theta}_n)\}^{-1}$.

3. Asymptotic Properties

Denote the true values of θ and Λ by θ_0 and Λ_0, respectively. To establish the asymptotic properties of the proposed estimators, we assume the following conditions:

Condition 1. The function $\Lambda_0(t)$ is strictly increasing and continuously differentiable with $\Lambda_0(\tau) < \infty$, and the parameters θ_0 lie in the interior of a compact set $\mathcal{C}$.

Condition 2. The covariates vectors $X_{j,k}$ are bounded in the sense that $P(|X_{j,k}| < m) = 1$ for some positive constant m, for any j, k, as n goes to ∞. In addition, if there exists a vector γ and a deterministic function $A(t)$ such that $A(t) + \gamma^T X_{j,k} = 0$ with probability one, then $\gamma = 0$ and $A(t) = 0$.

Condition 3. The information matrix $I(\theta_0)$ defined in the Appendix is finite and positive definite.

Note that Conditions 1-3 are commonly assumed in the literature for establishing the asymptotic properties of nonparametric maximum likelihood estimators in survival models (e.g. Zeng and Lin (2006)). In particular, the boundedness of covariates assumed in Condition 2 is satisfied when the following two conditions hold: (i) the baseline covariates Z are bounded; (ii) the number of friends of each node is bounded by a constant as the number of nodes n goes to infinity, i.e. the social network is very sparse. This ensures that the number of active friends will not diverge to infinity since the study period is finite. Such an assumption can facilitate the derivation of the asymptotic results. In network-based causal inference problems, the boundedness assumption on the node degrees is almost necessary. As studied in van der Laan (2014); Ogburn, Sofrygin, Díaz and van der Laan (2017), a denser network introduces stronger correlation so valid statistical inference is only possible for very sparse networks.

Theorem 1 (Consistency)**.** *Assumptions conditions 1-2 hold. We have, as n goes to ∞,*

$$\sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0 \ a.s. \quad \text{and} \ \ ||\hat{\theta}_n - \theta_0||_2 \to 0 \ a.s.$$

Theorem 2 (Asymptotic Normality)**.** *Assume conditions 1-3 hold. We have $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to a multivariate normal with mean 0 and variance $\{I(\theta_0)\}^{-1}$, as n goes to ∞.*

4. Simulation Studies

In this section, we illustrate the performance of the proposed estimators under several settings. Here, we consider a social network with different network structures for adjacency matrix W:

1. Random Graph (RG): $P(W_{i,j} = 1) = 0.1$ for $i \neq j$, and $n = 1000$;

2. Stochastic Block Model (SBM): consider 3 blocks with sample sizes $(200, 300, 500)$ within each block, and $P(W_{i,j} = 1|\text{within block}) = 0.1$ and $P(W_{i,j} = 1|\text{between block}) = 0.005$, for $i \neq j$;

3. The network from the mobile game data application with $n = 966$;

4. Degree-2 networks with $n = 900$:

   - a ring network in which all nodes have exactly degree 2;
   - a network of isolated triangles such that each node has degree 2;
   - a random regular-2 network.

Event times $T_{(k)}$'s are generated sequentially following the descriptions given in Section 2.1. Here, we consider a single baseline covariate Z generated from a standard normal distribution and a logarithm transformation of the time-varying covariate, $g(a) = \log(a + 1)$. We choose the regression parameters as $\beta = 0.5$ and $\beta_a = 0, 0.01, 0.05$ or 0.1. Here, β_a measures the magnitude of the social influence. In addition, we set $\Lambda(t) = \lambda t$ with $\lambda = 0.01$. We consider the link function $G(x) = \frac{1}{s}\log(1 + sx)$ with $s =$

$(0, 0.5, 1)$. The study duration τ is chosen to yield the total number of events $M_n = \alpha \times n$, where the censoring rate is chosen as $\alpha = 60\%$, or 80%.

We conduct 1000 replicates for each setting. The results of the RG model, SBM, the observed network in mobile game data application, and degree-2 networks are shown in Table 1, Table 2, Table 3 and Table 4, respectively. We observe that in all settings the proposed estimators are nearly unbiased, the standard error estimators are close to the standard deviations of the estimators, and the empirical coverage probabilities of the 95% Wald-type confidence intervals are close to the nominal level. In particular, the proposed estimators have very comparable variances under all three degree-2 networks. Since the analytic form of the asymptotic variance of the proposed estimators is very complicated, it is hard to tell how it depends on the network structure. However, based on the simulation results, the node degree distribution may play an important effect here, that is, the proposed estimators tend to have comparable variances when the node degrees of networks are comparable.

5. Analysis of Mobile Game Data

We apply our method to analyze a time-to-event data about playing a popular mobile game from one of the largest online social network platforms (the platform that provided us data has requested anonymity). The

Table 1: Simulation results for RG. SE, mean of estimated standard errors; SD, standard deviations of the estimates; CP, empirical coverage probability of 95% Wald-type confidence intervals.

		$M_n = 600$				$M_n = 800$			
$s = 0$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.500	0.043	0.044	0.950	0.500	0.038	0.039	0.943
β_a	0	0.002	0.179	0.185	0.946	0.002	0.172	0.178	0.942
β	0.5	0.500	0.043	0.044	0.947	0.500	0.038	0.039	0.946
β_a	0.01	0.012	0.180	0.185	0.945	0.012	0.173	0.178	0.943
β	0.5	0.500	0.043	0.044	0.949	0.500	0.038	0.039	0.941
β_a	0.05	0.051	0.180	0.186	0.945	0.050	0.173	0.178	0.946
β	0.5	0.500	0.043	0.044	0.946	0.500	0.038	0.039	0.946
β_a	0.1	0.101	0.181	0.187	0.944	0.100	0.174	0.179	0.945
$s = 0.5$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.502	0.053	0.053	0.954	0.502	0.050	0.050	0.958
β_a	0	0.005	0.196	0.151	0.970	0.008	0.191	0.148	0.965
β	0.5	0.502	0.054	0.053	0.957	0.502	0.051	0.050	0.958
β_a	0.01	0.016	0.197	0.153	0.970	0.019	0.191	0.149	0.962
β	0.5	0.502	0.054	0.054	0.956	0.502	0.051	0.051	0.959
β_a	0.05	0.058	0.198	0.154	0.968	0.060	0.193	0.151	0.965
β	0.5	0.503	0.055	0.056	0.954	0.503	0.052	0.052	0.956
β_a	0.1	0.110	0.200	0.160	0.964	0.111	0.194	0.156	0.961
$s = 1$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.506	0.063	0.067	0.948	0.505	0.062	0.064	0.954
β_a	0	0.017	0.207	0.207	0.935	0.015	0.203	0.197	0.942
β	0.5	0.506	0.064	0.067	0.949	0.505	0.062	0.064	0.955
β_a	0.01	0.027	0.208	0.208	0.936	0.026	0.204	0.199	0.939
β	0.5	0.506	0.064	0.068	0.948	0.506	0.063	0.065	0.955
β_a	0.05	0.068	0.209	0.211	0.936	0.068	0.205	0.201	0.936
β	0.5	0.507	0.066	0.070	0.951	0.506	0.064	0.067	0.955
β_a	0.1	0.120	0.212	0.214	0.927	0.119	0.207	0.204	0.936

Table 2: Simulation results for SBM. SE, mean of estimated standard errors; SD, standard deviations of the estimates; CP, empirical coverage probability of 95% Wald-type confidence intervals.

		$M_n = 600$				$M_n = 800$			
$s = 0$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.500	0.043	0.044	0.953	0.500	0.038	0.039	0.945
β_a	0	-0.001	0.008	0.008	0.945	0.000	0.005	0.005	0.951
β	0.5	0.501	0.043	0.044	0.957	0.500	0.038	0.039	0.950
β_a	0.01	0.009	0.007	0.008	0.948	0.010	0.005	0.005	0.952
β	0.5	0.498	0.043	0.043	0.956	0.498	0.038	0.038	0.947
β_a	0.05	0.050	0.006	0.006	0.948	0.050	0.004	0.004	0.954
β	0.5	0.500	0.043	0.043	0.958	0.500	0.038	0.038	0.952
β_a	0.1	0.100	0.006	0.006	0.955	0.100	0.004	0.004	0.955
$s = 0.5$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.500	0.052	0.052	0.960	0.500	0.049	0.048	0.960
β_a	0	-0.001	0.113	0.101	0.955	-0.002	0.107	0.093	0.963
β	0.5	0.500	0.052	0.052	0.964	0.500	0.049	0.048	0.960
β_a	0.01	0.010	0.113	0.102	0.957	0.008	0.107	0.093	0.963
β	0.5	0.500	0.053	0.052	0.963	0.499	0.049	0.049	0.961
β_a	0.05	0.052	0.113	0.101	0.960	0.049	0.107	0.094	0.962
β	0.5	0.500	0.053	0.052	0.959	0.499	0.049	0.049	0.956
β_a	0.1	0.100	0.113	0.100	0.955	0.099	0.107	0.094	0.967
$s = 1$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.502	0.060	0.061	0.948	0.501	0.058	0.059	0.947
β_a	0	0.006	0.127	0.0124	0.945	0.006	0.124	0.121	0.945
β	0.5	0.502	0.061	0.061	0.948	0.501	0.058	0.059	0.948
β_a	0.01	0.016	0.127	0.125	0.940	0.016	0.124	0.122	0.943
β	0.5	0.501	0.061	0.062	0.950	0.501	0.059	0.060	0.952
β_a	0.05	0.056	0.128	0.127	0.942	0.056	0.125	0.124	0.939
β	0.5	0.501	0.062	0.062	0.952	0.501	0.059	0.060	0.956
β_a	0.1	0.106	0.129	0.129	0.940	0.108	0.126	0.126	0.937

Table 3: Simulation results for the observed network in the mobile game data application. SE, mean of estimated standard errors; SD, standard deviations of the estimates; CP, empirical coverage probability of 95% Wald-type confidence intervals.

		$M_n/n = 60\%$				$M_n/n = 80\%$			
$s = 0$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.502	0.044	0.045	0.942	0.501	0.039	0.040	0.943
β_a	0	-0.001	0.011	0.011	0.946	0.000	0.007	0.007	0.948
β	0.5	0.503	0.044	0.045	0.948	0.501	0.039	0.041	0.946
β_a	0.01	0.009	0.010	0.010	0.944	0.010	0.007	0.007	0.937
β	0.5	0.503	0.044	0.045	0.946	0.502	0.039	0.041	0.936
β_a	0.05	0.051	0.009	0.009	0.950	0.051	0.007	0.007	0.955
β	0.5	0.502	0.044	0.044	0.955	0.501	0.039	0.040	0.949
β_a	0.1	0.101	0.009	0.009	0.946	0.101	0.008	0.008	0.951
$s = 0.5$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.503	0.053	0.053	0.953	0.501	0.049	0.049	0.949
β_a	0	-0.003	0.072	0.073	0.950	-0.002	0.063	0.064	0.947
β	0.5	0.503	0.053	0.053	0.953	0.501	0.049	0.049	0.946
β_a	0.01	0.006	0.072	0.073	0.948	0.007	0.063	0.064	0.949
β	0.5	0.503	0.053	0.053	0.953	0.501	0.049	0.050	0.951
β_a	0.05	0.047	0.072	0.074	0.951	0.048	0.063	0.065	0.945
β	0.5	0.503	0.053	0.053	0.953	0.501	0.049	0.050	0.953
β_a	0.1	0.097	0.072	0.073	0.952	0.097	0.064	0.066	0.939
$s = 1$	Trues	Estimates	SE	SD	CP	Estimates	SE	SD	CP
β	0.5	0.502	0.060	0.060	0.953	0.501	0.058	0.057	0.949
β_a	0	-0.003	0.088	0.091	0.941	-0.002	0.082	0.84	0.942
β	0.5	0.502	0.060	0.060	0.953	0.501	0.058	0.057	0.950
β_a	0.01	0.007	0.088	0.092	0.944	0.007	0.082	0.085	0.943
β	0.5	0.503	0.061	0.060	0.952	0.501	0.058	0.058	0.949
β_a	0.05	0.048	0.089	0.093	0.943	0.048	0.083	0.087	0.941
β	0.5	0.503	0.061	0.061	0.949	0.502	0.058	0.059	0.947
β_a	0.1	0.098	0.090	0.092	0.943	0.098	0.084	0.087	0.936

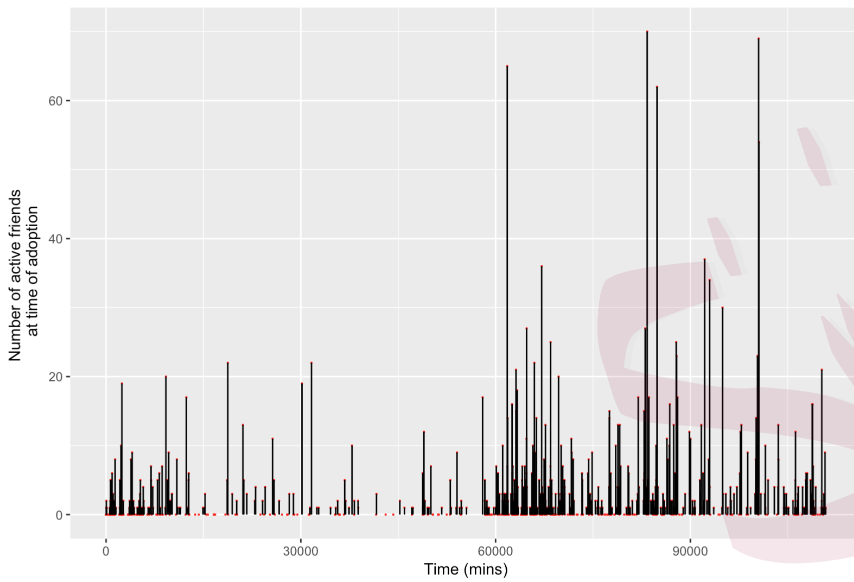
Table 4: Simulation results for degree-2 networks with $M_n = 800$ and $n = 900$. SE, mean of estimated standard errors; SD, standard deviations of the estimates; CP, empirical coverage probability of 95% Wald-type confidence intervals.

		ring net			all isolated triangles			random regular-2		
$s = 0$	Trues	Est	SE	CP	Est	SE	CP	Est	SE	CP
β	0.5	0.500	0.038	0.946	0.500	0.038	9.948	0.503	0.039	0.950
β_a	0	0.001	0.059	0.947	-0.002	0.059	0.954	-0.001	0.059	0.966
β	0.5	0.500	0.038	0.952	0.500	0.038	0.953	0.503	0.039	0.954
β_a	0.01	0.011	0.059	0.944	0.008	0.059	0.956	0.009	0.059	0.965
β	0.5	0.500	0.038	0.953	0.500	0.038	0.951	0.504	0.039	0.960
β_a	0.05	0.050	0.059	0.943	0.048	0.058	0.949	0.049	0.059	0.962
β	0.5	0.500	0.038	0.951	0.501	0.038	0.954	0.503	0.039	0.957
β_a	0.1	0.100	0.058	0.953	0.099	0.057	0.949	0.099	0.058	0.960
$s = 0.5$	Trues	Est	SE	CP	Est	SE	CP	Est	SE	CP
β	0.5	0.501	0.050	0.946	0.501	0.050	0.944	0.504	0.050	0.954
β_a	0	0.000	0.140	0.952	-0.006	0.140	0.956	-0.001	0.140	0.962
β	0.5	0.501	0.050	0.943	0.501	0.050	0.947	0.504	0.050	0.951
β_a	0.01	0.010	0.140	0.954	0.003	0.140	0.955	0.010	0.140	0.963
β	0.5	0.501	0.050	0.949	0.501	0.050	0.945	0.504	0.050	0.954
β_a	0.05	0.049	0.140	0.958	0.043	0.140	0.956	0.052	0.140	0.968
β	0.5	0.501	0.050	0.948	0.501	0.050	0.940	0.504	0.050	0.957
β_a	0.1	0.099	0.140	0.960	0.093	0.140	0.950	0.100	0.140	0.965
$s = 1$	Trues	Est	SE	CP	Est	SE	CP	Est	SE	CP
β	0.5	0.502	0.060	0.941	0.501	0.060	0.946	0.504	0.060	0.954
β_a	0	-0.001	0.174	0.957	-0.009	0.174	0.950	0.000	0.174	0.952
β	0.5	0.502	0.060	0.945	0.501	0.060	0.946	0.504	0.060	0.951
β_a	0.01	0.009	0.174	0.956	0.002	0.174	0.954	0.010	0.174	0.952
β	0.5	0.502	0.060	0.944	0.501	0.060	0.945	0.505	0.060	0.946
β_a	0.05	0.047	0.174	0.953	0.040	0.174	0.952	0.051	0.174	0.950
β	0.5	0.501	0.060	0.950	0.502	0.060	0.950	0.506	0.060	0.946
β_a	0.1	0.098	0.175	0.950	0.090	0.174	0.954	0.099	0.175	0.955

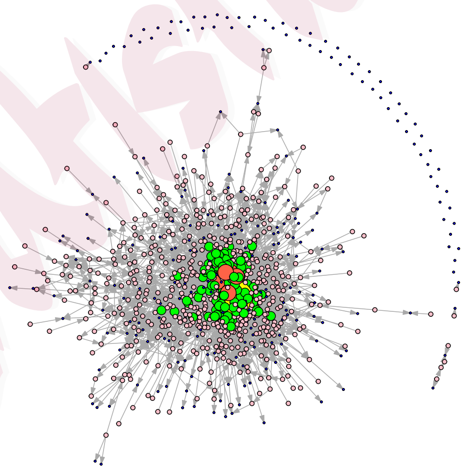
study involves 966 individuals over the 77 days duration. The friendship connection between individuals are known, which can be represented as the adjacency matrix W. The time at which each individual began to play the mobile game since it was launched are recorded. Figure 1a shows the number of active friends at the time of adoption. As expected, at the adoption times of later events, players tend to have more active friends. In addition, the baseline information, such as age, gender, location, and active level, are recorded. In total there are 241 isolated nodes in the network. We divide individuals into five groups based on the number of active friends at the end of study, and show them in different color in Figure 1b. The majority of individuals belong to the second group with the number of active friends greater than 0 and less than or equal to 10. Note that there is one individual with the number of active friends greater than 100, denoted by the yellow dot in Figure 1b.

We fit the proposed models with the scaled age and gender included as the baseline covariates. As in simulations, we consider the logarithm transformation of the time-varying covariate, $g(a_i) = \log(a_i + 1)$, and the link function $G(x) = \frac{1}{s}\log(1 + sx)$ with $s = (0, 0.5, 0.75, 1)$. In addition, we include the scale-free model using the logarithm transformation of the proportion of active friends, $h^*(a_i) = \log(a_i/\sum_j W_{i,j} + 1)$, for comparison. The estimation results of the fitted models are given in Table 5. We also report the log likelihood value of the fitted models. Based on the results of the fitted models, the corresponding estimated coefficients for the contagion-based network dependence parameter β_a are all positive, indicating that as the number/proportion of active friends increases, an individual is more likely to start playing the game soon. However, the results based on the number of active friends are more significant than those based on the proportion of active friends. In addition, the models based on the number of active friends with $s = 0.5, 0.75$ or 1 have the best fit in terms of the likelihood values (they are almost the same), which are larger than those of the models based on the proportion of active friends.

Next, we evaluate the prediction performance of our proposed network structure-based method and the standard Cox proportional hazard model without incorporating the network information. Specifically, we consider the estimated model based on the number of active friends with $s = 1$, in which the estimated coefficients are equal to -0.017, 0.059 and 0.343 for the scaled age, gender, and $g(a)$ respectively. To evaluate the prediction performance of a fitted model, we do the following three steps. (1) Simulate $T_{j,k}$'s for all $j \in N_k$ up to time $T_{(k)}$ based on the estimated model using our method and the standard Cox proportional hazard model fit, i.e. $T_{j,k}^{pro}$'s



(a) Plot of time by the number of active friends at the time of adoption.



(b) Number of active-friends at the end of study with blue dot$=0$; pink dot $\in (0,10]$; green dot $\in (10,50]$; orange dot $\in (50,100]$; yellow dot>100.

Figure 1: Network visualization for mobile game data

Table 5: Analysis of mobile game data. log-LH, log likelihood value of the fitted model; $g(a_i) = \log(a_i + 1)$, where a_i is the number of active friends for user i; $h^*(a_i) = \log(a_i / \sum_j W_{i,j} + 1)$.

Original scale		age	gender	$g(a)$	log-LH
	Estimation	0.018	0.024	0.108	
$s = 0$	SE	0.036	0.032	0.039	-6639.250
	Z statistics	0.501	0.731	2.803	
	Estimation	-0.017	0.059	0.342	
$s = 0.5$	SE	0.052	0.057	0.086	-6631.869
	Z statistics	-0.331	1.041	3.976	
	Estimation	-0.017	0.059	0.343	
$s = 0.75$	SE	0.052	0.057	0.086	-6631.869
	Z statistics	-0.331	1.045	3.980	
	Estimation	-0.017	0.059	0.343	
$s = 1$	SE	0.052	0.057	0.086	-6631.868
	Z statistics	-0.334	1.040	3.978	
Scale-free		**age**	**gender**	**$h^*(a)$**	**log-LH**
	Estimation	0.036	0.023	0.101	
$s = 0$	SE	0.034	0.032	0.131	-6642.786
	Z statistics	1.038	0.697	0.770	
	Estimation	0.020	0.056	0.536	
$s = 0.5$	SE	0.052	0.056	0.304	-6637.526
	Z statistics	0.381	0.992	1.764	
	Estimation	0.019	0.056	0.542	
$s = 0.75$	SE	0.052	0.056	0.304	-6637.527
	Z statistics	0.364	0.990	1.781	
	Estimation	0.019	0.056	0.540	
$s = 1$	SE	0.052	0.056	0.304	-6637.527
	Z statistics	0.365	0.998	1.776	

and $T_{j,k}^{cox}$'s. (2) Sort all the event times $T_{j,k}$'s of subjects in the at-risk set N_k, and find the rank of $T_{i_k,k}^{b}$ among the ordered event times $T_{j,k}^{b}$'s, where $b = pro$ or cox and i_k is the user index who actually adopted the action at time $T_{(k)}$ in the data. Let R_k^{pro} and R_k^{cox} denote the ranks correspondingly. (3) Compute the proportion of rank comparison: $M_n^{-1}\sum_{k=1}^{M_n} I(R_k^{pro} < R_k^{cox})$.

It is expected that a model with a better fit should tend to have smaller ranks at observed event times $T_{(k)}$, $k = 1, \ldots, M_n$. The reason is that if the rank R_k^b is small, the user who actually adopted the action at time $T_{(k)}$ will be predicted more likely to be the person taking the action at time $T_{(k)}$ based on the fitted model b. Therefore, if the proposed model has a better fit than the standard Cox model, the proportion computed in Step 3 should be much larger than 0.5. We conduct 1000 Monte Carlo replications of the above three-step procedure and find that the average proportion is 0.874 with standard deviation 0.010. Figure 2 shows the boxplot of the proportions over 1000 replications. In conclusion, the proposed model by incorporating network structure can improve the prediction performance than the standard Cox model ignoring the network structure.

In addition, we compare the performance of our proposed method with the stochastic actor-based model (Snijders, Van der Bunt and Steglich, 2010). To apply this model, certain assumptions need to be satisfied. One

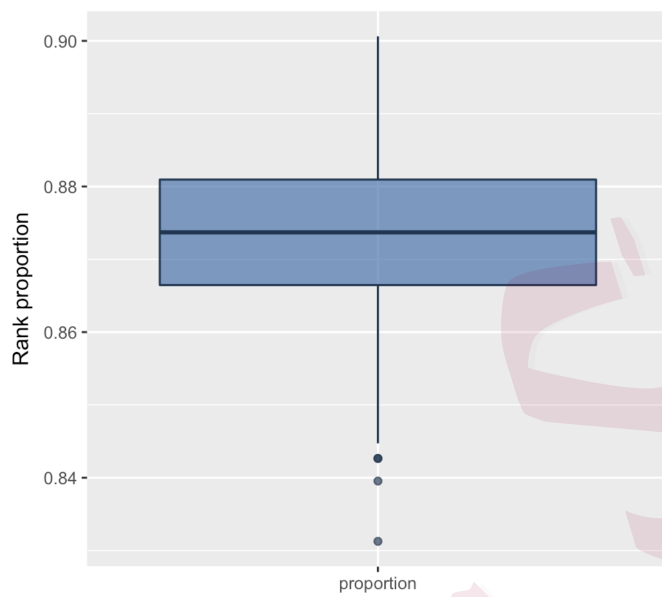


Figure 2: Boxplot of the proportions for rank comparison over 1000 replications.

of the important assumptions is the total number of changes between consecutive observations should be large enough to provide strong information for estimating the parameters, and the number of observation points is usually between 2 and 10. In the mobile game dataset, there are total 966 observation points, which means only one change between each consecutive pair of observations. Due to the expensive computation and limited memory storage, we only split the mobile game dataset into 3, 4, 5, and 8 observation points, and fit the stochastic actor-based models for the resulting dataset by using *RSiena* package (Ripley, Snijders, Boda and Vörös, 2011). Specifically, we consider the age and gender effects for the behavior dynamic process, and the total exposure effect (Greenan, 2015) for studying the contagion-based network dependence. Here, the total exposure refers to the total number of active friends that tied to each node and its effect measures the magnitude of the contagion-based network dependence. The results based on different number of observation points are shown in Table 6. The overall maximum convergence t-ratios are also reported, which are used for assessing the convergence of the algorithm. The results imply that the convergence is better in the model with 8 observation points. Different from our findings, the total exposure effects are not significant in all settings. One possible reason is that after splitting the data into several observation time points, it might loose some information and thus lead to less efficient estimators. Another reason is that it dose not make effective use of the survival model information for fitting the time-to-event data.

6. Conclusion and Discussion

In this work, we propose a new way of modeling and estimation for contagion-based social dependence with time-to-event data. It can be extended to accommodate multiple events, such as network-based recurrent event data, by incorporating both self-exciting and contagion-based social exciting processes. This warrants a thorough study for future research. Another interesting direction for future research would be extending our

Table 6: Analysis of mobile game data with stochastic actor-based models. n_obs, the number of observation points; total_exp, the total exposure; mc t-ratio, the overall maximum convergence t-ratio; time, the total runtime of the algorithm.

		age	gender	total_exp	mc t-ratio	time
	Estimation	-0.008	0.013	0.012		
n_obs = 3	SE	0.008	0.104	0.015	1.195	3 hours
	Z statistics	-0.904	0.124	0.801		
	Estimation	0.004	0.048	0.001		
n_obs = 4	SE	0.007	0.105	0.010	1.054	3.3 hours
	Z statistics	0.486	0.463	0.096		
	Estimation	-0.002	0.043	-0.002		
n_obs = 5	SE	0.008	0.088	0.009	0.918	4 hours
	Z statistics	-0.244	0.488	-0.213		
	Estimation	0.006	0.064	0.003		
n_obs = 8	SE	0.007	0.082	0.008	0.667	5 hours
	Z statistics	0.821	0.779	0.329		

method to dynamic social networks with evolving friendship connections. For example, individuals could become friends with others during the study period, which will change the value of the adjacency matrix. However, as long as the time-dependent adjacency matrices are correctly specified, the proposed estimation method can be easily modified to incorporate it.

In current work, it is assumed that the censuring only occurs at the end of the study, which is a reasonable assumption in the considered mobile game data application. However, if a subject can be censored in the middle of the study, it is generally not known whether the potential event of this subject occurred in the future will have an effect on his or her friends. If we can assume that the effect of a subject on connected nodes will disappear after he or she is censored, the proposed conditional survival model and its associated estimation method may still be valid. If not, it becomes much more complicated and may encounter an identifiability issue. One possible solution is to model the censoring distribution and develop an inverse probability of censoring weighted estimation method. This is an interesting topic that needs further investigation.

Acknowledgment

We thank the Editor, an Associate Editor, and both reviewers for their constructive comments to improve the paper.

Appendix

Our proofs follow similar steps as Murphy (1995); Scharfstein, Tsiatis and Gilbert (1998); Lu (2008). However, the difference lies on the fact that the event times of individuals are no longer independent and identically distributed as in classical survival data. We need to represent the data sequentially based on the ordered event times as in the data generation process and define the associated martingale processes. Specifically, define the martingale process as

$$M_{j,k}(t) = N_{j,k}(t) - \int_0^t Y_{j,k}(u)\dot{G}\left\{\Lambda(u)e^{\theta^T X_{j,k}}\right\} e^{\theta^T X_{j,k}} d\Lambda(u),$$

where $N_{j,k}(t) = \mathrm{I}\,(T_{(k-1)} < T_{j,k} \le t)$ and $Y_{j,k}(t) = \mathrm{I}\,(T_{j,k} \ge t > T_{(k-1)})$ for individual $j \in N_k$. For simplicity, we only consider the link function $G(x) = \frac{1}{s}\log(1+sx)$ in our proofs.

We can rewrite the log-likelihood as

$$\begin{aligned}\ell_n(\theta,\Lambda) &= \sum_{k=1}^{M_n}\sum_{j\in N_k} \ell_{j,k}(\theta,\Lambda)\\ &= \sum_{k=1}^{M_n}\sum_{j\in N_k}\left(\int_0^{T_{(k)}} \log\left[\lambda(t)e^{\theta^T X_{j,k}}\dot{G}\left\{\Lambda(t)e^{\theta^T X_{j,k}}\right\}\right] dN_{j,k}(t)\right.\\ &\qquad \left. - \int_0^{T_{(k)}} Y_{j,k}(t)e^{\theta^T X_{j,k}}\dot{G}\left\{\Lambda(t)e^{\theta^T X_{j,k}}\right\} d\Lambda(t)\right). \qquad \text{(A.1)}\end{aligned}$$

Next, consider one-dimensional submodel $\Lambda_d(t) = \int_0^t \{1+dh_1(u)\}d\hat{\Lambda}_n(u)$ and $\theta_d = dh_2 + \hat{\theta}_n$, where h_1 is a function and h_2 is a $(p+1)$-dimensional vector. Let $S_n(\hat{\Lambda}_n,\hat{\theta}_n)(h_1,h_2)$ denote the first derivative of $\ell_n(\theta_d,\Lambda_d)$ with respect to d and evaluated at $d=0$. Then, we have $S_n(\hat{\Lambda}_n,\hat{\theta}_n)(h_1,h_2) = 0$ for all (h_1,h_2), since $(\hat{\Lambda}_n,\hat{\theta}_n)$ maximizes $\ell_n(\theta,\Lambda)$. In addition, S_n can be written as $S_n = S_{n_1} + S_{n_2}$, where

$$S_{n_1}(\hat{\Lambda}_n, \hat{\theta}_n)(h_1) = \sum_{k=1}^{M_n} \sum_{j \in N_k} \int_0^{T_{(k)}} \left[h_1(t) + \frac{\ddot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\}}{\dot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\}} e^{\hat{\theta}_n^T X_{j,k}} \int_0^t h_1(v) d\hat{\Lambda}_n(v) \right] \times \left[dN_{j,k}(t) - Y_{j,k}(t) e^{\hat{\theta}_n^T X_{j,k}} \dot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\} d\hat{\Lambda}_n(t) \right], \tag{A.2}$$

$$S_{n_2}(\hat{\Lambda}_n, \hat{\theta}_n)(h_2) = \sum_{k=1}^{M_n} \sum_{j \in N_k} \int_0^{T_{(k)}} \left[h_2^T X_{j,k} + \frac{\ddot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\}}{\dot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\}} \hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}} h_2^T X_{j,k} \right] \times \left[dN_{j,k}(t) - Y_{j,k}(t) e^{\hat{\theta}_n^T X_{j,k}} \dot{G}\left\{\hat{\Lambda}_n(t) e^{\hat{\theta}_n^T X_{j,k}}\right\} d\hat{\Lambda}_n(t) \right], \tag{A.3}$$

where $\ddot{G}(u) = d\dot{G}(u)/du$.

After some calculations, we can show that the efficient score for θ is given by

$$S_{\text{eff}} = \sum_{k=1}^{M_n} \sum_{j \in N_k} \int_0^{T_{(k)}} \left\{ \frac{X_{j,k}}{1 + s\Lambda_0(t) e^{\theta_0^T X_{j,k}}} - w_{\text{eff}}(t) + \frac{s \int_0^t \lambda_0(v) e^{\theta_0^T X_{j,k}} w(v) dv}{1 + s\Lambda_0(t) e^{\theta_0^T X_{j,k}}} \right\} dM_{j,k}(t) \equiv \sum_{k=1}^{M_n} S_{\text{eff},k},$$

where $w_{\text{eff}}(t)$ is a solution to the following integral equation

$$w(t) - \int_0^\tau Q(t, v) w(v) d\Lambda_0(v) = f(t), \quad t \in [0, \tau],$$

and

$$
\begin{aligned}
Q(t,v) =& \left[E\left\{ \sum_{k=1}^{M_n} \sum_{j\in N_k} \mathrm{I}(t \le T_{(k)}) Y_{j,k}(t) \frac{e^{\theta_0 X_{j,k}}}{1+s\Lambda_0(t)e^{\theta_0 X_{j,k}}} \right\} \right]^{-1} \\
&\times \left(E\left[\sum_{k=1}^{M_n} \sum_{j\in N_k} \frac{s\mathrm{I}(v\vee t \le T_{(k)}) Y_{j,k}(t) e^{2\theta_0^T X_{j,k}}}{(1+s\Lambda_0(t)e^{\theta_0^T X_{j,k}})^2} \right] \right. \\
&- \left. E\left[\sum_{k=1}^{M_n} \sum_{j\in N_k} \int_{v\vee t}^{T_{(k)}} \frac{s^2 Y_{j,k}(u) e^{3\theta_0^T X_{j,k}}}{(1+s\Lambda_0(u)e^{\theta_0^T X_{j,k}})^3} d\Lambda_0(u) \right] \right), \\
f(t) =& \left[E\left\{ \sum_{k=1}^{M_n} \sum_{j\in N_k} \mathrm{I}(t \le T_{(k)}) Y_{j,k}(t) \frac{e^{\theta_0 X_{j,k}}}{1+s\Lambda_0(t)e^{\theta_0 X_{j,k}}} \right\} \right]^{-1} \\
&\times \left(E\left[\sum_{k=1}^{M_n} \sum_{j\in N_k} \frac{X_{j,k}\mathrm{I}(t \le T_{(k)}) Y_{j,k}(t) e^{\theta_0^T X_{j,k}}}{(1+s\Lambda_0(t)e^{\theta_0^T X_{j,k}})^2} \right] \right. \\
&- \left. E\left[\sum_{k=1}^{M_n} \sum_{j\in N_k} \int_{t}^{T_{(k)}} \frac{s X_{j,k} Y_{j,k}(u) e^{2\theta_0^T X_{j,k}}}{(1+s\Lambda_0(u)e^{\theta_0^T X_{j,k}})^3} d\Lambda_0(u) \right] \right).
\end{aligned}
$$

Here, $v\vee t = \max(v,t)$. Note that the terms $S_{\text{eff},k}$ and $S_{\text{eff},k}^T$ are uncorrelated for any $k \neq k'$. Then, the information matrix for θ_0 can be defined by $I(\theta_0) = \lim_{n\to\infty} n^{-1} \sum_{k=1}^{M_n} E(S_{\text{eff},k}, S_{\text{eff},k}^T)$.

S2: Proof of Theorem 1 The proof of consistency consists of three steps: first we show that the nonparametric maximum likelihood estimators $\hat{\Lambda}_n$ and $\hat{\theta}_n$ exist or that the jump sizes of $\hat{\Lambda}_n$ are finite; next we show that $\hat{\Lambda}_n$ is bounded almost surely so that, along a subsequence, $\hat{\Lambda}_{n_m}(t) \to \Lambda^*(t)$ for all $t \in [0,\tau]$ and $\hat{\theta}_{n_m} \to \theta^*$; finally we show that $\Lambda^* = \Lambda_0$ and $\beta^* = \beta_0$.

Step 1. By Condition 1, $\hat{\theta}_n$ is finite, and we have $\sup_{\theta \in \mathcal{C}} |\theta^T X_{j,k}| \leq M_0$ for some constant $M_0 > 0$ for all j and k. Therefore, the log-likelihood (2.3) is bounded above by

$$\begin{aligned}\ell_n(\theta, \Lambda) \; < & \sum_{k=1}^{M_n} \Bigg(\log \Lambda\{T_{(k)}\} + M_0 - \log\left\{1 + s\sum_{\ell=1}^{k-1} \Lambda\{T_{(\ell)}\} e^{-M_0}\right\} \\ & - |N_k| \left[G\left\{\sum_{\ell=1}^{k} \Lambda(T_{(\ell)}) e^{-M_0}\right\} - G\left\{\sum_{\ell=1}^{k-1} \Lambda(T_{(\ell)}) e^{-M_0}\right\}\right]\Bigg),\end{aligned} \tag{A.4}$$

where $|N_k|$ is the number of elements in N_k. The right-hand side in (A.4) diverges to $-\infty$ if $\Lambda\{T_{(k)}\}$ goes to infinity for some k, which contradicts to the property of log-likelihood.

Step 2. By the property $n^{-1}\{\ell(\hat{\Lambda}_n, \hat{\theta}_n) - \ell(\bar{\Lambda}_n, \hat{\theta}_n)\} \geq 0$ with $\bar{\Lambda}_n = \hat{\Lambda}_n / \hat{\Lambda}_n(\tau)$, and following the similar steps as Zeng and Lin (2006), we can show that $\sup_n \hat{\Lambda}_n(\tau) < \infty$.

Step 3. Define the following quantity

$$\tilde{\Lambda}_n(t) = \int_0^t \sum_{k=1}^{n} \frac{\mathrm{I}\ (T_{(k)} \leq u) \sum_{j \in N_k} dN_{j,k}(u)}{\sum_{j=1}^{n} \mathrm{I}\ (T_{(k)} \geq u) Y_{j,k}(u) e^{\theta_0^T X_{j,k}} / \{1 + s\Lambda_0(u-) e^{\theta_0^T X_{j,k}}\}}, \tag{A.5}$$

which is a step function with jumps at $T_{(k)}$'s and converges uniformly to Λ_0

by uniform weak law of large numbers.

By Helly's theorem, we know that there exists convergent subsequences $\{\hat{\theta}_{n_m}\}$ and $\{\hat{\Lambda}_{n_m}\}$ such that $\hat{\theta}_{n_m} \to \theta^*$ and $\hat{\Lambda}_{n_m}(t) \to \Lambda^*(t)$ for all $t \in [0, \tau]$. Furthermore, we have $n^{-1}\{\ell(\hat{\Lambda}_{n_m}, \hat{\theta}_{n_m}) - \ell(\tilde{\Lambda}_{n_m}, \theta_0)\} \geq 0$. By taking limits on both sides we obtain $E\{\ell(\Lambda^*, \theta^*)\} = E\{\ell(\Lambda_0, \theta_0)\}$, since the Kullback-Leibler information is negative.

Recall in term (A.1), we have

$$\ell_{j,k}(\Lambda, \theta) = \int_{T_{(k-1)}}^{T_{(k)}} \log\left\{\frac{\lambda(t)e^{\theta X_{j,k}}}{1 + s\Lambda(t-)e^{\theta X_{j,k}}}\right\} dN_{j,k}(t) \\ - \int_{T_{(k-1)}}^{T_{(k)}} Y_{j,k}(t) \frac{e^{\theta X_{j,k}}}{1 + s\Lambda(t)e^{\theta X_{j,k}}} d\Lambda(t).$$

Then, the above equality holds if and only if $E\{\ell_{j,k}(\Lambda^*, \theta^*)\} = E\{\ell_{j,k}(\Lambda_0, \theta_0)\}$ for all j and k. Next, for $k = 1, \ldots, M_n$ consider two cases: (1) $N_{j,k}(T_{(k)}) = 0$, $Y_{j,k}(T_{(k)}) = 1$ for some $j \in N_k$, and (2) $N_{j,k}(T_{(k)}) = 1$, $N_{j,k}(t-) = 0$, and $Y_{j,k}(T_{(k)}) = 1$ for some $j \in N_k$ and t is between time $T_{(k-1)}$ and $T_{(k)}$. By taking difference between the equalities from two cases above, for all $t \in [0, \tau]$ we conclude that

$$\frac{\lambda^*(t)e^{\theta^{*T} X_{j,k}}}{1 + s\Lambda^*(t)e^{\theta^{*T} X_{j,k}}} = \frac{\lambda_0(t)e^{\theta_0^T X_{j,k}}}{1 + s\Lambda_0(t)e^{\theta_0^T X_{j,k}}}.$$

Then, integrating from 0 to t on both sides of above equality and by some simple algebra, we have

$$\Lambda^*(t)/\Lambda_0(t) = e^{(\theta_0-\theta^*)^T X_{j,k}}, \quad \text{for all } t \in [T_{(k-1)}, T_{(k)}] \text{ and } k = 1, \ldots, M_n$$

By Condition 2, we have that $E\{\ell(\Lambda^*, \theta^*)\} = E\{\ell(\Lambda_0, \theta_0)\}$ if and only if $\Lambda^* = \Lambda_0$ and $\theta^* = \theta_0$. Therefore, we show that the subsequences $(\hat{\Lambda}_{n_m}, \hat{\theta}_{n_m}) \to (\Lambda_0, \theta_0)$. By Helly's theorem, we know that $(\hat{\Lambda}_n, \hat{\theta}_n)$ must also converge to (Λ_0, θ_0) almost surely. Since $\hat{\Lambda}_0$ and Λ_0 are bounded monotone function, the pointwise convergence can be strengthened to uniform convergence on $[0, \tau]$.

S3: Proof of Theorem 2 Here, we give an outline of the proof. Define $\psi_0 = (\Lambda_0, \theta_0)$, $\psi = (\Lambda, \theta)$ and $h = (h_1, h_2)$. Assume that the class of h belongs to the space $H = B \otimes R^{p+1}$, where B is the space of bounded variation functions defined on $[0, \tau]$. Define the norm $||h||_H = ||h_1||_v + |h_2|_1$, where $||h_1||_v$ is the total variation norm on $[0, \tau]$ and $|h_2|_1$ is the L_1-norm. In addition, define $H_m = \{h \in H : ||h||_H \le m\}$. Assume $\psi \subset \ell^\infty(H_m)$, where $\ell^\infty(H_m)$ is the space of bounded real-valued functions on H_m under the supremum norm $||A(h)|| = \sup_{h\in H_m} |A(h)|$. First, by the martingale central limit theorem, we can show that $n^{-1/2}S_n(\psi_0)(h)$ converges weakly to a tight

Gaussian process G on $\ell^\infty(H_m)$. Define $S(\psi)(h) = \lim_{n\to\infty} n^{-1}S_n(\psi)(h)$. We have $S(\psi_0)(h) = 0$. Then, following similar arguments in Scharfstein, Tsiatis and Gilbert (1998) and Lu (2008), we can show that $S(\psi)(h)$ is Fréchet differentiable, and its derivative $\dot{S}(\psi)(h)$ is a continuous linear operator and continuously invertible on its range. Finally, by the maximal inequality for martingales (Nishiyama, 1999, Theorem 2.3), we have

$$||n^{-1/2}\{(S_n - S)(\psi_n) - (S_n - S)(\psi_0)\}|| = o_{p^*}(1).$$

for any $||\psi_n - \psi_0|| = O_p(n^{-1/2})$. Therefore, $n^{-1/2}(\hat{\psi}_n - \psi_0)(h)$ converges weakly to the Gaussian process $-\{\dot{S}(\psi)\}^{-1}G$. Then, following similar arguments in Lu (2008), we can show that $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to a multivariate normal with mean 0 and variance $\{I(\theta_0)\}^{-1}$.

References

Anagnostopoulos, A., Kumar, R. and Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 7–15. ACM.

Aral, S. and Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* **57**, 1623–1639.

Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology* **92**, 1287–1335.

Cheng, S., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–845.

Dabrowska, D. M. and Doksum, K. A. (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics* **15**, 1–23.

Greenan, C. C. (2015). Diffusion of innovations in dynamic networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **178**, 147–166.

Iyengar, R., Van den Bulte, C. and Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**, 195–212.

Lee, L.-F. (2004). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* **72**, 1899–1925.

Lee, L.-F., Liu, X. and Lin X. (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal* **13**, 145–176.

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**, 545–574.

Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations. *Nervous and Mental Disease Publishing Co.*

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics* **23**, 182–198.

Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimators in a Gaussian white noise model. *The Annals of Statistics* **27**, 675–696.

Ogburn, E. L., Sofrygin, O., Díaz, I. and van der Laan, M. J. (2017). Causal inference for social network data. *arXiv preprint arXiv:1705.08527v4*.

Pacheco, J. (2012). The social contagion model: Exploring the role of public opinion on the diffusion of antismoking legislation across the American states. *The Journal of Politics* **74**, 187–202.

Ripley, R. M., Snijders, T. A. B., Boda, Z. and Vörös, A. and Preciado, P. (2011). Manual for RSIENA. *University of Oxford, Department of Statistics, Nuffield College* **1**.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference* **2**, 13–74.

Scharfstein, D. O., Tsiatis, A. A. and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis* **4**, 355–391.

Shalizi, C. R., and Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research* **40**, 211–239.

Snijders, T. A. B., Van de Bunt, G. G. and Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks* **32**, 44-60.

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

Zhao, Y. P. , Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* **40**, 2266–2292.

Zhou, J., Tu, Y., Chen, Y. D. and Wang, H. S. (2015). Estimating Spatial Autocorrelation with Sampled Network Data. *Journal of Business & Economic Statistics* **35**, 130–138.

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A.

E-mail: lyu9@ncsu.edu

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A.

E-mail: lu@stat.ncsu.edu

School of Statistics, Renmin University, Beijing, China 100872.

E-mail: dyhuang89@126.com