Biometrika (2019), *pp*. 1–8
Printed in Great Britain

Event History Analysis of Dynamic Networks

BY T. SIT,
Department of Statistics, The Chinese University of Hong Kong, Sha Tin, Hong Kong
tonysit@sta.cuhk.edu.hk

Z. YING,
Department of Statistics, Columbia University, New York, New York, 10027, U.S.A.
zying@stat.columbia.edu

AND Y. YU
Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.
yi.yu.2@warwick.ac.uk

SUMMARY

Statistical analysis on networks has received growing attention due to demand from various emerging applications. In dynamic networks, one of the key interests is to model the event history of time-stamped interactions among nodes. We model dynamic directed networks via multivariate counting processes. A pseudo partial likelihood approach is exploited to capture the network dependence structure. Asymptotic results are established. Numerical experiments are performed to demonstrate effectiveness of our proposal.

Some key words: Recurrent event; Survival analysis; Multivariate counting processes; Dynamic directed networks.

1. INTRODUCTION

In classical survival analysis, one would like to understand how specific covariates are affecting the lifetime distributions. One of the many powerful tools thereof is the Cox model (Cox, 1972), which stipulates the hazard function in the following form:

$$\lambda_T(t \mid Z) = \exp\{\beta^{o\top} Z\}\lambda_0(t), \tag{1}$$

where $\lambda_T(\cdot \mid Z)$ denotes the conditional hazard function given covariates $Z \in \mathbb{R}^p$ of the event time of interest T, $\beta^o \in \mathbb{R}^p$ denotes the coefficients, where p is a positive integer, and $\lambda_0(\cdot)$ is the baseline function.

There has been an extensive amount of research on theoretical, methodological, computational and application aspects of (1) since Cox (1972). Dynamic network data, which consist of a sequence of networks ordered by time, are ubiquitous in the Big Data era. Among many different possibilities in modelling dynamic networks, counting processes are also deployed. Existing works, including Butts (2008), Snijders et al. (2010), Krivitsky & Handcock (2014), Cauchemez et al. (2011) and Perry & Wolfe (2013), to name but a few, used the hazard functions involved models to analyze how certain covariates affect the networks forming.

At first glance, it is straightforward to extend the scope of (1) and apply it to model activities in networks. For example, one may model the interaction times between pairs of nodes and edge-specific information as event times and covariates, respectively. However, there are two hurdles.

(i) The most prominent feature of networked data is the dependence among observations. If our interest lies in the interactions among nodes and the observations are the interactions among nodes, the independence assumption among individuals may no longer be valid.

(ii) The Cox model was originally proposed for lifetime data, which implies that for each subject, there is at most one observation, i.e. death. However in many applications, for example international trades, there are multiple interactions between two countries, and these interactions are dependent across time. Rigorously speaking, this additional level of dependence handicaps the martingale foundation in the Cox model analysis (e.g. Andersen & Gill, 1982) and voids the inference thereof.

As for hurdle (ii), recurrent event data, e.g. cancer relapses data, share the same difficulty. In order to model multiple event times, Wei et al. (1989) developed a marginal approach, and Lin et al. (2000) established large sample theory based on empirical processes. Readers may also refer to Andersen et al. (1993) and Martinussen & Scheike (2006) for summaries.

Hurdle (i) is inherited from the relational nature of the networked data. The common practice in the statistical network literature is to capture the dependence structure by conditioning on a set of covariates or observations, which happens with latent state models (e.g. Hoff et al., 2002) and Markov-type models (e.g. Hunter et al., 2011; Hanneke et al., 2010). This strategy is handy for many problems, but it is challenging to adopt the same treatment if we are interested in the interaction times and hazard rates thereof. In order to guarantee that (1) is a valid hazard function, the covariates can only be external; in other words, the observed values of these covariates should not carry information of the failure times. In contrast to the external covariates, many network-dependence-related covariates, e.g. history information or common neighbours, are internal covariates, and adding them in (1) is an abuse of the hazard function. For detailed explanations, we refer the readers to Section 6.3 in Kalbfleisch & Prentice (1992).

Another intuitive remedy for (ii) is to exploit time series models. Unfortunately, in network data, the observations are not collected in a linear order like those in the time series data. Lacking a widely-recognized definition of distances, it is hard to directly adopt any time series method. Having said this, lacking an explicit definition of 'distance' does not mean that there is not a certain form of 'distance'. To understand this, we see that in a social network, one person might only have direct influences over a number of other people, and the influences decrease as a certain sense of 'closeness' decreases. This distance is possibly a combination of many factors.

In this paper, we provide a novel and theoretically-rigorous framework to study how the edge-specific covariates affect the edge formulation in dynamic and directed networks, allowing for an unspecified distance. We list our main contributions as follows.

Firstly, since it is hard to propose a suitable distance in network data, we opt for an implicit definition of distance. We borrow the idea of composite likelihood to capture the associated but unspecified dependence structure, to which, our method will be able to improve the covariate inference procedure.

Secondly, in order to derive asymptotic results, we establish a network version of the m-dependent central limit theorem using Stein's (Stein, 1972) method. We only need to assume that for any node in the network, there are at most m other dependent nodes. In the analysis procedure, we do not need to know which m nodes are dependent or how they depend on each other. A refined asymptotic tightness of stochastic processes result is derived by allowing m to grow with the sample size at a suitable rate. Convergence results of our proposed estimators are presented in Section 2·2.

In this paper, denote $\mathcal{S} = \{1, \ldots, n\}$ by the set of nodes and $\mathcal{S}^{\times 2} = \mathcal{S} \times \mathcal{S} \setminus \{(i,i),\, i \in \mathcal{S}\}$. For each pair $(i,j) \in \mathcal{S}^{\times 2}$, we have the edge-specific covariates $Z_{ij}(t) \in \mathbb{R}^p$, $t \in [0,T]$, $T > 0$, its corresponding event times $T_{ij,0}, \ldots, T_{ij,n_{ij}}$, where $T_{ij,0} = 0$. We also assume that there is no self-loop, i.e. for the pair (i,i), $i \in \mathcal{S}$, $n_{ii} = 0$. For a vector $v \in \mathbb{R}^p$, let $\|v\|_1$ and $\|v\|$ be the ℓ_1- and ℓ_2-norms of v, respectively. For a $p \times q$ matrix $A = (a_{ij})$, we adopt the notation of $\|A\|_\infty = \max_{i,j} |a_{ij}|$ and $\|A\|_1 = \max_i \sum_{j=1}^q |a_{ij}|$. For any set $\mathcal{B}$, we denote its cardinality by $|\mathcal{B}|$. For $f(n), g(n) > 0$, let $f(n) \asymp g(n)$ denote that $f(n) = O(g(n))$ and $g(n) = O(f(n))$. Let $\mathbb{1}(\cdot) \in \{0,1\}$ be the indicator function. The graph theory definitions are standard in this paper and are deferred to the supplementary materials.

2. METHODOLOGY

2·1. *Setup*

Recall that our interest lies in the effects of edge-specific covariates on the distributions of interaction times. For any $i \in \mathcal{S}$, we use multivariate counting processes to record its activities. Specifically, we adopt the notation

$$\mathcal{N}_i = \{N_i(t), t \in [0,T]\} = \{(N_{ij}(t),\, j \neq i),\, t \in [0,T]\},$$

where the univariate counting process $\{N_{ij}(t), t \in [0,T]\}$ encodes direct edges starting at i and ending at j. The corresponding mean function is given by

$$E\big\{dN_{ij}(t) \mid Z_{ij}(t)\big\} = \exp\big\{\beta^{o\mathrm{T}} Z_{ij}(t)\big\}\lambda_0(t)\,dt. \tag{2}$$

Perry & Wolfe (2013) assumed that each interaction in a network follows the Cox model and is conditionally independent with other previous events given covariates history. To be specific, in order to ensure that (1) is a valid hazard function, time-varying covariates involved in the regression have to be external (see Section 6.3 of Kalbfleisch & Prentice, 1992). In the context of dynamic networks, if one adds history events or network structures in the covariates in the hazard function, then one is implicitly assuming that history and network information are either pre-determined or their distributions do not involve the lifetime distribution specified by (1). This is not realistic and therefore one would not prefer adding history or network information in (1).

Different from Perry & Wolfe (2013), our formulation adopts the idea in Lin et al. (2000), which does not require accurate specifications of the dependence of sequential events within each pair. There is a subtle difference between (2) and the Cox model which essentially assumes

$$E\{dN_{ij}(t) \mid \mathcal{F}_{t-}\} = E\{dN_{ij}(t) \mid Z_{ij}(t)\} \tag{3}$$

in addition to (2). The notation $\mathcal{F}_{t-}$ is the natural σ-field generated by $\{N_{ij}(s),\, i,j \in \mathcal{S},\, 0 \leq s < t \leq T\}$.

The requirement specified in (3) implies that the covariates included can capture all the dependence between the future and past events. This is a valid assumption when at most one event occurs; however, when multiple events may happen over a certain period of time, it is challenging to capture all the dependence by a set of covariates. Our formulation does not require (3).

If we define $\{M_{ij}(\beta,t),\, t \in [0,T]\}$ as

$$M_{ij}(\beta,t) = N_{ij}(t) - \int_0^t \exp\{\beta^{\mathrm{T}} Z_{ij}(s)\}\lambda_0(s)\,ds = N_{ij}(t) - \Lambda_{ij}(\beta,t), \tag{4}$$

then, due to (2), each $\{M_{ij}(\beta^o,t),\, t \in [0,T]\}$ is a mean zero process but not a martingale difference process because (3) is no longer assumed.

To establish our inference procedure and introduce our proposed estimator, we borrow the idea of composite likelihood (e.g. Lindsay, 1987; Cox & Reid, 2004; Varin & Vidoni, 2005) and consider a pairwise pseudo partial likelihood. Since the observations are possibly dependent, this formulation is particularly useful when the full likelihood is too complicated to be expressed or optimized. The corresponding log pseudo partial likelihood is defined as follows,

$$\ell_n(\beta) = \sum_{i=1}^n \sum_{j\neq i} \int_0^T \left\{ \beta^{\mathrm{T}} Z_{ij}(t) - \log\left[\sum_{k=1}^n \sum_{l\neq k} \exp\big\{\beta^{\mathrm{T}} Z_{kl}(t)\big\} \right] \right\} dN_{ij}(t),$$

whose score function is given by

$$U_n(\beta) = \sum_{i=1}^n \sum_{j\neq i} U_{ij}(\beta) = \sum_{i=1}^n \sum_{j\neq i} \int_0^T \big\{ Z_{ij}(t) - \bar{Z}_n(\beta,t) \big\}\, dN_{ij}(t),$$

where

$$\bar{Z}_n(\beta, t) = \frac{\sum_{k=1}^n \sum_{l \neq k} Z_{kl}(t) \exp\{\beta^{\mathrm{T}} Z_{kl}(t)\}}{\sum_{k=1}^n \sum_{l \neq k} \exp\{\beta^{\mathrm{T}} Z_{kl}(t)\}}.$$

As shown in Section 2·2, the score function $U_n(\beta^0)$, when suitably normalized, is asymptotically normal with mean zero. This makes $U_n(\beta) = 0$ a valid consistent estimation equation. In the sequel, we define $\hat{\beta}_n$ as the solution of

$$U_n(\hat{\beta}_n) = 0. \tag{5}$$

2·2. *Theory*

As we have emphasized that it is restrictive to assume that the edges are independent and difficult to define a distance or neighbourhood in a network. To overcome these hurdles, we only assume that for every node in the network, there are at most m other nodes, the stochastic processes associated with which are dependent. In Condition 1, we formalize this dependence.

Condition 1. For any $i \in \mathcal{S}$, there exists $\mathcal{J}_i \subset \mathcal{S}$, such that for any $j \in \mathcal{S} \setminus \mathcal{J}_i$, $\mathcal{N}_i$ and $\mathcal{N}_j$ are independent, and $\mathcal{Z}_i$ and $\mathcal{Z}_j$ are independent, where $\mathcal{Z}_i = \{Z_{il}(t), l \in \mathcal{S}, t \in [0, T]\}$. Assume for any $i \in \mathcal{S}$, it holds that

$$|\mathcal{J}_i| \asymp m_n = o(n^{1/4}). \tag{6}$$

Condition 2. Assume for all $(i, j) \in \mathcal{S}^{\times 2}$, there exists a universal constant $K > 0$ such that $\|Z_{ij}(0)\|_1 + \int_0^T \|dZ_{ij}(t)\|_1 \leq K$. Let

$$\mu_n(\beta^o, t) = \frac{E\{\sum_{i=1}^n \sum_{j \neq i} Z_{ij}(t) \exp(\beta^{o\mathrm{T}} Z_{ij}(t))\}}{E\{\sum_{i=1}^n \sum_{j \neq i} \exp(\beta^{o\mathrm{T}} Z_{ij}(t))\}},$$

$$\Sigma_{1,n} = E\left[\frac{\sum_{i=1}^n \sum_{j \neq i}}{n(n-1)} \int_0^T \{Z_{ij}(t) - \mu_n(\beta^o, t)\}\{Z_{ij}(t) - \mu_n(\beta^o, t)\}^{\mathrm{T}} \exp(\beta^{o\mathrm{T}} Z_{ij}(t)) \lambda_0(t)\, dt\right]$$

and

$$\Sigma_{2,n} = E\Bigg(\frac{1}{n(n-1)} \left[\sum_{i=1}^n \sum_{j \neq i} \int_0^T \{Z_{ij}(t) - \mu_n(\beta^o, t)\}\, dM_{ij}(\beta^o, t)\right] \times \left[\sum_{i=1}^n \sum_{j \neq i} \int_0^T \{Z_{ij}(t) - \mu_n(\beta^o, t)\}\, dM_{ij}(\beta^o, t)\right]^{\mathrm{T}}\Bigg),$$

satisfying that $0 < \limsup_{n \to \infty} \rho_{\max}(\Sigma_{j,n}) / \rho_{\min}(\Sigma_{j,n}) < \infty$, $j = 1, 2$, where $\rho_{\min}(A)$ and $\rho_{\max}(A)$ are the minimum and maximum eigenvalues of matrix A, respectively. In addition, we assume that there exist a non-random vector $\mu(t)$ and matrices Σ_1, Σ_2 such that

$$\sup_{t \in [0,T]} \max\left\{\|\mu_n(\beta^o, t) - \mu(t)\|, \left\|\Sigma_{1,n} - \Sigma_1\right\|, \left\|\Sigma_{2,n} - \Sigma_2\right\|\right\} \xrightarrow{P} 0.$$

Condition 1 is the key assumption, but is mild. It restricts the growing rate of the number of dependent notes. The m-dependence assumption is commonly adopted in time series, where m-dependence stipulates that the observation at time point t is independent with the observations with lags more than m. To ensure that this assumption is realistic in our applications concerned, we also allow the dependence number m_n to diverge with n. It is a parallel version of long-range memory models for network data. Under Condition 1, each node in a network can be dependent with an increasing number of other nodes as the network grows, but we expect it to be of a much slower growth compared with that of the network size.

This assumption also covers conventional models including stochastic block models assuming that all edges are independent in which case we have $m_n = 1$. In order to show the asymptotic normality of the estimators in Theorem 1, we need to show a finite-dimensional central limit theorem, which requires $\max_{i\in\mathcal{S}} |\mathcal{J}_i| \le m_n = o(n^{1/4})$, and the tightness of relevant processes, which requires for any $i \in \mathcal{S}$, $|\mathcal{J}_i| \asymp m_n \le O(n^{1/3})$.

In fact, if an accepted definition of distance is available, then one would first estimate the neighbourhood based on the distance. Our method and theory can adapt correspondingly with reasonable convergence results. A well-defined distance will lead to a well-defined neighbourhood for each node, which in turns provides efficient estimators; see Chapter 3 of Guyon (1995).

Condition 2 requires a specific covariance-type structure that guarantees its positive definiteness. We define $\Sigma_{1,n}$ and $\Sigma_{2,n}$ in this way such that we can apply the self-normalizing version of the central limit theorem, which works under mild assumptions on the dependence structure. Condition 2 also implies that the mean processes $\{\Lambda_{ij}(\beta^o, t),\, t \in [0, T]\}$ are Lipschitz continuous. This fact will be repeatedly used in the proof in the Appendix.

THEOREM 1. *Under Conditions* 1 *and* 2*, with* $\hat{\beta}_n$ *defined in* (5) *and* $\Sigma_{1,n}, \Sigma_{2,n}$ *specified in Condition* 2*, we have that* $\hat{\beta}_n$ *is asymptotically normal with mean* β^o *and covariance matrix* $\Sigma_n = \Sigma_{1,n}^{-1}\Sigma_{2,n}\Sigma_{1,n}^{-1}$*. In particular, for any* $v \in \mathbb{R}^p$*,*

$$\frac{v^{\mathrm{T}}(\hat{\beta}_n - \beta^o)}{\left(v^{\mathrm{T}}\Sigma_n^{-1}v\right)^{1/2}} \to \mathcal{N}(0, 1), \tag{7}$$

in distribution, as $n \to \infty$*.*

Theorem 1 states that the convergence rate is related with m_n. It follows from (5) in the proof that the convergence rate of the variance $v^{\mathrm{T}}\Sigma_n^{-1}v$, $v \in \mathbb{R}^p$, which is $O(m_n^2 n^{-1/2})$. If $m_n = 1$ as assumed in the independent edges cases or $m_n = O(1)$ as in Schweinberger & Handcock (2015), then the convergence rate is $n^{-1/2}$, which is the same as the standard situation and is optimal. The convergence rate decreases as m_n increases. Schweinberger & Handcock (2015) also considered dependent network models, and the results developed therein are based on a general exponential random graph model in a Bayesian framework. The key differences between ours and theirs are summarized as follows: (i) the random variable associated with each edge is a Bernoulli random variable representing the presence of the edge in Schweinberger & Handcock (2015), while in our paper, each edge has its own counting process; (ii) although both papers allow m-dependence, in Schweinberger & Handcock (2015) m is assumed to be finite, while in our setting, m is allowed to be of order $o(n^{1/4})$; and (iii) since we are in the survival analysis framework, the dependence structure is assumed to be among the senders, while Schweinberger & Handcock (2015) investigated exponential random graph models, and therefore the dependence lies among edges.

It is also natural to estimate the mean function $\Lambda_0(t)$ for $i \in \mathcal{S}$ by the Aalen–Breslow-type estimator $\hat{\Lambda}_0(t) = \int_0^t \{d\sum_{i=1}^n \sum_{i\neq j} N_{ij}(s)\}/\{\sum_{i=1}^n \sum_{i\neq j} \exp\{\hat{\beta}_n^{\mathrm{T}} Z_{ij}(s)\}\}$, $t \in [0, T]$, which can be shown to be consistent, by proving that $\hat{\beta}_n$ is almost surely consistent in Step 1 in the proof in the supplementary materials. This requires a strong law of large numbers for dependent random variables (e.g. Korchevsky & Petrov, 2010). The result can also be generalized further so that individual can have different baseline hazards.

The result (7) has the same sandwich variance form which resembles those obtained via the composite likelihood inference (e.g. Lin et al., 2000). It degenerates to the efficient estimator case when $\Sigma_{1,n} = \Sigma_{2,n}$, i.e. all the directed edges are assumed to be independent.

There are a few key ingredients in the proof of Theorem 1. First, we exploit Stein's method, which has been extended to dependent cases (e.g. Baldi & Rinott, 1989). We have developed a new device of the central limit theorem which is designed for the semiparametric setting, integrating chaining arguments. Second, we extend the weak convergence proof in the independent case stated in Lin et al. (2000) to a dependent case. Since we allow m_n to diverge, this is a substantial improvement.

2·3. *Variance estimators*

We have established the asymptotic of $\hat{\beta}_n$ in Theorem 1, but it involves unknown population quantities $\Sigma_{1,n}$ and $\Sigma_{2,n}$. Usual estimators are based on the assumption of independent observations or independent innovations in the time series context. To tackle the unknown dependence structure, we adopt the jackknife sandwich estimator proposed in the composite likelihood literature (e.g. Varin et al., 2011) in an attempt to provide reasonable estimation for $\Sigma_{1,n}$ and $\Sigma_{2,n}$.

Let $\widehat{\Sigma}_{1,n}$ and $\widehat{\Sigma}_{2,n}$ be the estimators of $\Sigma_{1,n}$ and $\Sigma_{2,n}$, respectively, defined as follows:

$$\widehat{\Sigma}_{1,n} = \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n_{ij}} \left\{ Z_{ij}(T_{ij,k}) - \bar{Z}\big(\hat{\beta}_n, T_{ij,k}\big) \right\} \left\{ Z_{ij}(T_{ij,k}) - \bar{Z}\big(\hat{\beta}_n, T_{ij,k}\big) \right\}^{\mathrm{T}} \delta_{ij},$$

and

$$\widehat{\Sigma}_{2,n} = \frac{1}{n} \sum_{s=1}^{n} \left[\sum_{i \neq s} \sum_{j \neq i,s} \sum_{k=1}^{n_{ij}} \big\{ Z_{ij}(T_{ij,k}) - \bar{Z}(\hat{\beta}^{(-s)}, T_{ij,k}) \big\}^{\delta_{ij}} \right] \times \left[\sum_{i \neq s} \sum_{j \neq i,s} \sum_{k=1}^{n_{ij}} \big\{ Z_{ij}(T_{ij,k}) - \bar{Z}(\hat{\beta}^{(-s)}, T_{ij,k}) \big\}^{\delta_{ij}} \right]^{\mathrm{T}}, \tag{8}$$

where $\delta_{ij} = \mathbb{1}\{N_{ij}(T) > 0\}$, $\hat{\beta}^{(-s)}$ is the estimator to the estimating equation (5) after deleting the sth node and its corresponding data from the observations. The proposed jackknife procedure (8) offers one possible approach for variance estimation in Theorem 1. Although unless there is additional network and time dependence structural assumption, there is no single construction of consistent estimates for $\Sigma_{2,n}$ nor the variance, the jackknife procedure performs decently in terms of achieving empirical coverage probabilities that are very close to the nominal values, as we shall demonstrate in Section 3.

In fact, for a general composite likelihood inference problem, estimating the variability matrix is difficult. The essence of composite likelihood is to make use of the working independence and sandwich variance estimator to capture the dependence so that specific dependent structures need not be assumed. In practice, a good estimator of the variance is inevitably a function of the unknown structure.

Consequently, for $\alpha \in (0,1)$, a $(1-\alpha) \times 100\%$ percent confidence region for the true regression parameter β^o can, be expressed as $\big\{\beta : (\hat{\beta}_n - \beta^o)^{\mathrm{T}} \Big(\widehat{\Sigma}_{1,n} \Sigma_{2,n}^{-1} \widehat{\Sigma}_{1,n}\Big)^{-1} (\hat{\beta}_n - \beta^o) \leq \chi^2_{p,1-\alpha}\big\}$, where $\chi^2_{p,1-\alpha}$ is the corresponding quantile of a chi-square distribution with degrees of freedom p.

3. NUMERICAL ANALYSIS

3·1. *Simulation*

We generate recurrent events for each pair of nodes of size 150 in a connected network from the model $\lambda_{ij}(t; Z) = \eta_i \lambda_{0i}(t) \exp\{\beta_0^{\mathrm{T}} Z_{ij}\}$, where $\lambda_{0i}(t) = \mathbb{1}(i \leq n/2) + 1.2\mathbb{1}(i > n/2)$ denotes the baseline hazard. The sample size of 150 is chosen to resemble the Enron dataset analyzed in Section 2 in the Supplementary materials. The unobserved random variable η that introduces heterogeneity to this random-effect intensity model is assumed to follow $\Gamma(1, 1/16)$. Denote $N = n(n-1)$. We consider two sets of time-invariant covariates. The first set $\widetilde{Z} = (Z_1, Z_2, Z_3) = (Z_{ij})_{N \times 3}$ is generated as follows: $Z_1 = (Z_{11.}^{\mathrm{T}}, \ldots, Z_{1n.}^{\mathrm{T}})^{\mathrm{T}} = (\mathbb{1}(\Xi_1 \geq 0)^{\mathrm{T}}, \ldots, \mathbb{1}(\Xi_n \geq 0)^{\mathrm{T}})^{\mathrm{T}}$, where $\geq$ denotes the element-wise comparison, Ξ_j is an independent and identically distributed $\mathcal{N}(0_{n-1}, \Sigma^{(1)})$, with $\Sigma^{(1)} = (\mathbb{1}(i=j) + \rho\mathbb{1}(i \neq j))$ for $j = 1, \ldots, n$. Covariate Z_2's are generated from $\textsc{Unif}(0,1)$ and $Z_3 = (Z_{31.}^{\mathrm{T}}, \ldots, Z_{3n.}^{\mathrm{T}})^{\mathrm{T}}$ are independent and identically distributed as $\mathcal{N}(0, \Sigma^{(2)})$ with $\big(\Sigma^{(2)}\big)_{ij} = \mathbb{1}(i=j) + \rho\mathbb{1}(|i-j| = 1)$. Under this setup, when $\rho \neq 0$, we impose the dependence in the data.

For the jackknife procedure, both odd-1-out and odd-2-out procedures are examined: The odd-2-out procedure randomly removes two nodes from the network for estimating the corresponding variances based on 150 random draws.

Table 1. *Summary statistics for the simulation studies*

ρ	Parameters	Bias	SE	SEE (P&W)	SEE (JK)	SEE (JK2)	SEE	ECP (P&W)	ECP (JK)	ECP (JK2)	ECP
0·00	β_{10}	0·015	0·016	0·008	0·016	0·016	0·016	0·212	0·952	0·952	0·952
	β_{20}	-0·005	0·022	0·009	0·024	0·024	0·025	0·016	0·952	0·952	0·960
	β_{30}	0·002	0·007	0·004	0·007	0·007	0·007	0·040	0·920	0·920	0·936
0·30	β_{10}	0·013	0·019	0·020	0·019	0·019	0·019	0·624	0·962	0·962	0·872
	β_{20}	-0·004	0·026	0·010	0·025	0·025	0·025	0·000	0·936	0·936	0·956
	β_{30}	0·001	0·008	0·005	0·007	0·007	0·007	0·040	0·952	0·952	0·900
0·50	β_{10}	0·018	0·033	0·019	0·046	0·046	0·042	0·572	0·973	0·973	0·856
	β_{20}	-0·003	0·028	0·013	0·025	0·025	0·026	0·040	0·928	0·928	0·944
	β_{30}	0·002	0·008	0·004	0·007	0·007	0·007	0·024	0·924	0·928	0·944

Bias is the mean differences between parameter estimates and their corresponding true values, SE denotes the standard errors of the parameter estimates; SEE(JK2), SEE(JK), SEE are the means of the variance estimates under odd-two-out, odd-one-out jackknife and naïve standard error estimates, respectively; SEE (P&W) is the standard error of the estimates from Perry & Wolfe (2013); the corresponding empirical coverage probabilities are denoted as ECP(JK2), ECP(JK), ECP and ECP (P&W), respectively. The naïve standard error estimates are based on $\widehat{\Sigma}_{1,n}^{-1}$.

The results of the simulation studies are summarized in Table 1 based on 250 iterations. The estimates provided by the proposed method are virtually unbiased while the variance estimator also provides reasonably accurate estimation of the true variances of $\hat{\beta}$ upon which the confidence intervals constructed demonstrate empirical coverage probabilities that are close to their nominal values. As we can see from the standard errors and empirical coverage probabilities of estimates yielded from conventional approach, if one ignores the recurrent event nature of the problem as well as the possible dependence amongst individuals, the standard errors will be substantially underestimated, resulting in potentially erroneous inference conclusions as reflected in the less than par empirical coverage probabilities. It is worth mentioning that the proposals of Perry & Wolfe (2013) perform worse than the naïve estimators. This is potentially due to their abuse of the covariates added in the hazard functions. As we have emphasized, only external covariates can be added into the hazard functions.

4. Discussion

In this paper, we assume the same set of nodes across time. One way to relax this condition is to study dynamic networks in the branching processes (e.g. Banerjee et al., 2018). In addition, we assume bounded covariates, which can be relaxed by exploiting concentration inequalities and empirical processes arguments, and constructing a large probability event where the boundedness assumption holds. We also assume the dimension of the regression coefficient to be fixed, i.e. $\beta^o \in \mathbb{R}^p$, and p is fixed. One could incorporate high-dimensional inference techniques developed (e.g. Huang et al., 2013). The undirected network is a simpler model than the directed case, and one can apply our methodology to undirected networks straightforwardly by ignoring the directions of the interactions.

If for some real data sets, where a reasonable distance between pairs of edges is available and the dependence structure is known, one can also use the results in random fields to establish the limiting distribution of the estimators with efficient variance estimators. We would leave this as future work when data sets with definitions of distance properly defined are available.

Acknowledgement

We thank the Editor, the Associate Editor and the referee for constructive suggestions. The first author acknowledges the financial support of HKSAR GRF Grants 14317716 and 14301618. The second author is partially supported by U.S. NSF Grant IIS-1633360 and U.S. NIH Grant R01GM047845.

SUPPLEMENTARY MATERIAL

Supplementary material available at Biometrika online includes the proof of Theorem 1 and a real data analysis.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø, G. R. D. & KEIDING, N. (1993). *Statistical Models based on Counting Processes*. Springer-Verlag New York.

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of statistics* **10**, 1033–1311.

BALDI, P. & RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17**, 1646–1650.

BANERJEE, S., BHAMIDI, S. & CARMICHAEL, I. (2018). Fluctuation bounds for continuous time branching processes and nonparametric change point detection in growing networks. *arXiv preprint arXiv:1808.02439* .

BUTTS, C. T. (2008). 4. a relational event framework for social action. *Sociological Methodology* **38**, 155–200.

CAUCHEMEZ, S., BHATTARAI, A., MARCHBANKS, T. L., FAGAN, R. P., OSTROFF, S., FERGUSON, N. M., SWERDLOW, D. & THE PENNSYLVANIA H1N1 WORKING GROUP (2011). Role of social networks in shaping disease trasmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences (USA)* **108**, 2825–30.

COX, D. R. (1972). Regression models and life tables. *J R Stat Soc Series B Stat Methodol* **34**, 187–220.

COX, D. R. & REID, N. (2004). A note on pesudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–737.

GUYON, X. (1995). *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media.

HANNEKE, S., FU, W. & XING, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics* **4**, 585–605.

HOFF, P. D., RAFTERY, A. E. & HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098.

HUANG, J., SUN, T., YING, Z., YU, Y. & ZHANG, C.-H. (2013). Oracle inequalities for the lasso in the cox model. *Ann. Statist.* **41**, 1142–1165.

HUNTER, D., SMYTH, P., VU, D. Q. & ASUNCION, A. U. (2011). Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1992). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

KORCHEVSKY, V. M. & PETROV, V. V. (2010). On the strong law of large numbers for sequences of dependent random variables. *Vestnik St. Petersburg University: Mathematics* **43**, 143–147.

KRIVITSKY, P. N. & HANDCOCK, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 29–46.

LIN, D. Y., WEI, L.-J., YANG, I. & YING, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Series B Stat Methodol* **62**, 711–730.

LINDSAY, B. G. (1987). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, N. U. Prabhu, ed., vol. 80. American Mathematical Society.

MARTINUSSEN, T. & SCHEIKE, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer-Verlag New York.

PERRY, P. O. & WOLFE, P. J. (2013). Point process modelling for directed interaction networks. *J R Stat Soc Series B Stat Methodol* **75**, 821–849.

SCHWEINBERGER, M. & HANDCOCK, M. S. (2015). Local dependence in random graph models: characterization, properties and statistical inference. *J R Stat Soc Series B Stat Methodol* **77**, 647–676.

SNIJDERS, T. A., KOSKINEN, J. & SCHWEINBERGER, M. (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics* **4**, 567.

STEIN, C. (1972). A bound for the error in he normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **2**, 586–602.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

VARIN, C. & VIDONI, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–528.

WEI, L., LIN, D. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.