ESTIMATING LINKS OF A NETWORK FROM TIME TO EVENT DATA

By Tso-Jung Yen[*,1], Zong-Rong Lee[*], Yi-Hau Chen[*,2], Yu-Min Yen[†]
AND Jing-Shiang Hwang[*]

Academia Sinica[*] and National Chengchi University[†]

In this paper we develop a statistical method for identifying links of a network from time to event data. This method models the hazard function of a node conditional on event time of other nodes, parameterizing the conditional hazard function with the links of the network. It then estimates the hazard function by maximizing a pseudo partial likelihood function with parameters subject to a user-specified penalty function and additional constraints. To make such estimation robust, it adopts a pre-specified risk control on the number of false discovered links by using the Stability Selection method. Simulation study shows that under this hybrid procedure, the number of false discovered links is tightly controlled while the true links are well recovered. We apply our method to estimate a political cohesion network that drives donation behavior of 146 firms from the data collected during the 2008 Taiwanese legislative election. The results show that firms affiliated with elite organizations or firms of monopoly are more likely to diffuse donation behavior. In contrast, firms belonging to technology industry are more likely to act independently on donation.

1. Introduction. Network analysis aims to understand a network by exploring its links along with node attributes. Important network properties such as centrality [Freeman (1977)], transitivity and assortativity [Newman (2002)] are computed using information about links with additional information about node attributes. Links of a network are conventionally defined by nodes of the network or identified by researchers during the data collection process. However, information about links is not always explicit, sometimes it is even unobservable. When such information is not available, one way to obtain the links is to identify them statistically, by building a model for a network with the links as parameters and estimating these parameters from the data.

Currently there are several statistical approaches to identifying links of a network. One approach relies on estimating Gaussian graphical models with continuous data [Banerjee, El Ghaoui and d'Aspremont (2008), Chandrasekaran, Parrilo

and Willsky (2012), Danaher, Wang and Witten (2014), Friedman, Hastie and Tibshirani (2008), Khare, Oh and Rajaratnam (2015), Meinshausen and Bühlmann (2006), Peng, Zhou and Zhu (2009), Ren et al. (2015), Yuan and Lin (2007)]. The other approach focuses on estimating Ising models with binary data [Amhed and Xing (2009), Anandkumar et al. (2012), Loh and Wainwright (2013), Ravikumar, Wainwright and Lafferty (2010), Xue, Zou and Cai (2012)]. The third approach is to estimate graphical models using nonparametric techniques [Cai, Liu and Luo (2011), Lafferty, Liu and Wasserman (2012), Liu et al. (2012), Qiu et al. (2016)]. Although these methods are efficient and powerful, with excellent properties in estimation, they are mainly developed for dealing with non-censored continuous data or binary data. To deal with censored data, or time to event data in general, there exist several research works with a focus on estimation of network diffusion. One early study is Strang and Tuma (1993), who proposed a Cox proportional hazards model for peer effects on behavior diffusion. Recent research works in this field include Gomez-Rodriguez, Leskovec and Krause (2012), who proposed several probabilistic models and estimated these models with data collected from online social media, and Gomez-Rodriguez et al. (2016), who proposed a full likelihood approach to estimating network diffusion with censored data and investigated conditions that guarantee full recovery of the diffusion paths as sample size increases.

In this paper we propose a statistical method for estimating links of a network from time to event data. In particular we focus on the time to event data in that event times are vector-valued, and entries of each event time vector are either observed or right-censored. In Section 2 we provide two data sets: the synthetic diffusion data generated from a hierarchical model, and the campaign donation data collected during the 2008 Taiwanese legislative election, to illustrate the data structure and the problem of estimating links of a network from time to event data. In Section 3 we develop a hazard network model in that the hazard rate of a node is a function of event times of other nodes. This setting allows us to parameterize the hazard rate function in terms of links of the network underlying the dependence structure of the event occurrences. In Section 4 we develop a partial likelihood-based method for estimating parameters in the hazard network model. In Section 5 we evaluate our method by applying it to estimate the hazard network model with the synthetic diffusion data. In Section 6, we apply our method to identify links of a political cohesion network that drives donation behavior among 146 firms from the campaign donation data collected during the 2008 Taiwanese legislative election. In Section 7 we discuss possible extensions of our method and future research directions. In Supplementary Materials [Yen et al. (2017)] we describe a numerical algorithm for carrying out our method, provide details on data aggregation, and additional results on simulation experiments and real data application.

2. The problem setting and data sets. In this section we first describe the problem of estimating links of a network from time to event data. We then provide two data sets to illustrate the problem.

2.1. *The estimation problem.* Now consider a situation in which there are m subjects. Let $t_i = (t_{i1}, t_{i2}, \ldots, t_{im})$ denote the event time vector of the m subjects in instance i, and $\delta_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{im})$ denote the indicator vector of the m subjects such that $\delta_{ij} = 1$ if subject j experiences an event at time t_{ij} in instance i, and $\delta_{ij} = 0$ if j is censored at time t_{ij} in instance i.

We assume some subjects are more likely to experience the event following a certain group of subjects, but some others may not. This assumption implies there may exist a network underlying the dependence structure of event occurrences among the m subjects. Let H denote the links of the network. Now we have collected n independent instances $\{(t_1, \delta_1), (t_2, \delta_2), \ldots, (t_n, \delta_n)\}$. Can we use the n instances (t_i, δ_i)'s to estimate H, the links of the network underlying the dependence structure of the event occurrences?

2.2. *Synthetic diffusion data.* Below we describe a probabilistic model for generating diffusion data in which the dependence structure is determined by a user-specified network containing m nodes. In the diffusion data, each observation is an event time vector of the m nodes. Let $Y_k(t)$ denote the number of events occurred for node k in the time interval $[0, t]$. We assume $Y_k(t)$ follows a Poisson distribution with mean

$$(2.1) \qquad \Theta_k(t) = \int_0^t \exp\{\eta_k(u)\}\, du,$$

where

$$(2.2) \qquad \eta_k(u) = \sum_{j:(j,k)\in H} \gamma_{jk} \exp\{-(u - t_j)\}\mathbb{I}\{u > t_j\}.$$

In (2.2), γ_{jk} is a parameter that measures the impact of node j on node k, t_j is the event time of node j, and H is a set that contains links of the network underlying the dependence structure of the impacts among the nodes. For practical purposes we assume γ_{jk} is nonnegative-valued. Equation (2.2) implies that node j can only influence the value of $\eta_k(u)$ when there is a directed link from j to k, and when the event time of j is prior to the event time of k.

2.2.1. *The data set.* To generate the diffusion data, we first need to specify the network H. One may specify it by asking for expert knowledge, or draw it from some random graph model. After specifying H, we then determine the nonnegative weight γ_{jk} for each link in H. Again, one may consult experts for specifying the value of γ_{jk} or randomly drawing it from a gamma distribution. In our case, we draw it from Gamma(1, 1). Next we select a starting node according to a probability with the mass function proportional to the out-degree of the node. Under this probability mechanism, the more the out-links a node has, the more likely the node will be selected. We then go on to generate random variable $Y_k(t)$ for each

node from the Poisson hierarchical model in (2.1) and (2.2). We use the trapezoidal rule to numerically compute the mean function $\Theta_k(t)$. The event time for node k is defined as $t_k = \min\{t : Y_k(t-) = 0, Y_k(t) > 0\}$. In addition, during the data generating process, we also randomly generate censoring time for each node according to a pre-specified censoring probability. The event times and censoring times generated from the data generating process described above are collected to form the diffusion data.

Figure 1 shows three instances from the Poisson hierarchical model (2.1) and (2.2) with censoring probability equal to 0.1 at each time point. On the left hand side of Figure 1, black dots represent nodes who experienced an event during the observation period ($\delta_{ij} = 1$), while white dots represent nodes whose observations were right-censored ($\delta_{ij} = 0$). The network underlying the dependence structure for generating the three instances was generated from the Barabási–Albert model [Barabási and Albert (1999)] $BA(m, w)$ with m, the number of nodes, equal to 100, and w the value of the exponent of the degree distribution, equal to 1. We generated the network using R package "igraph" [Csardi and Nepusz (2006)].

2.2.2. *Motivation and aims.* The synthetic diffusion data contain event times and censoring indices of 100 nodes, which correspond to t_i and δ_i in our problem setting, respectively. Here we are interested in using these event times and censoring indices to recover the link set H in (2.2). In addition, the values of the layout coordinates shown on the right hand side of Figure 1 provide spatial information about the 100 nodes. Such exogenous information may be useful in estimating H. We will provide more details on modeling the synthetic diffusion data and estimating the link set H in Section 5.

2.3. *Campaign donation data.* In modern electoral politics, campaign donation plays an important role in influencing election results. While candidates need donations from firms to run their election campaigns, firms also have a strong interest to build connections and maintain their access to political influence through the donations. When different firms make donations to the same candidate, they become embedded in a network driven by the donations. Such a donation-driven network of firms reflects political cohesion among the firms [Burris (2005), Mizruchi (1989)]. Earlier scholarly efforts have been devoted mostly to identify the mechanisms that lead to such consensus. Important mechanisms include direct social interactions among business elites in the boardrooms, or competition among firms that may impose the conformity pressure and drive firms to imitate each other in their donations [Burris (2005), Mizruchi (1992)]. Empirical analyses have thus focused on how donation decisions of firms are influenced by networks generated by shared board members of firms or market competitions emerged from the same market niche.
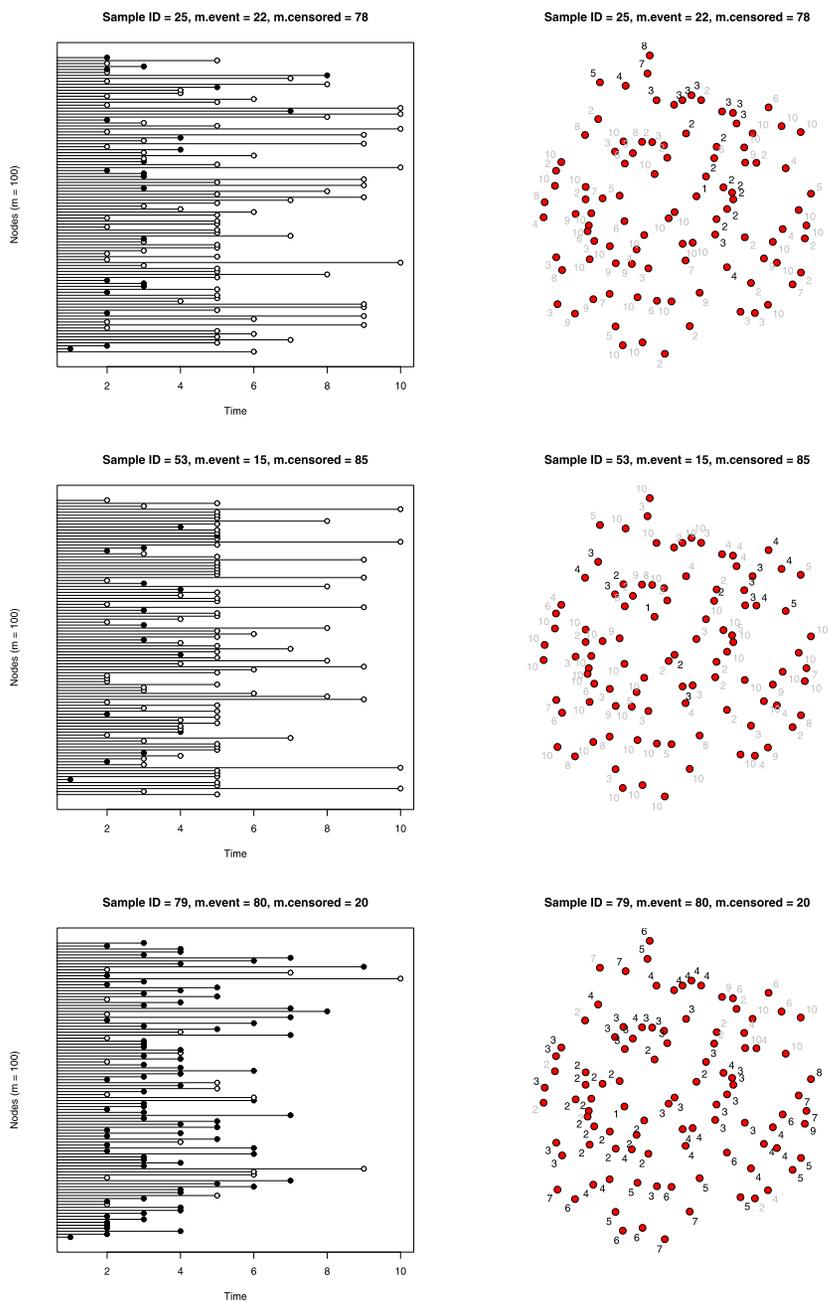
Sample ID = 25, m.event = 22, m.censored = 78

Sample ID = 25, m.event = 22, m.censored = 78

Sample ID = 53, m.event = 15, m.censored = 85

Sample ID = 53, m.event = 15, m.censored = 85

Sample ID = 79, m.event = 80, m.censored = 20

Sample ID = 79, m.event = 80, m.censored = 20

Fig. 1. *Three instances from the simulated data.* Left: *Event time plots of the nodes. The x-axis represents the time, and the y-axis represents the nodes. The black dot represents the time when an event occurred, and the white dot represents the time when the observation was censored.* Right: *Scatter plots of the node positions. The number in black color indicates the time when an event occurred, and the number in gray color indicates the time when the observation was censored.*

2.3.1. *The data set.*   The campaign donation records were collected during the 2008 Taiwanese legislative election. These records were provided by the Control Yuan of the Taiwanese government. Each record contains the time of the donation, the name of the firm, and the name of the candidate corresponding to a donation event. In the original records, some firms were subsidiaries of other firms. Prior studies [Burris (2005), Mizruchi (1992)] suggested such firms tended to make their donation decisions collectively. In this sense these individual firms can be seen as a business group. In Taiwan, such business groups were considered as the major agents in the local business and political activities [Chu (1994), Lee (2016), Numazaki (1986)]. Based on the above studies, we treated the business group as the basic unit in the subsequent analysis. To obtain the grouped donation records, we aggregated the original records, using the Unified Business Number, a business coding system provided by the China Credit Information Service (CCIS) to identify which business group a firm belongs to. In Supplementary Materials we provide further details on the data aggregation procedure. In addition, although by definition a business group is an organization aggregated from individual firms (or subsidiaries), as it serves as the basic unit in our analysis, it still can be seen as a firm when making donation decisions. Therefore to avoid confusion, we will call the business groups as firms in our subsequent analysis. The data aggregation led to 579 campaign donation records involving 146 firms (business groups) and 133 candidates. The first record occurred on April 20, 2007, and the final record occurred on January 28, 2008.

Among the 579 campaign donation records, a small fraction of them (33 events, or 5.7 percent) were recurrent events of donations between the same firm (business group) and the same candidate. Excluding these recurrent events leads to 546 campaign donation records. For practical purposes, we will only use records on the first donation for analysis. However, we will discuss issues on modeling the recurrent events in Section 7.1. Figure 2 shows event time plots of the 546 campaign contribution records. The plots suggest that events occurred sporadically during the first few months and became intensive during the final two months.

2.3.2. *Motivation and aims.*   Here we are interested in recovering the network underlying the dependence structure of the donation events of the 146 firms. We call this network the "event-time-driven political cohesion network of the firms", or simply the "political cohesion network" since it reflects the fact that donating money to a certain candidate is a kind of behavior collectively made by these firms. To investigate this question, we may treat the campaign donation records as the data set that contains 133 observed event time vectors corresponding to the 133 candidates. In addition, each event time vector is a 146-dimensional vector containing event times of donations from the 146 firms to the corresponding candidate. These event time vectors are shown in the plot at the bottom of Figure 2. We will discuss further modeling and estimation issues relating to the campaign donation data in Section 6.
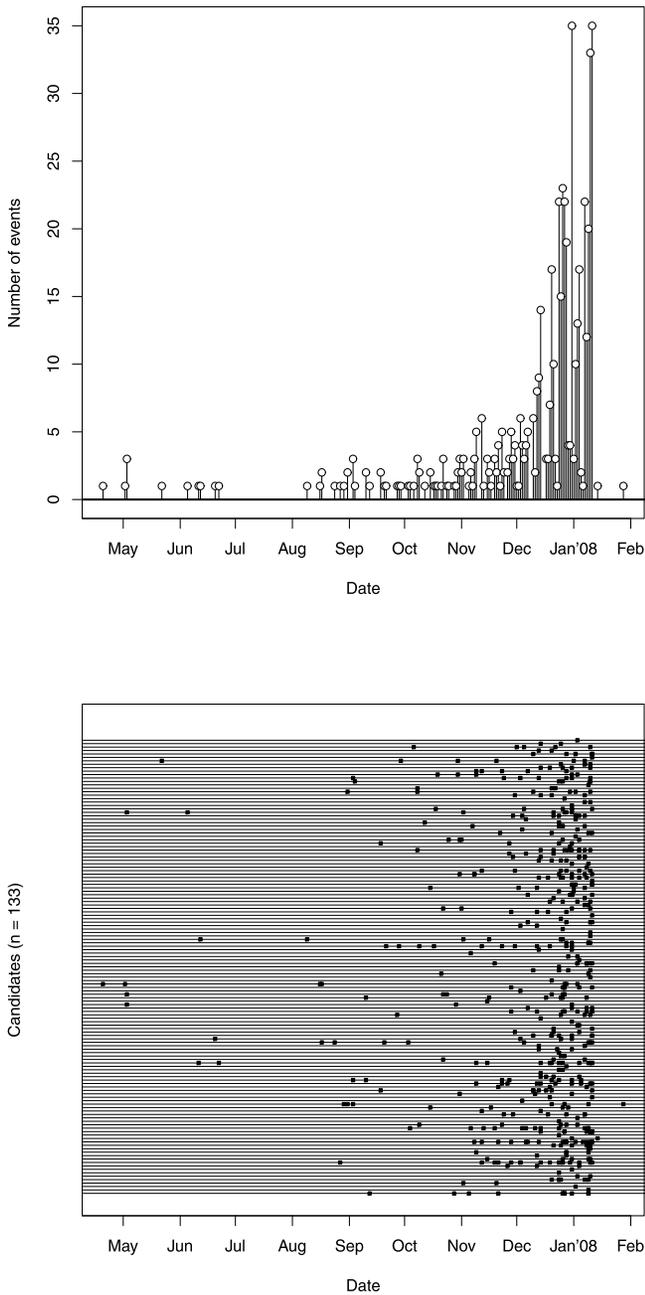
FIG. 2.   *Event time plots of the campaign donation records. Top: The plot of the number of donation events against time. The x-axis represents time, and the y-axis represents the number of events. Bottom: The plot of time when a candidate received a donation. The x-axis represent time, and the y-axis represents the candidate. Each square dot point represents the time when the corresponding candidate received a donation.*

3. The model. Let $\theta_{ik}(t)$ denote the hazard function of event time for node k in instance i. Since our aim is to estimate links of a network underlying the dependence structure of event occurrences among the m nodes, it is intuitive to model $\theta_{ik}(t)$ as a function of event times of other nodes. We model the hazard function $\theta_{ik}(t)$ by

$$\theta_{ik}(t) = \vartheta_i \exp\{\eta_{ik}(t; \alpha, \gamma)\},$$

(3.1)

$$\eta_{ik}(t; \alpha, \gamma) = \alpha_{kk} g_{ikk}(t) + \sum_{j \neq k} \gamma_{jk} g_{ijk}(t, t_{ij}),$$

where $\alpha = (\alpha_{11}, \alpha_{22}, \ldots, \alpha_{mm})$ and $\gamma = \{\gamma_{jk} : j \neq k, j, k = 1, 2, \ldots, m\}$, ϑ_i is a baseline hazard function specified for instance i, $g_{ijk}(t, t_{ij})$ is a user-specified covariates function that contains information about event time of node j in instance i, and γ_{jk} is a parameter that measures the impact of node j on node k. Note that although the covariate function $g_{ijk}(t, t_{ij})$ mainly contains information about event time of node j in instance i, one may modify it by including information about attributes of nodes k and j if such information is available in instance i at time t. We will discuss how to choose $g_{ijk}(t, t_{ij})$ in Section 3.1. In addition, it is the parameter set γ that quantifies the links of the network we are interested in. Formally we define the link set H as $H = \{(j, k) : \gamma_{jk} \neq 0 \text{ for } j \neq k, \text{ and } j, k = 1, 2, \ldots, m\}$. This definition implies that if dyadic pair $(j, k) \in H$, then $\gamma_{jk} \neq 0$, and event time of node j will have an impact on the hazard function for node k via the term $\gamma_{jk} g_{ijk}(t, t_{ij})$.

3.1. *Choices of the covariate function.* The covariate function $g_{ijk}(t, t_{ij})$ is a function of event time of node k and event time of node j. It should be context-dependent, relying on further assumptions from researchers and information relating to research issues. For example, consider a situation in which a disease is transmitted among the nodes. We want to know the network underlying such transmission. After knowning the properties of the disease, we may assume: (a) Only the infected node can transmit the disease to non-infected nodes; (b) Once a node is infected, the chance that it makes non-infected nodes get infected will decrease over time. Under these assumptions, we have several choices of $g_{ijk}(t, t_{ij})$. We may let

(3.2) $$g_{ijk}(t, t_{ij}) = \exp\{-(t - t_{ij})_+\} \mathbb{I}\{t > t_{ij}\} \mathbb{I}\{\delta_{ij} = 1\},$$

or

(3.3) $$g_{ijk}(t, t_{ij}) = \left[\frac{1}{(t - t_{ij})_+}\right] \mathbb{I}\{t > t_{ij}\} \mathbb{I}\{\delta_{ij} = 1\},$$

where $\mathbb{I}\{\mathcal{A}\}$ is an indicator function such that $\mathbb{I}\{\mathcal{A}\} = 1$ if $\mathcal{A}$ is true and $\mathbb{I}\{\mathcal{A}\} = 0$ otherwise, and δ_{ij} is an indicator function such that $\delta_{ij} = 1$ if node j experiences an event at time t_{ij} in instance i, and $\delta_{ij} = 0$ if j is censored at time t_{ij} in instance i.

The covariate functions (3.2) and (3.3) are non-increasing functions of t and only take nonnegative values. In addition, $g(t, t_{ij}) \to 0$ as $(t - t_{ij})_+ \to \infty$. If t_{ij} is censored, i.e. $\delta_{ij} = 0$, then $g(t, t_{ij}) = 0$, and the impact of node j on the hazard function of node k will vanish.

In practice, researchers should choose the covariate function $g_{ijk}(t, t_{ij})$ based on the research question they want to address. For example, if the transmitted disease is subject to a delay mechanism under which the infected node is unable to transmit the disease to other nodes until at least d days after the infection, then we may modify the covariate function as

$$g_{ijk}(t, t_{ij}) = \exp\{-(t - t_{ij} - d)_+\}\mathbb{I}\{(t - t_{ij})_+ > d\}\mathbb{I}\{\delta_{ij} = 1\}.$$

Under the above covariate function, we have $g_{ijk}(t, t_{ij}) = 0$ for $(t - t_{ij}) \le d$, which implies even node j has been infected with the disease at time t_{ij}, it will be unable to transmit the disease to node k within d days after the infection. This setting provides a delay mechanism for modeling disease transmission with an extra parameter d, which may be specified by researchers.

On the other hand, if we have information about node attributes, we may define

$$(3.4) \quad g_{ijk}(t, t_{ij}) = \exp\left\{-(t - t_{ij})_+ - \frac{\|x_k - x_j\|_2^2}{2\sigma^2}\right\}\mathbb{I}\{t > t_{ij}\}\mathbb{I}\{\delta_{ij} = 1\},$$

$$(3.5) \quad g_{ijk}(t, t_{ij}) = \left[\frac{1}{(t - t_{ij})_+}\right]\exp\left\{-\frac{\|x_k - x_j\|_2^2}{2\sigma^2}\right\}\mathbb{I}\{t > t_{ij}\}\mathbb{I}\{\delta_{ij} = 1\},$$

where x_k and x_j are covariates containing information about attributes of node j and node k, respectively, and σ^2 is a user-specified scaling parameter. Here the function $\exp\{-\|x_k - x_j\|_2^2/(2\sigma^2)\}$ is used to measure similarity between node k and node j. Such a formulation is based on the idea of homophily [McPherson, Smith-Lovin and Cook (2001)] in that a node may have a higher chance of getting the disease from another inflected node if the two nodes are similar to each other.

4. Estimation. Under model (3.1), parameters γ_{jk}'s quantify the link set H. Below we develop an estimation procedure for estimating γ_{jk}'s. This estimation procedure is based on maximization of a partial likelihood function. To derive the partial likelihood function, we first pool all n instances together and assume events in the n instances occurred at q time points $0 \le u_1 \le u_2 \le \cdots \le u_q < \infty$. For notation simplicity, let $\theta_{isk} = \theta_{ik}(u_s)$. The probability that node k will experience an event in instance i around the time point u_s is

$$(4.1) \qquad\qquad p_{isk} = \frac{\theta_{isk}}{\sum_{k' \in R_{is}} \theta_{isk'}},$$

where R_{is} is the risk set associated with instance i at time u_s. Note that in model (3.1) the hazard function $\theta_{ik}(u_s)$ is defined conditional on the event times of other nodes, and therefore the probability (4.1) is a probability conditional on

the event times of other nodes observed in instance i until the time point u_s. With p_{isk}, we can derive a conditional likelihood function of $\theta = \{\theta_{isk}\}_{i,s,k}$ given data observed in instance i until the time point u_s. This conditional likelihood function is

$$L_{is}(\theta) = \prod_{k=1}^{m} p_{isk}^{\mathbb{I}\{k \in D_{is}\}} = \prod_{k \in D_{is}} \frac{\theta_{isk}}{\sum_{k' \in R_{is}} \theta_{isk'}},$$

where D_{is} is the event set associated with instance i at the time point u_s. With $L_{is}(\theta)$, we can go further to derive the conditional likelihood function of θ given all observed data. Let $l(\theta)$ denote logarithm of the conditional likelihood function given all observed data. The conditional likelihood function is the multiplication of $L_{is}(\theta)$ over instances i's and event time points u_s's:

$$
\begin{aligned}
l(\theta) &= \log\left[\prod_{i=1}^{n} \prod_{s=1}^{q} L_{is}(\theta)\right] \\
&= \log\left\{\prod_{i=1}^{n} \prod_{s=1}^{q} \left(\prod_{k \in D_{is}} \frac{\theta_{isk}}{\sum_{k' \in R_{is}} \theta_{isk'}}\right)\right\}.
\end{aligned}
$$
(4.2)

Now let $\eta_{isk}(\alpha, \gamma) = \eta_{ik}(u_s; \alpha, \gamma)$. Following model (3.1) we have $\theta_{isk} = \vartheta_i \exp\{\eta_{isk}(\alpha, \gamma)\}$, and then (4.2) becomes

$$l(\alpha, \gamma) = \log\left\{\prod_{i=1}^{n} \prod_{s=1}^{q} \frac{\exp\{\sum_{k \in D_{is}} \eta_{isk}(\alpha, \gamma)\}}{[\sum_{k' \in R_{is}} \exp\{\eta_{isk'}(\alpha, \gamma)\}]^{|D_{is}|}}\right\},$$
(4.3)

where the instance-specific baseline hazard function ϑ_i is canceled out. We estimate α and γ by

$$(\widehat{\alpha}, \widehat{\gamma}) = \arg\max_{\alpha \in \mathbb{R}^m, \gamma \in \mathcal{C}} \left\{\frac{1}{n} l(\alpha, \gamma) - \lambda \sum_{k=1}^{m} \text{pen}(\gamma_k)\right\},$$
(4.4)

where $\mathcal{C}$ is a convex set, and $\text{pen}(\gamma_k)$ is a function used to regularize estimated values of $\gamma_k = \{\gamma_{jk}\}_{j \neq k}$, and λ is a tuning parameter used to control the impact of $\sum_{k=1}^{m} \text{pen}(\gamma_k)$ on the estimation. We then estimate the link set H by

$$\widehat{H} = \{(j, k) : \widehat{\gamma}_{jk} \neq 0\}.$$

In practice we may estimate γ by setting $\mathcal{C} = \mathbb{R}^{m(m-1)}$ or $\mathcal{C} = \{\gamma : \gamma \geq 0\}$, and $\text{pen}(\gamma_k) = \|\gamma_k\|_1$. We provide an algorithm for numerically solving the estimation problem (4.4) in Supplementary Materials.

4.1. *Tuning parameter selection.* We adopt the following steps to select tuning parameter λ. We first run estimation with different values of λ chosen from interval $[\lambda_{\min}, \lambda_{\max}]$. We evaluate an information criterion using estimates based

on different values of λ. We then select the one that minimizes the information criterion as the optimal value of λ. In practice we use

$$(4.5) \qquad \mathrm{BIC}(\lambda) = -2l\big(\widehat{\alpha}(\lambda), \widehat{\gamma}(\lambda)\big) + |\widehat{H}_\lambda| \log n$$

to select the value of λ. Here $\widehat{\alpha}(\lambda)$ and $\widehat{\gamma}(\lambda)$ are estimates of α and γ at λ, respectively, $\widehat{H}_\lambda = \{(j,k) : \widehat{\gamma}_{jk}(\lambda) \neq 0\}$, and $|\widehat{H}_\lambda|$ is the number of elements in $\widehat{H}_\lambda$.

4.2. *Stability selection.* Information-based model selection criteria like BIC are unable to provide meaningful risk control such as those on the number of false positives in the estimation. Currently there exist several methods for risk control in high-dimensional model estimation. One of the methods is the Stability Selection method proposed by Meinshausen and Bühlmann (2010) [see also Shah and Samworth (2013)]. The Stability Selection is a subsampling-based method that controls the number of false positives by applying an explicit upper bound for the expected number of false positives. It estimates link set H by

$$(4.6) \qquad \widehat{H}^{\mathrm{stable}} = \left\{ (j,k) : \max_{\lambda \in \Lambda} (\Pi_{\lambda,(j,k)}) \geq \pi_{\mathrm{thr}} \right\},$$

where $\Lambda = [\lambda_{\min}, \lambda_{\max}]$, π_{thr} is a threshold value specified by researchers, $\max_{\lambda \in \Lambda}(\Pi_{\lambda,(j,k)})$ is the selection probability corresponding to dyad (j,k), and $\Pi_{\lambda,(j,k)} = \mathbb{E}_I[\mathbb{P}\{(j,k) \in \widehat{H}_{\lambda,I}\}]$, where I is a random subsample with size $\lfloor n/2 \rfloor$, and $\widehat{H}_{\lambda,I}$ is an estimate of H using subsample I at λ.

To explain how the Stability Selection method works, first let $\widehat{H}^{\mathrm{FP,stable}}$ denote the index set of false positives in the estimator $\widehat{H}^{\mathrm{stable}}$. According to Meinshausen and Bühlmann (2010), under some regularity conditions, the expectation of the number of false positives $h = \mathbb{E}(|\widehat{H}^{\mathrm{FP,stable}}|)$ is bounded by

$$(4.7) \qquad h \leq \frac{1}{(2\pi_{\mathrm{thr}} - 1)} \frac{q_\Lambda^2}{[m(m-1)]},$$

where π_{thr} is the same as defined above, $q_\Lambda = \mathbb{E}_I(|\widehat{H}_{\Lambda,I}|)$, $\widehat{H}_{\Lambda,I} = \bigcup_{\lambda \in \Lambda} \widehat{H}_{\lambda,I}$, and $m(m-1)$ is the number of parameters in the estimation.

A main advantage of the Stability Selection method is that it does not require researchers to specify a value for the tuning parameter λ in the estimation. In addition, by fixing the number of false positives h and q_Λ, one can use inequality (4.7) to obtain a value for the threshold probability π_{thr} in terms of h and q_Λ. On the other hand, one can also fix π_{thr} and q_Λ, and use the inequality (4.7) to obtain an upper bound for h.

In practice, we estimate q_Λ and $\Pi_{\lambda,(j,k)}$ as follows. We first draw a sequence $\{\lambda_b\}_{b=1}^B$ such that $\lambda_{\min} \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_B \leq \lambda_{\max}$ from the interval $[\lambda_{\min}, \lambda_{\max}]$. We estimate the link set H at each λ_b using subsample I with size $\lfloor n/2 \rfloor$ from the original sample. We run such estimation for several times using different subsamples. Let $\mathcal{I}$ denote the set of the subsamples. We estimate q_Λ by $\widehat{q}_\Lambda =$

$|\mathcal{I}|^{-1} \sum_{I \in \mathcal{I}} |\bigcup_{\lambda \in \{\lambda_b\}_{b=1}^{B}} \widehat{H}_{\lambda,I}|$, and $\Pi_{\lambda,(j,k)}$ by $\widehat{\Pi}_{\lambda_b,(j,k)} = |\mathcal{I}|^{-1} \sum_{I \in \mathcal{I}} \mathbb{I}\{(j,k) \in \widehat{H}_{\lambda_b,I}\}$. Because we only use $\lambda \in \{\lambda_b\}_{b=1}^{B}$ to compute these estimates, therefore we replace Λ with $\{\lambda_b\}_{b=1}^{B}$ when computing the selection probability $\max_{\lambda \in \Lambda} \Pi_{\lambda,(j,k)}$ in (4.6). With $\widehat{\Pi}_{\lambda_b,(j,k)}$, the Stability Selection estimator of H in (4.6) becomes

$$\widehat{H}^{\text{stable}} = \left\{ (j,k) : \max_{b \in \{1,2,\dots,B\}} (\widehat{\Pi}_{\lambda_b,(j,k)}) \geq \pi_{\text{thr}} \right\}.$$

In later sections we will combine our method with the Stability Selection for model selection.

5. Application to the synthetic diffusion data. We first demonstrated our method by estimating the hazard network model with the synthetic diffusion data described in Section 2.2. We then conducted several simulation experiments to evaluate performance of our method when combining it with the Stability Selection method for controlling the number of false discovered links.

5.1. *Model estimation and the results.* We estimated model (3.1) with the synthetic diffusion data described in Section 2.2. We used (3.4) as the covariate function in our model, with $\sigma^2 = 50$, and x_j and x_k being the two-dimensional vectors corresponding to the coordinates of nodes j and k in the scatter plots on the right hand side of Figure 1. In these plots each node was positioned according to the layout coordinates computed by the Fruchterman–Reingold algorithm. In addition, we selected the tuning parameter by using the Bayesian information criterion (4.5). The results are shown in Figure 3. The plot on the top left hand side of Figure 3 shows the true network underlying the dependence structure of event occurrences. From these plots we can see that when λ decreases to a certain value, the method eventually recovers most of the true links (TPR $= 0.95$). However, such recovery is done at the expense of a rapid increase in the number of false discovered links (FDR $= 0.59$).

We further investigated whether using a different covariate function $g_{ij}(t, t_j)$ will have an impact on the estimation result. In addition to the covariate function (3.4), we also considered covariate functions (3.2), (3.3) and (3.5). Note that since we had used covariate function (3.2) to generate the data, and therefore the model with covariate function (3.2) can be viewed as the true model. We labeled this model as the "exponential time" model in our analysis. On the other hand, models with other covariate functions (3.4), (3.3) and (3.5) may be viewed as the candidate models. We labeled these models in our analysis as the "exponential time with covariate" model, "inverse time" model and "inverse time with covariate" model, respectively. We estimated these models with the synthetic diffusion data and plotted ROC curves for the corresponding estimation results. The ROC curves are shown on the bottom right hand side of Figure 3. From the ROC curve plot we can see the inverse time model has an ROC curve similar to the exponential
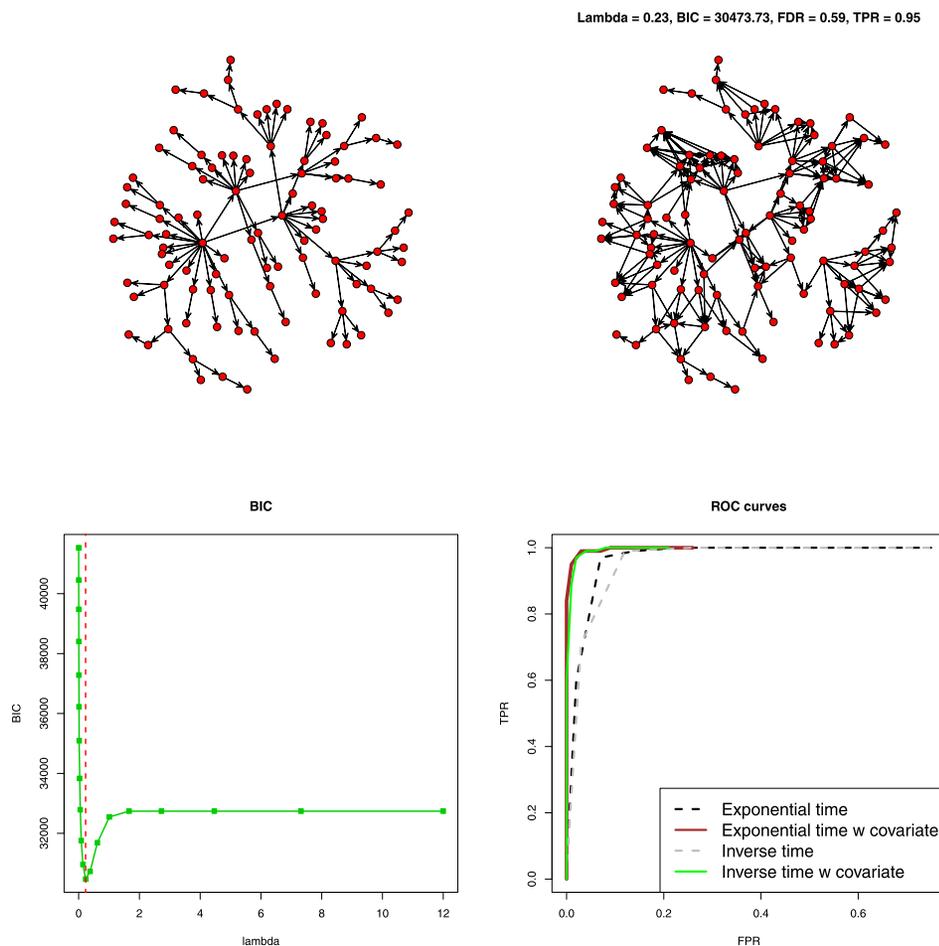
FIG. 3. *Estimation results from the synthetic diffusion data. Top left: The network underlying the dependent structure of the synthetic diffusion data. Nodes are positioned according to the layout coordinates computed by the Fruchterman–Reingold algorithm. Top right: The estimated links. Here BIC is the Bayesian information criterion, FDR is the false discovery rate, and TPR is the true positive rate. Bottom left: The BIC against the tuning parameter. The x-axis represents the value of tuning parameter, and the y-axis represents the value of the BIC. The red dash line indicates the BIC corresponding to the estimated links. Bottom right: ROC curves for estimations with the true model (the exponential time model) and candidate models (the exponential time with covariate model, the inverse time model and the inverse time with covariate model). The x-axis represents the false positive rate, and the y-axis represents the true positive rate.*

time model (the true model), while the exponential time with covariate model has an ROC curve similar to the inverse time with covariate model. The ROC curves imply that models with information about the layout coordinates performed better

than those without such information, suggesting our estimations may have bene-
fited from incorporating exogenous information in the covariate function.

5.2. *Further simulation experiments.*   We further evaluated our method under
two scenarios when the network underlying the dependence structure of event oc-
currences was generated from (a) the Watts–Strogatz model WS(m, w_1, w_2) [Watts
and Strogatz (1999)], where $m = 40$ is the number of nodes, $w_1 = 1$ is the num-
ber of neighbors on one side, and $w_2 = 0.04$ is the rewiring probability, and (b)
the Barabási–Albert model BA(m, w), where $m = 50$ is the number of nodes, and
$w = 1$ is the value of the exponent of the degree distribution. In each scenario, we
generated 100 networks. Each of the 100 networks was further used to generate
a sample of $n = 100$ instances of m-dimensional event time vectors according to
the Poisson hierarchical model (2.1) and (2.2). During the data generating process,
each node had censoring probability equal to 0.05.

Under each scenario, we considered three estimations. The first one was the
estimation using BIC for tuning parameter selection. The second one was the esti-
mation using the Stability Selection with the expected number of false discovered
links h controlled at a level less than or equal to $1/10$ of the number of the true
links. The last one was the estimation using stability selection with h controlled at
a level less than or equal to half of the number of the true links. We carried out the
first estimation for each network using samples with sizes n ranging from 10 to
100. For estimations with the Stability Selection, we only carried out them using
samples with sizes ranging from 50 to 100.

To evaluate the three estimations, we considered the following three measures:
(1) the true positive rate; (2) the false discovery rate; and (3) the ratio between the
number of realized false discovered links and the number of controlled false dis-
covered links h. We calculated the three performance measures for each estimation
under different sample sizes and then averaged the performance measures over the
100 estimations.

Figure 4 shows estimation results from data generated from the Watt-Strogatz
models WS(40, 1, 0.04) and the Barabási–Albert model BA(50, 1). These results
imply that our method could recover more and more true links as the sample size
increased. However, our method based on BIC yielded excessive numbers of false
discovered links. In contrast, when combined our method with the Stability Selec-
tion method, our method was able to find fewer false discovered links, and the true
positive rates increased steady as the sample size increased while the false discov-
ery rates remained at a low level. But it also came with a trade-off. Our method
yielded fewer true links when the number of false discovered links was controlled
at a low level. Nevertheless it was able to yield more and more true links as the
false discovery control was loosened. Results for other simulation experiments can
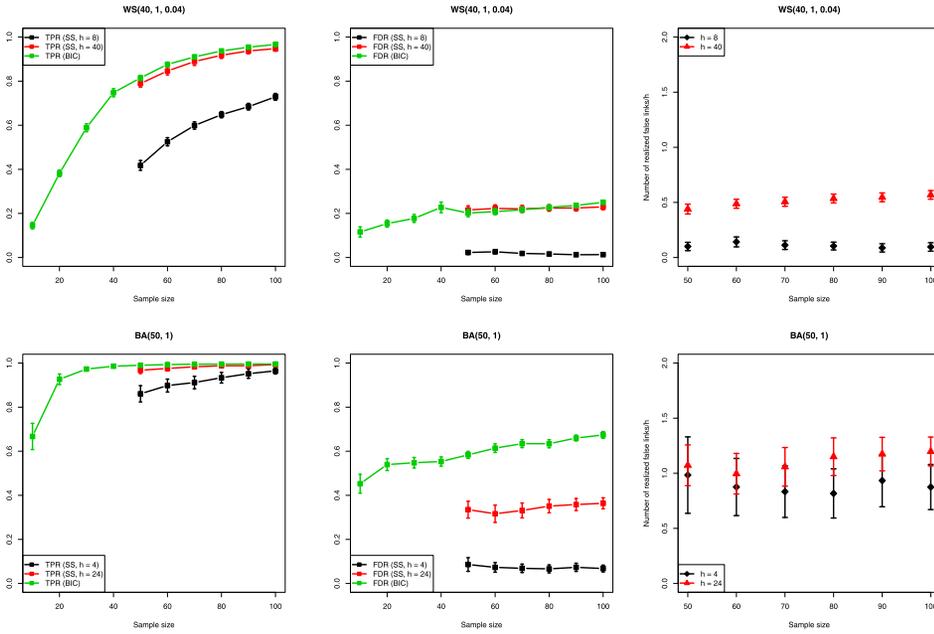be found in Supplementary Materials.

FIG. 4. *Estimation results from simulated data. The x-axis represents the sample size, and the y-axis represents the true positive rate or the false discovery rate. Top*: *Estimation results for* WS(40, 1, 0.04). *Bottom*: *Estimation results for BA*(50, 1). *Left*: *Plots of the true positive rate against the sample size. Center*: *Plots of the false discovery rate against the sample size. Right*: *Plots of the ratio between the realized number of false discovered links and the nominal number of false discovered links against the sample size.*

6. Application to the campaign donation data. Below we analyze the campaign donation data described in Section 2.3. We discuss the modeling strategy and the model fitting procedure before showing the results.

6.1. *Modeling the political cohesion network.* We were mainly interested in understanding the political cohesion network that drives the firms to donate following other firm's donations. To model links of the political cohesion network underlying the firms' donation behavior, we assumed that logarithm of the hazard function of donation made by firm k to candidate i at time t is a function of event times of donations made by other firms to candidate i prior to time t. We modeled the hazard function by

$$(6.1) \qquad \log \theta_{ik}(t) = \varsigma_i + \alpha_{kk} + \sum_{j \neq k} \gamma_{jk} g_{ijk}(t, t_{ij}),$$

where

$$g_{ijk}(t, t_{ij}) = \exp\left\{\frac{-(t - t_{ij})_+}{7}\right\} \mathbb{I}\left\{0 < \frac{(t - t_{ij})_+}{7} \leq d\right\} \mathbb{I}\{\delta_{ij} = 1\},$$

and t_{ij} is the event time of the donation made by firm j to candidate i. Model (6.1) has a candidate-specific term ς_i and a firm-specific term α_{kk}. From Figure 2 we can see that donation events occurred intensively in the last two month of year 2007, indicating that a firm's donation behavior might be influenced by factors such as the seasonal effect or those not generated from the network underlying firms' donation behavior, for example, candidate's attributes. Therefore in order to account behavioral differences between firms due to non-network effects, we included the terms ς_i and α_{kk} in the model. Note that model (6.1) is the same as model (3.1) with $\varsigma_i = \log \vartheta_i$ and $g_{ikk}(t) = 1$. In addition, the covariate function $\exp\{-(t - t_{ij})_+/7\}$ is a decreasing function of t, meaning that influences from other firms on firm k will decrease over time. Here the influences are measured on a weekly basis as the term $(t - t_{ij})_+$ has been scaled by 7. In addition, the quantity d controls the duration of such influences. Moreover, influences from other firms on firm k are restricted to those who had made donations to candidate i prior to the time when k makes a donation to candidate i. The setting of $g_{ijk}(t, t_{ij})$ implies that such influences can only exist up to d weeks and will decrease exponentially as time gap $(t - t_{ij})_+$ increases.

6.2. *Model estimation.*   In our analysis each candidate's donation record was treated as an instance, and therefore we had $n = 133$ instances. In addition, each firm was treated as a node, and therefore we had $m = 146$ nodes in our model. We let the influence limit $d = \infty$. This allows the impact of nodes j on the hazard function of node k diminishes to zero as the time gap $(t - t_{ij})_+$ increases to infinity. We defined January 31, 2008 as the censoring time as the last observed event occurred on January 28, 2008. We then estimated parameters in model (6.1) by applying our method with the Stability Selection method. To carry out the estimation, we first formed a subsample by randomly selecting $\lfloor n/2 \rfloor$ data points from the data. We estimated parameters in model (6.1) using the subsample with pre-specified tuning parameter values in interval $[\lambda_{\min}, \lambda_{\max}] = [0.01, 2]$. We then run the procedure for 100 times, collecting 100 subsample-based estimates. We used the 100 subsample estimates to calculate the selection probability of each link. Finally we set criteria to include links in our model according to each link's selection probability.

6.3. *The results.*   The top left plot of Figure 5 shows the selection probabilities of the estimated links of the political cohesion network of firms. This plot shows when the threshold value for the selection probability is 0.5, the number of estimated links in the network is 549. However, such a number decreases to 11 when the threshold value increases to 0.9. Figure 6 shows the graph of the political cohesion network of the firms based on the estimation result with the threshold value equal to 0.6, a value suggested by Meinshausen and Bühlmann (2010). We plotted nodes with different colors to show whether the corresponding firms are members of the National Association of Industry and Commerce (NAIC), an elite
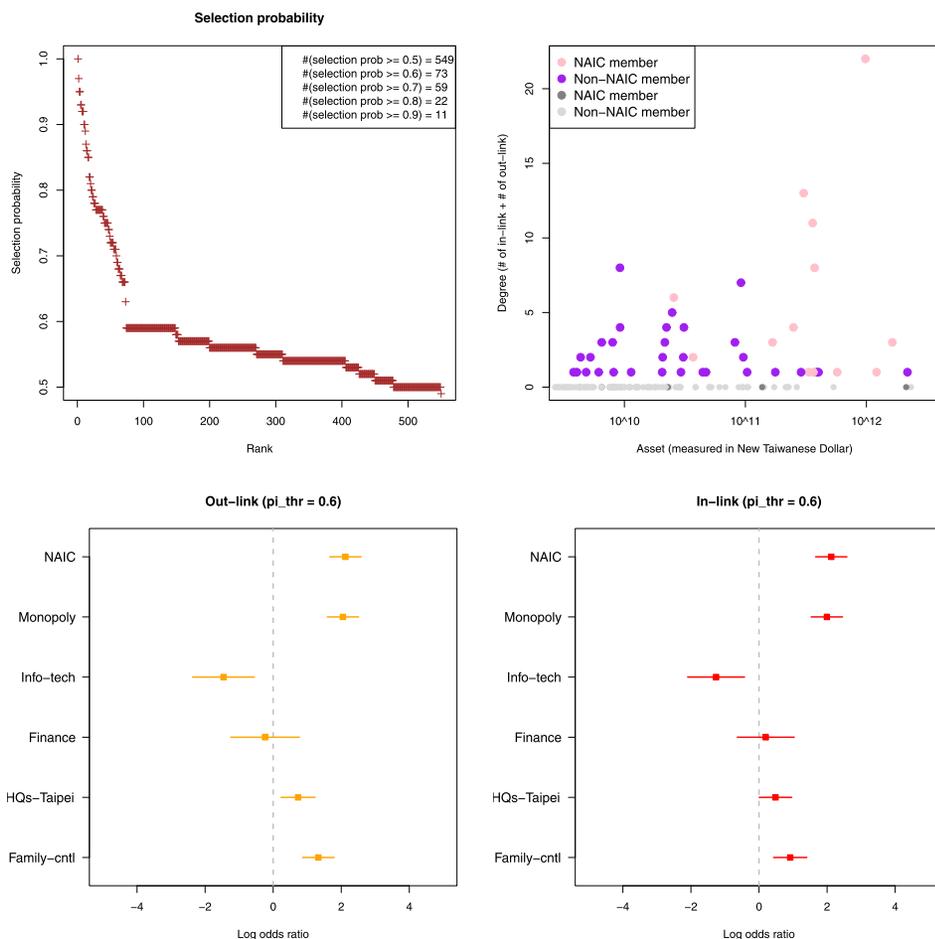
FIG. 5. *Estimation results for the campaign donation data. Top left*: *Selection probability of the link when $d = \infty$. Here we only show the estimated links with selection probabilities greater than or equal to* 0.5 *since* 0.5 *is the theoretical lower bound of the selection probability under the Stability Selection method. Top right*: *Scatter plot of degree of the firm against asset of the firm based on the estimation results when $d = \infty$ and $\pi_{\text{thr}} = 0.6$. Here pink and purple-colored nodes represent those with either in-degree or out-degree great than zero, while grey-colored nodes represent those with both in-degree and out-degree equal to zero. Bottom left*: *Plot of the logarithm of the odds ratio for out-links. Bottom right*: *Plot of the logarithm of the odds ratio for in-links. The x-axis represents the logarithm of the odds ratio, and the y-axis represent the attributes of firms.*

organization that holds a wide range of political influence on domestic politics in Taiwan. We also plotted nodes with different sizes to show how large the asset that the corresponding firm has. We provided more graphs based on estimation results with other threshold values in the Supplementary Materials.
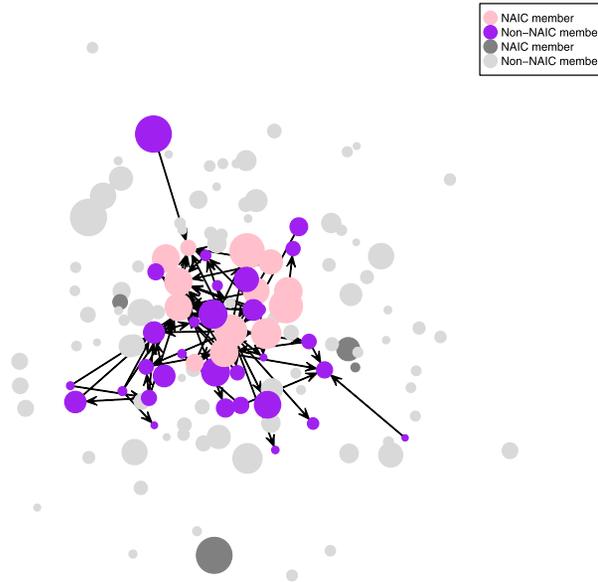
Estimated diffusion path (pi_thr = 0.6)

FIG. 6. *Graph based on the estimation results when $d = \infty$ and $\pi_{\text{thr}} = 0.6$. The size of the node represents the asset of the corresponding firm measured at the logarithm scale. In addition, pink and purple-colored nodes represent those with either in-degree or out-degree great than zero, while grey-colored nodes represent those with both in-degree and out-degree equal to zero.*

6.4. *Further analysis.* Based on the estimation result given above, we further investigated the relationship between a firm's asset and its degree in the political cohesion network. We provided a scatter plot of the two variables on the top right hand side of Figure 5. The Kendall rank correlation coefficient between the two variables is 0.264 (p-value $= 4.268 \times 10^{-5}$), suggesting a moderate dependence relationship between a firm's asset and its degree in the political cohesion network.

We were also interested in whether firms with a certain attribute would influence or be influenced by firms without such an attribute. To investigate this question, we computed logarithm of the odds ratio with respect to the numbers of links between firms with a certain attribute and firms without such an attribute. We considered the following six attributes: (1) whether a firms is a member of the National Association of Industry and Commerce (NAIC); (2) whether a firm belong to a monopoly industry; (3) whether a firm belongs to the information technology industry; (4) whether a firm belongs to the finance industry; (5) whether a firm locates its headquarters in Taipei; (6) whether a firm is family-controlled. We defined the probability that there is an out-link from a firm with attribute $\mathcal{K}$ by

(6.2) $\qquad p_{\mathcal{K}} = \mathbb{P}(\text{A firm has an out-link} \mid \text{The firm has attribute } \mathcal{K}).$

We estimated $p_{\mathcal{K}}$ by

$$\widehat{p}_{\mathcal{K}} = \frac{\#\{(j,k) : (j,k) \in \widehat{H}, j \text{ has attribute } \mathcal{K}\}}{m_{\mathcal{K}}(m-1)},$$

where $\widehat{H}$ is the estimated link set, $m_{\mathcal{K}}$ is the number of firms with attribute $\mathcal{K}$, and m is the number of firms in the network. The estimated logarithm of the odds ratio with respect to the numbers of out-links between firms with attribute $\mathcal{K}$ and firms without attribute $\mathcal{K}$ is

(6.3)
$$\widehat{\log \text{OR}}_{\mathcal{K}} = \log\left(\frac{\widehat{p}_{\mathcal{K}}(1 - \widehat{p}_{\mathcal{K}^c})}{\widehat{p}_{\mathcal{K}^c}(1 - \widehat{p}_{\mathcal{K}})}\right).$$

We computed the standard error of (6.3) using the following formula:

(6.4)
$$\text{Std err.}(\widehat{\log \text{OR}}_{\mathcal{K}}) = \left(\frac{1}{m_{\mathcal{K}}(m-1)\widehat{p}_{\mathcal{K}}} + \frac{1}{m_{\mathcal{K}}(m-1)(1 - \widehat{p}_{\mathcal{K}})}\right.$$
$$\left. + \frac{1}{m_{\mathcal{K}^c}(m-1)\widehat{p}_{\mathcal{K}^c}} + \frac{1}{m_{\mathcal{K}^c}(m-1)(1 - \widehat{p}_{\mathcal{K}^c})}\right)^{1/2}.$$

For the in-link cases, we followed a similar way to define the probability (6.2) and computed the corresponding logarithm of the odds ratio.

The results are shown in the bottom plots of Figure 5. The results imply that if firms are affiliated with the National Association of Industry and Commerce (NAIC), or belong to a monopoly industry, or are family-controlled, then they will be more likely to diffuse the donation behavior. On the other hand, firms with headquarters in Taipei are more likely to follow other firms to donate to the same candidate. In contrast, firms in the information technology industry are more likely to make independent decisions on campaign donations.

7. Discussion. We have developed a statistical method for estimating links of a network from time to event data under the right-censoring mechanism. When applying to the synthetic diffusion data, our method can reduce the number of false discovered links when combining with the Stability Selection method, and yield more and more true links as the false discovery control is loosened. When applying to the campaign data, our method has recovered the network underlying the dependence structure of donation events among 146 firms. By computing logarithm of the odds ratio with respect to the numbers of links between firms with a certain attribute and firms without such an attribute, we found that if firms are affiliated with the National Association of Industry and Commerce (NAIC), or belong to a monopoly industry, or are family-controlled, then they will be more likely to diffuse the donation behavior. On the other hand, firms with headquarters in Taipei are more likely to follow other firms to donate to the same candidate. In contrast, firms in the information technology industry are more likely to make independent decisions on campaign donations.

7.1. *Modeling recurrent events.*   When analyzing the campaign donation data, we only used the first campaign donation between a firm and a candidate and discarded subsequent donations involving the same firm and the same candidate from analysis. Such recurrent events are important since they may reflect the structure of the political cohesion network of the firms. However, modeling the recurrent events and assessing their impacts on the estimation result are challenging. One possible way to model the recurrent events is to allow firms to appear repeatedly in the risk set in the likelihood function, and at the same time assume each regression coefficient as a time-varying function. Estimating such a model may require large amounts of data since modeling each regression coefficient as a time-varying function needs extra parameters. Despite these difficulties, it is worth to including modeling recurrent events in the list of future research directions.

7.2. *Extensions.*   Although in this paper our method was applied to recover the political cohesion network of firms from the campaign donation data, it can also be applied to other research areas in the social network analysis. It has a potential to deal with various arrays of empirical social phenomenon. In contrast to earlier studies, which usually saw the firm-to-firm network as observable and paid attention to identify how political cohesion among firms are achieved by influences from the pre-existing links, our goal in this paper is different. Similar to the classical event-based strategy for network boundary delimitation [Homans (1950)], our method aims to identify the regularized network structure with information from event records. We can push one step further and assume that the pre-existing links are unobservable and only information about event activities (i.e., campaign donations of the firms) is available. With the statistical method we proposed, we are able to estimate the pivotal network structure of cohesion to which a social group or a community are able to achieve. Researchers have long noted when a network is not appropriately defined, research findings corresponding to the network will be seriously biased. Our method may provide a methodological solution to this classical boundary specification error that social network researchers may easily commit [Laumann, Marsden and Prensky (1982)].

Our method can also be used to deal with data in which the temporal order of events or behaviors is clearly documented [Rogers (1995), Strang and Soule (1998)]. For example, consider the online social media data in which users follow other users to click the "like" button on posts. In this situation, there is a time delay between the "like" stamps. Researchers may see the users as the firms and the posts as the candidates, and the goal is to identify links of the user-to-user network. With these links, researcher can further investigate issues such as who influences who, and who is the "hub" in the user-to-user network. Another example is the stock trading data in which a buyer may follow other buyers to buy a certain stock. In this setting, researchers may see the buyers as the firms and the stocks as the candidates, and the goal is to identify links of the buyer-to-buyer network.

Acknowledgments. We thank the Associate Editor and two anonymous referees, whose comments have greatly improved the paper.

SUPPLEMENTARY MATERIAL

Supplementary Materials for "Estimating links of a network from time to event data" (DOI: 10.1214/17-AOAS1032SUPP; .pdf). Supplementary Materials contain an numerical algorithm for obtaining estimator (4.4), further details on aggregation of the campaign donation data, additional results for simulation study and additional results for real data application.

REFERENCES

AMHED, A. and XING, E. (2009). Recovering time-varying networks of dependencies in social and biological studies. *Proc. Natl. Acad. Sci. USA* **106** 11878–11883.

ANANDKUMAR, A., TAN, V. Y. F., HUANG, F. and WILLSKY, A. S. (2012). High-dimensional structure estimation in Ising models: Local separation criterion. *Ann. Statist.* **40** 1346–1375. MR3015028

BANERJEE, O., EL GHAOUI, L. and D'ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. MR2417243

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

BURRIS, V. (2005). Interlocking directorates and political cohesion among corporate elites. *Am. J. Sociol.* **111** 249–283.

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

CHANDRASEKARAN, V., PARRILO, P. A. and WILLSKY, A. S. (2012). Latent variable graphical model selection via convex optimization. *Ann. Statist.* **40** 1935–1967. MR3059067

CHU, Y.-H. (1994). The realignment of business-government relations and regime transition in Taiwan. In *Business and Government in Industrialising Asia* (A. MacIntyre, ed.). Cornell Univ. Press, Ithaca, NY.

CSARDI, G. and NEPUSZ, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems* 1695.

DANAHER, P., WANG, P. and WITTEN, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 373–397. MR3164871

DANESHMAND, H., GOMEZ-RODRIGUEZ, M., SONG, L. and SCHÖLKOPF, B. (2014). Estimating diffusion networks: Recovery conditions, sample complexity & soft-thresholding algorithm. In *Proceedings of the* 31 *st International Conference on Machine Learning*. Beijing, China.

FREEMAN, L. (1977). A set of measures of centrality based on betweenness. *Sociometry* **40** 35–41.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

GOMEZ-RODRIGUEZ, M., LESKOVEC, J. and KRAUSE, A. (2012). Inferring networks of diffusion and influence. *ACM Trans. Knowl. Discov. Data* **5**.

HOMANS, G. C. (1950). *The Human Group*. Harcourt, New York.

KHARE, K., OH, S.-Y. and RAJARATNAM, B. (2015). A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 803–825. MR3382598

LAFFERTY, J., LIU, H. and WASSERMAN, L. (2012). Sparse nonparametric graphical models. *Statist. Sci.* **27** 519–537. MR3025132

LAUMANN, E. O., MARSDEN, P. V. and PRENSKY, D. (1982). The boundary specification problem in network analysis. In *Applied Network Analysis* (R. S. Burt, M. Minor, eds.) 18–34. Sage, Beverly Hills, CA.

LEE, Z.-R. (2016). Corporate power and democracy: An analysis of business groups' campaign contributions in the 2008 legislator election. *Taiwanese Sociology* **31** 43–83.

LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326. MR3059084

LOH, P.-L. and WAINWRIGHT, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Ann. Statist.* **41** 3022–3049. MR3161456

MCPHERSON, M., SMITH-LOVIN, L. and COOK, J. M. (2001). Bird of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27** 415–444.

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523

MIZRUCHI, M. S. (1989). Similarity of political behavior among large American corporations. *Am. J. Sociol.* **95** 401–424.

MIZRUCHI, M. S. (1992). *The Structure of Corporate Political Action*: *Interfirm Relations and Their Consequences*. Harvard Univ. Press, Cambridge, MA.

NEWMAN, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.* **89** 208701.

NUMAZAKI, I. (1986). Network of Taiwanese big business: A preliminary analysis. *Mod. China* **12** 487–534.

PENG, J., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591

QIU, H., HAN, F., LIU, H. and CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 487–504. MR3454206

RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343

REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695

ROGERS, E. M. (1995). *Diffusion of Innovations*. Free Press, New York.

SHAH, R. D. and SAMWORTH, R. J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. MR3008271

STRANG, D. and SOULE, S. A. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annu. Rev. Sociol.* **24** 265–290.

STRANG, D. and TUMA, N. B. (1993). Spatial and temporal heterogeneity in diffusion. *Am. J. Sociol.* **99** 614–639.

WATTS, D. J. and STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393** 440–442.

XUE, L., ZOU, H. and CAI, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *Ann. Statist.* **40** 1403–1429. MR3015030

YEN, T.-J, LEE, Z.-R, CHEN, Y.-H, YEN, Y.-M and HWANG, J.-S (2017). Supplement to "Estimating links of a network from time to event data." DOI:10.1214/17-AOAS1032SUPP.

YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824

T.-J. YEN
Y.-H. CHEN
J.-S. HWANG
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
128 ACADEMIA ROAD, SECTION 2, NANKANG
TAIPEI 11529
TAIWAN
E-MAIL: tjyen@stat.sinica.edu.tw
        yhchen@stat.sinica.edu.tw
        jshwang@stat.sinica.edu.tw

Z.-R. LEE
INSTITUTE OF SOCIOLOGY
ACADEMIA SINICA
128 ACADEMIA ROAD, SECTION 2, NANKANG
TAIPEI 11529
TAIWAN
E-MAIL: zlee@sinica.edu.tw

Y.-M. YEN
DEPARTMENT OF INTERNATIONAL BUSINESS
NATIONAL CHENGCHI UNIVERSITY
NO. 64, SECTION 2, ZHI-NAN ROAD, WENSHAN
TAIPEI 11605
TAIWAN
E-MAIL: yyu_min@nccu.edu.tw