Learning Network-Structured Dependence From Non-Stationary Multivariate Point Process Data

Muhong Gao, Chunming Zhang, and Jie Zhou

***Abstract*— Learning the sparse network-structured dependence among nodes from *multivariate point process* data $\{\boldsymbol{T}_i\}_{i\in\mathcal{V}}$ has wide applications in information transmission, social science, and computational neuroscience. This paper develops new continuous-time stochastic models of the *conditional intensity functions* $\{\lambda_i(t \mid \mathscr{F}_t) : t \geq 0\}_{i\in\mathcal{V}}$, dependent on past event counts of parent nodes, to uncover the network structure within an array of non-stationary *multivariate counting processes* $\{\boldsymbol{N}(t) : t \geq 0\}$ for $\{\boldsymbol{T}_i\}_{i\in\mathcal{V}}$. The stochastic mechanism is crucial for statistical inference of graph parameters relevant to structure recovery but does not satisfy the key assumptions of commonly used processes like the Poisson process, Cox process, Hawkes process, queuing model, and piecewise deterministic Markov process. We introduce a new *marked point process for intensity discontinuities*, derive compact representations of their conditional distributions, and demonstrate the cyclicity property of $\boldsymbol{N}(t)$ driven by recurrence time points. These new theoretical properties enable us to establish statistical consistency and convergence properties of the proposed penalized M-estimators for graph parameters under mild regularity conditions. Simulation evaluations demonstrate computational simplicity and increased estimation accuracy compared to existing methods. Real multiple neuron spike train recordings are analyzed to infer connectivity in neuronal networks.**

***Index Terms*— Consistency, generalized linear model, conditional intensity function, M-estimation, multivariate counting process, network structure.**

I. Introduction

STRUCTURED multivariate point process data, ranging from neuron multiple spike trains, file access patterns and failure events in server farms, queuing networks to social networks, has wide applications. Inference of the network structure underlying such multivariate point processes and addressing queries based on the learned structure are important issues. For example, learning the structure of cooperative activity between multiple neurons is an important task in understanding neural spike activity and identifying patterns of information transmission and storage in cortical circuits [1], [2], [3], [4], [5]. Analogously, learning the access patterns of files can be exploited for developing faster file access systems.

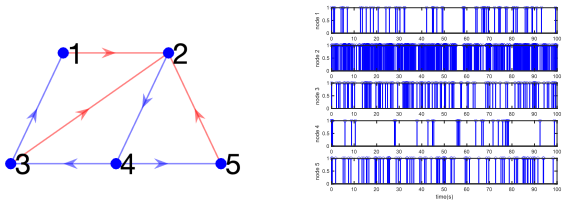

Fig. 1. Each node of the network graph in the left panel corresponds to a point process in the right panel. Arrows indicate interactions (red for excitatory and blue for inhibitory effects).

Typically, multivariate temporal *point processes* refer to random processes of occurrences of a particular event (such as neuron spike firing) in time, recorded at V nodes as $\{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_V\}$, where

$$\boldsymbol{T}_i = (T_{i,1}, \ \ldots, \ T_{i,\mathrm{N}_i})^\top \quad \text{with } 0 < T_{i,1} < \cdots < T_{i,\mathrm{N}_i} \leq \mathrm{T}, \quad \text{for } i \in \mathcal{V}. \tag{1}$$

These correspond to series of time points $T_{i,\ell}$ of the ℓth event, $\ell = 1, \ldots, \mathrm{N}_i$, arriving at the ith node in an experiment with time length T, where the superscript $\top$ denotes transpose, and $\mathcal{V} = \{1, \ldots, V\}$ is the node set. The corresponding *counting process*, $N_i(t) = \sum_{\ell\geq 1} \mathrm{I}(0 \leq T_{i,\ell} \leq t)$, tallies the number of events occurring up to time t at node $i \in \mathcal{V}$, where $\mathrm{I}(\cdot)$ denotes the indicator operator. An important objective is to extract the dependency structure among nodes within the network from V sequences of time series. This dependence network, also recognized as the 'local independence graph' [6], [7], visually represents the dependence relationship of historical events from parent nodes on the current events of child nodes. Figure 1 showcases the network-structured dependence (in the left panel) of multivariate point process data at 5 nodes (in the right panel).

Due to the stochastic nature of the point process data $\{T_{i,\ell}\}$ in (1) for event occurrences, two types of methods are relevant for modeling multivariate point process

Manuscript received 10 June 2023; revised 1 April 2024; accepted 24 April 2024. Date of publication 3 May 2024; date of current version 16 July 2024. The work of Muhong Gao was supported in part by China Postdoctoral Science Foundation under Grant E2909328. The work of Chunming Zhang was supported in part by the U.S. National Science Foundation under Grant DMS-2013486 and Grant DMS-1712418 and in part by the University of Wisconsin–Madison Office of the Vice Chancellor for Research and Graduate Education with funding from Wisconsin Alumni Research Foundation. The work of Jie Zhou was supported in part by the National Natural Science Foundation of China under Grant 12171329. *(Corresponding author: Chunming Zhang.)*

Muhong Gao is with the Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: gaomh@amss.ac.cn).

Chunming Zhang is with the Department of Statistics, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: cmzhang@stat.wisc.edu).

Jie Zhou is with the School of Mathematical Sciences, Capital Normal University, Beijing 100048, China (e-mail: zhoujie@amss.ac.cn).

Communicated by Y. Xie, Associate Editor for Statistics, Machine Learning, and Signal Processing.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3396778.

Digital Object Identifier 10.1109/TIT.2024.3396778

0018-9448 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

data. (a) The discrete-time modeling approach includes the dynamic Bayesian network [8], [9] and variants of generalized linear models (GLM) [3], [10], [11], [12]. This approach partitions the time axis into equally spaced time bins and transforms the series of event times into a sequence of event bin counts, empirically modelled by Poisson distributions. However, a major drawback is the tradeoff between discrete approximation error and the loss of information. (b) In contrast, the continuous-time approach aims to depict physical processes more accurately but faces substantial challenges in modeling both the time-varying part of *conditional intensity functions* (CIF) and the sparsity feature underlying the network structure. Several specific continuous-time point process models have been developed, such as the Cox process [13], [14], inhomogeneous Poisson process [15], the linear Hawkes process [16], [17], [18], and the non-linear Hawkes process [19], [20]. Other recent works analyzing point process data include [21], [22], [23], [24], with [24] focusing on spatiotemporal data (e.g., crime data) and [21], [22], [23] focusing on interaction data (e.g., E-mail/text messages). In particular, [22], [23] focus on identifying uniform effects (e.g., homophily, dyadic, and triadic effects) in a predetermined network.

To capture the unknown dependency structure between point process data represented in (1) both qualitatively and quantitatively, we aim to develop new network structure learning methods that integrate the utility of continuous-time and discrete-time modeling. Specifically, we build new continuous-time GLM-type stochastic models (12) for the *conditional intensity functions* $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$, where each $\lambda_i(t \mid \mathscr{F}_t)$ depends on short-term past events of all other nodes up to time t (in contrast to [13], [14], where the CIF is a separate stochastic process independent of the past events) and incorporates the magnitude and direction of interaction effects in graph parameters. By employing penalized M-estimation of parameters in the graph structure (as in (52)), we obtain a sparse network. In contrast, the consideration of sparsity was not present in [21], [22], [23], and [24]. Our method captures both excitatory and inhibitory effects between nodes, distinguishing itself from the linear Hawkes process [16], [17], specifically tailored for modeling excitatory effects to ensure a non-negative CIF. Furthermore, our method does not require partitioning the data into bins, making it partition-free and avoiding the subjective choice of bin width associated with the discrete-time approach.

Addressing the theoretical challenges arising from statistical learning procedures in continuous-time stochastic modeling remains a central issue. To the best of our knowledge, there are limited theoretical studies at the intersection of continuous-time point processes and network-structured learning methods. Traditional tools for establishing stochastic convergence and statistical consistency are not directly applicable in the context of statistical estimation from point process data. This is because the loss function (e.g., in (48)) for parameter estimation primarily relies on the non-standard dependence structure of counting processes $\{N_i(t)\}_{i\in\mathcal{V}}$ associated with the point process data $\{T_{i,\ell}\}$. While works for the non-linear Hawkes process [19], [25], queuing models [26], [27], and the piecewise deterministic Markov process [28] provide insights, they rely on specific assumptions and properties that do not hold for our model (12). Refer to Sections IV-A.3 and IV-B for more detailed discussions.

This paper aims to contribute to several aspects that are central to statistical inference for a wide array of non-stationary multivariate point process data encountered in various applications.

(i) We introduce a new tool called the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I}) = (\{\breve{T}_\ell\}_{\ell\geq 1}, \{I_\ell\}_{\ell\geq 1})$ for capturing intensity discontinuities (see Section IV-A). This tool involves compiling all the discontinuity points of the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ into a single sequence $\{\breve{T}_\ell\}_{\ell\geq 1}$, where each $\breve{T}_\ell$ is accompanied by a unique categorical mark $I_\ell \in \mathcal{V}\cup\{0\}$. We derive the probability distributions of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ and establish a series of probabilistic properties. These results for $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ also provide valuable insights into the probabilistic properties of the original counting processes $\{N_i(t)\}_{i\in\mathcal{V}}$ and enable the development of a new simulation algorithm for generating synthetic data.

(ii) We establish the cyclicity property of $\{N_i(t)\}_{i\in\mathcal{V}}$ driven by a sequence of recurrence time points $R_1 < R_2 < \cdots$ (see Section IV-B). This property demonstrates that our counting processes $\{N_i(t)\}_{i\in\mathcal{V}}$, upon reaching each recurrence time point $t = R_\ell$, initiate a renewed cyclic procedure independent of the event history, as illustrated in Figure 3 of Section IV-B. Building on this property, we further derive the asymptotic mean stationarity of $\{N_i(t)\}_{i\in\mathcal{V}}$.

(iii) All these probabilistic results are essential for deriving the statistical properties, such as the consistency of the proposed penalized M-estimation in structure learning, in Section V.

The validity of our proposed penalization method for inferring network-structured dependencies is supported by extensive simulation studies, and its practical utility in the analysis of real-world multivariate point process data is illustrated with a prefrontal cortex spike train dataset.

The rest of the paper is arranged as follows. Section II reviews the multivariate point process and outlines the proposed continuous-time modeling framework. Section III presents our proposed model for $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$, and Section IV investigates related probabilistic properties of $\{N_i(t)\}_{i\in\mathcal{V}}$. Section V addresses statistical properties related to the proposed network recovery procedure. Section VI illustrates simulation evaluations of the proposed method, and Section VII analyzes real spike train data. Section VIII briefly discusses and concludes the paper. Appendices A and B collect all supplementary simulations, technical details, and derivations.

II. Multivariate Point Process in Our Setup

We start with a brief review of the point process. For a more comprehensive discussion, refer to [29]. Denote by $\mathcal{V} = \{1, \ldots, V\}$ the set of nodes. Throughout the paper, we focus on the setting where the number V of nodes is a fixed constant. For each node $i \in \mathcal{V}$, we define the univariate *point process* as

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$\{T_{i,\ell}\}_{\ell\geq 1}$ on the probability space $(\Omega, \mathscr{F}, \mathrm{P})$, where the time-ordered sequence of event time points at node i is denoted as

$$0 < T_{i,1} < T_{i,2} < \cdots . \tag{2}$$

For $t \geq 0$, we use $N_i(t)$ to represent the event counts in the time interval $[0, t]$:

$$N_i(t) = \sum_{\ell\geq 1} \mathrm{I}(0 \leq T_{i,\ell} \leq t). \tag{3}$$

The term $\{N_i(t)\}_{t\geq 0}$ refers to the *counting process* of $\{T_{i,\ell}\}_{\ell\geq 1}$. More generally, we denote the event counts in any Borel set $\mathcal{T} \in \mathcal{B}(\mathbb{R})$ as:

$$N_i(\mathcal{T}) = \sum_{\ell\geq 1} \mathrm{I}(T_{i,\ell} \in \mathcal{T}), \tag{4}$$

which, for $\mathcal{T} = [0, t]$, reduces to $N_i(t)$ as defined in (3).

According to (3), a point process $\{T_{i,\ell}\}_{\ell\geq 1}$ uniquely defines a counting process $\{N_i(t)\}_{t\geq 0}$. Conversely, $\{N_i(t)\}_{t\geq 0}$ uniquely yields a point process, due to the identity $T_{i,\ell} = \inf\{t > T_{i,\ell-1} : N_i(t) > N_i(T_{i,\ell-1})\}$, where $T_{i,0} = 0$. Thus, the counting process $\{N_i(t)\}_{t\geq 0}$ and the point process $\{T_{i,\ell}\}_{\ell\geq 1}$ are equivalent to each other.

For the multivariate setting with V nodes, we define the vector $\boldsymbol{N}(t) = (N_1(t), \ldots, N_V(t))^\top$, and call $\{\boldsymbol{N}(t)\}_{t\geq 0}$ the *multivariate counting process*, corresponding to the *multivariate point process* $\{T_{i,\ell} : \ell \geq 1\}_{i\in\mathcal{V}}$. For each $t \geq 0$, let $\mathscr{F}_t \subseteq \mathscr{F}$ be the smallest sub σ-algebra that contains all the information of the multivariate counting process in the history up to time t, formally defined as:

$$\mathscr{F}_t = \sigma\big(\{N_i(s) : s \in [0, t],\ i \in \mathcal{V}\}\big), \tag{5}$$

where $\mathscr{F}_0 = \{\Omega, \varnothing\}$. From (5), it is seen that

$$\mathscr{F}_{t_1} \subseteq \mathscr{F}_{t_2} \subseteq \cdots, \quad \text{for any } 0 \leq t_1 \leq t_2 \leq \cdots . \tag{6}$$

We refer to the sequence of σ-algebras $\{\mathscr{F}_t\}_{t\geq 0}$ in (5), satisfying the property (6), as the *filtration generated by* $\{\boldsymbol{N}(t)\}_{t\geq 0}$, and call $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathrm{P})$ the corresponding *filtered probability space*.

A. Total Intensity Function of $\boldsymbol{N}(t)$

For a single node i, the stochastic character of a counting process $N_i(t)$ is captured by the corresponding CIF $\lambda_i(t \mid \mathscr{F}_t)$, which measures the instantaneous rate of event occurrence at node i. In this paper, we adopt the definition of the CIF from [30]:

$$\lambda_i(t \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}\big(N_i(t+\Delta) = N_i(t) + 1 \mid \mathscr{F}_t\big) \tag{7}$$

$$= \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}\big(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t\big) \tag{8}$$

almost surely (a.s.), for $i \in \mathcal{V}$ and $t \geq 0$.

For the multivariate case with V nodes, we similarly define the CIF of $\boldsymbol{N}(t)$ as:

$$\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}\big(\cup_{i\in\mathcal{V}}\{N_i(t+\Delta) = N_i(t) + 1\} \mid \mathscr{F}_t\big) \tag{9}$$

$$= \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}\big(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t\big), \quad \text{a.s.}, \tag{10}$$

where, for $s \neq t$, the event $\{\boldsymbol{N}(s) \neq \boldsymbol{N}(t)\}$ denotes $\cup_{i\in\mathcal{V}}\{N_i(s) \neq N_i(t)\}$.

Remark 1: Our definition of the CIF in (7) and (8), following [30], assumes that $\lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}(N_i(t+\Delta) = N_i(t)+1 \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0} \Delta^{-1}\mathrm{P}(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t)$ holds a.s. for every $i \in \mathcal{V}$ and $t \geq 0$. This assumption essentially means that simultaneous events from a single node are not allowed in our point process. As shown in Appendix B, any multivariate counting process with identical limits (7) and (8) also has identical limits (9) and (10) a.s. for every $t \geq 0$.

B. Orthogonality of Martingales of $\boldsymbol{N}(t)$

For the multivariate setting in our study, the structure of the counting process $\boldsymbol{N}(t)$ cannot be fully described by solely presenting the CIFs $\lambda_i(t \mid \mathscr{F}_t)$ at individual nodes i. Additionally, it is necessary to clarify how the increments of event counts, $N_i(t+\Delta) - N_i(t)$ and $N_j(t+\Delta) - N_j(t)$, are correlated between any pair of distinct nodes i and j. For $\boldsymbol{N}(t)$ in Definition 1 below, we introduce the notion of orthogonality of martingales (OM) which refers to the case where $N_i(t+\Delta) - N_i(t)$ and $N_j(t+\Delta) - N_j(t)$, conditional on $\mathscr{F}_t$, are asymptotically independent for all $i \neq j$.

*Definition 1 (Orthogonality of Martingales (*OM*)):* A multivariate counting process $\boldsymbol{N}(t)$ satisfies the OM condition if, for any two distinct nodes $i, j \in \mathcal{V}$ and any time $t \geq 0$:

$$\begin{aligned}
&\lim_{\Delta\downarrow 0} \Delta^{-2}\mathrm{P}\big(N_i(t+\Delta) = N_i(t) + 1,\\
&\qquad N_j(t+\Delta) = N_j(t) + 1 \mid \mathscr{F}_t\big)\\
&= \lim_{\Delta\downarrow 0} \Big\{\Delta^{-2}\mathrm{P}\big(N_i(t+\Delta) = N_i(t) + 1 \mid \mathscr{F}_t\big)\\
&\qquad \times \mathrm{P}\big(N_j(t+\Delta) = N_j(t) + 1 \mid \mathscr{F}_t\big)\Big\}\\
&= \lambda_i(t \mid \mathscr{F}_t)\, \lambda_j(t \mid \mathscr{F}_t), \quad \text{a.s.}.
\end{aligned} \tag{11}$$

Lemma 1 shows that for a multivariate counting process $\boldsymbol{N}(t)$ that satisfies the OM condition, the total CIF $\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t)$ in (9) and (10) equals the sum of all CIFs $\lambda_i(t \mid \mathscr{F}_t)$ over individual nodes $i \in \mathcal{V}$. In the remainder of the paper, we consistently assume the OM condition for the multivariate counting process $\boldsymbol{N}(t)$.

Lemma 1 (Total CIF *of the Multivariate Counting Process $\boldsymbol{N}(t)$):* Assume conditions A1 and A2 in Appendix B. Assume that $\mathrm{P}(\lambda_i(t \mid \mathscr{F}_t) < \infty) = 1$ for all $i \in \mathcal{V}$ and $t \geq 0$. If $\boldsymbol{N}(t)$ satisfies the OM condition, then for any $t \geq 0$, the total CIF $\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t)$ defined in (9) and (10) satisfies

$$\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t) = \sum_{i=1}^{V} \lambda_i(t \mid \mathscr{F}_t), \quad \text{a.s.}.$$

III. Statistical Model for $\lambda_i(t \mid \mathscr{F}_t)$ With Network Structure

We propose a continuous-time GLM-type model for $\lambda_i(t \mid \mathscr{F}_t)$:

$$\lambda_i(t \mid \mathscr{F}_t) = \exp\Big\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, x_j(t)\Big\}, \quad i \in \mathcal{V},\ t \geq 0. \tag{12}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Let $\lambda_i(0) = \lambda_i(0 \mid \mathscr{F}_0) = \exp(\beta_{0;i})$ denote the CIF of node i at time $t = 0$. The parameters $\beta_{0;i}$ and $\beta_{j,i}$, along with the history-dependent covariates $x_j(t)$, have the following interpretations:

Baseline intensity parameter $\beta_{0;i}$. Since the background intensity may vary over nodes, we include a bias term $\beta_{0;i}$ in (12) to associate the baseline intensity parameter with each node i.

Connection strength parameter $\beta_{j,i}$. The parameter $\beta_{j,i}$ in (12) quantifies the magnitude and direction of influence from parent node j on child node i, represented as $(j) \xrightarrow{\beta_{j,i}} (i)$. Specifically:

$$\begin{array}{lll} \beta_{j,i} > 0: & \text{Excitatory} & \text{effect from node } j \text{ to node } i; \\ \beta_{j,i} = 0: & \text{No} & \text{effect from node } j \text{ to node } i; \\ \beta_{j,i} < 0: & \text{Inhibitory} & \text{effect from node } j \text{ to node } i. \end{array}$$

For interpretability, we assume $\beta_{i,i} = 0$ for all $i \in \mathcal{V}$, meaning there is no self-effect. The network graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ can be obtained from all pairs of nodes (j, i) with non-zero connection parameters $\beta_{j,i}$ in the edge set:

$$\mathcal{E} = \{(j,i) \in \mathcal{V} \times \mathcal{V} : \beta_{j,i} \neq 0;\, j \neq i\} = \mathcal{E}_+ \cup \mathcal{E}_-. \quad (13)$$

This distinguishes the edge set for excitatory effects:

$$\mathcal{E}_+ = \{(j,i) \in \mathcal{V} \times \mathcal{V} : \beta_{j,i} > 0;\, j \neq i\}, \quad (14)$$

from the edge set for inhibitory effects:

$$\mathcal{E}_- = \{(j,i) \in \mathcal{V} \times \mathcal{V} : \beta_{j,i} < 0;\, j \neq i\}. \quad (15)$$

The configuration of this graph $\mathcal{G}$ reveals the interaction effects between nodes, and learning such a graph structure through statistical estimation methods is the main goal of this paper.

Regression covariates $x_j(t)$. Regression covariates $x_j(t)$ aim to represent the effect from other nodes $j \in \mathcal{V}$ on node i within a short period of time until t. We formulate $x_j(t)$ as follows:

$$x_j(t) = g\big(r_{j,\phi}(t)\big), \quad (16)$$

where $r_{j,\phi}(t)$ is the empirical rate during a short time interval of width $\phi \in (0, \infty)$:

$$r_{j,\phi}(t) = N_j\big((t-\phi, t]\big)/\phi. \quad (17)$$

Here, $g(\cdot) : [0,\infty) \to [0,\infty)$ is a non-linear shape-function that is continuous, non-negative, and monotonically increasing, with $g(0) = 0$. It is worth noting that within the modelling framework (12) for $\lambda_i(t \mid \mathscr{F}_t)$, the function g is not restricted to be bounded. Condition A5 assumes a bounded $g(\cdot)$ to facilitate the analytical derivation of theoretical results, such as probabilistic properties of $\boldsymbol{N}(t)$ and asymptotic properties of parameter estimators. However, this assumption may be relaxed in certain cases. For practical choices of the shape-function g and the time-lag constant ϕ, refer to Appendix A-A. Additionally, for empirical performances in data analysis, parameter estimation, and structure learning, see Sections VI–VII.

A. Connection of Model (12) With Other Models

The proposed model (12) employs the GLM-type framework to link the CIF $\lambda_i(t \mid \mathscr{F}_t)$ with both historical data and the network structure. This provides a novel continuous-time approach for modeling multivariate point process data. Note that the exponential link function in our model (12) is convex and twice-differentiable, aiding theoretical analysis and computational efficiency. This distinguishes it from other non-linear link functions such as ReLU [31], [32] or sigmoid functions [20], used in Hawkes processes for capturing both excitatory and inhibitory effects. As shown below, by selecting two specific choices of the shape-function g in (16) (combined with (17)), model (12) establishes connections with two existing models.

Example 1: $g(x) = x$. Then model (12) becomes:

$$\begin{aligned} \lambda_i(t \mid \mathscr{F}_t) &= \exp\Big\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, r_{j,\phi}(t)\Big\} \\ &= \exp\Big\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \int_{-\infty}^{t} \frac{1}{\phi}\, \beta_{j,i}\, \mathrm{I}(0 \le t-u < \phi)\, \mathrm{d}N_j(u)\Big\}, \end{aligned}$$

which is a special case of the general multivariate non-linear Hawkes process [33]:

$$\lambda_i(t \mid \mathscr{F}_t) = \varphi\Big(\beta_{0;i} + \sum_{j\in\mathcal{V}} \int_{-\infty}^{t} \omega_{j,i}(t-u)\, \mathrm{d}N_j(u)\Big),$$

when we set the non-linear link function $\varphi(\cdot) = \exp(\cdot)$, the interaction function $\omega_{j,i}(u) = \beta_{j,i}\, \mathrm{I}(0 \le u < \phi)/\phi$, and assume $\beta_{i,i} = 0$.

Example 2: $g(x) = \log(1 + x)$. Then model (12) becomes:

$$\begin{aligned} \lambda_i(t \mid \mathscr{F}_t) &= \exp\Big(\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, \log\{1 + r_{j,\phi}(t)\}\Big) \\ &= \exp(\beta_{0;i}) \prod_{j\in\mathcal{V}} \{1 + r_{j,\phi}(t)\}^{\beta_{j,i}}, \end{aligned}$$

which agrees with [15]. In comparison to Example 1, the shape-function $g(x) = \log(1 + x)$ in Example 2 is relatively flat. This moderates the steepness of the exponential link function and down-weights the influence of excessively large intensities. Therefore, Example 2 is expected to better represent the dynamics of multivariate point process data in real applications.

B. Distinction From Markov Processes

A general stochastic process is Markovian if, conditional on the past and present states, the probability of transitioning to a future state depends solely on the present state, but not on the past history ([34], p. 132). In our case, the counting process $\{\boldsymbol{N}(t)\}_{t\ge 0}$ associated with the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in model (12) (together with (16) and (17)), is *not* Markovian. This is because the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ depend not only on the current state of $\boldsymbol{N}(t)$ but also on the past states of $\boldsymbol{N}((t-\phi, t))$. This distinction highlights the clear difference between our model and other Markovian models of stochastic processes, such as the Markov multi-state model [35] commonly used in survival analysis, the versatile

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Markovian point process [36] used for modeling queuing systems, or the piecewise deterministic Markov process [28] used for modeling physical processes of particle motions.

IV. Properties of the Proposed Intensity Model

In this section, we investigate the probabilistic properties of the counting process $\boldsymbol{N}(t)$ associated with the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in model (12). These results are essentially required for deriving our statistical properties (Theorems 5–7 and Corollary 1 in Section V).

A distinctive feature of our CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in (12) is that they are piecewise-constant functions of time t (as to be shown in Section IV-A.1). In other words, unlike many other models, our $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ do not change continuously over time, yielding a countable number of discontinuity points in $(0,\infty)$ from all nodes. The discontinuity points of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ play an important role in characterizing the stochastic features of our CIFs. We begin by investigating the set of discontinuity points of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in Section IV-A.

A. Marked Point Process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ for Intensity Discontinuities

In this section, we conduct a step-by-step analysis based on the discontinuity points of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in (12). Section IV-A.1 demonstrates the piecewise-constant nature of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$. Section IV-A.2 defines the 'marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ for intensity discontinuities,' which proves to be equivalent (as shown in (24) and (25)) for studying the point process $\{T_{i,\ell}\}_{\ell\geq 1,\, i\in\mathcal{V}}$ and the counting process $\boldsymbol{N}(t)$. Section IV-A.3 derives the probability distribution of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ (in Theorem 1) and presents related properties (in Lemmas 5–7). By translating the results of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ into the analogues of $\boldsymbol{N}(t)$, Section IV-A.4 demonstrates the bounded variance and finiteness properties (in Theorem 2) for our counting process $\boldsymbol{N}(t)$.

1) Piecewise-Constant $\lambda_i(t \mid \mathscr{F}_t)$: Recall that, using (12), (16), and (17), the CIF at each node $i \in \mathcal{V}$ can be rewritten as $\lambda_i(t \mid \mathscr{F}_t) = \exp\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, g(N_j((t-\phi,t])/\phi)\}$, which forms a continuous function of $\{N_j((t-\phi,t])\}_{j\in\mathcal{V}}$, where

$$N_j\big((t-\phi,t]\big) \;=\; N_j(t) - N_j(t-\phi), \quad t \geq 0. \tag{18}$$

Consequently, the smoothness of $\lambda_i(t \mid \mathscr{F}_t)$ is directly reliant on the smoothness of $\{N_j((t-\phi,t])\}_{j\in\mathcal{V}}$.

For each node $j \in \mathcal{V}$ and event time points $\{T_{j,\ell}\}_{\ell\geq 1}$ in (2), we observe $N_j(\{t\}) = \sum_{\ell\geq 1} \mathrm{I}(T_{j,\ell} = t)$, which represents the jump size of $N_j(\cdot)$ at a single point t. It is clear that $N_j(\{t\}) \in \{0, 1\}$, and $N_j(\{t\}) = 1$ is equivalent to $t \in \{T_{j,\ell}\}_{\ell\geq 1}$. Moreover, two properties of $N_j((t-\phi,t])$ can be verified. First, $N_j((t-\phi,t])$ is non-negative, right-continuous, piecewise-constant, but not monotonically increasing in $t \in [0,\infty)$. Accordingly, $\lambda_i(t \mid \mathscr{F}_t)$ is also right-continuous and piecewise-constant. Second, the set of discontinuity points of $N_j((t-\phi,t])$ is given by

$$\begin{aligned}&\big\{t \geq 0 : N_j(\{t\}) - N_j(\{t-\phi\}) = +1\big\}\\ \cup&\big\{t \geq 0 : N_j(\{t\}) - N_j(\{t-\phi\}) = -1\big\},\end{aligned} \tag{19}$$

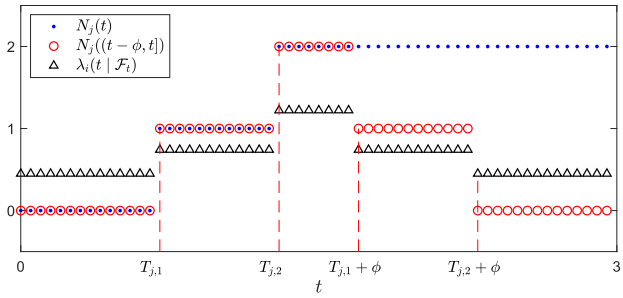

Fig. 2. Illustrative plot showing sample paths of stochastic processes $N_j(t)$ in (3), $N_j((t-\phi,t])$ in (18), and $\lambda_i(t \mid \mathscr{F}_t) = \exp\{-0.8 + 0.5 \cdot N_j((t-\phi,t])\}$ in (12), with $\mathcal{V} = \{1, 2\}$, $i = 1$, $j = 2$, and time-lag $\phi = 1$. Notice the overlap between $N_j(t)$ and $N_j((t-\phi,t])$ within the time interval $[0, 1.7)$. $\lambda_i(t \mid \mathscr{F}_t)$ is a piecewise-constant function with discontinuities identical to those of $N_j((t-\phi,t])$.

where

$$N_j(\{t\}) - N_j(\{t-\phi\}) = \begin{cases} 0, & \text{if } t \notin \{T_{j,\ell}\}_{\ell\geq 1}, \text{and } t \notin \{T_{j,k}+\phi\}_{k\geq 1},\\ +1, & \text{if } t \in \{T_{j,\ell}\}_{\ell\geq 1}, \text{and } t \notin \{T_{j,k}+\phi\}_{k\geq 1},\\ -1, & \text{if } t \notin \{T_{j,\ell}\}_{\ell\geq 1}, \text{and } t \in \{T_{j,k}+\phi\}_{k\geq 1},\\ 0, & \text{if } t \in \{T_{j,\ell}\}_{\ell\geq 1}, \text{and } t \in \{T_{j,k}+\phi\}_{k\geq 1}. \end{cases} \tag{20}$$

Following (20), for each node $j \in \mathcal{V}$, we can rewrite the set of discontinuity points in (19) as:

$$\begin{aligned}&\big\{t \geq 0 : t \in \{T_{j,\ell}\}_{\ell\geq 1};\, t \notin \{T_{j,k}+\phi\}_{k\geq 1}\big\}\\ \cup&\big\{t \geq 0 : t \notin \{T_{j,\ell}\}_{\ell\geq 1};\, t \in \{T_{j,k}+\phi\}_{k\geq 1}\big\},\end{aligned}$$

which belongs to the set

$$\{T_{j,\ell}\}_{\ell\geq 1} \cup \{T_{j,k}+\phi\}_{k\geq 1}.$$

Utilizing [37] (Theorem 2.4.7, p. 84) and the CIF $\lambda_j(t \mid \mathscr{F}_t) < \infty$ in (12), the event time points $\{T_{j,\ell}\}_{\ell\geq 1}$ are *totally inaccessible stopping times*, implying that $\mathrm{P}(\cup_{\ell\geq 1}\cup_{k\geq 1} \{T_{j,\ell} = T_{j,k}+\phi\}) = 0$. Thus, the right-continuous $\lambda_i(t \mid \mathscr{F}_t)$ is piecewise-constant in $t \in [0,\infty)$, with the set of discontinuity points specified in Lemma 2.

Lemma 2 (Piecewise-Constant $\lambda_i(t \mid \mathscr{F}_t)$ and Its Discontinuity Points): Assume conditions A1 and A2 in Appendix B. For each $i \in \mathcal{V}$, let $\mathrm{Pa}(i) = \{j \in \mathcal{V}\backslash\{i\} : \beta_{j,i} \neq 0\}$ denote the set of parent nodes for node i. If $\mathrm{Pa}(i) \neq \varnothing$, then $\lambda_i(t \mid \mathscr{F}_t)$ is a piecewise-constant function of $t \in [0,\infty)$, with all its discontinuity points listed in the set:

$$\cup_{j\in\mathrm{Pa}(i)} \big\{\{T_{j,\ell}\}_{\ell\geq 1} \cup \{T_{j,k}+\phi\}_{k\geq 1}\big\}.$$

If $\mathrm{Pa}(i) = \varnothing$, then $\lambda_i(t \mid \mathscr{F}_t) \equiv \exp(\beta_{0;i})$ is a constant, and $\{N_i(t)\}_{t\geq 0}$ reduces to a homogeneous Poisson process.

An illustration of $N_j(t)$, $N_j((t-\phi,t])$, and $\lambda_i(t \mid \mathscr{F}_t)$ is given in Figure 2. By aggregating the discontinuity points of all CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$, we directly obtain the following Lemma 3.

Lemma 3 (Discontinuity Points of All CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$): Assuming conditions A1 and A2 in Appendix B, the following results hold:

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

(i) The discontinuity points of all CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ are listed in the set:

$$\mathcal{T}^* = \cup_{i\in\mathcal{V}} \cup_{j\in \mathrm{Pa}(i)} \big\{\{T_{j,\ell}\}_{\ell\geq 1}\{T_{j,k}+\phi\}_{k\geq 1}\big\}. \quad (21)$$

(ii) Define the sequence of time points:

$$\{\breve{T}_1, \breve{T}_2, \ldots\} = \cup_{j\in\mathcal{V}}\big\{\{T_{j,\ell}\}_{\ell\geq 1}\{T_{j,k}+\phi\}_{k\geq 1}\big\}, \quad (22)$$

with $0 < \breve{T}_1 < \breve{T}_2 < \cdots$ arranged in increasing order. Then $\mathcal{T}^*$ is a subset of $\{\breve{T}_\ell\}_{\ell\geq 1}$.

Moreover, if $\mathrm{Ch}(i) \neq \varnothing$ for all $i \in \mathcal{V}$, where $\mathrm{Ch}(i) = \{k \in \mathcal{V}\setminus\{i\} : \beta_{i,k} \neq 0\}$ denotes the set of child nodes for node i, then we have $\mathcal{T}^* = \{\breve{T}_\ell\}_{\ell\geq 1}$.

Remark 2: Lemmas 2 and 3 demonstrate the close relationship between the fundamental characteristics of CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in model (12) and the properties of the network structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ in (13). For instance, if a node i has a parent node in the network $\mathcal{G}$, then $\lambda_i(t \mid \mathscr{F}_t)$ is non-constant and the point process on node i is not reduced to the trivial case of a homogeneous Poisson process. Additionally, if each node in the network $\mathcal{G}$ has at least one child node, then the set of all intensity discontinuities $\mathcal{T}^*$ in (21) is identical to the set of time points $\{\breve{T}_\ell\}_{\ell\geq 1}$ in (22). Since $\{\breve{T}_\ell\}_{\ell\geq 1}$ contains all the discontinuity points in $\mathcal{T}^*$ and has a simpler form than $\mathcal{T}^*$, we will focus our remaining analysis on $\{\breve{T}_\ell\}_{\ell\geq 1}$ and refer to it as 'the set of intensity discontinuities' with a slight abuse of terminology.

2) Marked Point Process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ for Studying Discontinuity Points of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$: To investigate $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$, we next introduce the concept of a 'marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I}) = (\{\breve{T}_\ell\}_{\ell\geq 1}, \{I_\ell\}_{\ell\geq 1})$ for intensity discontinuities' in Definition 2 below. A general marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ is a double sequence, where $\{\breve{T}_\ell\}_{\ell\geq 1}$ is a point process, and each $\breve{T}_\ell$ is associated with a *mark* I_ℓ, usually representing some additional features (such as labels or locations) related to the time point $\breve{T}_\ell$; refer to [29] and the references therein for further details.

Definition 2 (Marked Point Process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ for Discontinuity Points of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$): Assume conditions A1 and A2 in Appendix B. For the strictly increasing time points $\{\breve{T}_1, \breve{T}_2, \ldots\}$ defined in (22) and integers $\ell \geq 1$, let $I_\ell \in \mathcal{V}\cup\{0\}$ be the mark corresponding to $\breve{T}_\ell$, defined by:

$$I_\ell = \begin{cases} i \in \mathcal{V}, & \text{if } \breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1}, \\ 0, & \text{if } \breve{T}_\ell \in \cup_{j\in\mathcal{V}}\big\{\{T_{j,k}+\phi\}_{k\geq 1}\big\}. \end{cases} \quad (23)$$

We refer to the double sequence $(\breve{\boldsymbol{T}}, \boldsymbol{I}) = (\{\breve{T}_\ell\}_{\ell\geq 1}, \{I_\ell\}_{\ell\geq 1})$ as the 'marked point process for intensity discontinuities'.

The mark I_ℓ in (23) indicates the identity of $\breve{T}_\ell$: if the discontinuity point $\breve{T}_\ell$ of $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ is due to an event occurrence from some node i at that time point, then I_ℓ represents the index i of that node; otherwise, we set $I_\ell = 0$. Lemma 4 below guarantees the uniqueness of the mark I_ℓ defined in (23) for each $\breve{T}_\ell$.

Lemma 4 (Uniqueness of the Mark I_ℓ Corresponding to $\breve{T}_\ell$): Assume conditions A1 and A2 in Appendix B. Then, the mark I_ℓ in (23), corresponding to the discontinuity point $\breve{T}_\ell$, is uniquely defined a.s., i.e.,

(i) For any distinct $i, j \in \mathcal{V}$, $\mathrm{P}(I_\ell = i, I_\ell = j) = \mathrm{P}(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1}, \breve{T}_\ell \in \{T_{j,k}\}_{k\geq 1}) = 0$.

(ii) For any $i \in \mathcal{V}$, $\mathrm{P}(I_\ell = i, I_\ell = 0) = \mathrm{P}(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1}, \breve{T}_\ell \in \cup_{j\in\mathcal{V}}\big\{\{T_{j,k}+\phi\}_{k\geq 1}\big\}) = 0$.

As stated by Definition 2 and Lemma 4, a multivariate point process $\{T_{i,\ell}\}_{\ell\geq 1, i\in\mathcal{V}}$ uniquely defines a marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$. Conversely, $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ uniquely yields a multivariate point process $\{T_{i,\ell}\}_{\ell\geq 1, i\in\mathcal{V}}$, and accordingly, a multivariate counting process $\{\boldsymbol{N}(t)\}_{t\geq 0}$, due to the identities:

$$T_{i,\ell} = \inf\{\breve{T}_k > T_{i,\ell-1} : I_k = i,\ k \geq 1\},\ \ell \geq 1,\ i \in \mathcal{V}, \quad (24)$$

$$N_i(t) = \sum_{k\geq 1} \mathrm{I}(\breve{T}_k \leq t,\ I_k = i), \quad t \geq 0,\ i \in \mathcal{V}, \quad (25)$$

Hence, the point process $\{T_{i,\ell}\}_{\ell\geq 1, i\in\mathcal{V}}$, the counting process $\{\boldsymbol{N}(t)\}_{t\geq 0}$, and the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ can be deduced from each other. As shown in Theorem 1 below, the probability distribution of the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ has a closed-form expression, making $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ more convenient to analyze than $\{T_{i,\ell}\}_{\ell\geq 1, i\in\mathcal{V}}$ and $\{\boldsymbol{N}(t)\}_{t\geq 0}$.

3) Probabilistic Properties of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$: A non-negative random variable τ is called a *stopping time* with respect to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$ in (5), if $\{\tau \leq t\} \in \mathscr{F}_t$ holds for any $t \geq 0$. For each integer $\ell \geq 1$, the time point $\breve{T}_\ell$ in (22) is a stopping time, and let $\mathcal{F}_{\breve{T}_\ell} = \{A \in \mathscr{F} : A \cap \{\breve{T}_\ell \leq t\} \in \mathscr{F}_t \text{ for every } t > 0\}$ be the *stopping time σ-algebra* (defined as in [38]) with respect to $\breve{T}_\ell$, i.e., generated by the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ up to time $\breve{T}_\ell$. Denote by $\lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) = \lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}\big(N_i(\breve{T}_\ell+\Delta) = N_i(\breve{T}_\ell)+1 \mid \mathcal{F}_{\breve{T}_\ell}\big)$ the CIF of node i at stopping time $\breve{T}_\ell$. For $\ell = 0$, define $\mathcal{F}_{\breve{T}_0} = \mathscr{F}_0 = \{\Omega, \varnothing\}$, with $\breve{T}_0 = 0$ and $I_0 = 0$. Theorem 1 presents the probability distribution of the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ conditional on the filtration $\{\mathcal{F}_{\breve{T}_\ell}\}_{\ell\geq 0}$. For a random variable X, denote $\sigma(X)$ as the σ-field generated by X; for a σ-field $\mathscr{F}$, denote $\sigma(\mathscr{F}, X)$ as the smallest σ-field that contains all the events belonging to $\mathscr{F}\cup\sigma(X)$.

Theorem 1 (Conditional Distributions of $\breve{T}_{\ell+1}$ and $I_{\ell+1}$ Given $\mathcal{F}_{\breve{T}_\ell}$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. For each integer $\ell \geq 0$, define the set:

$$\mathcal{T}_\ell = \bigcup_{i\in\mathcal{V}}\big\{t \in (\breve{T}_\ell, \breve{T}_\ell+\phi] : N_i(\{t-\phi\}) = 1\big\} \quad (26)$$

and the $\mathcal{F}_{\breve{T}_\ell}$-measurable random variable:

$$T_\ell^* = \begin{cases} \min(\mathcal{T}_\ell), & \text{if } \mathcal{T}_\ell \neq \varnothing, \\ \infty, & \text{if } \mathcal{T}_\ell = \varnothing. \end{cases} \quad (27)$$

The following results hold:

(i) (Support of $\breve{T}_{\ell+1}$) $\mathrm{P}(\breve{T}_\ell < \breve{T}_{\ell+1} \leq T_\ell^*) = 1$.

(ii) (Conditional distribution of $\breve{T}_{\ell+1}$) If $T_\ell^* < \infty$, then $\breve{T}_{\ell+1}$, conditional on $\mathcal{F}_{\breve{T}_\ell}$, has a mixed-type probability distribution with a probability mass function (p.m.f.)

$$\mathrm{P}\big(\breve{T}_{\ell+1} = T_\ell^* \mid \mathcal{F}_{\breve{T}_\ell}\big) = \exp\{-\lambda^{\mathrm{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})\cdot(T_\ell^* - \breve{T}_\ell)\} \quad (28)$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

at the point T_ℓ^*, and a probability density function (p.d.f.)

$$f_{\breve{T}_{\ell+1}|\mathcal{F}_{\breve{T}_\ell}}(t \mid \breve{T}_\ell) = \lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) \times \exp\{-\lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) \cdot (t - \breve{T}_\ell)\} \tag{29}$$

for $t \in (\breve{T}_\ell, T_\ell^*)$, where $\lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) = \sum_{i=1}^V \lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})$. If $T_\ell^* = \infty$, then (28) and (29) reduce to $(\breve{T}_{\ell+1} - \breve{T}_\ell) \mid \mathcal{F}_{\breve{T}_\ell} \sim \text{Exp}(\lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}))$.

(iii) (Conditional distribution of $I_{\ell+1}$) If $T_\ell^* < \infty$, then for $i \in \mathcal{V}$,

$$\mathrm{P}\big(I_{\ell+1} = i \mid \sigma(\mathcal{F}_{\breve{T}_\ell}, \breve{T}_{\ell+1})\big) = \begin{cases} 0, & \text{if } \breve{T}_{\ell+1} = T_\ell^*, \\ \lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) / \lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}), & \text{if } \breve{T}_{\ell+1} \in (\breve{T}_\ell, T_\ell^*). \end{cases} \tag{30}$$

If $T_\ell^* = \infty$, then (30) reduces to $\mathrm{P}(I_{\ell+1} = i \mid \sigma(\mathcal{F}_{\breve{T}_\ell}, \breve{T}_{\ell+1})) = \lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) / \lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})$, for $i \in \mathcal{V}$ and $\breve{T}_{\ell+1} \in (\breve{T}_\ell, \infty)$.

The derivation of Theorem 1 primarily relies on the fact that the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ are constant within each interval $[\breve{T}_\ell, \breve{T}_{\ell+1})$. For instance, if $T_\ell^* < \infty$, (29) indicates that, conditional on $\mathcal{F}_{\breve{T}_\ell}$, the duration $\breve{T}_{\ell+1} - \breve{T}_\ell$ follows an exponential distribution with a rate $\lambda^{\text{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})$ before $\breve{T}_{\ell+1}$ reaches T_ℓ^*. Furthermore, $\breve{T}_{\ell+1} = T_\ell^*$ implies that $\breve{T}_{\ell+1} \in \{T_{i,k} + \phi\}_{i\in\mathcal{V},k\geq 1}$, while $\breve{T}_{\ell+1} < T_\ell^*$ indicates that the probability of the event $\{\breve{T}_{\ell+1} \in \{T_{i,k}\}_{k\geq 1}\}$, conditional on $\sigma(\mathcal{F}_{\breve{T}_\ell}, \breve{T}_{\ell+1})$, is proportional to the corresponding CIF $\lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})$ at node i. It is important to note that the V-dimensional CIF $\boldsymbol{\lambda}(t \mid \mathscr{F}_t) = (\lambda_1(t \mid \mathscr{F}_t), \ldots, \lambda_V(t \mid \mathscr{F}_t))^\top$ in model (12) is not a piecewise deterministic Markov process (PDMP) [28], and thus the general results for PDMP do not apply to the derivation of Theorem 1.

Theorem 1 has two important applications. Firstly, it provides a simulation algorithm to generate synthetic point process data $\{T_{i,\ell}\}_{i\in\mathcal{V},\ell\geq 1}$ with CIFs modeled by (12). By utilizing the conditional probability distributions of $(\breve{T}_{\ell+1}, I_{\ell+1})$ given in (28)–(30), one can sequentially generate the marked point $(\breve{T}_{\ell+1}, I_{\ell+1})$ for each $\ell \geq 0$, and then convert them into $\{T_{i,\ell}\}_{i\in\mathcal{V},\ell\geq 1}$ using (24). Secondly, Theorem 1 leads to probabilistic results of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$, as presented in Lemmas 5, 6, and 7, which are used to prove Theorem 2.

For clarity, the following notations are used: The *duration* τ_ℓ between two consecutive discontinuity time points $\breve{T}_\ell$ is given by

$$\tau_\ell = \breve{T}_\ell - \breve{T}_{\ell-1} \quad \text{for } \ell \geq 1. \tag{31}$$

The *event counts* $M_{i,\ell}$ at node $i \in \mathcal{V}$ are calculated as

$$M_{i,0} = 0, \quad M_{i,\ell} = \sum_{k=1}^{\ell} \mathrm{I}(I_k = i) \quad \text{for } \ell \geq 1. \tag{32}$$

The *piecewise-constant intensity* at node $i \in \mathcal{V}$ within the time interval $[\breve{T}_\ell, \breve{T}_{\ell+1})$ is denoted by

$$\lambda_{i,0} = \lambda_i(0), \quad \lambda_{i,\ell} = \lambda_i(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}) \quad \text{for } \ell \geq 1. \tag{33}$$

Lemma 5 (Expectation and Variance Related to $(\breve{\boldsymbol{T}}, \boldsymbol{I})$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Then for each integer $k \geq 1$, we have:

$$\mathrm{E}\big\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k \mid \mathcal{F}_{\breve{T}_{k-1}}\big\} = 0,$$
$$\mathrm{var}\big\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k \mid \mathcal{F}_{\breve{T}_{k-1}}\big\} = \mathrm{E}\big\{\mathrm{I}(I_k = i) \mid \mathcal{F}_{\breve{T}_{k-1}}\big\}. \tag{34}$$

Furthermore, for each integer $\ell \geq 1$,

$$\mathrm{E}\Big(M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\Big) = 0,$$
$$\mathrm{var}\Big(M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\Big) = \mathrm{E}(M_{i,\ell}). \tag{35}$$

Lemma 6 (Martingale Property Related to $(\breve{\boldsymbol{T}}, \boldsymbol{I})$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Then the random process $\{M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\}_{\ell\geq 1}$ is a martingale with respect to $\{\mathcal{F}_{\breve{T}_\ell}\}_{\ell\geq 1}$.

Lemma 7 (Upper Bound for Variance Related to t-Truncated $(\breve{\boldsymbol{T}}, \boldsymbol{I})$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. For a given deterministic time point $t \in (0, \infty)$, let

$$L_t = \sum_{\ell=1}^{\infty} \mathrm{I}(\breve{T}_\ell \leq t) \tag{36}$$

count the number of discontinuity points $\{\breve{T}_\ell\}_{\ell\geq 1}$ that occur up to t. For integers $\ell \geq 1$, let

$$\tau_\ell^{[t]} = \breve{T}_\ell \wedge t - \breve{T}_{\ell-1} \wedge t \tag{37}$$

be the duration between t-truncated $\breve{T}_\ell$ and $\breve{T}_{\ell-1}$, where $a \wedge b = \min(a, b)$. Let $\{X_\ell\}_{\ell\geq 0}$ be a sequence of random variables such that $X_\ell \geq 0$ is measurable with respect to $\mathcal{F}_{\breve{T}_\ell}$ for each $\ell \geq 0$, and $\sup_{\ell\geq 0} X_\ell \leq c_1$ a.s. for a constant $c_1 \in (0, \infty)$. Then, for each $i \in \mathcal{V}$, we have

$$\mathrm{E}\Big\{\sum_{k=1}^{L_t} X_{k-1}\,\mathrm{I}(I_k = i) - \sum_{k=1}^{L_t+1} X_{k-1}\,\lambda_{i,k-1} \cdot \tau_k^{[t]}\Big\} = 0, \tag{38}$$

and

$$\mathrm{var}\Big\{\sum_{k=1}^{L_t} X_{k-1}\,\mathrm{I}(I_k = i) - \sum_{k=1}^{L_t+1} X_{k-1}\,\lambda_{i,k-1} \cdot \tau_k^{[t]}\Big\} = \mathrm{E}\Big\{\sum_{k=1}^{L_t} X_{k-1}^2\,\mathrm{I}(I_k = i)\Big\} \leq C_i\, c_1^2\, t, \tag{39}$$

where the constant $C_i = \exp(\beta_{0;i} + C_0 \cdot \sum_{j\in\mathcal{V}} \beta_{j,i})$, with $C_0 \in (0, \infty)$ provided in Condition A5.

Remark 3: Derivations of Lemmas 5–7 are outlined as follows. Lemma 5 is obtained from direct calculations based on the probability distribution of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ in Theorem 1. Lemma 6 follows from (34) in Lemma 5. Lemma 7 is a non-trivial extension of Lemma 5, where a non-random index ℓ is replaced with a random index L_t; Lemma 7 aims to study the properties of the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ when truncated by a fixed time point $t \in (0, \infty)$, which is further used to translate these results into the forms of the counting process $\boldsymbol{N}(t)$.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

4) Translating Results of $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ Into Results of $\boldsymbol{N}(t)$: The equivalence verified in (25), between the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ and the counting process $\boldsymbol{N}(t)$, enables us to translate the results of Lemmas 5–7 into the counterparts of $\boldsymbol{N}(t)$ and directly obtain Theorem 2 below, which describes some useful properties of $\boldsymbol{N}(t)$.

Theorem 2 (Upper Bounds for Variances Related to $\boldsymbol{N}(t)$; Finiteness of $\boldsymbol{N}(t)$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Then, there exists a constant $c_1 \in (0, \infty)$ such that for any $i \in \mathcal{V}$ and any $t \in (0, \infty)$, we have

$$\operatorname{var}\Big\{N_i(t) - \int_0^t \lambda_i(u \mid \mathscr{F}_u)\, \mathrm{d}u\Big\} = \mathrm{E}\{N_i(t)\} \le c_1\, t, \quad (40)$$

which implies that the counting process $N_i(t)$ is finite a.s., i.e.,

$$\mathrm{P}(N_i(t) < \infty) = 1, \quad i \in \mathcal{V}. \quad (41)$$

Furthermore, for a random process $\{x(t)\}_{t\ge 0}$ such that $x(t)$ is $\mathscr{F}_t$-measurable, $0 \le \inf_{t\ge 0} x(t) \le \sup_{t\ge 0} x(t) \le c_2$ a.s. for a constant $c_2 \in (0, \infty)$, and $x(t)$ is constant in the interval $[\breve{T}_\ell, \breve{T}_{\ell+1})$ for each integer $\ell \ge 0$, it follows that for any $t \in (0, \infty)$,

$$\begin{aligned}&\operatorname{var}\Big[\int_0^t \big\{x(u-)\, \mathrm{d}N_i(u) - x(u)\, \lambda_i(u \mid \mathscr{F}_u)\, \mathrm{d}u\big\}\Big] \\ &\quad = \mathrm{E}\Big\{\int_0^t x^2(u)\, \lambda_i(u \mid \mathscr{F}_u)\, \mathrm{d}u\Big\} \le c_1\, c_2^2\, t, \end{aligned} \quad (42)$$

where $x(u-) = \lim_{t\uparrow u} x(t)$ denotes the left limit.

By considering the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$, we obtain Theorem 2, which ensures certain fundamental probabilistic properties of our counting process $\boldsymbol{N}(t)$. In the subsequent discussions in Section V, we will demonstrate the significance of these results in deriving the associated statistical properties, as presented in Theorems 5–7 and Corollary 1.

B. Cyclicity and Asymptotic Mean Stationarity of $\boldsymbol{N}(t)$

A counting process $\{N(t)\}_{t\ge 0}$ is considered *strict-sense stationary* if, for any time point $s \in [0, \infty)$, $N(t+s) - N(s) \overset{\mathrm{D}}{=} N(t)$ for every $t \ge 0$, where $X_1 \overset{\mathrm{D}}{=} X_2$ denotes that random quantities X_1 and X_2 have identical distributions (see [29] and references therein). In this paper, 'stationarity' exclusively refers to *strict-sense stationarity*, while 'non-stationarity' encompasses other cases. A strict-sense stationary counting process $\{N(t)\}_{t\ge 0}$ exhibits several well-known properties, including:

(P1) Invariant distribution of the CIF: The probability distribution of the CIF $\lambda(t \mid \mathscr{F}_t)$, as defined in (7) and (8), remains invariant for any $t \in [0, \infty)$.

(P2) Constant mean intensity: For any $t \in [0, \infty)$, the mean CIF satisfies $\mathrm{E}\{\lambda(t \mid \mathscr{F}_t)\} \equiv \lambda_0$ for some constant $\lambda_0 \in (0, \infty)$.

(P3) Expectation of increments: For any $t \in (0, \infty)$ and $s \in (0, \infty)$, $\mathrm{E}\{N(t+s) - N(s)\} = \lambda_0 \cdot t$. Furthermore, if $N(t)$ is ergodic, then $\lim_{t\to\infty} N(t)/t = \lambda_0$ a.s..

(P4) Finiteness of $N(t)$: For any $t \in (0, \infty)$, $\mathrm{P}(N(t) < \infty) = 1$.

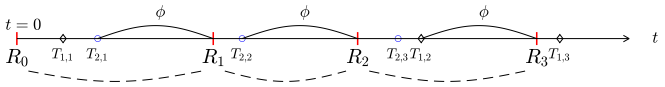

Fig. 3. Illustrative plot depicting recurrence time points $R_0, R_1, \ldots$, and event time points $\{T_{i,\ell}\}_{\ell\ge 1}$ of nodes $i \in \mathcal{V}$, where $\mathcal{V} = \{1, 2\}$. The cyclicity property in Theorem 3 denotes that after reaching each recurrence time point R_ℓ, $\boldsymbol{N}(t)$ enters a recurrence cycle $(R_\ell, R_{\ell+1}]$. Within this cycle, $\lambda_i(R_\ell \mid \mathcal{F}_{R_\ell}) = \lambda_i(R_0 \mid \mathcal{F}_{R_0}) = \lambda_i(0)$, initiating a renewed process $\{\boldsymbol{N}(t + R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\ge 0}$, independent of $\mathcal{F}_{R_\ell}$.

These properties resulting from the stationarity assumption significantly ease theoretical analysis. Hence, the stationarity assumption is commonly imposed in the relevant literature, such as [19], [39], and [40]. We refer to a multivariate counting process $\{\boldsymbol{N}(t)\}_{t\ge 0}$ as *strict-sense stationary* if $\{N_i(t)\}_{t\ge 0}$ is strict-sense stationary for each $i \in \mathcal{V}$.

However, the counting process $\boldsymbol{N}(t)$ associated with the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in model (12) is not strict-sense stationary. Lemma B.9 in Appendix B justifies that $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in (12) violates property (P1) of stationarity. Without possessing properties (P1)–(P4) listed above, a non-stationary point process poses significant challenges to theoretical analysis. Therefore, it becomes necessary to explore alternative properties for non-stationary point processes using a new approach.

Recall that a Poisson process assumes the independent increment property, resulting in the memoryless property [41]. In essence, for any constant $s \in (0, \infty)$, the time-shifted counting process $\{N(t+s) - N(s)\}_{t\ge 0}$ is independent of the history up to the time point s. In Theorem 3, we will demonstrate that our study establishes a relaxed version of this memoryless property for our $\boldsymbol{N}(t)$. Specifically, the counting process $\{\boldsymbol{N}(t + R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\ge 0}$ at certain *random* time points R_ℓ is independent of the σ-field $\mathcal{F}_{R_\ell}$, where R_ℓ and $\mathcal{F}_{R_\ell}$ are introduced in Definition 3.

Definition 3 (Recurrence Time Points R_ℓ, Recurrence Cycle of $\boldsymbol{N}(t)$, and $\mathcal{F}_{R_\ell}$): Let $R_0 = 0$. For each integer $\ell \ge 1$, R_ℓ is defined as the first time point, after $R_{\ell-1} + \phi$, such that no events occur at any node in the time interval $(R_\ell - \phi, R_\ell]$, i.e.,

$$R_\ell = \min\{t \ge R_{\ell-1} + \phi : \boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}\}. \quad (43)$$

We call R_ℓ the ℓth recurrence time point, and the interval $(R_{\ell-1}, R_\ell]$ the ℓth recurrence cycle. Define $\mathcal{F}_{R_0} = \mathscr{F}_0 = \{\Omega, \varnothing\}$. For $t \ge 0$ and integer $\ell \ge 0$, $t + R_\ell$ is a stopping time. Denote $\mathcal{F}_{t+R_\ell} = \{A \in \mathscr{F} : A \cap \{t + R_\ell \le u\} \in \mathscr{F}_u \text{ for every } u > 0\}$ as the stopping time σ-algebra with respect to $t + R_\ell$.

Figure 3 illustrates the recurrence time points R_ℓ. For our $\boldsymbol{N}(t)$, the existence of R_ℓ is verified by Lemma 8.

Lemma 8 (Existence of R_ℓ): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. For each integer $\ell \ge 1$, the recurrence time point R_ℓ in Definition 3 exists with probability one.

The memoryless property induced by R_ℓ can be intuitively explained as follows. In our model (12), the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ at the current time t primarily depend on the historical event counts $\boldsymbol{N}((t-\phi, t])$ in the lag window $(t-\phi, t]$ of a fixed length ϕ. Once the counting process $\boldsymbol{N}(t)$ reaches a recurrence time point $t = R_\ell$, all event counts in

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

the lag window $(t-\phi, t]$ become empty, which separates the dependence of the future CIFs $\{\lambda_i(t \mid \mathscr{F}_t) : t \geq R_\ell\}_{i\in\mathcal{V}}$ on the past event history up to that time point R_ℓ. At $t = R_\ell$, both the CIFs and the counting process 'reset,' becoming independent of $\mathcal{F}_{R_\ell}$, and initiating a renewed 'cyclic' process. Based on these considerations, we establish a new cyclicity property of $\boldsymbol{N}(t)$, formally presented in Theorem 3.

Theorem 3 (Cyclicity of $\boldsymbol{N}(t)$ Driven by R_ℓ): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let $\boldsymbol{N}(t)$ be the counting process with the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in (12). Then, for each recurrence time point R_ℓ in (43) with $\ell \geq 1$,

(i) both $\{\lambda_i(t+R_\ell \mid \mathcal{F}_{t+R_\ell})\}_{i\in\mathcal{V}}$ and $\boldsymbol{N}(t+R_\ell)-\boldsymbol{N}(R_\ell)$ are independent of $\mathcal{F}_{R_\ell}$, with $\boldsymbol{N}(t+R_\ell)-\boldsymbol{N}(R_\ell) \overset{\mathrm{D}}{=} \boldsymbol{N}(t)$ for each $t \geq 0$.

(ii) $\{\boldsymbol{N}((R_{\ell-1}, R_\ell]) : \ell \geq 1\}$ is a sequence of i.i.d. random vectors.

(iii) $\{R_\ell - R_{\ell-1} : \ell \geq 1\}$ is a sequence of i.i.d. random variables with finite second moment.

This cyclicity property of $\boldsymbol{N}(t)$ will be used to derive Theorem 4 below, as well as Theorem 5 in Section V-A. In comparison, our cyclicity property is analogous to the renewal property of the non-linear Hawkes process [25] or queuing models [26], [27]. However, tools for deriving the renewal property are not directly applicable to model (12), as it violates some basic assumptions underlying the non-linear Hawkes process and queuing models. For example, our point process (when $\mathcal{E} \neq \varnothing$) does not meet the assumption of a deterministic arrival rate required in M_t/G/∞ queues, and the non-linear Hawkes process does not allow for the general type of shape-function $g(\cdot)$ in model (12).

Theorem 4 (Asymptotic Mean Stationarity of $\boldsymbol{N}(t)$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Then, there exists a constant vector $\boldsymbol{c}_0 \in (0,\infty)^V$ such that the counting process $\boldsymbol{N}(t)$ associated with the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in (12) satisfies

$$\boldsymbol{N}(t)/t \xrightarrow{\mathrm{P}} \boldsymbol{c}_0, \quad \text{as } t \to \infty. \tag{44}$$

Theorem 4 verifies that the vector $\boldsymbol{N}(t)/t$ of average counts converges in probability to a constant vector as t approaches infinity. For a non-stationary counting process $\boldsymbol{N}(t)$, this type of property is called *asymptotic mean stationarity* (a notion used in [42]). It is noted that the constant $\boldsymbol{c}_0$ in (44) deterministically depends on the baseline parameters $\{\beta_{0;i}\}$ and network parameters $\{\beta_{j,i}\}$ in model (12), but a closed-form formulation of this dependence is not available due to the non-linearities of both the exponential link function and the shape-function $g(\cdot)$ in model (12). This is in contrast with the case of a linear Hawkes process [43], for which a closed-form moment equation (Equation (3) in [43]) could be constructed to relate the mean intensity with the network parameters. Nevertheless, without knowing the explicit value of $\boldsymbol{c}_0$, Theorem 4 suffices to assist in proving further useful statistical convergence properties, as will be shown in Section V.

In summary, Lemma B.9 states the fact that our counting process $\boldsymbol{N}(t)$ is not *strict-sense stationary*; nevertheless, we have verified that $\boldsymbol{N}(t)$ possesses some desirable properties similar to stationary processes. For example, Theorem 4 is similar to the ergodicity in property (P3); Theorem 2 verifies property (P4); and Theorem 3(i) indicates a feature similar to the shift invariance property of stationarity. Theorems 3 and 4 are crucial for deriving the related statistical asymptotic properties (in Theorems 5–7) in Section V. In comparison with existing results, technical tools we have developed are easier to interpret and utilize.

V. Parameter Estimation via Penalized M-Estimation

Our primary interest is to learn the network structure from the observed data $\{\boldsymbol{T}_i\}_{i\in\mathcal{V}}$ in (1) of the multivariate point process in the time interval $[0, \mathrm{T}]$, where $\mathrm{T} \in (0,\infty)$ is the total time length of the experiment. We denote the true values of the CIF (12) as

$$\lambda_i^*(t \mid \mathscr{F}_t) = \exp\left\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\right\}, \tag{45}$$

where $\widetilde{\boldsymbol{\beta}}_i^* = (\beta_{0;i}^*, \boldsymbol{\beta}_i^{*\top})^\top = (\beta_{0;i}^*, \beta_{1,i}^*, \ldots, \beta_{i-1,i}^*, \beta_{i+1,i}^*, \ldots, \beta_{V,i}^*)^\top \in \mathbb{R}^V$ is the vector of true parameters, and $\widetilde{\boldsymbol{x}}_i(t) = (1, \boldsymbol{x}_i(t)^\top)^\top = (1, x_1(t), \ldots, x_{i-1}(t), x_{i+1}(t), \ldots, x_V(t))^\top \in \mathbb{R}^V$ is the vector of regression covariates. Our statistical learning aims to estimate $\widetilde{\boldsymbol{\beta}}_i^*$ in (45) and recover the true network structure $\mathcal{G}^* = \{\mathcal{V}, \mathcal{E}^*\}$, where the true edge set $\mathcal{E}^* = \mathcal{E}_+^* \cup \mathcal{E}_-^*$ corresponds to $\mathcal{E} = \mathcal{E}_+ \cup \mathcal{E}_-$ in (13) with parameters $\beta_{j,i}$ replaced by $\beta_{j,i}^*$.

The existing parameter estimation methods can be categorized into two categories: (i) moment or correlation-based approaches [43], [44]; and (ii) intensity-based approaches [18], [19]. The moment or correlation-based approaches are typically applied to the linear models of $\lambda_i(t \mid \mathscr{F}_t)$ and are not suitable for our non-linear model (12). Therefore, we adopt the intensity-based approach, where parameter estimation is achieved through the minimization of a suitable loss function that measures the discrepancy between the true and estimated CIFs.

A. Loss Function

In the existing literature, there are different loss functions used for estimating parameters in a generic counting process $N(t)$ associated with a CIF $\lambda(t \mid \mathscr{F}_t)$, including the negative log-likelihood function [18], [30]:

$$\mathcal{L}(\boldsymbol{\beta}) = -\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\log\{\lambda(t- \mid \mathscr{F}_{t-})\}\,\mathrm{d}N(t) - \lambda(t \mid \mathscr{F}_t)\,\mathrm{d}t\Big], \tag{46}$$

and the squared loss [19], [40]:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big\{\lambda^2(t \mid \mathscr{F}_t)\,\mathrm{d}t - 2\lambda(t- \mid \mathscr{F}_{t-})\,\mathrm{d}N(t)\Big\}, \tag{47}$$

where $\lambda(t- \mid \mathscr{F}_{t-}) = \lim_{u\uparrow t}\lambda(u \mid \mathscr{F}_u)$ denotes the left limit.

The squared loss (47) is more suitable for linear models of $\lambda(t \mid \mathscr{F}_t)$, such as the linear Hawkes process [40], while the negative log-likelihood function (46) is typically used for non-linear cases, such as when using an exponential link function

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

in model (12). Therefore, we will focus our discussion on the use of (46). In our multi-dimensional setting, we choose to estimate $\widetilde{\boldsymbol{\beta}}_i^*$ at individual nodes i, and recover the network structure by aggregating estimators of $\{\widetilde{\boldsymbol{\beta}}_i^*\}_{i\in\mathcal{V}}$ using the following loss function:

$$\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) = -\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\Big[\widetilde{\boldsymbol{x}}_i(t-)^\top\widetilde{\boldsymbol{\beta}}_i\,\mathrm{d}N_i(t) - \exp\{\widetilde{\boldsymbol{x}}_i(t)^\top\widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}t\Big], \quad (48)$$

where $\widetilde{\boldsymbol{\beta}}_i = (\beta_{0;i}, \boldsymbol{\beta}_i^\top)^\top = (\beta_{0;i}, \beta_{1,i}, \ldots, \beta_{i-1,i}, \beta_{i+1,i}, \ldots, \beta_{V,i})^\top \in \mathbb{R}^V$ represents a vector of generic parameters.

In many application fields [21], [22], [23], [24], the number of recorded event time points could be large, often in the order of millions or more. This motivates us to study the behavior of our estimation approach for a large number $N_i(\mathrm{T})$ of event time points, or equivalently, a long total time length T. Theorem 5 presents the asymptotic convergence results for the gradient vector and the Hessian matrix of $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)$ as T approaches infinity. These results will be used to derive parameter estimation consistency (Theorems 6 and 7) in Section V-C.

Theorem 5 (Asymptotic Convergence Related to Loss Function $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ in (48)): Assume conditions A1, A2, A3, A4, A5, and A6 in Appendix B. For each $i \in \mathcal{V}$, denote $\nabla\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ and $\nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ as the gradient vector and Hessian matrix of $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ in (48) respectively. Then, we have the following results as $\mathrm{T}\to\infty$:

(i) $\nabla\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)$ converges to $\mathbf{0}$ in probability at a square-root rate, i.e.,

$$\nabla\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\Big[\widetilde{\boldsymbol{x}}_i(t)\,\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top\widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t - \widetilde{\boldsymbol{x}}_i(t-)\,\mathrm{d}N_i(t)\Big] = O_{\mathrm{P}}(\sqrt{1/\mathrm{T}}). \quad (49)$$

(ii) There exists a constant matrix $\mathbf{C}_i$ such that

$$\nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\widetilde{\boldsymbol{x}}_i(t)\,\widetilde{\boldsymbol{x}}_i(t)^\top\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top\widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t \xrightarrow{\mathrm{P}} \mathbf{C}_i. \quad (50)$$

Furthermore, the matrix $\mathbf{C}_i$ is positive definite with all entries positive.

Theorem 5 is derived from the probabilistic results of $\boldsymbol{N}(t)$ in Section IV. Specifically, (49) is attained from the bounded variance property (42) of $\boldsymbol{N}(t)$ in Theorem 2, which itself originates from the properties of the marked point process $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ (as in Theorem 1 and Lemmas 5–7). Result (50) is derived using the cyclicity property of $\boldsymbol{N}(t)$ from Theorem 3 and the asymptotic mean stationarity of $\boldsymbol{N}(t)$ as proven in Theorem 4.

Remark 4: Conventional tools for asymptotic results, such as the law of large numbers or central limit theorems, are not directly applicable to Theorem 5 due to the distinctive features of the stochastic processes $\boldsymbol{N}(t)$ and $\widetilde{\boldsymbol{x}}_i(t)$ in (49) and (50). Specifically, the non-stationary counting process $\boldsymbol{N}(t)$ is closely linked with the historical events up to t (via its associated CIFs $\{\lambda_i^*(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ modeled by (45)), resulting in a complicated dependence structure of $\boldsymbol{N}(t)$ across time t. Additionally, the stochastic process $\widetilde{\boldsymbol{x}}_i(t)$, defined as $\widetilde{\boldsymbol{x}}_i(t) = (1, x_1(t), \ldots, x_{i-1}(t), x_{i+1}(t), \ldots, x_V(t))^\top$, relies on the special type of stochastic process $N_j((t-\phi, t])$ (see (16), (17) and (18)), for which probabilistic properties are not available in the existing literature. The use of Theorems 1–4 enables us to prove Theorem 5, justifying the importance of our results in Section IV.

B. Penalized Estimation of Parameters

Sparsity assumptions are commonly imposed on the true network structure in various real-world applications (e.g., [11], [12], [19]). To promote a sparse network structure with the most significant interactions, we employ the weighted L_1-penalty:

$$\mathcal{P}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) = \sum_{j\in\mathcal{V}\setminus\{i\}} w_{j,i,\mathrm{T}}\,|\beta_{j,i}|, \quad (51)$$

where $\{w_{j,i,\mathrm{T}} : j \in \mathcal{V}\setminus\{i\}\}$ represent non-negative weights. We estimate the true parameter vector $\widetilde{\boldsymbol{\beta}}_i^*$ using the *penalized M-estimator*, minimizing the sum of the loss function (48) and the penalty function (51):

$$\begin{aligned}\widehat{\widetilde{\boldsymbol{\beta}}}_i &= \arg\min_{\widetilde{\boldsymbol{\beta}}_i\in\mathbb{R}^V}\big\{\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) + \mathcal{P}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)\big\}\\ &= \arg\min_{\widetilde{\boldsymbol{\beta}}_i\in\mathbb{R}^V}\Big\{\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}}\Big[\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top\widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}t - \widetilde{\boldsymbol{x}}_i(t-)^\top\widetilde{\boldsymbol{\beta}}_i\,\mathrm{d}N_i(t)\Big] + \sum_{j\in\mathcal{V}\setminus\{i\}} w_{j,i,\mathrm{T}}\,|\beta_{j,i}|\Big\}, \quad (52)\end{aligned}$$

where the vector $\widehat{\widetilde{\boldsymbol{\beta}}}_i = (\widehat{\beta}_{0;i}, \widehat{\beta}_{1,i}, \ldots, \widehat{\beta}_{i-1,i}, \widehat{\beta}_{i+1,i}, \ldots, \widehat{\beta}_{V,i})^\top$ collects the estimates $\widehat{\beta}_{0;i}$ and all $\{\widehat{\beta}_{j,i} : j \in \mathcal{V}\setminus\{i\}\}$. The estimated network is obtained as follows:

$$\widehat{\mathcal{E}} = \{(j,i) \in \mathcal{V}\times\mathcal{V} : \widehat{\beta}_{j,i} \neq 0;\ j \neq i\}.$$

Furthermore, considering that the sign of an estimator indicates the type of effect, we estimate the sets of excitatory and inhibitory effects separately:

$$\widehat{\mathcal{E}}_+ = \{(j,i) \in \mathcal{V}\times\mathcal{V} : \widehat{\beta}_{j,i} > 0;\ j \neq i\}, \quad (53)$$

$$\widehat{\mathcal{E}}_- = \{(j,i) \in \mathcal{V}\times\mathcal{V} : \widehat{\beta}_{j,i} < 0;\ j \neq i\}. \quad (54)$$

C. Asymptotic Results for Structure Learning

For continuous-time point process data, the total time length T is roughly proportional to the number of the observed data points and typically serves as the sample size (e.g., in [19], [20]). Therefore, in this section, we establish the asymptotic properties of the penalized M-estimator $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ in (52) with respect to T approaching infinity. To establish estimation consistency, we first provide the following conditions for the weights $w_{j,i,\mathrm{T}}$ in (51):

$$\max_{j\in\mathrm{Pa}^*(i)} w_{j,i,\mathrm{T}} = O_{\mathrm{P}}(\sqrt{1/\mathrm{T}}); \quad (55)$$

$$\max_{j\in\mathrm{Pa}^*(i)} w_{j,i,\mathrm{T}} = o_{\mathrm{P}}(\sqrt{1/\mathrm{T}}); \quad (56)$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$\min_{j\in\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}} \sqrt{\mathrm{T}}\, w_{j,i,\mathrm{T}} \xrightarrow{\mathrm{P}} \infty, \quad \text{as } \mathrm{T}\to\infty. \tag{57}$$

Here, $\mathrm{Pa}^*(i)=\{j\in\mathcal{V}\setminus\{i\}:\beta^*_{j,i}\neq 0\}$ denotes the nodes that have a true non-zero effect on node i; condition (56) employed in Theorem 7 and Corollary 1 is stronger than condition (55) used in Theorem 6. An example of weights $\{w_{j,i,\mathrm{T}}\}$ that satisfy (56) and (57) is the adaptive lasso penalty [45], in which $w_{j,i,\mathrm{T}}=\eta_{\mathrm{T}}\,|\breve{\beta}_{j,i}|^{\gamma}$, with $\eta_{\mathrm{T}}=O(1/\mathrm{T}^a)$ for $1/2<a<3/2$, $\gamma=-2$, and $\breve{\boldsymbol{\beta}}_i=(\breve{\beta}_{0;i},\breve{\beta}_{1,i},\ldots,\breve{\beta}_{i-1,i},\breve{\beta}_{i+1,i},\ldots,\breve{\beta}_{V,i})^\top$ denoting the minimizer of $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$.

Theorem 6 guarantees the existence of a $\sqrt{1/\mathrm{T}}$-consistent estimator $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ in (52).

Theorem 6 (Existence of a Consistent Penalized M-Estimator): Assume conditions A1, A2, A3, A4, A5, A6, and A7 in Appendix B. Assume (55) for the weights $w_{j,i,\mathrm{T}}$. Then, there exists a local minimizer $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ in (52) such that $\|\widehat{\widetilde{\boldsymbol{\beta}}}_i-\widetilde{\boldsymbol{\beta}}^*_i\|=O_{\mathrm{P}}(\sqrt{1/\mathrm{T}})$, as $\mathrm{T}\to\infty$.

Following Theorem 6, the sparsistency of the penalized M-estimator is given in Theorem 7 below. Before stating it, we introduce some notations. We partition the true parameter vector as $\widetilde{\boldsymbol{\beta}}^*_i=(\beta^*_{0;i},\boldsymbol{\beta}^{*\top}_i)^\top=(\beta^*_{0;i},\boldsymbol{\beta}^{*(\mathrm{I})\top}_i,\boldsymbol{\beta}^{*(\mathrm{II})\top}_i)^\top=(\widetilde{\boldsymbol{\beta}}^{*(\mathrm{I})\top}_i,\boldsymbol{\beta}^{*(\mathrm{II})\top}_i)^\top$, where $\boldsymbol{\beta}^{*(\mathrm{II})}_i=\mathbf{0}$, and $\boldsymbol{\beta}^{*(\mathrm{I})}_i$ collects all the non-zero components in $\boldsymbol{\beta}^*_i$. Similarly, for the estimator $\widehat{\widetilde{\boldsymbol{\beta}}}_i$, we adopt the partition $\widehat{\widetilde{\boldsymbol{\beta}}}_i=(\widehat{\beta}_{0;i},\widehat{\boldsymbol{\beta}}^{(\mathrm{I})\top}_i,\widehat{\boldsymbol{\beta}}^{(\mathrm{II})\top}_i)^\top=(\widehat{\widetilde{\boldsymbol{\beta}}}^{(\mathrm{I})\top}_i,\widehat{\boldsymbol{\beta}}^{(\mathrm{II})\top}_i)^\top$, with index sets I and II corresponding to those of $\boldsymbol{\beta}^{*(\mathrm{I})}_i$ and $\boldsymbol{\beta}^{*(\mathrm{II})}_i$, respectively.

Theorem 7 (Sparsistency of the Penalized M-Estimator): Assume conditions A1, A2, A3, A4, A5, A6, and A7 in Appendix B. Assume that the weights $w_{j,i,\mathrm{T}}$ satisfy (56) and (57). Then, any $\sqrt{1/\mathrm{T}}$-consistent local minimizer $\widehat{\widetilde{\boldsymbol{\beta}}}_i=(\widehat{\widetilde{\boldsymbol{\beta}}}^{(\mathrm{I})\top}_i,\widehat{\boldsymbol{\beta}}^{(\mathrm{II})\top}_i)^\top$ in (52) satisfies

$$\mathrm{P}\big(\widehat{\boldsymbol{\beta}}^{(\mathrm{II})}_i=\mathbf{0}\big)\to 1, \quad \text{as } \mathrm{T}\to\infty. \tag{58}$$

The sparsistency result in (58) immediately yields the network recovery consistency stated in Corollary 1.

Corollary 1 (Network Recovery Consistency): Assume the same conditions as in Theorem 7. Then, the network structure estimators $\widehat{\mathcal{E}}_+$ in (53) and $\widehat{\mathcal{E}}_-$ in (54), based on $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ in (52), are consistent with the true edges $\mathcal{E}^*_+$ and $\mathcal{E}^*_-$, respectively. In other words, $\mathrm{P}(\widehat{\mathcal{E}}_+=\mathcal{E}^*_+,\widehat{\mathcal{E}}_-=\mathcal{E}^*_-)\to 1$ as $\mathrm{T}\to\infty$.

Corollary 1 demonstrates that our method can consistently recover the true network structure as the total time length T increases. This provides theoretical support for the utility of our proposed statistical learning procedure.

Remark 5: Standard results regarding parameter estimation consistency for general M-estimators are extensively established in the existing statistical literature, as discussed in Chapter 5 of [46]. However, the general theory does not directly apply to prove the consistency results presented in Theorems 6 and 7 within our specific context. This arises from our complex loss function (48), which relies on a complicated stochastic integral, while the typical loss function in [46] is often constrained to a simpler form involving basic summation statistics. The proofs of Theorems 6 and 7 rely on the asymptotic convergence of $\nabla\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}^*_i)$ and $\nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}^*_i)$ in Theorem 5, derived from the probabilistic results of $\boldsymbol{N}(t)$ in Theorems 1–4.

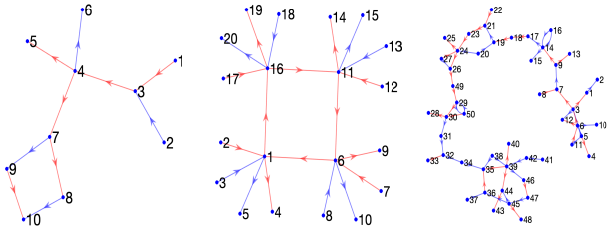

Fig. 4. **(Simulation Study: True** Network-1**,** Network-2**, and** Network-3**)** The left panel: Network-1, a simple network with 10 nodes; the middle panel: Network-2, a medium-complexity network with 20 nodes; and the right panel: Network-3, a complex network with 50 nodes. Red arrows indicate excitatory effects, while blue arrows indicate inhibitory effects.

VI. Simulation Study

In this section, we conduct numerical experiments to demonstrate the practical utility of our continuous-time modeling approach and estimation procedure.

A. Types of Network Structures

The simulation studies consider three simulated networks, as depicted in Figure 4, representing networks with varying degrees of complexity. Network-1 is a simple network of 10 nodes, including 6 excitatory and 4 inhibitory effects. It is designed to resemble a directed acyclic graph, aiming to capture the information flow from sensory neurons to motor neurons. Network-2, adapted from [12], is a moderately complex network comprising 20 nodes, with 12 excitatory and 8 inhibitory effects. This network is intended to mimic the potential *hub* and *leaf* structures observed in neuron ensembles. Specifically, nodes 1, 6, 11, and 16 are hub nodes with a degree of 6, while the remaining nodes are leaves with a degree of 1. Network-3 is a complex network consisting of 50 nodes, with 30 excitatory and 30 inhibitory effects. It shares the same design motivation as Network-1.

The synthetic multivariate point process data were generated using the simulation algorithm induced by Theorem 1. The CIFs in model (12) employ

$$x_j(t)=g\big(r_{j,\phi}(t)\big),\ j\in\mathcal{V},\ t\in[0,\mathrm{T}], \qquad \text{where } \phi=1,\ \text{and } g(x)=\log(1+x\wedge 10). \tag{59}$$

B. Comparison of Estimation Methods

We compare the following estimation procedures:

(i) Continuous-time modeling (our proposed method): This method estimates the parameters using the penalized M-estimator in (52) with two scenarios for the penalty (51): the L_1-penalty $\mathcal{P}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)=\eta\sum_{j\in\mathcal{V}\setminus\{i\}}|\beta_{j,i}|$; the weighted-$L_1$ penalty $\mathcal{P}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)=\sum_{j\in\mathcal{V}\setminus\{i\}}\eta_{\mathrm{T}}|\breve{\beta}_{j,i}|^{\gamma}\cdot|\beta_{j,i}|$, with $\gamma=-2$, and the M-estimator $\breve{\beta}_{j,i}$ of $\beta^*_{j,i}$, where the tuning parameters η and η_{T} are selected using the Bayesian Information Criterion (BIC) [47].

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

TABLE I
METHOD DESCRIPTIONS FOR SIMULATION STUDIES

abbreviation of method		description
Discrete_L1	bin=0.5	method (ii) with L_1-penalty, and bin width $= 0.5$.
	bin=0.25	method (ii) with L_1-penalty, and bin width $= 0.25$.
	bin=0.1	method (ii) with L_1-penalty, and bin width $= 0.1$.
Continuous_L1		method (i) with L_1-penalty.
Discrete_wL1	bin=0.5	method (ii) with weighted-L_1 penalty, and bin width $= 0.5$.
	bin=0.25	method (ii) with weighted-L_1 penalty, and bin width $= 0.25$.
	bin=0.1	method (ii) with weighted-L_1 penalty and bin width $= 0.1$.
Continuous_wL1		method (i) with weighted-L_1 penalty.
Zhao_2012	bin=0.5	method (iii) with bin width $= 0.5$.
	bin=0.25	method (iii) with bin width $= 0.25$.
	bin=0.1	method (iii) with bin width $= 0.1$.
SIE-GLM	bin=0.5	method (iv) with bin width $= 0.5$.
	bin=0.25	method (iv) with bin width $= 0.25$.
	bin=0.1	method (iv) with bin width $= 0.1$.
Raj_2005	parent=3	method (v) with maximum parent number = 3.
	parent=2	method (v) with maximum parent number = 2.

(ii) Discrete-time approximation modeling: Method (ii) utilizes the discrete-time approximation. The entire time interval $[0, \mathrm{T}]$ is divided into n equally-spaced time bins $\{(t_{k-1}, t_k] : k = 1, \ldots, n\}$, each of length T/n. The observed point process $\{\boldsymbol{T}_i\}_{i\in\mathcal{V}}$ is transformed into sequences of bin counts $\{N_{i,k}\}_{i\in\mathcal{V};k=1,\ldots,n}$. The interaction parameters are estimated using a penalized M-estimation similar to (52). However, in this case, a Poisson distribution with rate $\lambda_i(t_{k-1})\,\mathrm{T}/n$ is assumed for $N_{i,k}$ at node i.

(iii) Discrete-time modeling with groups of connection parameters in [12]: Method (iii) is similar to method (ii), differing in that the effect from node j to i is modeled by a group of parameters $\{\beta_{j,i,q} : q = 1, \ldots, Q\}$ instead of a single parameter $\beta_{j,i}$. Here, Q is determined by the integer part of $\phi/(\mathrm{T}/n)$.

(iv) SIE-GLM method in [11]: Method (iv) is an extension of method (iii) that incorporates structural information in the parameter space. It employs the sparse group lasso penalty for parameter estimation.

(v) Bayesian method in [15]: Method (v) is a continuous-time modeling approach that uses a default shape-function $\log(1 + x)$ and the same loss function as our proposed method (i). This method explores all subsets of components in $\widetilde{\boldsymbol{\beta}}_i$ and selects the best subset with the maximum Bayesian posterior density. To ensure a fair comparison, method (v) assumes a uniform prior (i.e., no prior information) for $\widetilde{\boldsymbol{\beta}}_i$.

To facilitate further discussion, we categorize all methods in Table I. All methods which involve ϕ and $g(\cdot)$ for parameter estimation adopt our empirical choices: $\phi = 1$ and $g(x) = \log(1+x\wedge c)$, where the data-driven choice c is given in (62), unless stated otherwise. The coordinate descent algorithm [48] is utilized to solve (52).

C. Simulation Results

We consider three different total time lengths: $\mathrm{T} \in \{500, 1000, 2000\}$. For each $i \in \mathcal{V}$, the true baseline intensity parameter is $\beta^*_{0;i} = -0.8$, resulting in a base rate of approximately $\exp(-0.8) \approx 0.45$. The true graph parameters $\{\beta^*_{j,i} : i, j \in \mathcal{V};\ j \neq i\}$ are set as follows: $\beta^*_{j,i}$ is β for the excitatory effect, $-\beta$ for the inhibitory effect, and 0 for no effect from node j to node i. Here, $\beta \in \{0.4, 0.5\}$ reflects the magnitude of the connection strength.

The performance of each method is evaluated using the following criterion measures: Corret_All (correctly detected number of excitatory and inhibitory effects), Detected_A (correctly detected number of excitatory effects), Detected_B (correctly detected number of inhibitory effects), and Correct_NC (correctly detected number of non-effects). For comparison, Corret_All, Detected_A, and Detected_B reflect the sensitivity level, which is defined as the percentage of correctly identified effects. It measures how sensitive each method is in detecting excitatory or inhibitory effects. Additionally, Correct_NC indicates the specificity level, defined as the percentage of correctly identified non-effects. It represents the ability of the method to correctly identify the absence of an effect.

1) Complex Network: For complex network Network-3, we first compare the performance of each method under different connection strengths $\beta \in \{0.4, 0.5\}$ in Table II. Most methods are successful in detecting the sparse structure of the network, correctly identifying most true non-effects and achieving a good level of specificity. However, the sensitivity results are relatively worse compared to specificity. All methods with $\beta = 0.5$ exhibit better sensitivity results compared to $\beta = 0.4$. This is expected since a larger connection strength parameter implies stronger interaction between nodes, making detection easier. In both strength parameter settings, continuous-time methods (Continuous_L1 and Continuous_wL1) outperform the discrete-time approximation methods (Discrete_L1, Discrete_wL1, bin $\in \{0.5, 0.25, 0.1\}$) in terms of sensitivity. It is worth noting that for Discrete_L1 and Discrete_wL1, a smaller bin width yields better results but does not surpass the corresponding continuous-time methods Continuous_L1 and Continuous_wL1. This observation suggests that continuous-time modeling can be considered as a limiting case of discrete-time modeling when the bin width approaches zero, thus providing the most accurate results. Regarding the penalty choices in methods (i) and (ii), consistently using the weighted-L_1 penalty yields better results than using the L_1-penalty when the same loss function is employed. Methods (iii) (Zhao_2012, bin $\in \{0.5, 0.25\}$) and (iv) (SIE-GLM, bin $\in \{0.5, 0.25\}$) exhibit relatively reduced sensitivity performance compared to other methods, with Correct_All being less than 38 out of 60. As for method (v) (Raj_2005, parent $= 2$), to reduce the computational cost of searching all possible subsets of parents for each node, only subsets with a maximum size of 2 are considered. Since the true network is sparse with a degree no greater than 2 for each node, this setting is most favorable for method (v). Nevertheless, method (v) only performs well in terms of sensitivity and significantly underperforms in terms of specificity compared to other methods. In summary, our proposed continuous-time method (Continuous_wL1) with the weighted-L_1 penalty demonstrates the best overall performance across $\beta \in \{0.4, 0.5\}$.

We next present Table III to compare the results using different values of the total time length $\mathrm{T} \in \{1000, 2000\}$. It is evident that $\mathrm{T} = 2000$ outperforms $\mathrm{T} = 1000$ for all methods. This aligns with expectations since larger datasets

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

TABLE II

(SIMULATION STUDY: Network-3 WITH CONNECTION STRENGTH $\beta \in \{0.4, 0.5\}$)THE TIME LENGTH IS $T = 2000$.RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All		Detected_A		Detected_B		Correct_NC	
strength $\beta =$		0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5
Discrete_L1	bin=0.5	21.18 (0.42)	42.55 (0.33)	13.71 (0.26)	24.92 (0.17)	7.47 (0.25)	17.63 (0.26)	2385.87 (0.26)	2379.92 (0.33)
	bin=0.25	36.21 (0.45)	53.39 (0.25)	21.29 (0.25)	28.83 (0.10)	14.92 (0.30)	24.56 (0.21)	2382.13 (0.33)	2376.73 (0.42)
	bin=0.1	44.04 (0.41)	56.95 (0.17)	24.76 (0.18)	29.67 (0.05)	19.28 (0.29)	27.28 (0.15)	2379.40 (0.38)	2375.46 (0.38)
Continuous_L1		51.86 (0.33)	58.44 (0.14)	27.33 (0.15)	29.87 (0.03)	24.53 (0.26)	28.57 (0.12)	2368.23 (0.60)	2367.52 (0.43)
Discrete_wL1	bin=0.5	40.02 (0.39)	54.05 (0.25)	22.69 (0.22)	28.86 (0.11)	17.32 (0.29)	25.19 (0.20)	2380.52 (0.39)	2379.92 (0.36)
	bin=0.25	50.60 (0.28)	58.55 (0.11)	27.05 (0.17)	29.83 (0.03)	23.55 (0.21)	28.72 (0.09)	2380.32 (0.39)	2381.44 (0.34)
	bin=0.1	54.81 (0.22)	59.44 (0.07)	28.51 (0.12)	29.97 (0.01)	26.30 (0.18)	29.47 (0.06)	2380.07 (0.36)	2382.73 (0.30)
Continuous_wL1		56.80 (0.21)	59.72 (0.05)	29.08 (0.11)	30.00 (0.00)	27.72 (0.17)	29.72 (0.05)	2380.71 (0.34)	2383.44 (0.25)
Zhao_2012	bin=0.5	10.08 (0.29)	26.42 (0.37)	6.36 (0.20)	16.46 (0.24)	3.72 (0.16)	9.96 (0.19)	2388.66 (0.13)	2385.15 (0.29)
	bin=0.25	3.62 (0.17)	9.41 (0.26)	3.28 (0.16)	8.08 (0.22)	0.34 (0.06)	1.33 (0.11)	2389.66 (0.05)	2388.86 (0.14)
SIE-GLM	bin=0.5	19.88 (0.40)	39.28 (0.33)	14.00 (0.26)	24.54 (0.18)	5.88 (0.22)	14.74 (0.25)	2387.55 (0.18)	2384.09 (0.28)
	bin=0.25	17.43 (0.35)	37.24 (0.31)	13.29 (0.23)	24.28 (0.20)	4.13 (0.20)	12.96 (0.22)	2388.23 (0.13)	2384.84 (0.25)
Raj_2005	parent=2	57.34 (0.08)	57.91 (0.02)	29.37 (0.07)	29.45 (0.06)	27.97 (0.09)	28.46 (0.06)	2347.34 (0.08)	2347.91 (0.02)
true		60		30		30		2390	

TABLE III

(SIMULATION STUDY: Network-3 WITH TIME LENGTH $T \in \{1000, 2000\}$)THE CONNECTION STRENGTH IS $\beta = 0.5$.RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All		Detected_A		Detected_B		Correct_NC	
time length $T =$		1000	2000	1000	2000	1000	2000	1000	2000
Discrete_L1	bin=0.5	16.75 (0.39)	42.55 (0.33)	11.74 (0.26)	24.92 (0.17)	5.01 (0.22)	17.63 (0.26)	2385.84 (0.23)	2379.92 (0.33)
	bin=0.25	28.24 (0.40)	53.39 (0.25)	18.19 (0.24)	28.83 (0.10)	10.05 (0.25)	24.56 (0.21)	2383.26 (0.30)	2376.73 (0.42)
	bin=0.1	35.75 (0.41)	56.95 (0.17)	21.91 (0.24)	29.67 (0.05)	13.84 (0.26)	27.28 (0.15)	2379.78 (0.34)	2375.46 (0.38)
Continuous_L1		48.79 (0.33)	58.44 (0.14)	27.84 (0.12)	29.87 (0.03)	20.95 (0.28)	28.57 (0.12)	2359.78 (0.81)	2367.52 (0.43)
Discrete_wL1	bin=0.5	33.18 (0.45)	54.05 (0.25)	19.69 (0.26)	28.86 (0.11)	13.49 (0.30)	25.19 (0.20)	2376.73 (0.42)	2379.92 (0.36)
	bin=0.25	44.57 (0.34)	58.55 (0.11)	25.04 (0.19)	29.83 (0.03)	19.53 (0.26)	28.72 (0.09)	2377.19 (0.42)	2381.44 (0.34)
	bin=0.1	49.88 (0.30)	59.44 (0.07)	27.24 (0.15)	29.97 (0.01)	22.64 (0.25)	29.47 (0.06)	2376.29 (0.41)	2382.73 (0.30)
Continuous_wL1		52.61 (0.26)	59.72 (0.05)	28.06 (0.12)	30.00 (0.00)	24.55 (0.21)	29.72 (0.05)	2375.96 (0.42)	2383.44 (0.25)
Zhao_2012	bin=0.5	6.15 (0.25)	26.42 (0.37)	4.33 (0.20)	16.46 (0.24)	1.82 (0.13)	9.96 (0.19)	2388.78 (0.12)	2385.15 (0.29)
	bin=0.25	2.84 (0.17)	9.41 (0.26)	2.60 (0.16)	8.08 (0.22)	0.24 (0.04)	1.33 (0.11)	2389.59 (0.08)	2388.86 (0.14)
SIE-GLM	bin=0.5	15.07 (0.36)	39.28 (0.33)	11.77 (0.27)	24.54 (0.18)	3.30 (0.18)	14.74 (0.25)	2387.59 (0.18)	2384.09 (0.28)
	bin=0.25	12.24 (0.36)	37.24 (0.31)	10.30 (0.29)	24.28 (0.20)	1.94 (0.15)	12.96 (0.22)	2388.46 (0.12)	2384.84 (0.25)
Raj_2005	parent=2	55.44 (0.16)	57.91 (0.02)	28.96 (0.09)	29.45 (0.06)	26.48 (0.13)	28.46 (0.06)	2345.44 (0.16)	2347.91 (0.02)
true		60		30		30		2390	

provide more information and lead to more accurate estimations. This finding is also consistent with our theoretical result of network recovery consistency stated in Corollary 1 of Section V-C, which indicates that the detected network becomes closer to the true network as the time length T increases. Under each T setting, the pattern of results is similar to that in Table II. Continuous_wL1 maintains the best overall performance.

To assess the robustness of our method to misspecified time-lags for the true time-lag ϕ (equal to 1), we utilize specified time-lag values, $\phi_a \in \{0.5, 1, 1.5\}$, in the estimation procedure. The results are provided in Table IV. As anticipated, $\phi_a = 1$ exhibits the most favorable performance. The misspecified ϕ_a values of $\{0.5, 1.5\}$ do not significantly impact specificity but do decrease sensitivity across most of the listed methods. Among the listed methods, the continuous-time approaches (Continuous_L1 and Continuous_wL1) still demonstrate the most robust overall performance. Particularly, the sensitivity of Continuous_L1 diminishes by less than 15% under both misspecified time-lag scenarios, indicating a degree of resilience of our method against this type of misspecification.

To assess the robustness of our methods against misspecified models for CIFs $\lambda_i(t \mid \mathscr{F}_t)$, we conducted a separate simulation study on data generated from the non-linear Hawkes model, with the true CIF:

$$\lambda_i^*(t \mid \mathscr{F}_t) = \exp\{\beta_{0;i}^* + \sum_{j\in\mathcal{V}} \int_{-\infty}^{t} \beta_{j,i}^* \, \mathrm{I}(0 \le t-u < 1) \, \mathrm{d}N_j(u)\}, \tag{60}$$

for $i \in \mathcal{V}$. In this model, we set $\beta_{0;i}^* = -0.8$, $\beta_{j,i}^* = 3$ for the excitatory effect, $\beta_{j,i}^* = -3$ for the inhibitory effect, and $\beta_{j,i}^* = 0$ for no effect from node j to node i. The results in Table V indicate that the performances of each method largely agrees with the results reflected in Tables II, III, and IV. Our proposed method, Continuous_wL1, continues to exhibit the best overall performance. In summary, this simulation result demonstrates the robustness of our estimation method against model misspecification. Our method performs well even when the shape-function g is unbounded in the true model. This indicates that our estimation method and theoretical results are applicable to a broader range of models beyond the non-linear Hawkes process.

2) Simple and Medium-Complex Networks: For Network-1 and Network-2, we conducted the same simulation evaluation as for Network-3. The results of the two networks, comparing connection strength, time length and time-lag width, resemble those obtained for Network-3 and have been omitted for brevity. Among all the methods, Continuous_wL1 consistently exhibits the best overall performance in each setting.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

(SIMULATION STUDY: Network-3 PARAMETER ESTIMATION USING SPECIFIED TIME-LAGS $\phi_a \in \{0.5, 1, 1.5\}$ FOR THE TRUE TIME-LAG $\phi = 1$) THE CONNECTION STRENGTH IS $\beta = 0.5$, AND THE TIME LENGTH IS $\mathrm{T} = 2000$. RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All			Detected_A			Detected_B			Correct_NC		
time-lag ϕ_a =		0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Discrete_L1	bin=0.5	37.10 (0.36)	42.55 (0.33)	23.29 (0.45)	22.57 (0.19)	24.92 (0.17)	15.42 (0.28)	14.53 (0.25)	17.63 (0.26)	7.87 (0.27)	2382.26 (0.31)	2379.92 (0.33)	2384.75 (0.25)
	bin=0.25	37.11 (0.42)	53.39 (0.25)	37.15 (0.43)	22.70 (0.25)	28.83 (0.10)	22.39 (0.25)	14.41 (0.25)	24.56 (0.21)	14.76 (0.27)	2381.75 (0.30)	2376.73 (0.42)	2381.04 (0.32)
	bin=0.1	36.77 (0.42)	56.95 (0.17)	44.89 (0.38)	22.26 (0.21)	29.67 (0.05)	25.88 (0.19)	14.51 (0.29)	27.28 (0.15)	19.01 (0.28)	2382.06 (0.29)	2375.46 (0.38)	2379.39 (0.34)
Continuous_L1		43.90 (0.32)	58.44 (0.14)	54.26 (0.26)	25.70 (0.17)	29.87 (0.03)	28.06 (0.14)	18.20 (0.25)	28.57 (0.12)	26.20 (0.19)	2362.46 (0.60)	2367.52 (0.43)	2369.17 (0.60)
Discrete_wL1	bin=0.5	51.83 (0.25)	54.05 (0.25)	40.63 (0.36)	27.70 (0.13)	28.86 (0.11)	23.83 (0.21)	24.13 (0.19)	25.19 (0.20)	16.80 (0.27)	2380.07 (0.42)	2379.92 (0.36)	2380.05 (0.36)
	bin=0.25	51.52 (0.29)	58.55 (0.11)	50.86 (0.28)	27.60 (0.15)	29.83 (0.03)	27.85 (0.11)	23.92 (0.24)	28.72 (0.09)	23.01 (0.23)	2379.61 (0.34)	2381.44 (0.34)	2380.12 (0.40)
	bin=0.1	51.44 (0.28)	59.44 (0.07)	54.72 (0.23)	27.56 (0.13)	29.97 (0.01)	29.10 (0.08)	23.88 (0.22)	29.47 (0.06)	25.62 (0.20)	2380.21 (0.34)	2382.73 (0.30)	2380.28 (0.35)
Continuous_wL1		51.05 (0.28)	59.72 (0.05)	56.49 (0.17)	27.58 (0.15)	30.00 (0.00)	29.43 (0.07)	23.47 (0.21)	29.72 (0.05)	27.06 (0.15)	2380.23 (0.36)	2383.44 (0.25)	2381.25 (0.35)
Zhao_2012	bin=0.5	34.61 (0.35)	26.42 (0.37)	23.51 (0.37)	20.62 (0.20)	16.46 (0.24)	14.86 (0.25)	14.00 (0.23)	9.96 (0.19)	8.65 (0.18)	2383.59 (0.33)	2385.15 (0.29)	2386.07 (0.21)
	bin=0.25	10.59 (0.31)	9.41 (0.26)	8.53 (0.26)	8.76 (0.25)	8.08 (0.22)	7.41 (0.23)	1.83 (0.13)	1.33 (0.11)	1.12 (0.11)	2388.87 (0.12)	2388.86 (0.14)	2388.94 (0.12)
SIE-GLM	bin=0.5	36.04 (0.32)	39.28 (0.33)	35.00 (0.32)	22.71 (0.18)	24.54 (0.18)	22.95 (0.20)	13.33 (0.22)	14.74 (0.25)	12.05 (0.23)	2385.73 (0.21)	2384.09 (0.28)	2384.92 (0.25)
	bin=0.25	25.84 (0.39)	37.24 (0.31)	29.83 (0.35)	18.60 (0.25)	24.28 (0.20)	21.09 (0.22)	7.24 (0.22)	12.96 (0.22)	8.74 (0.21)	2387.31 (0.17)	2384.84 (0.25)	2386.44 (0.21)
Raj_2005	parent=2	55.34 (0.14)	57.91 (0.02)	57.16 (0.08)	28.99 (0.09)	29.45 (0.06)	29.44 (0.06)	26.35 (0.15)	28.46 (0.06)	27.72 (0.09)	2345.34 (0.14)	2347.91 (0.02)	2347.16 (0.08)
true		60			30			30			2390		

TABLE V

(SIMULATION STUDY: Network-3 FOR DATA FROM A NON-LINEAR HAWKES MODEL WITH CIF IN (60)) THE TIME LENGTH IS $\mathrm{T} = 1000$. RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All	Detected_A	Detected_B	Correct_NC
Discrete_L1	bin=0.5	19.32 (0.37)	14.27 (0.28)	5.05 (0.21)	2385.11 (0.27)
	bin=0.25	29.47 (0.41)	20.03 (0.25)	9.44 (0.28)	2382.94 (0.36)
	bin=0.1	36.04 (0.37)	23.17 (0.21)	12.87 (0.26)	2380.29 (0.33)
Continuous_L1		48.70 (0.28)	28.13 (0.11)	20.57 (0.25)	2359.48 (0.88)
Discrete_wL1	bin=0.5	35.11 (0.37)	22.21 (0.22)	12.91 (0.28)	2376.48 (0.43)
	bin=0.25	45.04 (0.34)	26.27 (0.17)	18.77 (0.27)	2375.78 (0.45)
	bin=0.1	49.84 (0.27)	27.96 (0.13)	21.88 (0.24)	2376.30 (0.44)
Continuous_wL1		52.58 (0.23)	28.70 (0.10)	23.88 (0.22)	2376.53 (0.42)
Zhao_2012	bin=0.5	9.41 (0.32)	7.27 (0.24)	2.14 (0.14)	2388.17 (0.16)
	bin=0.25	4.95 (0.20)	4.69 (0.19)	0.26 (0.05)	2389.42 (0.07)
SIE-GLM	bin=0.5	19.03 (0.31)	15.65 (0.25)	3.38 (0.15)	2387.44 (0.20)
	bin=0.25	17.70 (0.33)	15.43 (0.28)	2.27 (0.13)	2387.73 (0.18)
Raj_2005	parent=2	55.33 (0.15)	29.19 (0.08)	26.14 (0.15)	2345.33 (0.15)
true		60	30	30	2390

This finding indicates that our conclusions are consistent across different types of true networks.

In summary, all of these simulation results confirm the superiority of our proposed continuous-time method over the other methods, regardless of the complexity level of the true network structure.

VII. REAL DATA ANALYSIS

In this section, we apply our method to real-world multivariate point process data, obtained from the prefrontal cortex spike train dataset 'pfc-6' on CRCNS, accessible at https://crcns.org/data-sets/pfc/pfc-6/about-pfc-6. This dataset comprises neuronal ensemble recordings from the medial prefrontal cortex, primarily the prelimbic cortex, of freely moving rats using tetrodes. The recordings were conducted during the rats' execution of a behavioral contingency task, as well as during sleep periods before and after the task. This dataset consists of 90 sessions, each representing an experiment. We choose the specific session folder 181020 for our analysis. Within this selected session, we have spike train data collected from 55 neurons spanning a duration of 6500 seconds. This data is stored in the file '181020_SpikeData.dat,' encompassing a total of $1,309,619$ spikes from the 55 neurons.

We apply our continuous-time modeling method, Continuous_wL1, to this dataset, with the tuning parameter selected using the BIC. Similar to the simulation studies, our estimation procedure adopts the empirical choices of $\phi = 1$ and $g(x) = \log(1 + x \wedge c)$, where the data-driven choice for c is given in (62). Previous studies in neuroscience [5], [10] have indicated that a neuron's spiking activity may influence other neurons primarily within a short period, often less than 1 second, known as the refractory-recovery period. Taking this into account, we empirically choose $\phi = 1$ to capture short-term interactions among neurons while considering the refractory-recovery period. The estimated network structure is presented in Figure 5 (left panel). We identify a total of 579 connections, including 352 excitatory effects and 227 inhibitory effects. Several interesting findings emerge from this study. For instance, pairs of neurons $\{6, 7\}$, $\{24, 34\}$, $\{38, 42\}$, $\{25, 27\}$, $\{21, 23\}$ demonstrate strong mutual excitatory effects, suggesting close functional connectivity and similarity within these pairs. Neuron 13 exhibits 34 excitatory effects on other neurons, which is significantly higher than any other neuron, while it does not impose any inhibitory effect. This suggests that neuron 13 may potentially serve as a hub neuron, playing a crucial role in triggering the activities of the entire neuron ensemble. To compare with BIC, we also incorporate the Generalized Information Criterion (GIC) [47] with a penalty perm $a_{\mathrm{T}} = V \log(\mathrm{T})$ to select the tuning parameter. The resulting network, shown in Figure 5 (right panel), is sparser and includes a number of isolated neurons that are disconnected with others. In this regard, GIC fails to capture all potential interactions compared to BIC in our experiment. It is important to note that the recovered connections, obtained through either the BIC or GIC method, represent the estimated statistical dependencies between neurons. However, these estimated connections do not necessarily imply the existence of real neuronal connections in the brain. Nevertheless, our results are valuable in assisting further neurological research.

We also employ two additional modeling methods for this dataset: (a) the Discrete_wL1 method with the BIC criterion and a bin size of 0.25; and (b) the linear Hawkes process (HK) modeling method [18]. The resulting estimated networks are displayed in Figure 6. The Discrete_wL1 method identifies a total of 479 effects, among which 448 are also identified by the Continuous_wL1 method with BIC. This reveals a significant overlap between the network estimated by the Discrete_wL1 and Continuous_wL1 methods, as expected since the Discrete_wL1 method approximates

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

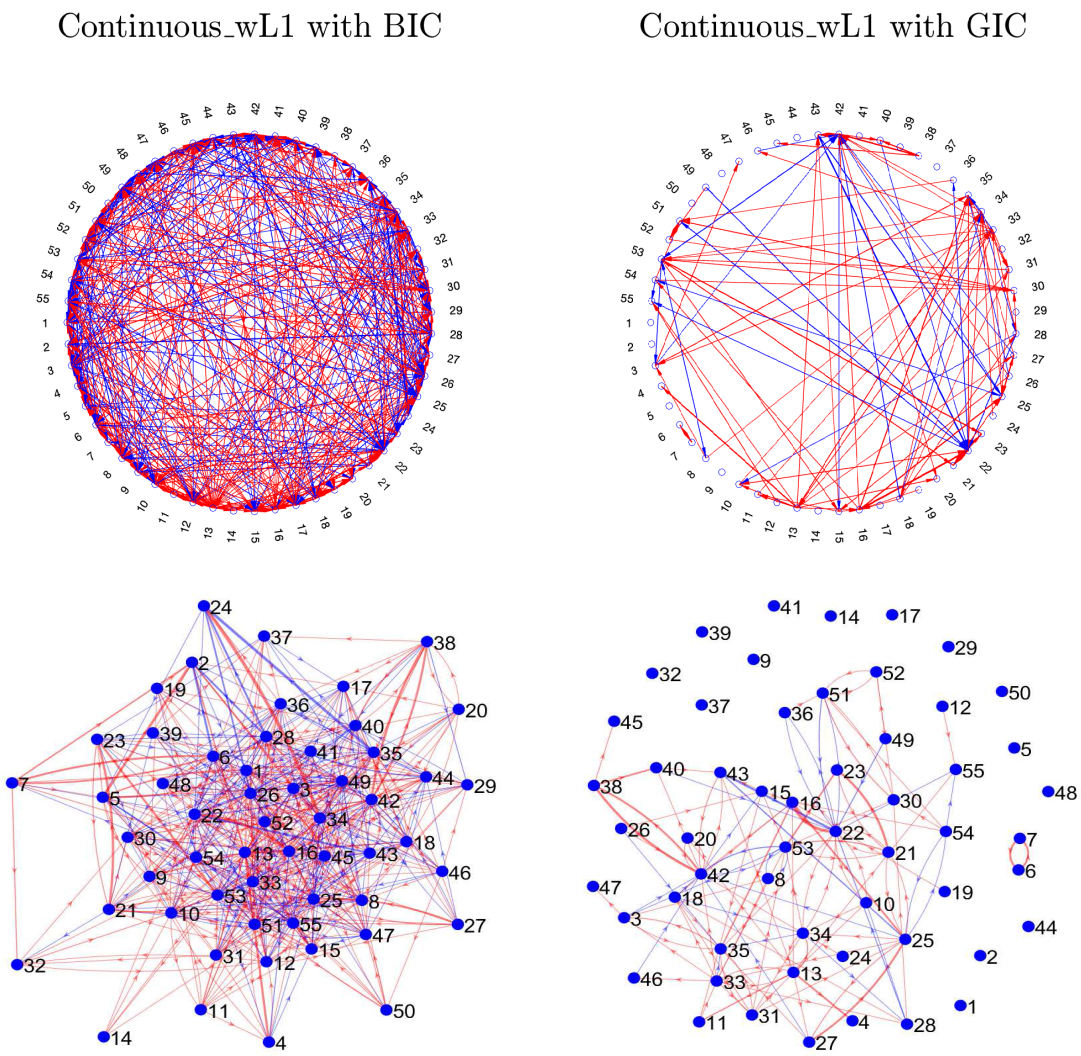

Fig. 5. **(Real Data: Estimated Networks Using the Continuous-Time Modeling Method with Weighted-L_1 Penalty)** Red arrows denote excitatory effects, while blue arrows indicate inhibitory effects. Thicker arrows represent stronger interactions. Top-left panel: BIC criterion (in circular layout); bottom-left panel: BIC criterion (in equilibrium layout); top-right panel: GIC criterion (in circular layout); bottom-right panel: GIC criterion (in equilibrium layout).

the Continuous_wL1 method when the bin size is sufficiently small. However, we assert that the Continuous_wL1 method is more accurate than the Discrete_wL1 method, particularly when the real physical intensity of neuronal spikes evolves in continuous-time. This claim finds support in our simulation results in Section VI. Regarding the HK method, it detected only 353 excitatory effects and no inhibitory effects. This limitation arises from the inherent nature of the linear Hawkes process, which is primarily self-exciting and does not accommodate negative parameterizations in its kernel function. In contrast, our Continuous_wL1 method can identify both excitatory and inhibitory interactions among neurons, providing a more comprehensive estimation of the potential network of functional connections among this group of neurons.

VIII. Discussion

Motivated by the crucial task of inferring neural connectivity from ensemble neural spike train data in neuroscience research, this paper aims to uncover the network-structured dependence underlying a class of non-stationary multivariate point process models. To achieve this goal, we propose a novel continuous-time stochastic model for the CIFs. We formulate the associated theoretical framework and derive probabilistic properties that are essential for learning the statistical properties of the proposed penalized M-estimator for graph parameters. These parameters are crucial for identifying the causal relationships among nodes in the network. In our approach, we develop new technical tools, including the marked point process with explicit conditional distributions, recurrence time points, and cyclicity property. These tools prove instrumental in analyzing the probabilistic properties of a wide range of continuous-time models for point processes. Furthermore, they play a central role in the statistical learning of network structure.

Our proposed framework extends beyond the learning of interaction effects among nodes. It has the flexibility to incorporate other factors, such as autoregressive effects, experimental units, and other extrinsic conditions, into model (12).

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

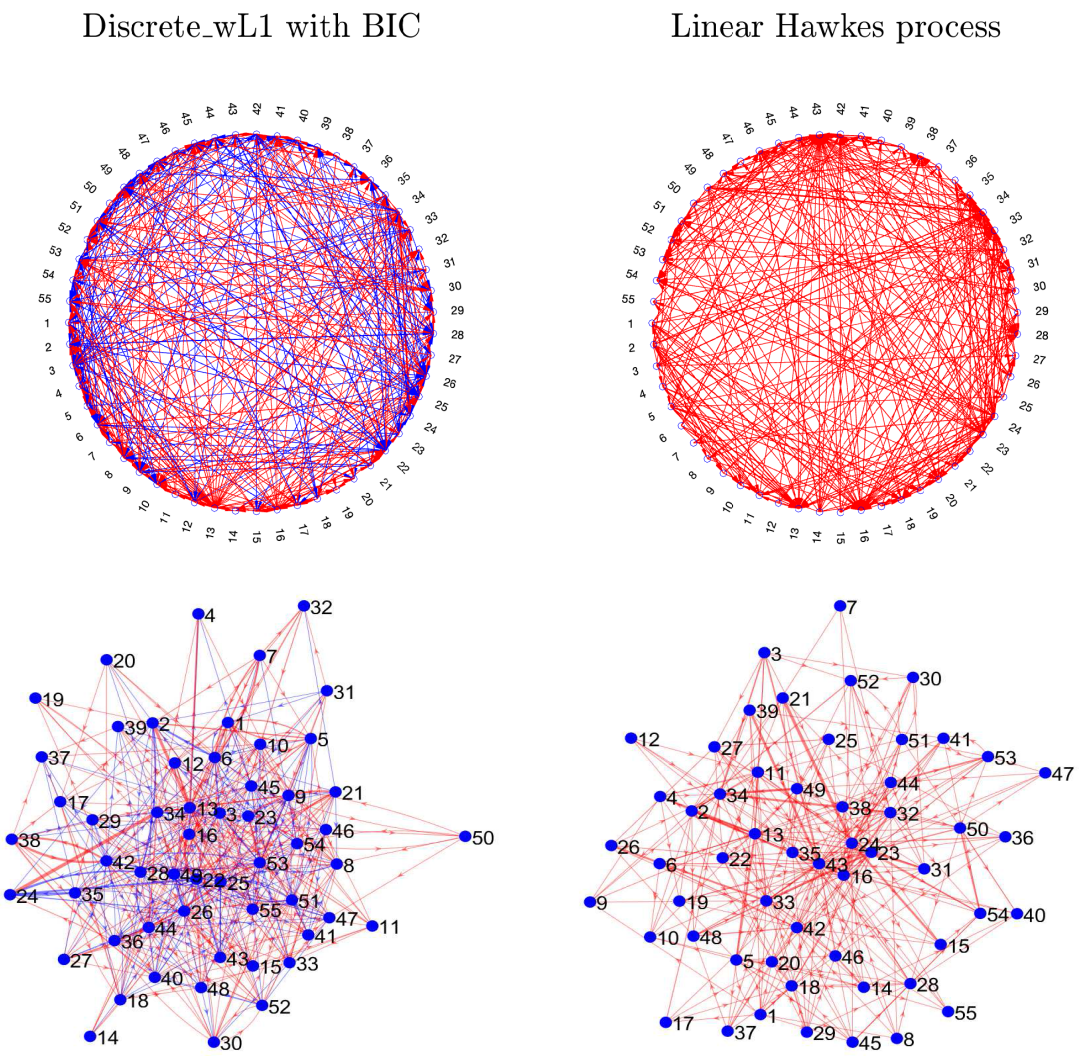

Fig. 6. **(Real Data: Estimated Networks Using Discrete-Time Modeling and Linear Hawkes Process Modeling Methods)** Red arrows represent excitatory effects, while blue arrows signify inhibitory effects; thicker arrows indicate stronger interactions. Top-left panel: Discrete-time modeling (plot in circular layout); bottom-left panel: Discrete-time modeling (plot in equilibrium layout); top-right panel: Linear Hawkes process (plot in circular layout); bottom-right panel: Linear Hawkes process (plot in equilibrium layout).

Furthermore, acknowledging potential variations in interaction time-lags among nodes, we can also allow the lag-width ϕ in the covariate $x_j(t)$, as illustrated in (16)–(17), to vary according to the node j. While these additional extensions hold potential for enhancing our understanding of complex systems, a comprehensive exploration of these aspects is beyond the scope of this paper. However, investigating these factors in future research would be valuable and could provide further insights into the dynamics of network structures.

APPENDIX A
PRACTICAL ISSUES AND SUPPLEMENTARY SIMULATIONS

A. Practical Issues on Selecting ϕ and $g(\cdot)$ in (16) and (17)

In practice, there are various methods for choosing the time-lag ϕ and the shape-function $g(\cdot)$, employing either prior knowledge or data-driven methods. Below, we provide some suggestions.

1) Selection of ϕ: The time-lag ϕ can be chosen in line with the number n of time bins with the bin-width T/n. Alternatively, our empirical choices fix $\phi = 1$ (a unit of time, also used in [15]). Moreover, ϕ could be selected using a data-driven algorithm as outlined below.

(1). Choose a sufficiently large $\phi_{\max}$, guided by domain knowledge or prior information.

(2). Iterate for each $k = 0, 1, 2, \ldots, k_{\max}$ using $\phi_k = \phi_{\max}\zeta^k$, where $\zeta \in (0, 1)$ represents a step length, and $k_{\max} > 1$ is the maximum iteration number. Obtain penalized M-estimators $\{\widehat{\widehat{\boldsymbol{\beta}}}_i^{(\phi_k)}\}_{i\in\mathcal{V}}$ by minimizing (52) with ϕ replaced by ϕ_k.

(3). Compute the joint negative log-likelihood $L_k = \sum_{i\in\mathcal{V}} \mathcal{L}_{i,\mathrm{T}}^{(k)}(\widehat{\widehat{\boldsymbol{\beta}}}_i^{(\phi_k)})$, where $\mathcal{L}_{i,\mathrm{T}}^{(k)}(\cdot)$ resembles $\mathcal{L}_{i,\mathrm{T}}(\cdot)$ in (48), replacing ϕ with ϕ_k.

(4). For $k = 0, 1, \ldots, k_{\max} - 1$, identify the first $k = \widehat{k}$ where $L_{\widehat{k}+1} > L_{\widehat{k}}$. Terminate the algorithm upon finding $\widehat{k}$. If $\widehat{k}$ is nonexistent, let $\widehat{k} = k_{\max}$. Our selected lag-width is $\phi_{\widehat{k}}$, and the corresponding estimators are $\{\widehat{\widehat{\boldsymbol{\beta}}}_i^{(\phi_{\widehat{k}})}\}_{i\in\mathcal{V}}$.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

TABLE VI

(SIMULATION STUDY: Network-1 WITH DATA-DRIVEN $\phi_{\widehat{k}}$) THE CONNECTION STRENGTH IS $\beta = 0.5$. THE CONTINUOUS_L1 AND CONTINUOUS_WL1 METHODS USE THE DATA-DRIVEN TIME-LAG $\phi_{\widehat{k}}$ FOLLOWING THE ALGORITHM IN APPENDIX A-A, WITH $\phi_{\max} = 3$, $\zeta = 0.7$, AND $k_{\max} = 8$. PARAMETER ESTIMATION INVOLVES $g(x) = \log(1 + x \wedge c)$, WITH DATA-DRIVEN c FROM (62). RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All		Detected_A		Detected_B		Correct_NC	
time length T =		500	1000	500	1000	500	1000	500	1000
Discrete_L1	bin=0.5	2.39 (0.15)	5.69 (0.16)	1.81 (0.12)	3.88 (0.11)	0.57 (0.07)	1.81 (0.08)	79.48 (0.08)	78.94 (0.11)
	bin=0.25	3.63 (0.16)	7.46 (0.14)	2.69 (0.11)	4.88 (0.09)	0.94 (0.08)	2.58 (0.09)	79.18 (0.10)	78.48 (0.14)
	bin=0.1	4.63 (0.17)	8.40 (0.12)	3.32 (0.11)	5.37 (0.07)	1.31 (0.09)	3.03 (0.08)	78.91 (0.14)	78.47 (0.13)
Continuous_L1		5.85 (0.19)	9.41 (0.08)	3.80 (0.13)	5.77 (0.05)	2.04 (0.10)	3.64 (0.06)	75.87 (0.24)	76.41 (0.21)
Discrete_wL1	bin=0.5	3.92 (0.15)	7.34 (0.14)	2.75 (0.12)	4.76 (0.10)	1.17 (0.08)	2.58 (0.09)	78.98 (0.10)	78.95 (0.11)
	bin=0.25	5.34 (0.16)	8.69 (0.11)	3.65 (0.11)	5.51 (0.06)	1.69 (0.09)	3.18 (0.08)	78.91 (0.12)	79.11 (0.10)
	bin=0.1	6.51 (0.15)	9.25 (0.08)	4.34 (0.10)	5.76 (0.04)	2.17 (0.10)	3.49 (0.07)	78.80 (0.12)	79.18 (0.09)
Continuous_wL1		6.50 (0.22)	9.38 (0.08)	4.18 (0.14)	5.82 (0.03)	2.31 (0.11)	3.56 (0.07)	78.19 (0.15)	79.01 (0.11)
Zhao_2012	bin=0.5	0.83 (0.08)	2.84 (0.16)	0.67 (0.07)	1.92 (0.12)	0.16 (0.03)	0.92 (0.07)	79.83 (0.04)	79.52 (0.07)
	bin=0.25	0.63 (0.06)	1.19 (0.09)	0.60 (0.06)	1.06 (0.08)	0.03 (0.01)	0.13 (0.03)	79.95 (0.01)	79.91 (0.03)
	bin=0.1	0.20 (0.04)	0.32 (0.05)	0.20 (0.04)	0.32 (0.05)	0.00 (0.00)	0.00 (0.00)	79.95 (0.01)	79.95 (0.02)
SIE-GLM	bin=0.5	2.00 (0.13)	5.12 (0.17)	1.63 (0.11)	3.67 (0.11)	0.38 (0.05)	1.45 (0.09)	79.69 (0.06)	79.20 (0.11)
	bin=0.25	1.81 (0.11)	4.63 (0.15)	1.61 (0.10)	3.51 (0.10)	0.20 (0.04)	1.12 (0.09)	79.68 (0.06)	79.39 (0.08)
	bin=0.1	0.82 (0.08)	2.50 (0.13)	0.80 (0.08)	2.27 (0.11)	0.02 (0.01)	0.22 (0.04)	79.94 (0.02)	79.68 (0.10)
Raj_2005	parent=3	9.68 (0.04)	9.97 (0.01)	5.85 (0.03)	5.99 (0.00)	3.83 (0.04)	3.98 (0.01)	60.15 (0.08)	63.78 (0.20)
true		10		6		4		80	

The algorithm primarily focuses on backtracking to determine ϕ, aiming to attain the highest likelihood value among all ϕ's for the corresponding estimators $\{\widehat{\widetilde{\boldsymbol{\beta}}}_i^{(\phi)}\}_{i\in\mathcal{V}}$. Table VI presents simulation results on Network-1 employing data-driven $\phi_{\widehat{k}}$ for estimation. It's evident that our methods, Continuous_L1 and Continuous_wL1, consistently outperform other methods, even without precise knowledge of the true time-lag ϕ.

2) Selection of $g(\cdot)$: We define $g(\cdot)$ as bounded functions, e.g.,

$$g(x) = \log(1 + x \wedge c), \quad \text{or} \quad g(x) = x \wedge c, \tag{61}$$

with a constant $c \in (0, \infty)$. For practical applications, we recommend the data-driven selection of c using:

$$c = \text{the 90th percentile of } \Big\{ \max_{t\in[0,\mathrm{T}]} \{N_j((t-\phi, t])/\phi\} : j \in \mathcal{V} \Big\}, \tag{62}$$

This approach ensures that the covariates $\{x_j(t)\}_{j\in\mathcal{V};t\in[0,\mathrm{T}]}$ closely reflect the empirical rates, maintaining numerical stability without increasing computational costs.

In practical demonstrations, we've included a simulation scenario in Table VII using the unbounded function $g(x) = \log(1 + x)$ for both generating synthetic data and estimating model parameters. These results demonstrate that even with an unbounded g, the proposed network modeling and recovery method remains effective.

Additionally, a simulation scenario has been added where both ϕ and $g(\cdot)$ are misspecified; the results are presented in Table VIII. These outcomes demonstrate that our proposed methods, Continuous_L1 and Continuous_wL1, consistently outperform other approaches, showcasing a certain level of robustness against misspecified ϕ and $g(\cdot)$.

B. Explicit Procedure for Implementing BIC Criterion in Simulation

In Method (i) of the simulation, we determine the tuning parameter η (or η_{T}) by minimizing the BIC function:

$$\mathrm{BIC}(\widehat{\widetilde{\boldsymbol{\beta}}}_i) = 2\mathcal{L}_{i,\mathrm{T}}(\widehat{\widetilde{\boldsymbol{\beta}}}_i) + \mathrm{df}(\widehat{\boldsymbol{\beta}}_i) \cdot \log(\mathrm{T})/\mathrm{T}.$$

Here, $\mathcal{L}_{i,\mathrm{T}}(\cdot)$ is defined in (48), and $\mathrm{df}(\widehat{\boldsymbol{\beta}}_i) = \sum_{j\in\mathcal{V};j\neq i} \mathrm{I}(\widehat{\beta}_{j,i} \neq 0)$ represents the count of non-zero elements in $\widehat{\boldsymbol{\beta}}_i$. Notably, $\mathrm{BIC}(\widehat{\widetilde{\boldsymbol{\beta}}}_i)$ is viewed as a function of η, given that $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ depends on η, denoted as $\widehat{\widetilde{\boldsymbol{\beta}}}_i = \widehat{\widetilde{\boldsymbol{\beta}}}_i^{(\eta)}$. We identify the minimizing value η of $\mathrm{BIC}(\widehat{\widetilde{\boldsymbol{\beta}}}_i)$ by exploiting the set of grid points $\{\eta_{\max} h^k : k = 0, 1, \ldots, 12\}$, where $\eta_{\max} = \sup\{\eta : \mathrm{df}(\widehat{\widetilde{\boldsymbol{\beta}}}_i^{(\eta)}) > 0\}$, and $h \in (0, 1)$ represents a constant. Specifically, we employ $h = 0.7$ across all our numerical experiments.

C. Incorporating DAG Constraint

As previously mentioned in Section VI-A, our Network-1 is a directed acyclic graph (DAG) intended to depict the information flow transferred among neurons. From a causal inference standpoint, this DAG configuration naturally represents a Granger causal graph [49], and learning this DAG plays a vital role in causal recovery. Motivated by this aspect, we further develop a DAG-informed network learning method by adding a DAG structural constraint into our penalized M-estimation scheme (52). Let $B = (\beta_{j,i})_{V\times V} \in \mathbb{R}^{V\times V}$ be the weighted adjacency matrix, and let $\boldsymbol{\beta}_0 = (\beta_{0;1}, \ldots, \beta_{0;V})^\top \in \mathbb{R}^V$ be the vector of baseline parameters. A DAG $\widehat{B}$ is recovered by solving the constrained minimization problem:

$$(\widehat{B}, \widehat{\boldsymbol{\beta}}_0) = \arg\min_{B\in\mathbb{R}^{V\times V};\boldsymbol{\beta}_0\in\mathbb{R}^V} \{\mathcal{L}_{\mathrm{T}}(B, \boldsymbol{\beta}_0) + \mathcal{P}_{\mathrm{T}}(B)\}, \quad \text{subject to } B \text{ representing a DAG},$$

where $\mathcal{L}_{\mathrm{T}}(B, \boldsymbol{\beta}_0) = \sum_{i\in\mathcal{V}} \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ is the loss function, and $\mathcal{P}_{\mathrm{T}}(B) = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{V}\setminus\{i\}} w_{j,i,\mathrm{T}} |\beta_{j,i}|$ is the

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

TABLE VII

(SIMULATION STUDY: Network-1 WITH UNBOUNDED $g(\cdot)$) THE CONNECTION STRENGTH IS $\beta = 0.5$. WE USE $g(x) = \log(1 + x)$ IN BOTH SYNTHETIC DATA GENERATION AND PARAMETER ESTIMATION. RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All		Detected_A		Detected_B		Correct_NC	
time length T =		500	1000	500	1000	500	1000	500	1000
Discrete_L1	bin=0.5	2.39 (0.15)	5.69 (0.16)	1.81 (0.12)	3.88 (0.11)	0.57 (0.07)	1.81 (0.08)	79.48 (0.08)	78.94 (0.11)
	bin=0.25	3.63 (0.16)	7.46 (0.14)	2.69 (0.11)	4.88 (0.09)	0.94 (0.08)	2.58 (0.09)	79.18 (0.10)	78.48 (0.14)
	bin=0.1	4.63 (0.17)	8.40 (0.12)	3.32 (0.11)	5.37 (0.07)	1.31 (0.09)	3.03 (0.08)	78.91 (0.14)	78.47 (0.13)
Continuous_L1		6.56 (0.14)	9.44 (0.07)	4.24 (0.09)	5.80 (0.04)	2.31 (0.09)	3.65 (0.05)	76.58 (0.21)	76.48 (0.21)
Discrete_wL1	bin=0.5	3.92 (0.15)	7.34 (0.14)	2.75 (0.12)	4.76 (0.10)	1.17 (0.08)	2.58 (0.09)	78.98 (0.10)	78.95 (0.11)
	bin=0.25	5.34 (0.16)	8.69 (0.11)	3.65 (0.11)	5.51 (0.06)	1.69 (0.09)	3.18 (0.08)	78.91 (0.12)	79.11 (0.10)
	bin=0.1	6.51 (0.15)	9.25 (0.08)	4.34 (0.10)	5.76 (0.04)	2.17 (0.10)	3.49 (0.07)	78.80 (0.12)	79.18 (0.09)
Continuous_wL1		7.44 (0.14)	9.52 (0.06)	4.80 (0.09)	5.90 (0.03)	2.64 (0.08)	3.62 (0.06)	78.67 (0.11)	79.03 (0.10)
Zhao_2012	bin=0.5	0.83 (0.08)	2.84 (0.16)	0.67 (0.07)	1.92 (0.12)	0.16 (0.03)	0.92 (0.07)	79.83 (0.04)	79.52 (0.07)
	bin=0.25	0.63 (0.06)	1.19 (0.09)	0.60 (0.06)	1.06 (0.08)	0.03 (0.01)	0.13 (0.03)	79.95 (0.01)	79.91 (0.03)
	bin=0.1	0.20 (0.04)	0.32 (0.05)	0.20 (0.04)	0.32 (0.05)	0.00 (0.00)	0.00 (0.00)	79.95 (0.01)	79.95 (0.02)
SIE-GLM	bin=0.5	2.00 (0.13)	5.12 (0.17)	1.63 (0.11)	3.67 (0.11)	0.38 (0.05)	1.45 (0.09)	79.69 (0.06)	79.20 (0.11)
	bin=0.25	1.81 (0.11)	4.63 (0.15)	1.61 (0.10)	3.51 (0.10)	0.20 (0.04)	1.12 (0.09)	79.68 (0.06)	79.39 (0.08)
	bin=0.1	0.82 (0.08)	2.50 (0.13)	0.80 (0.08)	2.27 (0.11)	0.02 (0.01)	0.22 (0.04)	79.94 (0.02)	79.68 (0.10)
Raj_2005	parent=3	9.68 (0.04)	9.97 (0.01)	5.85 (0.03)	5.99 (0.00)	3.83 (0.04)	3.98 (0.01)	60.15 (0.08)	63.78 (0.20)
true		10		6		4		80	

TABLE VIII

(SIMULATION STUDY: Network-1 WITH MISSPECIFIED ϕ AND $g(\cdot)$)THE CONNECTION STRENGTH IS $\beta = 0.5$. IN THE TRUE MODEL, THE TIME-LAG IS$\phi = 1$, AND THE SHAPE-FUNCTION IS $g(x) = \log(1 + x \wedge 10)$. IN THE ESTIMATION PROCESS, A MISSPECIFIED $\phi_a = 0.5$ IS USED ALONGSIDE $g_a(x) = x \wedge c$, INCORPORATING THE DATA-DRIVEN c FROM (62). RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

		Correct_All		Detected_A		Detected_B		Correct_NC	
time length T =		500	1000	500	1000	500	1000	500	1000
Discrete_L1	bin=0.5	2.12 (0.13)	4.68 (0.17)	1.67 (0.11)	3.40 (0.12)	0.45 (0.06)	1.28 (0.09)	79.19 (0.11)	79.59 (0.07)
	bin=0.25	2.00 (0.12)	4.73 (0.16)	1.57 (0.10)	3.40 (0.11)	0.44 (0.05)	1.33 (0.09)	79.36 (0.09)	79.67 (0.05)
	bin=0.1	2.08 (0.12)	4.97 (0.17)	1.60 (0.10)	3.54 (0.12)	0.48 (0.05)	1.43 (0.09)	79.25 (0.10)	79.63 (0.06)
Continuous_L1		4.30 (0.15)	6.75 (0.17)	2.99 (0.11)	4.22 (0.12)	1.32 (0.09)	2.52 (0.09)	76.90 (0.21)	75.56 (0.21)
Discrete_wL1	bin=0.5	3.43 (0.15)	6.47 (0.15)	2.44 (0.11)	4.28 (0.11)	0.99 (0.08)	2.19 (0.09)	78.98 (0.10)	79.19 (0.09)
	bin=0.25	3.28 (0.16)	6.40 (0.15)	2.24 (0.12)	4.23 (0.11)	1.04 (0.08)	2.17 (0.10)	79.09 (0.10)	79.06 (0.09)
	bin=0.1	3.37 (0.15)	6.43 (0.16)	2.29 (0.11)	4.21 (0.10)	1.08 (0.08)	2.22 (0.10)	79.23 (0.08)	78.93 (0.11)
Continuous_wL1		3.37 (0.16)	6.37 (0.16)	2.20 (0.12)	4.16 (0.11)	1.17 (0.07)	2.21 (0.09)	79.06 (0.09)	78.91 (0.11)
Zhao_2012	bin=0.5	1.50 (0.12)	3.99 (0.18)	1.12 (0.10)	2.73 (0.13)	0.38 (0.05)	1.26 (0.09)	79.28 (0.09)	79.70 (0.06)
	bin=0.25	0.66 (0.07)	1.35 (0.09)	0.61 (0.07)	1.11 (0.08)	0.05 (0.02)	0.24 (0.04)	79.92 (0.03)	79.94 (0.02)
	bin=0.1	0.23 (0.04)	0.27 (0.04)	0.23 (0.04)	0.27 (0.04)	0.00 (0.00)	0.00 (0.00)	79.95 (0.02)	79.92 (0.03)
SIE-GLM	bin=0.5	1.89 (0.13)	4.47 (0.18)	1.51 (0.11)	3.25 (0.13)	0.38 (0.05)	1.22 (0.09)	79.55 (0.07)	79.78 (0.04)
	bin=0.25	1.22 (0.10)	3.22 (0.14)	1.09 (0.09)	2.54 (0.11)	0.13 (0.03)	0.68 (0.07)	79.75 (0.06)	79.84 (0.04)
	bin=0.1	0.53 (0.06)	1.49 (0.09)	0.53 (0.06)	1.38 (0.09)	0.00 (0.00)	0.11 (0.03)	79.84 (0.03)	79.89 (0.03)
Raj_2005	parent=3	8.39 (0.11)	9.68 (0.05)	5.20 (0.07)	5.84 (0.03)	3.19 (0.07)	3.84 (0.04)	64.48 (0.21)	59.41 (0.15)
true		10		6		4		80	

weighted-L_1 penalty. The above optimization problem can be effectively solved by using the method in [50]. Table IX shows some preliminary simulation results for Network-1, comparing the DAG-constrained method with the non-constrained continuous-modeling Method (i). For both methods, we adopt two scenarios for the penalty functions: the L_1-penalty and weighted-L_1 penalty. Under each setting of penalty function, the DAG-constrained method achieves better overall performance than the non-constrained method. This indicates that adding the DAG constraint can effectively enhance the network recovery accuracy if it is known that the true network is a DAG.

TABLE IX

(SIMULATION STUDY: Network-1 WITH COMPARISONS TO DAG-CONSTRAINED METHOD)THE TIME LENGTH IS T = 1000. THE CONNECTION STRENGTH IS 0.5. FOR ALL METHODS, PARAMETER ESTIMATION INVOLVES $g(x) = \log(1 + x \wedge c)$, WITH DATA-DRIVEN c FROM (62). RESULTS ARE AVERAGED OVER 100 REPLICATIONS, WITH STANDARD ERRORS DENOTED IN PARENTHESES

	Correct_All	Detected_A	Detected_B	Correct_NC
Continuous_L1	9.49 (0.06)	5.81 (0.04)	3.68 (0.05)	76.47 (0.21)
Continuous_L1+DAG	9.56 (0.07)	5.91 (0.03)	3.65 (0.06)	79.01 (0.12)
Continuous_wL1	9.53 (0.06)	5.90 (0.03)	3.64 (0.05)	79.03 (0.10)
Continuous_wL1+DAG	9.66 (0.05)	5.91 (0.02)	3.75 (0.05)	79.02 (0.10)
true	10	6	4	80

APPENDIX B
PROOFS OF MAIN RESULTS

A. Notations in the Proof

For an event A in the sample space Ω, the event $\overline{A}$ denotes the complement of A. For two events A and B, the event $A \setminus B$ denotes $A \cap \overline{B}$. For an event A, we write $\sigma(\mathscr{F}, A) = \sigma(\mathscr{F}, \mathrm{I}(A))$. Let $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Let $\mathbf{C} \succ 0$ denote a positive definite matrix $\mathbf{C}$. For a vector $\boldsymbol{b} = (b_1, \ldots, b_d)^\top$, $\|\boldsymbol{b}\|_1 = \sum_{j=1}^d |b_j|$, and $\|\boldsymbol{b}\| = \|\boldsymbol{b}\|_2 = (\sum_{j=1}^d b_j^2)^{1/2}$.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

B. Conditions

The conditions aren't the weakest possible but are conducive to the derivations.

A1. The number of nodes $V \geq 2$ is a fixed integer. In the multivariate point process, event time points $\{T_{i,\ell}\}_{i\in\mathcal{V},\ell\geq 1}$ satisfy $0 < T_{i,1} < T_{i,2} < \cdots$ for each $i \in \mathcal{V}$.

A2. The multivariate counting process satisfies $\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(N_i(t+\Delta) = N_i(t)+1 \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t)$ a.s. for every $i \in \mathcal{V}$ and $t \geq 0$.

A3. The multivariate counting process $\boldsymbol{N}(t)$ satisfies the OM condition as defined in Definition 1.

A4. There exists a random variable $Z > 0$ with $\mathrm{E}(Z) < \infty$, such that for any constant $\Delta \in (0, c_0)$ where $c_0 \in (0,1)$, and any $t \geq 0$, $\mathrm{P}(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t)/\Delta \leq Z$ a.s..

A5. In (16), the shape-function $g(\cdot) : [0,\infty) \to [0,\infty)$ is continuous, non-negative, monotonically increasing, bounded above, with $g(0) = 0$, and $\sup_{x\in[0,\infty)} g(x) \leq C_0$ for some constant $C_0 \in (0,\infty)$.

A6. For all $i \in \mathcal{V}$, the true self-effect parameter $\beta^*_{i,i} = 0$.

A7. The true edge set $\mathcal{E}^* \neq \varnothing$.

A8. The edge set $\mathcal{E}$ in (13) satisfies $\mathcal{E} \neq \varnothing$.

Condition A1 pertains to the basic definition of a multivariate point process. Condition A2 aligns with the regular point process as defined in [30] and is explicitly discussed in our Remark 1. Condition A3 is explicitly presented in our Definition 1. Condition A4 resembles conditions (2)–(3) in [30], ensuring the applicability of the dominated convergence theorem. Condition A5 guarantees the boundedness property of the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i\in\mathcal{V}}$ in our model (12). Condition A6 excludes self-effects in model (12), preventing the presence of a 'self-loop' in the corresponding network structure $\mathcal{G}$. Conditions A7 and A8 are imposed to ensure that our multivariate point process does not reduce to the trivial case of a homogeneous Poisson process.

C. Proof of the Statement in Remark 1

We aim to prove the statement: any *multivariate regular point process* also has identical limits (9) and (10).

From the definition of a multivariate regular point process $\boldsymbol{N}(t)$, we have: $\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(N_i(t+\Delta) = N_i(t)+1 \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t)$, a.s., for any $i \in \mathcal{V}$ and $t \geq 0$. Since $\{N_i(t+\Delta) = N_i(t)+1\} \subseteq \{N_i(t+\Delta) \neq N_i(t)\}$, we further get:

$$\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}\Big(\{N_i(t+\Delta) \neq N_i(t)\} \setminus \{N_i(t+\Delta) = N_i(t)+1\} \,\Big|\, \mathscr{F}_t\Big) = 0, \quad \text{a.s..} \tag{63}$$

Thus,

$$\begin{aligned} 0 &\leq \lim_{\Delta\downarrow 0}\Delta^{-1}\Big[\mathrm{P}\big(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t\big) - \mathrm{P}\Big(\bigcup_{i\in\mathcal{V}}\{N_i(t+\Delta) = N_i(t)+1\} \,\Big|\, \mathscr{F}_t\Big)\Big] \\ &\leq \sum_{i\in\mathcal{V}}\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}\Big(\{N_i(t+\Delta) \neq N_i(t)\} \setminus \{N_i(t+\Delta) = N_i(t)+1\} \,\Big|\, \mathscr{F}_t\Big) = 0, \quad \text{a.s.,} \end{aligned}$$

where the last equality is from (63). Hence, we obtain: $\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(\cup_{i\in\mathcal{V}}\{N_i(t+\Delta) = N_i(t)+1\} \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t)$, a.s.. This completes the proof. □

D. Proof of Lemma 1

Define the events $A_{i,\Delta} = \{N_i(t+\Delta) = N_i(t)+1\}$. Then:

$$\mathrm{P}\Big(\bigcup_{i\in\mathcal{V}}\{N_i(t+\Delta) = N_i(t)+1\} \,\Big|\, \mathscr{F}_t\Big) = \mathrm{P}\Big(\bigcup_{i\in\mathcal{V}} A_{i,\Delta} \,\Big|\, \mathscr{F}_t\Big).$$

By the inclusion-exclusion formula, we have that

$$\begin{aligned} &\mathrm{P}\Big(\bigcup_{i\in\mathcal{V}} A_{i,\Delta} \,\Big|\, \mathscr{F}_t\Big) \\ &= \sum_{i=1}^{V}\mathrm{P}(A_{i,\Delta} \mid \mathscr{F}_t) - \sum_{k=2}^{V}(-1)^k \sum_{\{i_1,\ldots,i_k\}\subseteq\mathcal{V}} \mathrm{P}\Big(\bigcap_{j\in\{i_1,\ldots,i_k\}} A_{j,\Delta} \,\Big|\, \mathscr{F}_t\Big). \end{aligned} \tag{64}$$

For mutually distinct $\{i_1,\ldots,i_k\}$ with $k \geq 2$, the OM condition (11) implies:

$$\begin{aligned} &\mathrm{P}\Big(\bigcap_{j\in\{i_1,\ldots,i_k\}} A_{j,\Delta} \,\Big|\, \mathscr{F}_t\Big) \leq \mathrm{P}(A_{i_1,\Delta} \cap A_{i_2,\Delta} \mid \mathscr{F}_t) \\ &= \Delta^2\{\lambda_{i_1}(t \mid \mathscr{F}_t)\,\lambda_{i_2}(t \mid \mathscr{F}_t) + o(1)\}, \quad \text{a.s.} \end{aligned} \tag{65}$$

as $\Delta \downarrow 0$. Plugging (65) into (64), we obtain:

$$\begin{aligned} \frac{1}{\Delta}\mathrm{P}\Big(\bigcup_{i\in\mathcal{V}} A_{i,\Delta} \,\Big|\, \mathscr{F}_t\Big) &= \frac{1}{\Delta}\sum_{i=1}^{V}\mathrm{P}(A_{i,\Delta} \mid \mathscr{F}_t) + O(\Delta) \\ &= \sum_{i=1}^{V}\lambda_i(t \mid \mathscr{F}_t) + o(1), \quad \text{a.s.} \end{aligned}$$

as $\Delta \downarrow 0$. It follows that $\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t) = \lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(\cup_{i\in\mathcal{V}} A_{i,\Delta} \mid \mathscr{F}_t) = \sum_{i=1}^{V}\lambda_i(t \mid \mathscr{F}_t)$. This completes the proof. □

E. Proof of Lemma 4

Let's first establish part (i). Utilizing [37] (Theorem 2.4.7, p. 84) and given that $\lambda_j(t \mid \mathscr{F}_t)$ in (12) is finite, our event time points $\{T_{j,\ell}\}_{j\in\mathcal{V},\ell\geq 1}$ are *totally inaccessible stopping times*. This, combined with [37] (Proposition 2.4.6, p. 83), indicates that $\mathrm{P}(T_{i,k} = T_{j,r}) = 0$, for any distinct nodes $i, j \in \mathcal{V}$, and any integers $k \geq 1$ and $r \geq 1$. Consequently:

$$\begin{aligned} &\mathrm{P}\big(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1},\, \breve{T}_\ell \in \{T_{j,k}\}_{k\geq 1}\big) \\ &\leq \mathrm{P}\big(\cup_{k\geq 1}\cup_{r\geq 1}\{T_{i,k} = T_{j,r}\}\big) = 0. \end{aligned} \tag{66}$$

Now, to prove part (ii), we use a similar reasoning as in (66). For any $i \in \mathcal{V}$, we have:

$$\begin{aligned} &\mathrm{P}\big(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1},\, \breve{T}_\ell \in \{T_{j,k}+\phi\}_{k\geq 1}\big) \\ &\leq \mathrm{P}\big(\cup_{k\geq 1}\cup_{r\geq 1}\{T_{i,k} = T_{j,r}+\phi\}\big) = 0 \end{aligned}$$

for any $j \in \mathcal{V}$, and thus

$$\begin{aligned} &\mathrm{P}\big(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1},\, \breve{T}_\ell \in \{T_{j,k}+\phi\}_{j\in\mathcal{V},\,k\geq 1}\big) \\ &\leq \sum_{j\in\mathcal{V}}\mathrm{P}\big(\breve{T}_\ell \in \{T_{i,k}\}_{k\geq 1},\, \breve{T}_\ell \in \{T_{j,k}+\phi\}_{k\geq 1}\big) = 0. \end{aligned}$$

The proof is completed. □

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

F. Proof of Theorem 1

Before proving Theorem 1, we first establish Lemmas B.1 and B.2 based on Definition 4.

Definition 4 ($\mathrm{E}(X \parallel \mathscr{F}, A)$): Let $(\Omega, \mathscr{G}, \mathrm{P})$ represent a probability space. Consider $\mathscr{F} \subseteq \mathscr{G}$ as a sub σ-field, and $A \in \mathscr{G}$ as an event where $A \notin \mathscr{F}$ and $\mathrm{P}(A \mid \mathscr{F}) > 0$ almost surely. For any random variable X in $(\Omega, \mathscr{G}, \mathrm{P})$, define

$$\mathrm{E}(X \parallel \mathscr{F}, A) = \mathrm{E}\{X\,\mathrm{I}(A) \mid \mathscr{F}\}/\mathrm{P}(A \mid \mathscr{F}). \tag{67}$$

(Remark: when $\mathscr{F} = \{\Omega, \varnothing\}$, the expression $\mathrm{E}(X \parallel \mathscr{F}, A)$ in (67) simplifies to $\mathrm{E}(X \mid A) = \mathrm{E}\{X\,\mathrm{I}(A)\}/\mathrm{P}(A)$; when X is independent of both $\mathscr{F}$ and A, the term $\mathrm{E}(X \parallel \mathscr{F}, A)$ in (67) simplifies to $\mathrm{E}(X)$.)

Lemma B.1 (Conditional Probability $\mathrm{P}(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s)$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Then for $t \geq s \geq 0$,

$$\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big) = \exp\Big\{ - \int_s^t \lambda^{\mathrm{sum}}(u; \mathscr{F}_s)\,\mathrm{d}u \Big\}, \tag{68}$$

with

$$\lambda^{\mathrm{sum}}(t; \mathscr{F}_s) = \mathrm{E}\{\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t) \parallel \mathscr{F}_s, \boldsymbol{N}(t) = \boldsymbol{N}(s)\}, \tag{69}$$

where $\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t)$ denotes the total CIF in (9) and (10). (Remark: $\lambda^{\mathrm{sum}}(t; \mathscr{F}_s)$ in (69) is inspired by the definition $\lambda_{N(t)}(t, \mathscr{B}_s)$ in Lemma 1 of [30], and (68) is motivated by Corollary 1 in [30].)

Proof: For $t \geq s \geq 0$ and $\Delta > 0$, note that

$$\begin{aligned}\{\boldsymbol{N}(t+\Delta) = \boldsymbol{N}(s)\} &= \{\boldsymbol{N}((s, t+\Delta]) = \boldsymbol{0}\} \\ &\subseteq \{\boldsymbol{N}((s,t]) = \boldsymbol{0}\} = \{\boldsymbol{N}(t) = \boldsymbol{N}(s)\},\end{aligned}$$

which implies

$$\begin{aligned}&\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big) - \mathrm{P}\big(\boldsymbol{N}(t+\Delta) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big) \\ &\quad = \mathrm{P}\big(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t),\, \boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big).\end{aligned}$$

Combining this with the fact that $\mathscr{F}_s \subseteq \mathscr{F}_t$, and using (10), (67), (69), we obtain

$$\begin{aligned}&\partial\,\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big)/\partial t \\ &\quad = -\lim_{\Delta \downarrow 0} \Delta^{-1} \mathrm{E}\big\{\mathrm{P}\big(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t\big) \\ &\qquad\qquad \times \mathrm{I}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s)\big) \mid \mathscr{F}_s\big\} \\ &\quad = -\mathrm{E}\Big\{ \lim_{\Delta \downarrow 0} \Delta^{-1} \mathrm{P}\big(\boldsymbol{N}(t+\Delta) \neq \boldsymbol{N}(t) \mid \mathscr{F}_t\big) \\ &\qquad\qquad \times \mathrm{I}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s)\big) \mid \mathscr{F}_s \Big\} \qquad (70) \\ &\quad = -\lambda^{\mathrm{sum}}(t; \mathscr{F}_s) \cdot \mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big),\end{aligned}$$

where the interchange of limit and expectation in (70) follows from the dominated convergence theorem and condition A4. Solving the resulting differential equation completes the proof of (68). □

Lemma B.2 ($\mathrm{E}\{\lambda_i(S \mid \mathcal{F}_S) \parallel \mathscr{F}_s, \boldsymbol{N}(S) = \boldsymbol{N}(s)\}$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Define $\mathcal{T}_s = \cup_{j \in \mathcal{V}} \{t \in (s, s+\phi] : N_j(\{t-\phi\}) = 1\}$, and

$$T_s^{\mathrm{o}} = \begin{cases} \min(\mathcal{T}_s), & \text{if } \mathcal{T}_s \neq \varnothing, \\ \infty, & \text{if } \mathcal{T}_s = \varnothing, \end{cases} \tag{71}$$

for a fixed time point $s \in [0, \infty)$. Then, for any stopping time S satisfying $s \leq S < T_s^{\mathrm{o}}$, we have

$$\mathrm{E}\big\{\lambda_i(S \mid \mathcal{F}_S) \parallel \mathscr{F}_s, \boldsymbol{N}(S) = \boldsymbol{N}(s)\big\} = \lambda_i\big(s \mid \mathscr{F}_s\big), \tag{72}$$

for all $i \in \mathcal{V}$.

Proof: For $S = s$, (72) obviously holds. It suffices to prove (72) for $s < S < T_s^{\mathrm{o}}$. First, we establish:

$$\begin{aligned}&\text{for } s < S < T_s^{\mathrm{o}},\ \boldsymbol{N}(S) = \boldsymbol{N}(s) \text{ implies} \\ &\quad \cup_{j \in \mathcal{V}} \big\{t \in (s, S] : N_j(\{t\}) = 1\big\} = \varnothing \\ &\text{and } \cup_{j \in \mathcal{V}} \big\{t \in (s, S] : N_j(\{t-\phi\}) = 1\big\} = \varnothing.\end{aligned} \tag{73}$$

Let $B_1 = \cup_{j \in \mathcal{V}} \{t \in (s, S] : N_j(\{t\}) = 1\}$ and $B_2 = \cup_{j \in \mathcal{V}} \{t \in (s, S] : N_j(\{t-\phi\}) = 1\}$. Note that $\boldsymbol{N}(S) = \boldsymbol{N}(s)$ directly implies $B_1 = \varnothing$. To prove (73), it suffices to show $B_2 = \varnothing$, which is divided into cases whether $\mathcal{T}_s \neq \varnothing$ or not.

If $\mathcal{T}_s \neq \varnothing$, then (71) implies $s < \min(\mathcal{T}_s)$, indicating $\cup_{j \in \mathcal{V}} \{t \in (s, \min(\mathcal{T}_s)) : N_j(\{t-\phi\}) = 1\} = \varnothing$. This, along with $s < S < \min(\mathcal{T}_s)$, concludes $B_2 = \varnothing$.

If $\mathcal{T}_s = \varnothing$, we obtain:

$$\begin{aligned}B_2 &\subseteq \big\{ \cup_{j \in \mathcal{V}} \{t \in (s, s+\phi] : N_j(\{t-\phi\}) = 1\}\big\} \\ &\quad \cup \big\{ \cup_{j \in \mathcal{V}} \{t \in ((s+\phi) \wedge S, S] : N_j(\{t-\phi\}) = 1\}\big\} \\ &\subseteq \mathcal{T}_s \cup \{B_1 + \phi\} \\ &= \varnothing,\end{aligned}$$

where $B_1 + \phi = \{t + \phi : t \in B_1\}$.

Combining both cases confirms (73).

Next, we prove Lemma B.2. If $\boldsymbol{N}(S) = \boldsymbol{N}(s)$, then (73) indicates that the CIFs $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i \in \mathcal{V}}$ are continuous in $(s, S]$. For piecewise-constant functions $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i \in \mathcal{V}}$, if $\boldsymbol{N}(S) = \boldsymbol{N}(s)$, then $\lambda_i(S \mid \mathcal{F}_S) = \lambda_i(s \mid \mathscr{F}_s)$, which implies:

$$\lambda_i(S \mid \mathcal{F}_S) \cdot \mathrm{I}(A) = \lambda_i(s \mid \mathscr{F}_s) \cdot \mathrm{I}(A), \quad i \in \mathcal{V},$$

where A denotes the event $\{\boldsymbol{N}(S) = \boldsymbol{N}(s)\}$. Combining this with (67), we obtain:

$$\begin{aligned}\mathrm{E}\big\{\lambda_i(S \mid \mathcal{F}_S) \parallel \mathscr{F}_s,\, A\big\} &= \frac{\mathrm{E}\{\lambda_i(S \mid \mathcal{F}_S) \cdot \mathrm{I}(A) \mid \mathscr{F}_s\}}{\mathrm{P}(A \mid \mathscr{F}_s)} \\ &= \frac{\mathrm{E}\{\lambda_i(s \mid \mathscr{F}_s) \cdot \mathrm{I}(A) \mid \mathscr{F}_s\}}{\mathrm{P}(A \mid \mathscr{F}_s)} \\ &= \lambda_i(s \mid \mathscr{F}_s).\end{aligned}$$

This completes the proof. □

Now, we prove Theorem 1. We start by demonstrating part (i). As per the definition in (22), $\breve{T}_\ell < \breve{T}_{\ell+1}$ holds for any integer $\ell \geq 1$. It suffices to prove $\breve{T}_{\ell+1} \leq T_\ell^*$. If $\mathcal{T}_\ell = \varnothing$, then $T_\ell^* = \infty$ in (27), which completes the proof. If $\mathcal{T}_\ell \neq \varnothing$, then any $t_\ell \in \mathcal{T}_\ell$ in (26) indicates that $t_\ell = T_{i,k} + \phi$ for some integers $i \in \mathcal{V}$ and $k \geq 1$, and $\breve{T}_\ell < t_\ell \equiv T_{i,k} + \phi \leq \breve{T}_\ell + \phi$. Also, $t_\ell \equiv T_{i,k} + \phi \in \{\breve{T}_1, \breve{T}_2, \ldots\}$. This, combined with $\breve{T}_\ell < \breve{T}_{\ell+1}$, implies $\breve{T}_{\ell+1} \leq t_\ell$. Hence, $\breve{T}_{\ell+1} \leq \min\{t_\ell : t_\ell \in \mathcal{T}_\ell\} = \min(\mathcal{T}_\ell) = T_\ell^*$.

Before proving parts (ii) and (iii), let's prepare (a), (b), (c), and (d) below.

(a) Since $\mathcal{F}_{\breve{T}_\ell} = \sigma\{(\breve{T}_0, I_0), \ldots, (\breve{T}_\ell, I_\ell)\}$, it suffices to demonstrate that (28)–(30) hold conditional on each realization

$$\begin{aligned}&\Big\{\{(\breve{T}_0, I_0), \ldots, (\breve{T}_\ell, I_\ell)\} = \{(\breve{t}_0, i_0), \ldots, (\breve{t}_\ell, i_\ell)\}\Big\} \\ &\quad = \bullet \in \mathcal{F}_{\breve{T}_\ell}.\end{aligned} \tag{74}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Note that the realization $\bullet$ in (74) is known from the history up to time $\breve{t}_\ell$. Following the notation $\mathscr{F}_t$ in (5), we also have $\bullet \in \mathscr{F}_{\breve{t}_\ell}$, and thus for a random variable $X : \Omega \to \mathbb{R}$, which is measurable with respect to either $\mathcal{F}_{\breve{T}_\ell}$ or $\mathscr{F}_{\breve{t}_\ell}$, denote by $X(\bullet)$ the value of X at the realization $\bullet$. For example, we can write $\breve{T}_\ell(\bullet) = \breve{t}_\ell$ and $I_\ell(\bullet) = i_\ell$.

(b) Comparing the random variables T_ℓ^* in (27) and $T^{\mathrm{o}}_{\breve{t}_\ell}$ in (71) (with $s = \breve{t}_\ell$), we observe that they have the same value t_ℓ^* at the realization $\bullet$, i.e.,

$$t_\ell^* = T_\ell^*(\bullet) = T^{\mathrm{o}}_{\breve{t}_\ell}(\bullet), \qquad \text{and} \quad \breve{t}_\ell < t_\ell^*.$$

(c) Also, we verify the following equation for $t \in (\breve{t}_\ell, t_\ell^*)$:

$$\begin{aligned}&\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell}\big)(\bullet)\\ &\quad = \exp\{-\lambda^{\mathrm{sum}}(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell})(\bullet)\cdot(t-\breve{t}_\ell)\}. \end{aligned} \tag{75}$$

Using (68) (with $s = \breve{t}_\ell$) yields:

$$\begin{aligned}&\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell}\big)\\ &\quad = \exp\Big\{-\int_{\breve{t}_\ell}^{t} \lambda^{\mathrm{sum}}\big(u; \mathscr{F}_{\breve{t}_\ell}\big)\,\mathrm{d}u\Big\}, \text{ for } t > \breve{t}_\ell. \end{aligned} \tag{76}$$

Since both sides of (76) are $\mathscr{F}_{\breve{t}_\ell}$-measurable random variables, it follows that:

$$\begin{aligned}&\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell}\big)(\bullet)\\ &\quad = \exp\Big\{-\int_{\breve{t}_\ell}^{t} \lambda^{\mathrm{sum}}(u; \mathscr{F}_{\breve{t}_\ell})(\bullet)\,\mathrm{d}u\Big\}, \text{ for } t > \breve{t}_\ell. \end{aligned} \tag{77}$$

For $t \in (\breve{t}_\ell, t_\ell^*)$, define a random variable S such that $S(\bullet) = t$ and $\breve{t}_\ell \le S < T^{\mathrm{o}}_{\breve{t}_\ell}$ (e.g., if $T_\ell^* < \infty$, then let $S = \eta\breve{t}_\ell + (1-\eta)T^{\mathrm{o}}_{\breve{t}_\ell}$, with $\eta = (t_\ell^* - t)/(t_\ell^* - \breve{t}_\ell) \in (0,1)$; if $T_\ell^* = \infty$, then let $S = \breve{t}_\ell \cdot \mathrm{I}(T^{\mathrm{o}}_{\breve{t}_\ell} < \infty) + t \cdot \mathrm{I}(T^{\mathrm{o}}_{\breve{t}_\ell} = \infty)$, where $S(\bullet) = t$ due to $T^{\mathrm{o}}_{\breve{t}_\ell}(\bullet) = t_\ell^*$, and $\breve{t}_\ell \le S < T^{\mathrm{o}}_{\breve{t}_\ell}$ holds due to $\breve{t}_\ell < T^{\mathrm{o}}_{\breve{t}_\ell}$). Applying Lemma B.2 (with $s = \breve{t}_\ell$), we have:

$$\begin{aligned}&\mathrm{E}\big\{\lambda_i(S \mid \mathcal{F}_S) \parallel \mathscr{F}_{\breve{t}_\ell}, \boldsymbol{N}(S) = \boldsymbol{N}(\breve{t}_\ell)\big\}\\ &\quad = \lambda_i(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell}), \text{ for all } i \in \mathcal{V}. \end{aligned} \tag{78}$$

Similar to (69), for $i \in \mathcal{V}$ and $t \ge s \ge 0$, define:

$$\lambda_i(t; \mathscr{F}_s) = \mathrm{E}\{\lambda_i(t \mid \mathscr{F}_t) \parallel \mathscr{F}_s, \boldsymbol{N}(t) = \boldsymbol{N}(s)\}. \tag{79}$$

For $t \in (\breve{t}_\ell, t_\ell^*)$, using the definition (79) (with $s = \breve{t}_\ell$), the fact $S(\bullet) = t$, and (78), we obtain:

$$\begin{aligned}&\lambda_i(t; \mathscr{F}_{\breve{t}_\ell})(\bullet)\\ &\quad = \mathrm{E}\big\{\lambda_i(t \mid \mathscr{F}_t) \parallel \mathscr{F}_{\breve{t}_\ell}, \boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell)\big\}(\bullet)\\ &\quad = \mathrm{E}\big\{\lambda_i(S \mid \mathcal{F}_S) \parallel \mathscr{F}_{\breve{t}_\ell}, \boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell)\big\}(\bullet)\\ &\quad = \lambda_i(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell})(\bullet) \text{ for all } i \in \mathcal{V}. \end{aligned} \tag{80}$$

Summing over $i \in \mathcal{V}$ on both sides of (80) gives us:

$$\lambda^{\mathrm{sum}}(t; \mathscr{F}_{\breve{t}_\ell})(\bullet) = \lambda^{\mathrm{sum}}(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell})(\bullet). \tag{81}$$

By (81), we simplify (77) to be:

$$\begin{aligned}&\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell}\big)(\bullet)\\ &\quad = \exp\Big\{-\int_{\breve{t}_\ell}^{t} \lambda^{\mathrm{sum}}(u \mid \mathscr{F}_u)(\bullet)\,\mathrm{d}u\Big\}\\ &\quad = \exp\{-\lambda^{\mathrm{sum}}(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell})(\bullet)\cdot(t-\breve{t}_\ell)\}, \end{aligned}$$

which proves (75).

(d) We demonstrate the following statement:

$$\begin{aligned}&\text{for } t \in (\breve{T}_\ell, T_\ell^*),\\ &\{\boldsymbol{N}(t) = \boldsymbol{N}(\breve{T}_\ell)\} \text{ is equivalent to } \{\breve{T}_{\ell+1} > t\}. \end{aligned} \tag{82}$$

The sufficiency part's proof resembles that of (73). For the necessity part, if $\boldsymbol{N}(t) \neq \boldsymbol{N}(\breve{T}_\ell)$, then there exists $T_{i,k} \in (\breve{T}_\ell, t]$ for some integers $i \in \mathcal{V}$ and $k \ge 1$. Also, $\breve{T}_\ell$ and $\breve{T}_{\ell+1}$ are two consecutive discontinuity points in the set (22), implying that $\cup_{i\in\mathcal{V}} \cup_{k\ge1} \{T_{i,k} : \breve{T}_\ell < T_{i,k} < \breve{T}_{\ell+1}\} = \varnothing$. Combining this with the fact that $T_{i,k} \in (\breve{T}_\ell, t]$, we deduce $\breve{T}_\ell < \breve{T}_{\ell+1} \le T_{i,k} \le t$, thus $\breve{T}_{\ell+1} \le t$.

We'll now proceed to prove parts (ii) and (iii) of Theorem 1.

Proof of Part (ii)*:* For $T_\ell^* < \infty$, using the similar proof as (82), we establish that $\boldsymbol{N}((\breve{T}_\ell, T_\ell^*)) = \boldsymbol{0}$ is equivalent to $\breve{T}_{\ell+1} \ge T_\ell^*$. Also, the result $\breve{T}_{\ell+1} \in (\breve{T}_\ell, T_\ell^*]$ in part (i) implies that $\breve{T}_{\ell+1} \ge T_\ell^*$ is equivalent to $\breve{T}_{\ell+1} = T_\ell^*$. Consequently,

$$\begin{aligned}\mathrm{P}\big(\breve{T}_{\ell+1} = T_\ell^* \,\big|\, \mathcal{F}_{\breve{T}_\ell}\big) &= \mathrm{P}\big(\breve{T}_{\ell+1} \ge T_\ell^* \,\big|\, \mathcal{F}_{\breve{T}_\ell}\big)\\ &= \mathrm{P}\big(\boldsymbol{N}\big((\breve{T}_\ell, T_\ell^*)\big) = \boldsymbol{0} \mid \mathcal{F}_{\breve{T}_\ell}\big). \end{aligned} \tag{83}$$

Evaluating both sides of (83) given the realization $\bullet$ and using $t_\ell^* = T_\ell^*(\bullet)$, we derive:

$$\begin{aligned}&\mathrm{P}\big(\breve{T}_{\ell+1} = T_\ell^* \mid \mathcal{F}_{\breve{T}_\ell}\big)(\bullet)\\ &\quad = \lim_{t\uparrow t_\ell^*} \mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell}\big)(\bullet)\\ &\quad = \lim_{t\uparrow t_\ell^*} \exp\{-\lambda^{\mathrm{sum}}(\breve{t}_\ell \mid \mathscr{F}_{\breve{t}_\ell})(\bullet)\cdot(t-\breve{t}_\ell)\} \qquad (84)\\ &\quad = \exp\big\{-\lambda^{\mathrm{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})\cdot(T_\ell^* - \breve{T}_\ell)\big\}(\bullet), \end{aligned}$$

where (84) is derived from (75) with $t \in (\breve{t}_\ell, t_\ell^*)$. This proves (28). By using $\breve{t}_\ell = \breve{T}_\ell(\bullet)$, (82), and (75), for $t \in (\breve{T}_\ell, T_\ell^*)$, we have:

$$\begin{aligned}&f_{\breve{T}_{\ell+1}|\mathcal{F}_{\breve{T}_\ell}}(t)(\bullet)\\ &= \frac{-\partial\,\mathrm{P}(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell) \mid \mathscr{F}_{\breve{t}_\ell})(\bullet)}{\partial\, t}\\ &= \lambda^{\mathrm{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})(\bullet)\cdot\exp\{-\lambda^{\mathrm{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell})\cdot(t-\breve{T}_\ell)\}(\bullet), \end{aligned} \tag{85}$$

which verifies (29). For $T_\ell^* = \infty$, following the same proof as that of (85), we have that $(\breve{T}_{\ell+1} - \breve{T}_\ell) \mid \mathcal{F}_{\breve{T}_\ell} \sim \mathrm{Exp}(\lambda^{\mathrm{sum}}(\breve{T}_\ell \mid \mathcal{F}_{\breve{T}_\ell}))$.

Proof of part (iii)*:* For $T_\ell^* < \infty$, if $\breve{T}_{\ell+1} = T_\ell^*$, then (26) and (27) imply that $\mathcal{T}_\ell \neq \varnothing$ and $\breve{T}_{\ell+1} = T_\ell^* \in \mathcal{T}_\ell$. In other words, $\breve{T}_{\ell+1} = T_{i,k} + \phi$ for some integers $i \in \mathcal{V}$ and $k \ge 1$. This, combined with (23), leads to $I_{\ell+1} = 0$, and thus $\mathrm{P}(I_{\ell+1} = i \mid \sigma(\mathcal{F}_{\breve{T}_\ell}, \breve{T}_{\ell+1})) = 0$ for $i \in \mathcal{V}$. If $\breve{T}_{\ell+1} \in (\breve{T}_\ell, T_\ell^*)$, then for $i \in \mathcal{V}$ and $t \in (\breve{T}_\ell, T_\ell^*)$, using $\breve{t}_\ell = \breve{T}_\ell(\bullet)$, $\mathscr{F}_{\breve{t}_\ell} \subseteq \mathscr{F}_t$, (8), (79), (80), and (75), we obtain:

$$\begin{aligned}&\frac{\partial\,\mathrm{P}\big(\breve{T}_{\ell+1} \le t,\, I_{\ell+1} = i \mid \mathcal{F}_{\breve{T}_\ell}\big)(\bullet)}{\partial\, t}\\ &= \lim_{\Delta\downarrow0} \Delta^{-1}\mathrm{E}\big\{\mathrm{P}(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t)\\ &\qquad \cdot \mathrm{I}(\boldsymbol{N}(t) = \boldsymbol{N}(\breve{t}_\ell)) \mid \mathscr{F}_{\breve{t}_\ell}\big\}(\bullet)\\ &= \mathrm{E}\Big\{\lim_{\Delta\downarrow0} \Delta^{-1}\mathrm{P}(N_i(t+\Delta) \neq N_i(t) \mid \mathscr{F}_t)\end{aligned}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$\times \mathrm{I}(\boldsymbol{N}(t)=\boldsymbol{N}(\breve{t}_\ell))\mid \mathscr{F}_{\breve{t}_\ell}\Big\}(\bullet) \tag{86}$$
$$=\lambda_i(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})(\bullet)\cdot\exp\{-\lambda^{\mathrm{sum}}(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})\cdot(t-\breve{T}_\ell)\}(\bullet),$$

where the interchange of limit and expectation in (86) follows from the dominated convergence theorem and condition A4. This implies:

$$\frac{\partial\,\mathrm{P}\big(\breve{T}_{\ell+1}\le t,\,I_{\ell+1}=i\mid\mathcal{F}_{\breve{T}_\ell}\big)}{\partial\,t}$$
$$=\lambda_i(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})\exp\{-\lambda^{\mathrm{sum}}(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})\cdot(t-\breve{T}_\ell)\}. \tag{87}$$

Let $t\in(\breve{t}_\ell,t^*_\ell)$, where $t^*_\ell=T^*_\ell(\bullet)$ is defined in preparation (b). Similarly to (74), define the realization

$$\begin{aligned}\circ&=\bullet\cap\{\breve{T}_{\ell+1}=t\}\\&=\big\{\{(\breve{T}_0,I_0),\ldots,(\breve{T}_\ell,I_\ell),\breve{T}_{\ell+1}\}=\{(\breve{t}_0,i_0),\ldots,(\breve{t}_\ell,i_\ell),t\}\big\}\\&\in\sigma(\mathcal{F}_{\breve{T}_\ell},\breve{T}_{\ell+1}).\end{aligned}$$

Combining the fact $\breve{T}_{\ell+1}(\circ)=t$, (87), and (85), for $i\in\mathcal{V}$, we have:

$$\begin{aligned}&\mathrm{P}\big(I_{\ell+1}=i\mid\sigma(\mathcal{F}_{\breve{T}_\ell},\breve{T}_{\ell+1})\big)(\circ)\\&=\frac{\mathrm{P}(I_{\ell+1}=i,\,\breve{T}_{\ell+1}=t\mid\mathcal{F}_{\breve{T}_\ell})(\circ)}{\mathrm{P}(\breve{T}_{\ell+1}=t\mid\mathcal{F}_{\breve{T}_\ell})(\circ)}\\&=\frac{\lambda_i(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})(\circ)}{\lambda^{\mathrm{sum}}(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})(\circ)}.\end{aligned}\tag{88}$$

This proves (30). For $T^*_\ell=\infty$, following the same proof as (88), we obtain: $\mathrm{P}(I_{\ell+1}=i\mid\sigma(\mathcal{F}_{\breve{T}_\ell},\breve{T}_{\ell+1}))=\lambda_i(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})/\lambda^{\mathrm{sum}}(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})$, for $i\in\mathcal{V}$ and $\breve{T}_{\ell+1}\in(\breve{T}_\ell,\infty)$. The proof is completed. □

G. Proofs of Lemmas 5–7, and Theorem 2

The proofs are divided into three parts as follows:

- Part 1 introduces Definition 5, which defines a class of marked point processes, called the *Exponential Marked Point Process* (EMPP). This generalizes the 'marked point process $(\breve{\boldsymbol{T}},\boldsymbol{I})$ for intensity discontinuities' in Definition 2 and facilitates derivations. Lemmas B.3–B.5 will present the probabilistic properties of the EMPP.
- Part 2 presents Definition 6, which introduces the concept of *t-truncated* EMPP. Lemmas B.6–B.8 will present the probabilistic properties of the *t-truncated* EMPP.
- Part 3 uses the results from Parts 1 and 2 to the 'marked point process $(\breve{\boldsymbol{T}},\boldsymbol{I})$ for intensity discontinuities' to provide the proofs of Lemmas 5–7 and Theorem 2.

1) Part 1: EMPP *and Its Probabilistic Properties:*

*Definition 5 (Exponential Marked Point Process (*EMPP*) $(\boldsymbol{T},\boldsymbol{I})$):* Let the node set $\mathcal{V}=\{1,\ldots,V\}$, and let the mark set be $\mathcal{V}\cup\{0\}$. Set $T_0=0$, $I_0=0$, and $\mathcal{F}_{T_0}=\{\Omega,\varnothing\}$. Define $\{\mathcal{F}_{T_\ell}\}_{\ell\ge0}$ as the filtration generated by a marked point process $(\boldsymbol{T},\boldsymbol{I})=(\{T_\ell\}_{\ell\ge0},\{I_\ell\}_{\ell\ge0})\in([0,\infty),\mathcal{V}\cup\{0\})$, where $0<T_1<T_2<\cdots$. We term $(\boldsymbol{T},\boldsymbol{I})$ an *Exponential Marked Point Process* (EMPP) if, for each integer $\ell\ge0$, there exist $\mathcal{F}_{T_\ell}$-measurable random variables $\Delta_\ell\in[0,\infty]$ and $\{\lambda_{i,\ell}\}_{i\in\mathcal{V}}\in(0,\infty)$ such that the distributions of $T_{\ell+1}$ and $I_{\ell+1}$ satisfy the following conditions:

(i) (Support of $T_{\ell+1}$) $\mathrm{P}(T_\ell<T_{\ell+1}\le T_\ell+\Delta_\ell)=1$.

(ii) (Conditional distribution of $T_{\ell+1}$) If $\Delta_\ell<\infty$, then $T_{\ell+1}$ conditional on $\mathcal{F}_{T_\ell}$ has a mixed-type distribution with the p.m.f.

$$\mathrm{P}(T_{\ell+1}=T_\ell+\Delta_\ell\mid\mathcal{F}_{T_\ell})=\exp(-\lambda^{\mathrm{sum}}_\ell\cdot\Delta_\ell) \tag{89}$$

at $T_\ell+\Delta_\ell$, and the p.d.f.

$$f_{T_{\ell+1}\mid\mathcal{F}_{T_\ell}}(x\mid\mathcal{F}_{T_\ell})=\lambda^{\mathrm{sum}}_\ell\cdot\exp\{-\lambda^{\mathrm{sum}}_\ell\cdot(x-T_\ell)\}, \tag{90}$$

for $x\in(T_\ell,T_\ell+\Delta_\ell)$, where $\lambda^{\mathrm{sum}}_\ell=\sum_{i=1}^V\lambda_{i,\ell}$. If $\Delta_\ell=\infty$, then $(T_{\ell+1}-T_\ell)\mid\mathcal{F}_{T_\ell}\sim\mathrm{Exp}(\lambda^{\mathrm{sum}}_\ell)$.

(iii) (Conditional distribution of $I_{\ell+1}$) If $\Delta_\ell<\infty$, then $I_{\ell+1}$ has the conditional distribution: for $i\in\mathcal{V}$,

$$\mathrm{P}(I_{\ell+1}=i\mid\sigma(\mathcal{F}_{T_\ell},T_{\ell+1}))=\begin{cases}0, & \text{if } T_{\ell+1}-T_\ell=\Delta_\ell,\\ \lambda_{i,\ell}/\lambda^{\mathrm{sum}}_\ell, & \text{if } 0<T_{\ell+1}-T_\ell<\Delta_\ell.\end{cases} \tag{91}$$

If $\Delta_\ell=\infty$, then (91) reduces to $\mathrm{P}(I_{\ell+1}=i\mid\sigma(\mathcal{F}_{T_\ell},T_{\ell+1}))=\lambda_{i,\ell}/\lambda^{\mathrm{sum}}_\ell$, for $i\in\mathcal{V}$ and $T_{\ell+1}\in(T_\ell,\infty)$.

Remark 6: The *Exponential Marked Point Process* (EMPP) $(\boldsymbol{T},\boldsymbol{I})$ in Definition 5 generalizes the class of 'marked point process $(\breve{\boldsymbol{T}},\boldsymbol{I})$ for intensity discontinuities' in Definition 2 which follow the distribution in Theorem 1, according to $(\boldsymbol{T},\boldsymbol{I})=(\breve{\boldsymbol{T}},\boldsymbol{I})$, $\mathcal{F}_{T_\ell}=\mathcal{F}_{\breve{T}_\ell}$, $\lambda_{i,\ell}=\lambda_i(\breve{T}_\ell\mid\mathcal{F}_{\breve{T}_\ell})=\exp\{\beta_{0;i}+\sum_{j\in\mathcal{V}}\beta_{j,i}\,x_j(\breve{T}_\ell)\}$, $\Delta_\ell=T^*_\ell-\breve{T}_\ell$. To prove Lemmas 5–7, we will first show probabilistic properties of EMPP $(\boldsymbol{T},\boldsymbol{I})$ which then apply to our $(\breve{\boldsymbol{T}},\boldsymbol{I})$ *for intensity discontinuities*.

For clarity, we introduce some notations similar to (31) and (32). The duration τ_ℓ between consecutive time points is defined as:

$$\tau_\ell=T_\ell-T_{\ell-1},\quad \ell\ge1. \tag{92}$$

The event counts $M_{i,\ell}$ at node $i\in\mathcal{V}$ are calculated as:

$$M_{i,0}=0,\quad M_{i,\ell}=\textstyle\sum_{k=1}^{\ell}\mathrm{I}(I_k=i),\quad \ell\ge1. \tag{93}$$

Lemma B.3 presents the conditional expectation and variance of τ_k and $\mathrm{I}(I_k=i)$.

Lemma B.3 (Conditional Expectation and Variance Related to an EMPP $(\boldsymbol{T},\boldsymbol{I})$*):* Consider an EMPP $(\boldsymbol{T},\boldsymbol{I})$ as defined in Definition 5. For integers $k\ge1$ and $i\in\mathcal{V}$, we have

$$\begin{aligned}&\mathrm{var}\big\{\mathrm{I}(I_k=i)-\lambda_{i,k-1}\cdot\tau_k\mid\mathcal{F}_{T_{k-1}}\big\}\\&\quad=\mathrm{E}\big\{\mathrm{I}(I_k=i)\mid\mathcal{F}_{T_{k-1}}\big\}=\lambda_{i,k-1}\mathrm{E}(\tau_k\mid\mathcal{F}_{T_{k-1}}).\end{aligned}\tag{94}$$

Proof: For $\Delta_{k-1}<\infty$, the conditional distribution of τ_k is given by (89) and (90), and the conditional distribution of $\mathrm{I}(I_k=i)$ is given by (91). Direct calculations yield the following equations (95)–(97):

$$\begin{aligned}&\mathrm{E}(\lambda_{i,k-1}\,\tau_k\mid\mathcal{F}_{T_{k-1}})\\&\quad=\lambda_{i,k-1}\Big\{\Delta_{k-1}\mathrm{P}(\tau_k=\Delta_{k-1}\mid\mathcal{F}_{T_{k-1}})\\&\qquad+\int_0^{\Delta_{k-1}}x\,f_{\tau_k\mid\mathcal{F}_{T_{k-1}}}(x\mid\mathcal{F}_{T_{k-1}})\,\mathrm{d}x\Big\}\end{aligned}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$= \frac{\lambda_{i,k-1}}{\lambda_{k-1}^{\text{sum}}} \cdot (1 - e^{-\lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1}})$$
$$= \mathrm{E}\{\mathrm{I}(I_k = i) \mid \mathcal{F}_{T_{k-1}}\}, \tag{95}$$

together with

$$\begin{aligned}&\mathrm{E}\{\lambda_{i,k-1}\,\tau_k\,\mathrm{I}(I_k = i) \mid \mathcal{F}_{T_{k-1}}\}\\ &= \lambda_{i,k-1}\int_0^{\Delta_{k-1}} \frac{\lambda_{i,k-1}}{\lambda_{k-1}^{\text{sum}}}\, x\, f_{\tau_k|\mathcal{F}_{T_{k-1}}}(x \mid \mathcal{F}_{T_{k-1}})\,\mathrm{d}x\\ &= \frac{\lambda_{i,k-1}^2}{(\lambda_{k-1}^{\text{sum}})^2} \cdot \{1 - (1 + \lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1})\, e^{-\lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1}}\},\end{aligned} \tag{96}$$

and

$$\begin{aligned}&\mathrm{E}(\lambda_{i,k-1}^2\,\tau_k^2 \mid \mathcal{F}_{T_{k-1}})\\ &= \lambda_{i,k-1}^2 \cdot \Big\{\Delta_{k-1}^2\, e^{-\lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1}}\\ &\qquad + \int_0^{\Delta_{k-1}} x^2\, \lambda_{k-1}^{\text{sum}} \cdot e^{-\lambda_{k-1}^{\text{sum}} \cdot x}\,\mathrm{d}x\Big\}\\ &= \frac{\lambda_{i,k-1}^2}{(\lambda_{k-1}^{\text{sum}})^2}\{2 - (2\,\lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1} + 2)e^{-\lambda_{k-1}^{\text{sum}} \cdot \Delta_{k-1}}\}.\end{aligned} \tag{97}$$

Combining (96) and (97), we have:

$$\mathrm{E}\{\lambda_{i,k-1}^2 \cdot \tau_k^2 - 2\,\lambda_{i,k-1} \cdot \tau_k\,\mathrm{I}(I_k = i) \mid \mathcal{F}_{T_{k-1}}\} = 0. \tag{98}$$

For $\Delta_{k-1} = \infty$, the conditional distributions of τ_k and $\mathrm{I}(I_k = i)$ are given in parts (ii) and (iii) of Definition 5. Using this alongside similar calculations as in (95)–(98), we verify that (95) and (98) hold for $\Delta_{k-1} = \infty$ as well. From (95) and (98), we derive:

$$\begin{aligned}&\mathrm{var}\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k \mid \mathcal{F}_{T_{k-1}}\}\\ &\quad = \mathrm{E}\{\mathrm{I}(I_k = i) \mid \mathcal{F}_{T_{k-1}}\}.\end{aligned} \tag{99}$$

Combining (95) and (99) completes the proof of (94). □

Lemma B.4 follows directly from Lemma B.3.

Lemma B.4 (Martingale Property for EMPP*):* In an EMPP $(\boldsymbol{T}, \boldsymbol{I})$, for each $i \in \mathcal{V}$, the random process $\{M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\}_{\ell \geq 1}$ is a martingale with respect to $\{\mathcal{F}_{T_\ell}\}_{\ell \geq 1}$.

Proof: Using (94), for each integer $k \geq 1$, we have:

$$\mathrm{E}\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k \mid \mathcal{F}_{T_{k-1}}\} = 0. \tag{100}$$

This completes the proof. □

Lemma B.5 derives the variance of the martingale.

Lemma B.5 (Variance of the Martingale $\{M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\}_{\ell \geq 1}$*):* In an EMPP $(\boldsymbol{T}, \boldsymbol{I})$, for integers $i \in \mathcal{V}$ and $\ell \geq 1$, we have $\mathrm{var}(M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k) = \mathrm{E}(M_{i,\ell})$.

Proof: Using (100), for any indices r and k such that $1 \leq r < k$, we get:

$$\begin{aligned}&\mathrm{E}[\{\mathrm{I}(I_r = i) - \lambda_{i,r-1} \cdot \tau_r\}\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k\}]\\ &= \mathrm{E}[\{\mathrm{I}(I_r = i) - \lambda_{i,r-1} \cdot \tau_r\}\\ &\qquad \times \mathrm{E}\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k \mid \mathcal{F}_{T_{k-1}}\}]\\ &= 0.\end{aligned} \tag{101}$$

For any index $\ell \geq 1$, applying (101), (94), and (99) yields:

$$\begin{aligned}&\mathrm{E}\Big\{\Big(M_{i,\ell} - \sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau_k\Big)^2\Big\}\\ &= \sum_{k=1}^{\ell} \mathrm{E}\big[\{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k\}^2\big]\\ &\quad + \sum_{1 \leq k \neq r \leq \ell} \sum \mathrm{E}\big[\{\mathrm{I}(I_r = i) - \lambda_{i,r-1} \cdot \tau_r\}\\ &\qquad \times \{\mathrm{I}(I_k = i) - \lambda_{i,k-1} \cdot \tau_k\}\big]\\ &= \sum_{k=1}^{\ell} \mathrm{E}\{\mathrm{I}(I_k = i)\} = \mathrm{E}(M_{i,\ell}).\end{aligned}$$

This completes the proof. □

2) Part 2: t-Truncated EMPP *and Its Probabilistic Properties:* Next, we derive the probabilistic results of the EMPP $(\boldsymbol{T}, \boldsymbol{I})$ when the point process $\boldsymbol{T} = \{T_0, T_1, \ldots\}$ reaches a pre-specified time point $t \in (0, \infty)$. Definition 6 introduces the notion of *t-truncated* EMPP.

Definition 6 (t-Truncated EMPP*):* Consider an EMPP $(\boldsymbol{T}, \boldsymbol{I}) = (\{T_\ell\}_{\ell \geq 0}, \{I_\ell\}_{\ell \geq 0})$. Let $t \in (0, \infty)$ be a given deterministic time point. Define the marked point process $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]}) = (\{T_\ell^{[t]}\}_{\ell \geq 0}, \{I_\ell^{[t]}\}_{\ell \geq 0})$, where

$$T_\ell^{[t]} = T_\ell \wedge t, \qquad I_\ell^{[t]} = I_\ell\,\mathrm{I}(T_\ell \leq t). \tag{102}$$

We call $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ the *t-truncated* EMPP *derived from* $(\boldsymbol{T}, \boldsymbol{I})$.

Lemma B.6 states that any double sequence $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ defined in (102) is an EMPP.

Lemma B.6: Let $(\boldsymbol{T}, \boldsymbol{I}) = (\{T_\ell\}_{\ell \geq 0}, \{I_\ell\}_{\ell \geq 0})$ be an EMPP defined in Definition 5 associated with $\{\lambda_{i,\ell}\}_{i \in \mathcal{V}}$ and Δ_ℓ in (89)–(91). For $t \in (0, \infty)$, let $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ be the corresponding t-truncated EMPP as in Definition 6. Then the probability distributions of $\boldsymbol{T}^{[t]}$ and $\boldsymbol{I}^{[t]})$ meet the conditions (i) and (ii) in Definition 5 associated with $\{\lambda_{i,\ell}\}_{i \in \mathcal{V}}$ and $\Delta_{\ell,t}$ (instead of Δ_ℓ), where

$$\Delta_{\ell,t} = \Delta_\ell \wedge (t - T_\ell^{[t]}), \tag{103}$$

and thus $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ is an EMPP.

Proof: To prove Lemma B.6, it suffices to show that for each integer $\ell \geq 0$, $(T_{\ell+1}^{[t]}, I_{\ell+1}^{[t]})$ follows the conditional distribution in (89)–(91) with T_ℓ and Δ_ℓ replaced by $T_\ell^{[t]}$ and $\Delta_{\ell,t}$. We proceed by cases of $T_\ell^{[t]}$.

Case (i): $T_\ell^{[t]} < t - \Delta_\ell$. From (102) and (103), we observe that $T_\ell^{[t]} = T_\ell$ and $\Delta_{\ell,t} = \Delta_\ell$. Also, as per (89) and (90), we know that $T_{\ell+1} \leq T_\ell + \Delta_\ell \leq t$. Thus, $(T_{\ell+1}^{[t]},\ I_{\ell+1}^{[t]}) = (T_{\ell+1} \wedge t,\ I_{\ell+1}\mathrm{I}(T_{\ell+1} \leq t)) = (T_{\ell+1},\ I_{\ell+1})$, following the conditional distribution in (89)–(91).

Case (ii): $t - \Delta_\ell \leq T_\ell^{[t]} < t$. Using (102) and (103), we have $T_\ell^{[t]} = T_\ell$ and $\Delta_{\ell,t} = t - T_\ell$. Employing (89) and (90), we get

$$\mathrm{P}((T_{\ell+1} \wedge t) = t \mid \mathcal{F}_{T_\ell}) = \exp\{-\lambda_\ell^{\text{sum}} \cdot (t - T_\ell)\},$$

and

$$\begin{aligned}&f_{(T_{\ell+1} \wedge t)|\mathcal{F}_{T_\ell}}(x \mid \mathcal{F}_{T_\ell})\\ &= \lambda_\ell^{\text{sum}} \cdot \exp\{-\lambda_\ell^{\text{sum}} \cdot (x - T_\ell)\}, \quad \text{for } x \in (T_\ell, t).\end{aligned}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

This implies that $T^{[t]}_{\ell+1} = T_{\ell+1} \wedge t$ follows the distribution in (89) and (90) with T_ℓ and Δ_ℓ replaced by $T^{[t]}_\ell = T_\ell$ and $\Delta_{\ell,t} = t - T_\ell$. Also, by checking (91), we have that $I^{[t]}_{\ell+1} = I_{\ell+1}\mathrm{I}(T_{\ell+1} \le t)$ conditional on $T^{[t]}_{\ell+1}$, follows the distribution in (91).

Case (iii): $T^{[t]}_\ell = t$. Considering (103), we know $\Delta_{\ell,t} = 0$. Then $(T^{[t]}_{\ell+1}, I^{[t]}_{\ell+1}) = (t, 0)$ follows the distribution in (89)–(91) with T_ℓ and Δ_ℓ replaced by $T^{[t]}_\ell = t$ and $\Delta_{\ell,t} = 0$.

Summing up these cases completes the proof. □

Similar to (92) and (93), we employ notations $\tau^{[t]}_\ell$ and $M^{[t]}_{i,\ell}$ for the 'duration' and 'event counts', respectively, of the t-truncated EMPP $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$, denoted as follows:

$$\tau^{[t]}_\ell = T^{[t]}_\ell - T^{[t]}_{\ell-1}, \quad \ell \ge 1, \tag{104}$$

$$M^{[t]}_{i,0} = 0, \qquad M^{[t]}_{i,\ell} = \sum_{k=1}^{\ell} \mathrm{I}(I^{[t]}_k = i), \quad \ell \ge 1. \tag{105}$$

We define $M^{[t]}_{i,\infty} = \lim_{\ell\to\infty} M^{[t]}_{i,\ell}$. Let $L_t = \sum_{\ell=1}^{\infty} \mathrm{I}(T_\ell \le t)$. From (102),

$$M^{[t]}_{i,\infty} = \sum_{k=1}^{\infty} \mathrm{I}(I^{[t]}_k = i) = \sum_{k=1}^{L_t} \mathrm{I}(I_k = i) = M_{i,L_t} \tag{106}$$

which represents the total event counts of node i in the time interval $[0, t]$. Lemma B.7 establishes an upper bound for $\mathrm{E}(M^{[t]}_{i,\infty})$.

Lemma B.7 (Upper Bound for $\mathrm{E}(M^{[t]}_{i,\infty})$*):* Let $(\boldsymbol{T}, \boldsymbol{I})$ be an EMPP, and $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ be the corresponding t-truncated EMPP from $(\boldsymbol{T}, \boldsymbol{I})$ as defined in Definition 6. Assume that $\sup_{i\in\mathcal{V},\ell\ge0} \lambda_{i,\ell} \le c$ for a constant $c \in (0, \infty)$. Then:

$$\mathrm{E}(M^{[t]}_{i,\infty}) \;\le\; c\,t. \tag{107}$$

Proof: Lemma B.6 has verified that $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ is an EMPP. Applying Lemma B.4 to $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ yields that for each integer $\ell \ge 1$, $\mathrm{E}(M^{[t]}_{i,\ell}) = \mathrm{E}(\sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau^{[t]}_k)$. Note that $M^{[t]}_{i,\ell}$ and $\sum_{k=1}^{\ell} \lambda_{i,k-1} \cdot \tau^{[t]}_k$ are monotonically increasing in ℓ. By the Monotone Convergence Theorem, we obtain:

$$\begin{aligned}\mathrm{E}(M^{[t]}_{i,\infty}) &= \mathrm{E}\Big(\sum_{k=1}^{\infty} \lambda_{i,k-1} \cdot \tau^{[t]}_k\Big) \\ &\le c\,\mathrm{E}\Big(\sum_{k=1}^{\infty} \tau^{[t]}_k\Big) \le c\,t.\end{aligned} \tag{108}$$

This completes the proof. □

Lemma B.8 (Upper Bound for $\mathrm{var}\{\sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\}$*):* Let $(\boldsymbol{T}, \boldsymbol{I})$ be an EMPP, and $(\boldsymbol{T}^{[t]}, \boldsymbol{I}^{[t]})$ be the corresponding t-truncated EMPP from $(\boldsymbol{T}, \boldsymbol{I})$ as in Definition 6. Assume that $\sup_{i\in\mathcal{V},\ell\ge0} \lambda_{i,\ell} \le c$ for a constant $c \in (0, \infty)$. Let $\{X_\ell\}_{\ell\ge0}$ be a sequence of random variables such that $X_\ell \ge 0$ is measurable with respect to $\mathcal{F}_{T_\ell}$ for each integer $\ell \ge 0$, and $\sup_{\ell\ge0} X_\ell \le c_2$ a.s. for a constant $c_2 \in (0, \infty)$. Then:

$$\mathrm{E}\Big\{\sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big\} = 0, \tag{109}$$

and

$$\begin{aligned}&\mathrm{var}\Big\{\sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big\} \\ &\quad= \mathrm{E}\Big\{\sum_{k=1}^{\infty} X^2_{k-1}\,\mathrm{I}(I^{[t]}_k = i)\Big\} \le c\,c_2^2\,t.\end{aligned} \tag{110}$$

Proof: An argument similar to (108) gives us:

$$\begin{aligned}\mathrm{E}\Big\{\sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i)\Big\} &= \lim_{\ell\to\infty} \mathrm{E}\Big\{\sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i)\Big\} \\ &= \lim_{\ell\to\infty} \mathrm{E}\Big(\sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1}\,\tau^{[t]}_k\Big) \\ &= \mathrm{E}\Big(\sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1}\,\tau^{[t]}_k\Big),\end{aligned}$$

which proves (109). Using a similar proof as Lemma B.5, we have:

$$\begin{aligned}&\mathrm{E}\Big\{\Big(\sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big)^2\Big\} \\ &\quad= \mathrm{E}\Big\{\sum_{k=1}^{\ell} X^2_{k-1}\,\mathrm{I}(I^{[t]}_k = i)\Big\}.\end{aligned} \tag{111}$$

Note that

$$\lim_{\ell\to\infty} \sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) = \sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i), \text{ a.s.}$$

and

$$\begin{aligned}\lim_{\ell\to\infty} \sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k &= \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k \\ &\le \sum_{k=1}^{\infty} c\,c_2\,\tau^{[t]}_k \\ &= c\,c_2\,t \;<\; \infty, \quad \text{a.s.}.\end{aligned}$$

It follows that

$$\begin{aligned}&\sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k \\ &\overset{\text{a.s.}}{\to} \sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k,\end{aligned} \tag{112}$$

as $\ell \to \infty$. Also, for each integer $\ell \ge 1$, we have

$$\begin{aligned}&\Big\{\sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big\}^2 \\ &\le \Big\{\sum_{k=1}^{\ell} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i)\Big\}^2 + \Big(\sum_{k=1}^{\ell} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big)^2 \\ &\le (c_2\,M^{[t]}_{i,\infty})^2 + (c\,c_2\,t)^2.\end{aligned} \tag{113}$$

By (111), (112), (113), and the dominated convergence theorem, we have:

$$\mathrm{E}\Big\{\Big(\sum_{k=1}^{\infty} X_{k-1}\,\mathrm{I}(I^{[t]}_k = i) - \sum_{k=1}^{\infty} X_{k-1}\,\lambda_{i,k-1} \cdot \tau^{[t]}_k\Big)^2\Big\}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$= \mathrm{E}\Big\{ \sum_{k=1}^{\infty} X_{k-1}^2 \, \mathrm{I}(I_k^{[t]} = i) \Big\}$$
$$\leq c_2^2 \, \mathrm{E}(M_{i,\infty}^{[t]}) \; \leq \; c \, c_2^2 \, t,$$

where the last inequality is from (107). Thus, (110) is proved. □

3) Part 3: Proofs of Lemmas 5–7 and Theorem 2: Remark 6 verifies that the *marked point process* $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ *for intensity discontinuities* is an EMPP. Let $(\breve{\boldsymbol{T}}^{[t]}, \boldsymbol{I}^{[t]})$ be the *t-truncated* EMPP from $(\breve{\boldsymbol{T}}, \boldsymbol{I})$ as defined in Definition 6, i.e., for each integer $\ell \geq 1$, $(\breve{T}_\ell^{[t]}, I_\ell^{[t]}) = (\breve{T}_\ell \wedge t \,, \; I_\ell \, \mathrm{I}(\breve{T}_\ell \leq t))$. Recall that $M_{i,\ell}^{[t]}$ defined in (105) represents the event counts corresponding to $(\breve{\boldsymbol{T}}^{[t]}, \boldsymbol{I}^{[t]})$, and $M_{i,\infty}^{[t]} = \lim_{\ell \to \infty} M_{i,\ell}^{[t]}$. Recall L_t defined in (36). Using (25), (32), and (106), $N_i(t)$ has the equivalent expressions:

$$N_i(t) \; = \; M_{i,L_t} \; = \; \sum_{k=1}^{L_t} \mathrm{I}(I_k = i) \; = \; M_{i,\infty}^{[t]}. \tag{114}$$

Following (104), let $\tau_\ell^{[t]} = \breve{T}_\ell^{[t]} - \breve{T}_{\ell-1}^{[t]}$ be the duration between two consecutive time points $\breve{T}_{\ell-1}^{[t]}$ and $\breve{T}_\ell^{[t]}$. It is easily verified that this $\tau_\ell^{[t]}$ is identical to that defined in (37). Using (33) and (37), the integral $\int_0^t \lambda_i(u \mid \mathscr{F}_u) \, \mathrm{d}u$ has the following equivalent expressions:

$$\int_0^t \lambda_i(u \mid \mathscr{F}_u) \, \mathrm{d}u \; = \; \sum_{k=1}^{L_t+1} \lambda_{i,k-1} \cdot \tau_k^{[t]} \; = \; \sum_{k=1}^{\infty} \lambda_{i,k-1} \cdot \tau_k^{[t]}. \tag{115}$$

After clarifying the above facts, we will proceed to prove Lemmas 5–7 and Theorem 2.

4) Proof of Lemma 5: Lemma 5 is directly obtained from Lemmas B.3 and B.5. □

5) Proof of Lemma 6: Lemma 6 is directly derived from Lemma B.4. □

6) Proof of Lemma 7: From (102), (105), and (115), we have $\sum_{k=1}^{L_t} X_{k-1} \, \mathrm{I}(I_k = i) = \sum_{k=1}^{\infty} X_{k-1} \, \mathrm{I}(I_k^{[t]} = i)$, and $\sum_{k=1}^{L_t+1} X_{k-1} \, \lambda_{i,k-1} \cdot \tau_k^{[t]} = \sum_{k=1}^{\infty} X_{k-1} \, \lambda_{i,k-1} \cdot \tau_k^{[t]}$. Consequently, all the results in Lemma 7 could be directly derived from Lemma B.8. □

7) Proof of Theorem 2: Recall $N_i(t) = M_{i,L_t}$ in (114) and $\int_0^t \lambda_i(u \mid \mathscr{F}_u) \, \mathrm{d}u = \sum_{k=1}^{L_t+1} \lambda_{i,k-1} \cdot \tau_k^{[t]}$ in (115). For each integer $\ell \geq 0$, consider $X_\ell = x(\breve{T}_\ell)$. This leads us to $\int_0^t x(u-) \, \mathrm{d}N_i(u) = \sum_{k=1}^{L_t} X_{k-1} \, \mathrm{I}(I_k = i)$ and $\int_0^t x(u) \, \lambda_i(u \mid \mathscr{F}_u) \, \mathrm{d}u \; = \; \sum_{k=1}^{L_t+1} X_{k-1} \, \lambda_{i,k-1} \cdot \tau_k^{[t]}$. Consequently, (40) and (42) in Theorem 2 directly stem from (38) and (39), respectively, in Lemma 7. Additionally, the finiteness of $\boldsymbol{N}(t)$ in (41) directly follows from (40). □

H. Proof of Non-Stationarity of $\boldsymbol{N}(t)$ in Section IV-B

Lemma B.9 ($\boldsymbol{N}(t)$ Is Not Strict-Sense Stationary): Under conditions A1, A2, A3, A4, A5, and A8 in Appendix B, there exists a node $i_0 \in \mathcal{V}$ such that $N_{i_0}(t)$ is not *strict-sense stationary*. Hence, the multivariate counting process $\boldsymbol{N}(t)$ is not *strict-sense stationary*.

Before proving Lemma B.9, we first present Lemma B.10.

Lemma B.10 (Probabilistic Inequalities for $\lambda_i(t \mid \mathscr{F}_t)$ in Our (12)): Assume conditions A1, A2, A3, A4, A5, and A8 in Appendix B. Let $\{\lambda_i(t \mid \mathscr{F}_t)\}_{i \in \mathcal{V}}$ denote the CIFs defined in (12). For any distinct $i, j \in \mathcal{V}$, there exist constants c_0, c_1, c_2, $c_3 \in (0, \infty)$ such that for any $t \in (0, \phi)$, the following assertions hold:

$$\mathrm{P}(\lambda_i(t \mid \mathscr{F}_t) = \exp(\beta_{0;i})) \geq \exp(-c_0 \, t), \tag{116}$$
$$\mathrm{P}\big(\lambda_i(t \mid \mathscr{F}_t) = \exp\{\beta_{0;i} + \beta_{j,i} \cdot g(1/\phi)\}\big) \geq c_1 \cdot \exp(-c_2 \, t) \cdot \{1 - \exp(-c_3 \, t)\}. \tag{117}$$

Proof: For any $t \in (0, \phi)$, equations (12), (16), and (17) imply that $\{\boldsymbol{N}(t) = \boldsymbol{0}\} \subseteq \{\lambda_i(t \mid \mathscr{F}_t) = \exp(\beta_{0;i})\}$ and $\{N_j(t) = 1 \text{ and } N_k(t) = 0 \text{ for all } k \in \mathcal{V} \setminus \{j\}\} \subseteq \{\lambda_i(t \mid \mathscr{F}_t) = \exp(\beta_{0;i} + \beta_{j,i} \cdot g(1/\phi))\}$. To verify (116) and (117), it suffices to show that for any $t \in (0, \phi)$:

$$\mathrm{P}(\boldsymbol{N}(t) = \boldsymbol{0}) \; = \; \exp(-c_0 \, t), \tag{118}$$
$$\mathrm{P}(N_j(t) = 1, \text{ and } N_k(t) = 0 \text{ for all } k \in \mathcal{V} \setminus \{j\}) \; \geq \; c_1 \cdot \exp(-c_2 \, t) \cdot \{1 - \exp(-c_3 \, t)\}. \tag{119}$$

Recall that in Theorem 1, we have the following facts: $\breve{T}_0 = 0$, $\mathcal{F}_{\breve{T}_0} = \mathscr{F}_0 = \{\Omega, \varnothing\}$, $\mathcal{T}_0 = \varnothing$ and $T_0^* = \infty$. By Theorem 1(ii) and the fact that $T_0^* = \infty$, we have that $\breve{T}_1$ is a continuous random variable with the p.d.f.:

$$f_{\breve{T}_1}(x) = \lambda^{\mathrm{sum}}(0) \cdot \exp\{-\lambda^{\mathrm{sum}}(0) \cdot x\}, \quad x \in (0, \infty), \tag{120}$$

where $\lambda^{\mathrm{sum}}(0) = \lambda^{\mathrm{sum}}(0 \mid \mathscr{F}_0) = \sum_{i=1}^V \lambda_i(0) = \sum_{i=1}^V \exp(\beta_{0;i})$. On the other hand, it is easy to verify from (22) that the first discontinuity point $\breve{T}_1 = \min\{T_{j,\ell} : j \in \mathcal{V}, \ell \geq 1\}$ is exactly the first event time point, i.e.,

$$\boldsymbol{N}(t) = \boldsymbol{0} \quad \text{if and only if} \quad 0 \leq t < \breve{T}_1. \tag{121}$$

Combining (120) and (121), we have $\mathrm{P}(\boldsymbol{N}(t) = \boldsymbol{0}) = \exp\{-\lambda^{\mathrm{sum}}(0) \cdot t\}$. This proves (118) with a constant $c_0 = \lambda^{\mathrm{sum}}(0) = \sum_{i=1}^V \exp(\beta_{0;i})$.

Similarly, parts (ii) and (iii) of Theorem 3 yield that:

$$\begin{aligned} &\mathrm{P}\big(N_j(t) = 1, \text{ and } N_k(t) = 0 \text{ for all } k \in \mathcal{V} \setminus \{j\}\big) \\ &= \int_0^t \lambda^{\mathrm{sum}}(0) \exp\{-\lambda^{\mathrm{sum}}(0) \cdot x\} \cdot \frac{\lambda_j(0)}{\lambda^{\mathrm{sum}}(0)} \\ &\quad \times \Big[1 - \int_x^t \lambda^{\mathrm{sum}}(x \mid \mathscr{F}_x) \\ &\qquad \times \exp\{-\lambda^{\mathrm{sum}}(x \mid \mathscr{F}_x) \cdot (u - x)\} \, \mathrm{d}u\Big] \, \mathrm{d}x \\ &= \begin{cases} \frac{\lambda_1 \cdot \{\exp(-\lambda_3 \cdot t) - \exp(-\lambda_2 \cdot t)\}}{\lambda_2 - \lambda_3}, & \text{if } \lambda_2 \neq \lambda_3, \\ \lambda_1 \cdot t \cdot \exp(-\lambda_2 \cdot t), & \text{if } \lambda_2 = \lambda_3, \end{cases} \end{aligned} \tag{122}$$

where $\lambda_1 = \lambda_j(0) = \exp(\beta_{0;j})$, $\lambda_2 = \lambda^{\mathrm{sum}}(0) = \sum_{i=1}^V \exp(\beta_{0;i})$, and $\lambda^{\mathrm{sum}}(x \mid \mathscr{F}_x)$ reduces to $\lambda_3 = \sum_{i=1}^V \exp\{\beta_{0;i} + \beta_{j,i} \, g(1/\phi)\}$.

For $\lambda_2 \neq \lambda_3$, if $\lambda_2 - \lambda_3 = \delta > 0$, then (122) gives that:

$$\mathrm{P}\big(N_j(t) = 1, \text{ and } N_k(t) = 0 \text{ for all } k \in \mathcal{V} \setminus \{j\}\big) = \lambda_1/\delta \cdot \exp(-\lambda_3 \cdot t)\{1 - \exp(-\delta \, t)\}.$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Thus, (119) holds with $c_1 = \lambda_1/\delta$, $c_2 = \lambda_3$, and $c_3 = \delta$. Due to the symmetry between λ_2 and λ_3 in (122), the similar argument holds when $\lambda_2 - \lambda_3 < 0$.

For $\lambda_2 = \lambda_3$, (122) yields that:

$$\begin{aligned}&\mathrm{P}\big(N_j(t) = 1, \text{ and } N_k(t) = 0 \text{ for all } k \in \mathcal{V} \setminus \{j\}\big)\\ &= \lambda_1 \cdot t \cdot \exp(-\lambda_2 \cdot t) \geq \lambda_1 \cdot \exp(-\lambda_2 \cdot t) \cdot \{1 - \exp(-t)\}.\end{aligned}$$

Hence, (119) holds with $c_1 = \lambda_1$, $c_2 = \lambda_2$, and $c_3 = 1$. This completes the proof. □

Next, we prove Lemma B.9 using proof by contradiction. For any $(j_0, i_0) \in \mathcal{E}$ with $\beta_{j_0,i_0} \neq 0$, if $\{N_{i_0}(t)\}_{t\geq 0}$ is stationary, then property (P1) implies that the CIF $\lambda_{i_0}(t \mid \mathscr{F}_t)$ has the same distribution for all $t \in (0, \phi)$. Thus, for the two possible values $\exp(\beta_{0;i_0})$ and $\exp\{\beta_{0;i_0} + \beta_{j_0,i_0} \cdot g(1/\phi)\}$ of $\lambda_{i_0}(t \mid \mathscr{F}_t)$, there exist constants c_4 and c_5 in the range $[0, 1]$ such that $\mathrm{P}(\lambda_{i_0}(t \mid \mathscr{F}_t) = \exp(\beta_{0;i_0})) \equiv c_4$ and $\mathrm{P}(\lambda_{i_0}(t \mid \mathscr{F}_t) = \exp\{\beta_{0;i_0} + \beta_{j_0,i_0} \cdot g(1/\phi)\}) \equiv c_5$ hold for any $t \in (0, \phi)$. Combining this with (116) and (117), for any $t \in (0, \phi)$, we have:

$$\begin{aligned}&\mathrm{P}\big(\lambda_{i_0}(t \mid \mathscr{F}_t) = \exp(\beta_{0;i_0}),\\ &\qquad \text{or } \lambda_{i_0}(t \mid \mathscr{F}_t) = \exp\{\beta_{0;i_0} + \beta_{j_0,i_0} \cdot g(1/\phi)\}\big)\\ &\equiv c_4 + c_5\\ &\geq \sup_{t\in(0,\phi)} \big\{\exp(-c_0\, t)\big\}\\ &\quad + \sup_{t\in(0,\phi)} \big\{c_1 \cdot \exp(-c_2\, t) \cdot \{1 - \exp(-c_3\, t)\}\big\}\\ &\geq 1 + c_1 \cdot \exp(-c_2 \cdot \phi/2) \cdot \{1 - \exp(-c_3 \cdot \phi/2)\} > 1.\end{aligned}$$

This obviously contradicts. This completes the proof. □

I. Proof of Lemma 8

From (43), for each integer $\ell \geq 1$, we have $R_\ell = \min(\mathcal{U}_\ell)$, where $\mathcal{U}_\ell = \{t \geq R_{\ell-1} + \phi : \boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}\}$. Thus, R_ℓ exists if and only if the following two conditions hold:

(i) $\mathcal{U}_\ell \neq \varnothing$. (Since $\mathcal{U}_\ell$ is bounded below, this indicates $\inf(\mathcal{U}_\ell)$ exists.)

(ii) $\inf(\mathcal{U}_\ell) \in \mathcal{U}_\ell$. (This indicates $\min(\mathcal{U}_\ell) = \inf(\mathcal{U}_\ell)$.)

We start by proving the existence of R_1. We first prove that condition (i) holds with probability one for $\mathcal{U}_1$. Note that $\mathcal{U}_1 \neq \varnothing$ if and only if there exists $t \geq \phi$ such that $\boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}$. It suffices to show that

$$\mathrm{P}\Big(\bigcup_{t\geq\phi} \big\{\boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}\big\}\Big) = 1. \tag{123}$$

By condition A5, there exists some constant $c \in (0, \infty)$ such that $\lambda^{\mathrm{sum}}(t \mid \mathscr{F}_t) \leq c$. This together with (69) gives that $\lambda^{\mathrm{sum}}(t; \mathscr{F}_s) \leq c$. Then, by (68), for $t > s \geq 0$, we have:

$$\begin{aligned}\mathrm{P}\big(\boldsymbol{N}(t) = \boldsymbol{N}(s) \mid \mathscr{F}_s\big) &= \exp\Big\{-\int_s^t \lambda^{\mathrm{sum}}(u; \mathscr{F}_s)\, \mathrm{d}u\Big\}\\ &\geq \exp\{-c \cdot (t - s)\}.\end{aligned} \tag{124}$$

For each integer $k \geq 1$, plugging $s = (k-1)\,\phi$ and $t = k\,\phi$ into (124), we obtain

$$\mathrm{P}(A_k^* \mid \mathscr{F}_{(k-1)\,\phi}) \geq \exp(-c\,\phi), \tag{125}$$

where the event

$$\begin{aligned}A_k^* &= \big\{\boldsymbol{N}\big(((k-1)\,\phi, k\,\phi]\big) = \boldsymbol{0}\big\}\\ &= \big\{\boldsymbol{N}\big((k-1)\,\phi\big) = \boldsymbol{N}(k\,\phi)\big\}.\end{aligned}$$

Letting $k = 1$ in (125) yields:

$$\mathrm{P}(A_1^*) \geq \exp(-c\,\phi). \tag{126}$$

Also, for integers $k \geq 2$, by (125) and the fact that $\{\bigcap_{m=1}^{k-1} \overline{A_m^*}\} \in \mathscr{F}_{(k-1)\,\phi}$, we have: $\mathrm{P}(A_k^* \mid \cap_{m=1}^{k-1} \overline{A_m^*}) \geq \exp(-c\,\phi)$. Combining this with (126), for any integer $\ell \geq 2$, we have:

$$\begin{aligned}\mathrm{P}\Big(\bigcap_{k=1}^{\ell} \overline{A_k^*}\Big) &= \mathrm{P}(\overline{A_1^*}) \cdot \prod_{k=2}^{\ell} \mathrm{P}\Big(\overline{A_k^*} \,\Big|\, \bigcap_{m=1}^{k-1} \overline{A_m^*}\Big)\\ &\leq \big\{1 - \exp(-c\,\phi)\big\}^{\ell}.\end{aligned} \tag{127}$$

This gives:

$$\mathrm{P}\Big(\bigcup_{k=1}^{\ell} A_k^*\Big) = 1 - \mathrm{P}\Big(\bigcap_{k=1}^{\ell} \overline{A_k^*}\Big) \geq 1 - \{1 - \exp(-c\,\phi)\}^{\ell}.$$

Letting $\ell \to \infty$ in the above inequality yields $\mathrm{P}(\bigcup_{k=1}^{\infty} A_k^*) = 1$. It follows that:

$$\begin{aligned}&\mathrm{P}\Big(\bigcup_{t\geq\phi} \big\{\boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}\big\}\Big)\\ &\quad \geq \mathrm{P}\Big(\bigcup_{k\geq 1} \big\{\boldsymbol{N}(((k-1)\,\phi, k\,\phi]) = \boldsymbol{0}\big\}\Big)\\ &\quad = \mathrm{P}\Big(\bigcup_{k=1}^{\infty} A_k^*\Big) = 1,\end{aligned} \tag{128}$$

which proves (123).

We then prove that condition (ii) holds for $\mathcal{U}_1$ using a proof by contradiction. Let $R_1^* = \inf(\mathcal{U}_1)$. If $R_1^* \notin \mathcal{U}_1$, then there exists a sequence of time points $\{u_k\}_{k\geq 1} \in \mathcal{U}_1$ such that $u_1 > u_2 > \cdots$ and $\lim_{k\to\infty} u_k = R_1^*$. Note that $u_k \in \mathcal{U}_1$ implies that $\boldsymbol{N}(u_k) - \boldsymbol{N}(u_k - \phi) = \boldsymbol{N}((u_k - \phi, u_k]) = \boldsymbol{0}$. Using this and the fact that $\boldsymbol{N}(t)$ is right-continuous in $t \geq 0$, we have:

$$\begin{aligned}\boldsymbol{N}((R_1^* - \phi, R_1^*]) &= \boldsymbol{N}(R_1^*) - \boldsymbol{N}(R_1^* - \phi)\\ &= \lim_{k\to\infty} \{\boldsymbol{N}(u_k) - \boldsymbol{N}(u_k - \phi)\} = \boldsymbol{0},\end{aligned}$$

which contradicts $R_1^* \notin \mathcal{U}_1$.

Next, for each integer $\ell \geq 2$, we prove that R_ℓ exists with probability one. Following the same proof of condition (ii) for the case of $\mathcal{U}_1$, we can verify that condition (ii) also holds for $\mathcal{U}_\ell$ with integers $\ell \geq 2$. Now, we prove condition (i) holds with probability one for $\mathcal{U}_\ell$ with $\ell \geq 2$. Similar to (123), it suffices to show that

$$\mathrm{P}\Big(\bigcup_{t\geq R_{\ell-1}+\phi} \big\{\boldsymbol{N}((t-\phi, t]) = \boldsymbol{0}\big\}\Big) = 1. \tag{129}$$

For each integer $k \geq 1$ and real number $r > 0$, define the event

$$\begin{aligned}A_{k,r}^* &= \big\{\boldsymbol{N}\big(((k-1)\phi + r, k\phi + r]\big) = \boldsymbol{0}\big\}\\ &= \big\{\boldsymbol{N}((k-1)\phi + r) = \boldsymbol{N}(k\phi + r)\big\}.\end{aligned}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Using the same proof as that of (125)–(128), we obtain $\mathrm{P}(\bigcup_{k=1}^{\infty} A^*_{k,r}) = 1$, and

$$\begin{aligned}&\mathrm{P}\Big(\bigcup_{t\geq r+\phi}\big\{\boldsymbol{N}((t-\phi,t])=\mathbf{0}\big\}\Big)\\&\quad\geq \mathrm{P}\Big(\bigcup_{k\geq 1}\big\{\boldsymbol{N}\big(((k-1)\phi+r,k\phi+r]\big)=\mathbf{0}\big\}\Big)\\&\quad=\mathrm{P}\Big(\bigcup_{k=1}^{\infty}A^*_{k,r}\Big)=1.\end{aligned}$$

It follows that for each realization $R_{\ell-1}=r$, we have

$$\mathrm{P}\Big(\bigcup_{t\geq R_{\ell-1}+\phi}\big\{\boldsymbol{N}((t-\phi,t])=\mathbf{0}\big\}\,\Big|\,R_{\ell-1}=r\Big)=1,$$

which proves (129). □

J. Proof of Theorem 3

1) Proof of Part (i)*:* Before proving part (i), we first present Lemma B.11.

Lemma B.11: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let $R\in[\phi,\infty)$ be a stopping time with respect to the filtered probability space $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathrm{P})$, satisfying $\boldsymbol{N}((R-\phi,R])=\mathbf{0}$. Then for each $t\geq 0$, $\boldsymbol{N}(t+R)-\boldsymbol{N}(R)$ is independent of $\mathcal{F}_R$, where $\mathcal{F}_R$ is the stopping time σ-algebra with respect to R.

Proof: Recall the sequence of discontinuity points $\{\breve{T}_\ell\}_{\ell\geq 0}$ defined in (22). Let $\ell_R=\max\{\ell\geq 0:\breve{T}_\ell\leq R\}$. Define the *time-shifted-by-*R marked point process $(\{\vec{T}_\ell\}_{\ell\geq 0},\{\vec{I}_\ell\}_{\ell\geq 0})$ as follows:

$$\begin{aligned}&\vec{T}_0=R,\qquad \vec{I}_0=0,\\ \text{and}\quad&\vec{T}_\ell=\breve{T}_{\ell_R+\ell},\quad \vec{I}_\ell=I_{\ell_R+\ell},\quad \text{for } \ell\geq 1.\end{aligned}$$

The CIFs $\{\lambda_i(t\mid\mathscr{F}_t)\}_{i\in\mathcal{V}}$ are clearly piecewise-constant within the time interval $t\in[R,\infty)$, with their discontinuity points belonging to the set $\{\vec{T}_\ell\}_{\ell\geq 1}$. For $\ell\geq 0$, define $\mathscr{F}_{\vec{T}_\ell}=\{A\in\mathscr{F}:A\cap\{\vec{T}_\ell\leq t\}\in\mathscr{F}_t \text{ for every } t>0\}$ as the stopping time σ-algebra with respect to $\vec{T}_\ell$. Then we have the following fact.

(Theorem 1′) The statements in Theorem 1 still hold if $(\breve{T}_\ell,I_\ell,\mathcal{F}_{\breve{T}_\ell})$ is replaced by $(\vec{T}_\ell,\vec{I}_\ell,\mathcal{F}_{\vec{T}_\ell})$ for each integer $\ell\geq 0$.

For simplicity, we'll refer to this new version of Theorem 1 (modified for $(\vec{T}_\ell,\vec{I}_\ell,\mathcal{F}_{\vec{T}_\ell})$) as Theorem 1′. The proof of Theorem 1′ is similar to that of Theorem 1. We just need to replace (74) by $\{\{(\vec{T}_0,\vec{I}_0),\ldots,(\vec{T}_{\ell_R},\vec{I}_{\ell_R})\}=\{(\vec{t}_0,\vec{\imath}_0),\ldots,(\vec{t}_\ell,\vec{\imath}_\ell)\}\}=\bullet\in\mathcal{F}_{\vec{T}_\ell}$, and the remaining proof of Theorem 1′ follows the similar arguments as those in (75)–(88).

Next, we prove that

$$\begin{aligned}&(\vec{T}_{\ell+1}-\vec{T}_\ell,\vec{I}_{\ell+1},\boldsymbol{\lambda}(\vec{T}_{\ell+1}\mid\mathcal{F}_{\vec{T}_{\ell+1}}))\\&\qquad\text{is independent of }\mathcal{F}_R,\text{ for each integer }\ell\geq 0,\end{aligned}\quad(130)$$

where $\boldsymbol{\lambda}(t\mid\mathscr{F}_t)=(\lambda_1(t\mid\mathscr{F}_t),\ldots,\lambda_V(t\mid\mathscr{F}_t))^\top$ denotes the vector of CIFs. We start by proving the case of $\ell=0$. Using the facts that $\vec{T}_0=R$ and $\boldsymbol{N}((R-\phi,R])=\mathbf{0}$, we have $\vec{T}_0:=\bigcup_{i\in\mathcal{V}}\{t\in(\vec{T}_0,\vec{T}_0+\phi]:N_i(\{t-\phi\})=1\}=\varnothing$, implying that $\vec{T}_0^*=\infty$ (where $\vec{T}_\ell^*$ is the modified version of T_ℓ^* in (27)). Following result (ii) in Theorem 1′, we have that $(\vec{T}_1-\vec{T}_0)\mid\mathcal{F}_{\vec{T}_0}\sim\mathrm{Exp}(\lambda^{\mathrm{sum}}(\vec{T}_0\mid\mathcal{F}_{\vec{T}_0}))$. Combining (12), (16), (17), and the fact that $\boldsymbol{N}((R-\phi,R])=\mathbf{0}$ gives that $\lambda^{\mathrm{sum}}(\vec{T}_0\mid\mathcal{F}_{\vec{T}_0})=\sum_{i\in\mathcal{V}}\exp(\beta_{0;i})$ is a non-random constant. It follows that $\vec{T}_1-\vec{T}_0\sim\mathrm{Exp}(\sum_{i\in\mathcal{V}}\exp(\beta_{0;i}))$ is independent of $\mathcal{F}_{\vec{T}_0}$. Using result (iii) in Theorem 1′, we have $\mathrm{P}(\vec{I}_1=i\mid\sigma(\mathcal{F}_{\vec{T}_0},\vec{T}_1))=\lambda_i(\vec{T}_0\mid\mathcal{F}_{\vec{T}_0})/\lambda^{\mathrm{sum}}(\vec{T}_0\mid\mathcal{F}_{\vec{T}_0})=\exp(\beta_{0;i})/\sum_{j\in\mathcal{V}}\exp(\beta_{0;j})$, for each $i\in\mathcal{V}$. This implies that $(\vec{T}_1-\vec{T}_0,\vec{I}_1)$ is independent of $\mathcal{F}_{\vec{T}_0}$. Using $\boldsymbol{N}((R-\phi,R])=\mathbf{0}$, it is easy to show that $\boldsymbol{\lambda}(\vec{T}_1\mid\mathcal{F}_{\vec{T}_1})$ deterministically depends on $(\vec{T}_1-\vec{T}_0,\vec{I}_1)$. Therefore, we prove that $(\vec{T}_1-\vec{T}_0,\vec{I}_1,\boldsymbol{\lambda}(\vec{T}_1\mid\mathcal{F}_{\vec{T}_1})$ is independent of $\mathcal{F}_{\vec{T}_0}=\mathcal{F}_R$.

Suppose that (130) holds for $0,1,\ldots,\ell-1$, with $\ell\geq 1$. By induction, it suffices to prove the case of ℓ for (130). Using the fact that $\boldsymbol{N}((R-\phi,R])=\mathbf{0}$, we have that $\vec{T}_\ell^*-R$ deterministically depends on $\{(\vec{T}_k-\vec{T}_{k-1},\vec{I}_k)\}_{k=1,\ldots,\ell}$, which is independent of $\mathcal{F}_R$. Also, $\boldsymbol{\lambda}(\vec{T}_\ell\mid\mathcal{F}_{\vec{T}_\ell})$ is independent of $\mathcal{F}_R$. By results (ii) and (iii) in Theorem 1′, the distribution of $(\vec{T}_{\ell+1}-\vec{T}_\ell,\vec{I}_{\ell+1})$ conditional on $\mathcal{F}_{\vec{T}_\ell}$ deterministically depends on $(\boldsymbol{\lambda}(\vec{T}_\ell\mid\mathcal{F}_{\vec{T}_\ell}),\vec{T}_\ell^*-R)$, and thus is independent of $\mathcal{F}_R$. Finally, since $\boldsymbol{\lambda}(\vec{T}_{\ell+1}\mid\mathcal{F}_{\vec{T}_{\ell+1}})$ deterministically depends on $\{(\vec{T}_k-\vec{T}_{k-1},\vec{I}_k)\}_{k=1,\ldots,\ell+1}$, we prove that $(\vec{T}_{\ell+1}-\vec{T}_\ell,\vec{I}_{\ell+1},\boldsymbol{\lambda}(\vec{T}_{\ell+1}\mid\mathcal{F}_{\vec{T}_{\ell+1}}))$ is independent of $\mathcal{F}_R$. This proved (130).

Using similar arguments as in (24) and (25), we know that the counting process $\{\boldsymbol{N}(t+R)-\boldsymbol{N}(R)\}_{t\geq 0}$ and the marked point process $\{(\vec{T}_{\ell+1}-\vec{T}_\ell,\vec{I}_{\ell+1})\}_{\ell\geq 0}$ are equivalent to each other. Therefore, from (130), we conclude that $\boldsymbol{N}(t+R)-\boldsymbol{N}(R)$ is independent of $\mathcal{F}_R$ for each $t\geq 0$. This completes the proof. □

Now, we prove part (i) of Theorem 3. Recall the random processes $\lambda_i(t\mid\mathscr{F}_t)$ in (7) and $r_{j,\phi}(t)$ in (17). Define $\boldsymbol{\lambda}(t\mid\mathscr{F}_t)=(\lambda_1(t\mid\mathscr{F}_t),\ldots,\lambda_V(t\mid\mathscr{F}_t))^\top$ and $\boldsymbol{r}_\phi(t)=(r_{1,\phi}(t),\ldots,r_{V,\phi}(t))^\top$ as the vectors of these random processes. For $t\geq 0$, define the following *time-shifted-by-*R_ℓ random processes:

$$\begin{aligned}\vec{\boldsymbol{N}}(t)&=\boldsymbol{N}(t+R_\ell)=(\vec{N}_1(t),\ldots,\vec{N}_V(t))^\top,\\ \vec{\boldsymbol{\lambda}}(t\mid\vec{\mathscr{F}}_t)&=\boldsymbol{\lambda}(t+R_\ell\mid\mathcal{F}_{t+R_\ell})\\&=(\vec{\lambda}_1(t\mid\vec{\mathscr{F}}_t),\ldots,\vec{\lambda}_V(t\mid\vec{\mathscr{F}}_t))^\top,\\ \vec{\boldsymbol{r}}_\phi(t)&=\boldsymbol{r}_\phi(t+R_\ell)=(\vec{r}_{1,\phi}(t),\ldots,\vec{r}_{V,\phi}(t))^\top.\end{aligned}\quad(131)$$

Here, $\vec{\mathscr{F}}_t=\mathcal{F}_{t+R_\ell}=\{A\in\mathscr{F}:A\cap\{t+R_\ell\leq u\}\in\mathscr{F}_u \text{ for every } u>t\}$. Also, for $s<t$, let $\vec{N}_i((s,t])=\vec{N}_i(t\vee 0)-\vec{N}_i(s\vee 0)$. We have the three facts below:

Fact (a): $\vec{\lambda}_i(t\mid\vec{\mathscr{F}}_t)$ is the CIF of $\vec{N}_i(t)$. This is because

$$\begin{aligned}&\vec{\lambda}_i(t\mid\vec{\mathscr{F}}_t)\\&\quad=\lambda_i(t+R_\ell\mid\mathcal{F}_{t+R_\ell})\\&\quad=\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(N_i(t+R_\ell+\Delta)=N_i(t+R_\ell)+1\mid\mathcal{F}_{t+R_\ell})\\&\quad=\lim_{\Delta\downarrow 0}\Delta^{-1}\mathrm{P}(\vec{N}_i(t+\Delta)=\vec{N}_i(t)+1\mid\vec{\mathscr{F}}_t),\quad t\geq 0,\end{aligned}$$

agreeing with the definition in (7) of a CIF.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Fact (b): $\vec{\lambda}_i(t \mid \vec{\mathscr{F}}_t)$, $\vec{r}_{j,\phi}(t)$, and $\vec{N}_j(t)$ follow the same model as in (12), (16), and (17). This could be seen from the following identities:

$$\begin{aligned}\vec{\lambda}_i(t \mid \vec{\mathscr{F}}_t) &= \exp\Big\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, g\big(r_{j,\phi}(t+R_\ell)\big)\Big\} \\ &= \exp\Big\{\beta_{0;i} + \sum_{j\in\mathcal{V}} \beta_{j,i}\, g\big(\vec{r}_{j,\phi}(t)\big)\Big\}, \quad t \geq 0,\end{aligned} \tag{132}$$

which is identical to (12) and (16), and

$$\begin{aligned}\vec{r}_{j,\phi}(t) &= N_j\big((t+R_\ell-\phi,\, t+R_\ell]\big)/\phi \\ &= N_j\big(((t+R_\ell-\phi)\vee R_\ell,\, t+R_\ell]\big)/\phi \\ &= \vec{N}_j\big((t-\phi,\, t]\big)/\phi, \qquad t \geq 0,\end{aligned} \tag{133}$$

which is identical to (17). Here, (133) follows from the fact that $N_j((R_\ell-\phi, R_\ell]) = 0$, which is implied by (43).

Fact (c): The following mappings are deterministic, where $\mathrm{M1} = \mathrm{M1}'$, $\mathrm{M2} = \mathrm{M2}'$, $\mathrm{M3} = \mathrm{M3}'$, $\mathrm{M4} = \mathrm{M4}'$, and $\mathrm{M5} = \mathrm{M5}'$:

(M1) from $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0}$ to $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}((t+R_\ell-\phi)\vee R_\ell)\}_{t\geq 0}$

(M2) from $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0}$ to $\{\vec{\boldsymbol{r}}_\phi(t)\}_{t\geq 0}$.

(M3) from $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0}$ to $\{\vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t)\}_{t\geq 0}$.

(M4) from $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0}$ to $R_{\ell+1} - R_\ell$.

(M5) from $\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0}$ to $\boldsymbol{N}((R_\ell, R_{\ell+1}])$.

And

(M1′) from $\{\boldsymbol{N}(t)\}_{t\geq 0}$ to $\{\boldsymbol{N}(t) - \boldsymbol{N}(t-\phi)\}_{t\geq 0}$.

(M2′) from $\{\boldsymbol{N}(t)\}_{t\geq 0}$ to $\{\boldsymbol{r}_\phi(t)\}_{t\geq 0}$.

(M3′) from $\{\boldsymbol{N}(t)\}_{t\geq 0}$ to $\{\boldsymbol{\lambda}(t \mid \mathscr{F}_t)\}_{t\geq 0}$.

(M4′) from $\{\boldsymbol{N}(t)\}_{t\geq 0}$ to R_1.

(M5′) from $\{\boldsymbol{N}(t)\}_{t\geq 0}$ to $\boldsymbol{N}(R_1)$.

To show this, note that for any $t \geq 0$, the following two identities hold:

$$\begin{aligned}&\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}((t+R_\ell-\phi)\vee R_\ell) \\ &\quad= \{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\} \\ &\qquad - \{\boldsymbol{N}((t+R_\ell-\phi)\vee R_\ell) - \boldsymbol{N}(R_\ell)\} \\ &\quad= \{\boldsymbol{N}(u+R_\ell) - \boldsymbol{N}(R_\ell)\}|_{u=t} \\ &\qquad - \{\boldsymbol{N}(u+R_\ell) - \boldsymbol{N}(R_\ell)\}|_{u=(t-\phi)\vee 0},\end{aligned}$$

$$\boldsymbol{N}(t) - \boldsymbol{N}(t-\phi) = \boldsymbol{N}(u)|_{u=t} - \boldsymbol{N}(u)|_{u=(t-\phi)\vee 0},$$

which implies that the mapping M1 is deterministic, and $\mathrm{M1} = \mathrm{M1}'$. This, combined with (133) and (17), gives that the mappings $\mathrm{M2} = \mathrm{M2}'$ are deterministic. By using (132), (12), (16), and the fact that the mappings $\mathrm{M2} = \mathrm{M2}'$ are deterministic, we conclude that the mappings $\mathrm{M3} = \mathrm{M3}'$ are deterministic. From (43), we observe that

$$R_1 = \min\{t \geq \phi : \boldsymbol{N}(t) - \boldsymbol{N}(t-\phi) = \boldsymbol{0}\},$$

$$\begin{aligned}R_{\ell+1} - R_\ell = \min\{t \geq \phi :\ &\\ &\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}((t+R_\ell-\phi)\vee R_\ell) = \boldsymbol{0}\}.\end{aligned}$$

This, together with the fact that $\mathrm{M1} = \mathrm{M1}'$ are deterministic, yields that the mappings $\mathrm{M4} = \mathrm{M4}'$ are deterministic. From

$$\boldsymbol{N}((R_\ell, R_{\ell+1}]) = \boldsymbol{N}((R_{\ell+1} - R_\ell) + R_\ell) - \boldsymbol{N}(R_\ell),$$

we infer that the mapping from $(\{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\geq 0},\, R_{\ell+1} - R_\ell)$ to $\boldsymbol{N}((R_\ell, R_{\ell+1}])$ is deterministic and identical to that from $(\{\boldsymbol{N}(t)\}_{t\geq 0},\, R_1)$ to $\boldsymbol{N}(R_1)$. This, along with $\mathrm{M4} = \mathrm{M4}'$, concludes that $\mathrm{M5} = \mathrm{M5}'$ are deterministic.

Fact (a) confirms that $\vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t)$ represents the vector of CIFs of $\vec{\boldsymbol{N}}(t)$. Leveraging Fact (c), which establishes deterministic mappings $\mathrm{M3} = \mathrm{M3}'$, we deduce that $(\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell), \vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t)) \overset{\mathrm{D}}{=} (\boldsymbol{N}(t), \boldsymbol{\lambda}(t \mid \mathscr{F}_t))$ for any $t \geq 0$. This demonstrates $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell) \overset{\mathrm{D}}{=} \boldsymbol{N}(t)$. Moreover, from (132) and (133), it's evident that for any $t \geq 0$, $\vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t)$ is a deterministic function of $\{\boldsymbol{N}(s+R_\ell) - \boldsymbol{N}(R_\ell)\}_{0\leq s\leq t}$. Utilizing Lemma B.11, we establish that $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)$ is independent of $\mathcal{F}_{R_\ell}$ for any $t \geq 0$. This, coupled with the deterministic nature of the mapping M3, validates that $\vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t)$ is also independent of $\mathcal{F}_{R_\ell}$. This concludes the proof of part (i). □

2) Proof of Part (ii)*:* Utilizing the deterministic nature of the mappings $\mathrm{M5} = \mathrm{M5}'$ and the result from Theorem 3 (i) stating $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell) \overset{\mathrm{D}}{=} \boldsymbol{N}(t)$ for any $t \geq 0$, we derive $\boldsymbol{N}((R_\ell, R_{\ell+1}]) \overset{\mathrm{D}}{=} \boldsymbol{N}(R_1)$. Thus, $\{\boldsymbol{N}((R_\ell, R_{\ell+1}])\}_{\ell\geq 0}$ forms a sequence of identically distributed random vectors. On the other hand, given that $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)$ is independent of $\mathcal{F}_{R_\ell}$ and the mapping M5 is deterministic, we infer that $\boldsymbol{N}((R_\ell, R_{\ell+1}])$ is independent of $\mathcal{F}_{R_\ell}$. Furthermore, $\boldsymbol{N}((R_k, R_{k+1}])$ is $\mathcal{F}_{R_\ell}$-measurable for any $0 \leq k \leq \ell-1$. As a result, $\boldsymbol{N}((R_\ell, R_{\ell+1}])$ is independent of $\{\boldsymbol{N}((R_k, R_{k+1}])\}_{0\leq k\leq \ell-1}$. Hence, $\{\boldsymbol{N}((R_\ell, R_{\ell+1}])\}_{\ell\geq 0}$ constitutes a sequence of independent random vectors. This completes the proof. □

3) Proof of Part (iii)*:* Before proving part (iii), we'll establish Lemmas B.12 and B.13.

Lemma B.12: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let R_1 be the first *recurrence time point* as defined in (43) with $\ell = 1$. Then $\mathrm{E}(R_1^2) < \infty$.

Proof: Consider the random variable $L^* = \min\{k \geq 1 : \mathrm{I}(A_k^*) = 1\}$, where the event $A_k^* = \{\boldsymbol{N}(((k-1)\,\phi, k\,\phi]) = \boldsymbol{0}\}$ is defined in (125). According to (127), for any integer $\ell > 2$, we have:

$$\mathrm{P}(L^* \geq \ell) = \mathrm{P}\Big(\bigcap_{k=1}^{\ell-1} \overline{A_k^*}\Big) \leq \big\{1 - \exp(-c\,\phi)\big\}^{\ell-1}.$$

Thus, the tail probability of L^* is bounded by that of the geometric distribution with a constant success probability of $\exp(-c\,\phi)$. As the geometric distribution has finite first and second moments, we derive $\mathrm{E}\{(L^*)^2\} < \infty$. Moreover, $\mathrm{I}(A_{L^*}^*) = 1$ implies that $\boldsymbol{N}(((L^*-1)\,\phi, L^*\,\phi]) = \boldsymbol{0}$. This, combined with (43), yield $R_1 \leq L^*\,\phi$. Consequently:

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$\mathrm{E}(R_1^2) \le \mathrm{E}\{(L^*\phi)^2\} = \phi^2\,\mathrm{E}\{(L^*)^2\} < \infty.$$

This completes the proof. □

Lemma B.13: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let $h(\cdot) : \mathbb{R}^{2V} \to \mathbb{R}$ be a continuous function. For each integer $\ell \ge 1$, define the random variable:

$$\mathrm{S}_\ell = \int_{R_{\ell-1}}^{R_\ell} h(\boldsymbol{\lambda}(t \mid \mathscr{F}_t), \boldsymbol{r}_\phi(t))\,\mathrm{d}t. \tag{134}$$

Then $\{\mathrm{S}_\ell\}_{\ell\ge1}$ is a sequence of i.i.d. random variables.

Proof: Following the notations in (131), for $\ell \ge 1$, we get:

$$\begin{aligned}\mathrm{S}_{\ell+1} &= \int_{R_\ell}^{R_{\ell+1}} h(\boldsymbol{\lambda}(t \mid \mathscr{F}_t), \boldsymbol{r}_\phi(t))\,\mathrm{d}t \\ &= \int_0^{R_{\ell+1}-R_\ell} h(\vec{\boldsymbol{\lambda}}(t \mid \vec{\mathscr{F}}_t), \vec{\boldsymbol{r}}_\phi(t))\,\mathrm{d}t. \end{aligned}\tag{135}$$

By comparing (135) with $\mathrm{S}_1 = \int_0^{R_1} h(\boldsymbol{\lambda}(t \mid \mathscr{F}_t), \boldsymbol{r}_\phi(t))\,\mathrm{d}t$ and using the fact that the mappings $\mathrm{M2} = \mathrm{M2}'$, $\mathrm{M3} = \mathrm{M3}'$, and $\mathrm{M4} = \mathrm{M4}'$ are deterministic, we deduce that the mappings

$$\begin{aligned}&(\mathrm{M6}) \quad \text{from } \{\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)\}_{t\ge0} \text{ to } \mathrm{S}_{\ell+1},\\ &(\mathrm{M6}') \quad \text{from } \{\boldsymbol{N}(t)\}_{t\ge0} \text{ to } \mathrm{S}_1,\end{aligned}$$

are both deterministic, with $\mathrm{M6} = \mathrm{M6}'$. Combining this with the fact that $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell) \overset{\mathrm{D}}{=} \boldsymbol{N}(t)$ for any $t \ge 0$ from Theorem 3 (i), we obtain $\mathrm{S}_{\ell+1} \overset{\mathrm{D}}{=} \mathrm{S}_1$. Thus, $\{\mathrm{S}_\ell\}_{\ell\ge1}$ is a sequence of identically distributed random variables. On the other hand, using the facts that $\boldsymbol{N}(t+R_\ell) - \boldsymbol{N}(R_\ell)$ is independent of $\mathcal{F}_{R_\ell}$ and the mapping M6 is deterministic, we conclude that $\mathrm{S}_{\ell+1}$ is independent of $\mathcal{F}_{R_\ell}$. Also, S_k is $\mathcal{F}_{R_\ell}$-measurable for $1 \le k \le \ell$. Hence, $\{\mathrm{S}_\ell\}_{\ell\ge1}$ is a sequence of independent variables. This completes the proof. □

Now, we prove part (iii) of Theorem 3. Note that $D_\ell = R_\ell - R_{\ell-1}$ is a special case of S_ℓ in equation (134) with $h(\cdot) \equiv 1$. By applying Lemma B.13, we deduce that $\{D_\ell\}_{\ell\ge1}$ forms a sequence of i.i.d. random variables. Furthermore, Lemma B.12 proved that $D_1 = R_1$ has a finite second moment. This finalizes the proof. □

K. Proof of Theorem 4

Before proving Theorem 4, we establish Lemma B.14.

Lemma B.14 (Asymptotic Convergence of $\Lambda_i(t) = \int_0^t \lambda_i(u \mid \mathscr{F}_u)\,\mathrm{d}u$): Assume conditions A1, A2, A3, A4, and A5 in Appendix B. For each $i \in \mathcal{V}$, consider the random process $\Lambda_i(t) = \int_0^t \lambda_i(u \mid \mathscr{F}_u)\,\mathrm{d}u$ for $t > 0$. Then there exists a constant $c_i \in (0,\infty)$ such that

$$\Lambda_i(t)/t \xrightarrow{\mathrm{P}} c_i, \quad \text{as } t \to \infty. \tag{136}$$

Proof: Let the increment of $\Lambda_i(t)$ in the ℓth *recurrence cycle* $(R_{\ell-1}, R_\ell]$ be denoted as

$$\mathrm{S}_{i,\ell} = \Lambda_i(R_\ell) - \Lambda_i(R_{\ell-1}) = \int_{R_{\ell-1}}^{R_\ell} \lambda_i(t \mid \mathscr{F}_t)\,\mathrm{d}t.$$

Applying $h(\boldsymbol{\lambda}(t \mid \mathscr{F}_t), \boldsymbol{r}_\phi(t)) = \lambda_i(t \mid \mathscr{F}_t)$ to (134) in Lemma B.13 indicates that $\{\mathrm{S}_{i,\ell}\}_{\ell\ge1}$ is a sequence of i.i.d. random variables.

Under condition A5, there exists a constant $c \in (0,\infty)$ such that:

$$\mathrm{S}_{i,\ell} = \int_{R_{\ell-1}}^{R_\ell} \lambda_i(t \mid \mathscr{F}_t)\,\mathrm{d}t \le \int_{R_{\ell-1}}^{R_\ell} c\,\mathrm{d}t = c\,D_\ell, \tag{137}$$

where $D_\ell = R_\ell - R_{\ell-1}$. Combining Lemma B.12 and equation (137) implies that the second moments of D_ℓ and $\mathrm{S}_{i,\ell}$ are finite. Applying the strong law of large numbers, we derive:

$$\frac{1}{\ell}\sum_{k=1}^{\ell} \mathrm{S}_{i,k} \overset{\text{a.s.}}{\to} \mathrm{E}(\mathrm{S}_{i,1}), \quad \frac{1}{\ell}\sum_{k=1}^{\ell} D_k \overset{\text{a.s.}}{\to} \mathrm{E}(D_1), \quad \text{as } \ell \to \infty.$$

Thus, for arbitrarily small $\delta > 0$ and $\epsilon > 0$, there exists a sufficiently large C_1 such that:

$$\begin{aligned}\mathrm{P}\Big(\sup_{\ell > C_1} \max\Big\{ &\Big|\frac{1}{\ell}\sum_{k=1}^{\ell} \mathrm{S}_{i,k} - \mathrm{E}(\mathrm{S}_{i,1})\Big|, \\ &\Big|\frac{1}{\ell}\sum_{k=1}^{\ell} D_k - \mathrm{E}(D_1)\Big|\Big\} > \epsilon\Big) < \delta. \end{aligned}\tag{138}$$

On the other hand, for any time point $t > 0$, let $L_t = \sup\{\ell \ge 0 : R_\ell \le t\}$ be the number of recurrence time points up to t. We have:

$$\sum_{k=1}^{L_t} \mathrm{S}_{i,k} \le \Lambda_i(t) \le \sum_{k=1}^{L_t+1} \mathrm{S}_{i,k}, \quad \sum_{k=1}^{L_t} D_k \le t \le \sum_{k=1}^{L_t+1} D_k,$$

which directly yields:

$$\frac{\sum_{k=1}^{L_t} \mathrm{S}_{i,k}}{\sum_{k=1}^{L_t+1} D_k} \le \frac{\Lambda_i(t)}{t} \le \frac{\sum_{k=1}^{L_t+1} \mathrm{S}_{i,k}}{\sum_{k=1}^{L_t} D_k}. \tag{139}$$

Since $\mathrm{E}(D_1^2) < \infty$, it's straightforward to show that $L_t \xrightarrow{\mathrm{P}} \infty$ as $t \to \infty$. Thus, for arbitrarily small $\delta_2 > 0$ and large $C_2 > C_1$, there exists $t_0 > 0$ such that for all $t > t_0$, $\mathrm{P}(L_t > C_2) > 1 - \delta_2$. Combining (138) and (139), the following (140) holds with probability at least $1 - \delta - \delta_2$ for $t > t_0$:

$$\frac{C_2\,\{\mathrm{E}(\mathrm{S}_{i,1}) - \epsilon\}}{(C_2+1)\,\{\mathrm{E}(D_1) + \epsilon\}} \le \frac{\Lambda_i(t)}{t} \le \frac{(C_2+1)\,\{\mathrm{E}(\mathrm{S}_{i,1}) + \epsilon\}}{C_2\,\{\mathrm{E}(D_1) - \epsilon\}}. \tag{140}$$

Since ϵ, δ, and δ_2 are arbitrarily small and C_2 is arbitrarily large, (140) implies that

$$\Lambda_i(t)/t \xrightarrow{\mathrm{P}} \mathrm{E}(\mathrm{S}_{i,1})/\mathrm{E}(D_1), \quad \text{as } t \to \infty.$$

We complete the proof by setting $c_i = \mathrm{E}(\mathrm{S}_{i,1})/\mathrm{E}(D_1)$ in (136). □

Now, we prove Theorem 4. According to Theorem 2, for each $i \in \mathcal{V}$ and $t \in (0,\infty)$, we have $\mathrm{var}\{(N_i(t) - \Lambda_i(t))/t\} \le c_1/t$. This, together with Lemma 7, implies that

$$\{N_i(t) - \Lambda_i(t)\}/t \xrightarrow{\mathrm{P}} 0, \quad \text{as } t \to \infty. \tag{141}$$

Utilizing Lemma B.14 and (141), we derive

$$\boldsymbol{N}(t)/t \xrightarrow{\mathrm{P}} \boldsymbol{c}_0, \quad \text{as } t \to \infty,$$

where $\boldsymbol{c}_0 = (\mathrm{E}(\mathrm{S}_{1,1}), \ldots, \mathrm{E}(\mathrm{S}_{V,1}))^\top/\mathrm{E}(D_1)$. This completes the proof. □

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

L. Proof of Theorem 5

Before proving Theorem 5, we establish Lemmas B.15, B.16, and B.17.

Lemma B.15: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let $f(\cdot) : \mathbb{R}^V \to [0, \infty)$ be a non-negative continuous function bounded above by $c_0 \in (0, \infty)$. For $t \geq 0$, let $Y(t) = f(\boldsymbol{r}_\phi(t))$, where $\boldsymbol{r}_\phi(t)$ is defined above (131). Let R_1 be the first recurrence time point defined in (43) with $\ell = 1$. Assume that $\mathrm{E}\{\int_0^{R_1} Y(t)\,\mathrm{d}t\} > 0$. Then, there exists a constant $c_i \in (0, \infty)$ such that:

$$\int_0^t \lambda_i(u \mid \mathscr{F}_u)\, Y(u)\,\mathrm{d}u/t \xrightarrow{\mathrm{P}} c_i, \quad \text{as } t \to \infty.$$

Proof: For each integer $\ell \geq 1$, define

$$\mathrm{S}^*_{i,\ell} = \int_{R_{\ell-1}}^{R_\ell} \lambda_i(t \mid \mathscr{F}_t)\, Y(t)\,\mathrm{d}t.$$

Applying Lemma B.13 with $h(\boldsymbol{\lambda}(t \mid \mathscr{F}_t), \boldsymbol{r}_\phi(t \mid \mathscr{F}_t)) = \lambda_i(t \mid \mathscr{F}_t) f(\boldsymbol{r}_\phi(t)) = \lambda_i(t \mid \mathscr{F}_t)\, Y(t)$, we have that $\{\mathrm{S}^*_{i,\ell}\}_{\ell \geq 1}$ is a sequence of i.i.d. random variables. By condition A5, there exist constants c_2 and c_3 in $(0, \infty)$ such that $c_2 \leq \lambda_i(t \mid \mathscr{F}_t) \leq c_3$ for any $t \in [0, \infty)$. We obtain the following moment inequalities:

$$\begin{aligned}
\mathrm{E}(\mathrm{S}^*_{i,1}) &= \mathrm{E}\Big\{ \int_0^{R_1} \lambda_i(t \mid \mathscr{F}_t)\, Y(t)\,\mathrm{d}t \Big\} \\
&\geq c_2\, \mathrm{E}\Big\{ \int_0^{R_1} Y(t)\,\mathrm{d}t \Big\} > 0, \\
\mathrm{E}\{(\mathrm{S}^*_{i,1})^2\} &= \mathrm{E}\Big[\Big\{ \int_0^{R_1} \lambda_i(t \mid \mathscr{F}_t)\, Y(t)\,\mathrm{d}t \Big\}^2 \Big] \\
&\leq c_3^2\, c_0^2\, \mathrm{E}(R_1^2) < \infty.
\end{aligned}$$

Applying a similar proof as Lemma B.14 with $\mathrm{S}_{i,\ell} = \mathrm{S}^*_{i,\ell}$, one can demonstrate that:

$$\begin{aligned}
\frac{\int_0^t \lambda_i(u \mid \mathscr{F}_u)\, Y(u)\,\mathrm{d}u}{t} &\xrightarrow{\mathrm{P}} \frac{\mathrm{E}\{\int_0^{R_1} \lambda_i(u \mid \mathscr{F}_u)\, Y(u)\,\mathrm{d}u\}}{\mathrm{E}(R_1)} \\
&= \frac{\mathrm{E}(\mathrm{S}^*_{i,1})}{\mathrm{E}(R_1)} > 0, \quad \text{as } t \to \infty.
\end{aligned}$$

This completes the proof. □

Lemma B.16: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. For $i \in \mathcal{V}$, let $x_i(t) = g(r_{i,\phi}(t))$ be the covariate defined in (16), and $x_0(t) \equiv 1$. Then, for any $i, j \in \mathcal{V} \cup \{0\}$ (not necessarily distinct), we have:

$$\mathrm{E}\Big\{ \int_0^{R_1} x_i(u)\,\mathrm{d}u \Big\} > 0, \tag{142}$$

$$\mathrm{E}\Big\{ \int_0^{R_1} x_i(u)\, x_j(u)\,\mathrm{d}u \Big\} > 0. \tag{143}$$

Proof: If $i = 0$, then (142) obviously holds. If either i or j is zero, then (143) reduces to (142). Thus, to prove Lemma B.16, it suffices to verify (142) and (143) for the case of $i, j \in \mathcal{V}$.

By a proof similar to that below (121), for any $t > 0$, we have:

$$\begin{aligned}
\mathrm{P}(N_i(t) \geq 1) &= 1 - \mathrm{P}(N_i(t) = 0) \\
&= 1 - \exp\{-\lambda_i(0) \cdot t\} > 0.
\end{aligned} \tag{144}$$

The OM condition (11) implies that:

$$\mathrm{P}(N_i(t)\, N_j(t) \geq 1) = \lambda_i(0)\, \lambda_j(0)\, t^2 + o(t^2), \text{ as } t \to 0,$$

and thus, there exists $t_0 \in (0, \phi)$ such that:

$$\mathrm{P}(N_i(t_0)\, N_j(t_0) \geq 1) > 0. \tag{145}$$

For $t \in (0, \phi)$, observing that $r_{i,\phi}(t) = N_i((t - \phi, t])/\phi = N_i(t)/\phi$, we deduce that $r_{i,\phi}(t)$ is increasing in $t \in (0, \phi)$. This implies $x_i(t) = g(r_{i,\phi}(t))$ is also increasing in $t \in (0, \phi)$. Along with (144), (145), and considering $R_1 \geq \phi > t_0$, we derive:

$$\begin{aligned}
\mathrm{E}\Big\{ \int_0^{R_1} x_i(u)\,\mathrm{d}u \Big\} &\geq \mathrm{E}\{x_i(t_0) \cdot (\phi - t_0)\} > 0, \\
\mathrm{E}\Big\{ \int_0^{R_1} x_i^2(u)\,\mathrm{d}u \Big\} &\geq \mathrm{E}\{x_i^2(t_0) \cdot (\phi - t_0)\} > 0, \\
\mathrm{E}\Big\{ \int_0^{R_1} x_i(u)\, x_j(u)\,\mathrm{d}u \Big\} &\geq \mathrm{E}\{x_i(t_0)\, x_j(t_0) \cdot (\phi - t_0)\} > 0.
\end{aligned}$$

These complete the proof. □

Lemma B.17: Assume conditions A1, A2, A3, A4, and A5 in Appendix B. Let $\widetilde{\boldsymbol{x}}(t) = (1, x_1(t), x_2(t), \ldots, x_V(t))^\top$ be the vector of covariates defined in (16). Then, for any $\widetilde{\boldsymbol{u}} \in \mathbb{R}^{V+1}$ with $\|\widetilde{\boldsymbol{u}}\| > 0$,

$$\mathrm{E}\Big[\int_0^{R_1} \{\widetilde{\boldsymbol{x}}(t)^\top \widetilde{\boldsymbol{u}}\}^2\,\mathrm{d}t \Big] > 0.$$

Proof: Let $\widetilde{\boldsymbol{u}} = (u_0, u_1, \ldots, u_V)^\top$. We proceed by cases of u_0.

Case (i): $u_0 \neq 0$. Consider $t_0 \in (0, \phi)$. By (124), we have

$$\mathrm{P}(\boldsymbol{N}(t_0) = \boldsymbol{0}) \geq \exp\{-c \cdot (t_0 - 0)\} > 0. \tag{146}$$

Note that $\boldsymbol{N}(t_0) = \boldsymbol{0}$ implies that $\widetilde{\boldsymbol{x}}(t)^\top \widetilde{\boldsymbol{u}} = u_0$ for $t \in [0, t_0]$. Combining (146) and the fact that $R_1 \geq \phi > t_0$, we obtain:

$$\begin{aligned}
&\mathrm{E}\Big[\int_0^{R_1} \{\widetilde{\boldsymbol{x}}(t)^\top \widetilde{\boldsymbol{u}}\}^2\,\mathrm{d}t \Big] \\
&\quad \geq \mathrm{E}\Big[\int_0^{t_0} \{\widetilde{\boldsymbol{x}}(t)^\top \widetilde{\boldsymbol{u}}\}^2 \cdot \mathrm{I}(\boldsymbol{N}(t_0) = \boldsymbol{0})\,\mathrm{d}t \Big] \\
&\quad = t_0\, u_0^2\, \mathrm{P}(\boldsymbol{N}(t_0) = \boldsymbol{0}) > 0.
\end{aligned}$$

Case (ii): $u_0 = 0$. Since $\|\widetilde{\boldsymbol{u}}\| > 0$ and $u_0 = 0$, there exists $i \in \mathcal{V}$ such that $u_i \neq 0$. We have:

$$\begin{aligned}
&\mathrm{P}(N_j(t) = 0 \text{ for all } j \in \mathcal{V} \setminus \{i\}, \quad N_i(t) = 1) \\
&\quad \geq \mathrm{P}(N_i(t) = 1) - \sum_{j \in \mathcal{V} \setminus \{i\}} \mathrm{P}(N_i(t) = 1, N_j(t) = 1) \\
&\qquad - \sum_{j \in \mathcal{V} \setminus \{i\}} \mathrm{P}(N_j(t) > 1) \\
&\quad = \lambda_i(0)\, t + o(t) - \sum_{j \in \mathcal{V} \setminus \{i\}} \{\lambda_i(0)\, \lambda_j(0)\, t^2 + o(t^2)\} \\
&\qquad - o(t) \\
&\quad = \lambda_i(0)\, t + o(t), \text{ as } t \to 0,
\end{aligned} \tag{147}$$

where (147) is derived from (7), (8), and (11). Hence, there exists $t_0 \in (0, \phi)$ such that

$$\mathrm{P}(N_j(t_0) = 0 \text{ for all } j \in \mathcal{V} \setminus \{i\},\ N_i(t_0) = 1) > 0. \tag{148}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

Let $t_1 \in (t_0, \phi)$. From (124) and (148), we have:

$$\begin{aligned}
&\mathrm{P}(N_j(t_1)=0 \text{ for all } j \in \mathcal{V}\setminus\{i\},\ N_i(t_0)=N_i(t_1)=1)\\
&= \mathrm{P}(\boldsymbol{N}(t_1)=\boldsymbol{N}(t_0)\mid \\
&\qquad N_j(t_0)=0 \text{ for all } j\in\mathcal{V}\setminus\{i\},\ N_i(t_0)=1)\\
&\quad \times \mathrm{P}(N_j(t_0)=0 \text{ for all } j\in\mathcal{V}\setminus\{i\},\ N_i(t_0)=1)\\
&\geq \exp\{-c\cdot(t_1-t_0)\}\\
&\quad \times \mathrm{P}(N_j(t_0)=0 \text{ for all } j\in\mathcal{V}\setminus\{i\},\ N_i(t_0)=1)\\
&> 0. \qquad (149)
\end{aligned}$$

Note that the event $\{N_j(t_1) = 0 \text{ for all } j \in \mathcal{V}\setminus\{i\}, \text{ and } N_i(t_0) = N_i(t_1) = 1\}$ implies that $\boldsymbol{x}(t)^\top \widetilde{\boldsymbol{u}} = x_i(t)\, u_i = g(1/\phi)\, u_i$ for $t \in (t_0, t_1)$. Along with (149) and the fact that $R_1 \geq \phi > t_1$, we obtain:

$$\begin{aligned}
&\mathrm{E}\Big[\int_0^{R_1} \{\widetilde{\boldsymbol{x}}(t)^\top \widetilde{\boldsymbol{u}}\}^2\,\mathrm{d}t\Big]\\
&\quad \geq (t_1-t_0)\, g^2(1/\phi)\, u_i^2 \times \mathrm{P}\big(N_j(t_1)=0\\
&\qquad \text{for all } j\in\mathcal{V}\setminus\{i\}, N_i(t_0)=N_i(t_1)=1\big) > 0.
\end{aligned}$$

Combining the results of Cases (i) and (ii) completes the proof. □

Now, we prove Theorem 5. Recall (48):

$$\begin{aligned}
&\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)\\
&= \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}t - \{\widetilde{\boldsymbol{x}}_i(t-)^\top \widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}N_i(t)\Big].
\end{aligned}$$

With some algebra, we obtain the gradient vector and Hessian matrix of $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$:

$$\begin{aligned}
\nabla\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) &= \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\widetilde{\boldsymbol{x}}_i(t)\,\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}t\\
&\qquad - \widetilde{\boldsymbol{x}}_i(t-)\,\mathrm{d}N_i(t)\Big], \qquad (150)\\
\nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) &= \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \widetilde{\boldsymbol{x}}_i(t)\,\widetilde{\boldsymbol{x}}_i(t)^\top \times \exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i\}\,\mathrm{d}t. \qquad (151)
\end{aligned}$$

Let $\widetilde{x}_{i,j}(t)$ denote the jth component of $\widetilde{\boldsymbol{x}}_i(t)$:

$$\widetilde{x}_{i,j}(t) = \begin{cases} 1, & \text{if } j=1,\\ x_{j-1}(t), & \text{if } 1<j\leq i,\\ x_j(t), & \text{if } i<j\leq V. \end{cases} \qquad (152)$$

For each $j \in \mathcal{V}$, employing (38) in Lemma 7 yields:

$$\begin{aligned}
&\mathrm{E}\Big(\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\widetilde{x}_{i,j}(t)\,\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t - \widetilde{x}_{i,j}(t-)\,\mathrm{d}N_i(t)\Big]\Big)\\
&= 0. \qquad (153)
\end{aligned}$$

Also, using (42) in Theorem 2, we have:

$$\begin{aligned}
&\mathrm{var}\Big(\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\widetilde{x}_{i,j}(t)\,\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t - \widetilde{x}_{i,j}(t-)\,\mathrm{d}N_i(t)\Big]\Big)\\
&\leq c_1/\mathrm{T}, \qquad (154)
\end{aligned}$$

where $c_1 \in (0,\infty)$ is a constant. By Chebyshev's inequality, (153), and (154), we conclude:

$$\begin{aligned}
&\frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \Big[\widetilde{\boldsymbol{x}}_i(t)\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t - \widetilde{\boldsymbol{x}}_i(t-)\,\mathrm{d}N_i(t)\Big]\\
&= O_{\mathrm{P}}(\sqrt{1/\mathrm{T}}). \qquad (155)
\end{aligned}$$

This confirms (49) in part (i).

Next, we prove part (ii). Define $\mathbf{H}_{i,\mathrm{T}} = \nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)$. For any $j,k \in \mathcal{V}$, from (151) and (152), the (j,k)th entry of $\mathbf{H}_{i,\mathrm{T}}$ is given by:

$$\mathbf{H}_{i,\mathrm{T}}(j,k) = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \widetilde{x}_{i,j}(t)\,\widetilde{x}_{i,k}(t)\,\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t.$$

Let $Y_1(t) = \widetilde{x}_{i,j}(t)\,\widetilde{x}_{i,k}(t)$. Lemma B.16 proved that $\mathrm{E}\{\int_0^{R_1} Y_1(t)\,\mathrm{d}t\} > 0$. This allows us to apply Lemma B.15 with $Y(t) = Y_1(t)$, yielding:

$$\mathbf{H}_{i,\mathrm{T}}(j,k) = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \widetilde{x}_{i,j}(t)\,\widetilde{x}_{i,k}(t)\,\lambda_i^*(t\mid\mathscr{F}_t)\,\mathrm{d}t \xrightarrow{\mathrm{P}} c_{j,k},$$

as $\mathrm{T}\to\infty$, where $c_{j,k} \in (0,\infty)$ is some constant. Denote by $\mathbf{C}_i = (c_{j,k})_{V\times V}$, a matrix with all positive entries. It follows that:

$$\nabla^2\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) = \mathbf{H}_{i,\mathrm{T}} \xrightarrow{\mathrm{P}} \mathbf{C}_i, \quad \text{as } \mathrm{T}\to\infty. \qquad (156)$$

This proves the asymptotic convergence in (50).

Next, we aim to demonstrate $\mathbf{C}_i \succ 0$ using a proof by contradiction. Suppose $\mathbf{C}_i$ is not positive definite, implying the existence of a vector $\widetilde{\boldsymbol{u}}$ with $\|\widetilde{\boldsymbol{u}}\| > 0$ such that $\widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}} \leq 0$. Then, from (156), we have:

$$\widetilde{\boldsymbol{u}}^\top \mathbf{H}_{i,\mathrm{T}} \widetilde{\boldsymbol{u}} \xrightarrow{\mathrm{P}} \widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}} \leq 0, \quad \text{as } \mathrm{T}\to\infty. \qquad (157)$$

Let $Y_2(t) = \{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{u}}\}^2$. Lemma B.17 confirms $\mathrm{E}\{\int_0^{R_1} Y_2(t)\,\mathrm{d}t\} > 0$. Consequently, Lemma B.15 implies the existence of a constant $c_i \in (0,\infty)$ such that:

$$\widetilde{\boldsymbol{u}}^\top \mathbf{H}_{i,\mathrm{T}} \widetilde{\boldsymbol{u}} = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} \{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{u}}\}^2\,\lambda_i^*(t\mid\mathscr{F}_t)\,\mathrm{d}t \xrightarrow{\mathrm{P}} c_i > 0,$$

as $\mathrm{T}\to\infty$, which contradicts (157). This concludes the proof. □

M. Proof of Theorem 6

Before proving Theorem 6, we first present Lemma B.18.

Lemma B.18 (Consistency of M-Estimator): Assume conditions A1, A2, A3, A4, A5, A6, and A7 in Appendix B. As $\mathrm{T}\to\infty$, there exists a local minimizer $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ of the loss function $\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ in (48) such that $\|\widehat{\widetilde{\boldsymbol{\beta}}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| = O_{\mathrm{P}}(\sqrt{1/\mathrm{T}})$.

Proof: Let $r_{\mathrm{T}} = \sqrt{1/\mathrm{T}}$ and $\widetilde{\boldsymbol{u}} \in \mathbb{R}^V$. Following the arguments of Theorem 1 in [51], it suffices to show that for any given $\epsilon > 0$, there exists a sufficiently large constant $C_\epsilon \in (0,\infty)$ such that for sufficiently large T, the following holds:

$$\mathrm{P}\Big(\inf_{\|\widetilde{\boldsymbol{u}}\|=C_\epsilon} \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^* + r_{\mathrm{T}}\,\widetilde{\boldsymbol{u}}) > \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)\Big) \geq 1-\epsilon.$$

Let $\widetilde{\boldsymbol{\beta}}_i = r_{\mathrm{T}}\,\widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{\beta}}_i^*$ and $\|\widetilde{\boldsymbol{u}}\| = C_\epsilon$. By Taylor's expansion of $\mathcal{L}_{i,\mathrm{T}}(\cdot)$ at $\widetilde{\boldsymbol{\beta}}_i^*$, we get:

$$\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) - \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) \equiv I_{1,1} + I_{1,2} + I_{1,3}, \qquad (158)$$

with

$$\begin{aligned}
I_{1,1} = \frac{1}{\mathrm{T}}\int_0^{\mathrm{T}} &\big[\widetilde{\boldsymbol{x}}_i(t)^\top r_{\mathrm{T}}\,\widetilde{\boldsymbol{u}}\exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\,\mathrm{d}t\\
&- \widetilde{\boldsymbol{x}}_i(t-)^\top r_{\mathrm{T}}\,\widetilde{\boldsymbol{u}}\,\mathrm{d}N_i(t)\big],
\end{aligned}$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$I_{1,2} = \frac{1}{2\mathrm{T}} \int_0^{\mathrm{T}} \{\widetilde{\boldsymbol{x}}_i(t)^\top r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}\}^2 \exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\, \mathrm{d}t,$$
$$I_{1,3} = \frac{1}{6\mathrm{T}} \int_0^{\mathrm{T}} \{\widetilde{\boldsymbol{x}}_i(t)^\top r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}\}^3 \exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^{**}\}\, \mathrm{d}t,$$

where $\widetilde{\boldsymbol{\beta}}_i^{**}$ lies between $\widetilde{\boldsymbol{\beta}}_i^*$ and $\widetilde{\boldsymbol{\beta}}_i$.

For the term $I_{1,1}$, using (49) from Theorem 5:

$$\begin{aligned}|I_{1,1}| &\le \Big\| \frac{1}{\mathrm{T}} \int_0^{\mathrm{T}} \big[\widetilde{\boldsymbol{x}}_i(t) \exp\{\widetilde{\boldsymbol{x}}_i(t)^\top \widetilde{\boldsymbol{\beta}}_i^*\}\, \mathrm{d}t \\ &\qquad - \widetilde{\boldsymbol{x}}_i(t-)\, \mathrm{d}N_i(t)\big] \Big\| \|r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}\| \\ &= O_{\mathrm{P}}(\sqrt{1/\mathrm{T}})\, r_{\mathrm{T}}\, \|\widetilde{\boldsymbol{u}}\|. \end{aligned} \tag{159}$$

For the term $I_{1,2}$, (50) from Theorem 5 implies:

$$\begin{aligned} I_{1,2} &= r_{\mathrm{T}}^2\, \widetilde{\boldsymbol{u}}^\top \nabla^2 \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) \widetilde{\boldsymbol{u}} \\ &= r_{\mathrm{T}}^2\, \widetilde{\boldsymbol{u}}^\top \{\mathbf{C}_i + o_{\mathrm{P}}(1)\} \widetilde{\boldsymbol{u}}. \end{aligned} \tag{160}$$

For the term $I_{1,3}$, under condition A5, each component of $\widetilde{\boldsymbol{x}}_i(t)$ is bounded above by a positive constant. Hence:

$$|I_{1,3}| \le c\, r_{\mathrm{T}}^3 \|\widetilde{\boldsymbol{u}}\|^3, \tag{161}$$

where $c \in (0, \infty)$ is a constant.

Combining (158), (159), (160), and (161), we obtain:

$$\begin{aligned} &\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) - \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) \\ &\quad = O_{\mathrm{P}}(\sqrt{1/\mathrm{T}})\, r_{\mathrm{T}}\, \|\widetilde{\boldsymbol{u}}\| \\ &\qquad + r_{\mathrm{T}}^2 \{\widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}} + o_{\mathrm{P}}(1)\, \|\widetilde{\boldsymbol{u}}\|^2\} + c\, r_{\mathrm{T}}^3\, \|\widetilde{\boldsymbol{u}}\|^3 \\ &\quad = \frac{1}{\mathrm{T}} \{O_{\mathrm{P}}(1)\, \|\widetilde{\boldsymbol{u}}\| + \widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}} \\ &\qquad + o_{\mathrm{P}}(1)\, \|\widetilde{\boldsymbol{u}}\|^2 + o_{\mathrm{P}}(1)\, \|\widetilde{\boldsymbol{u}}\|^3\}. \end{aligned} \tag{162}$$

By (162), we can choose a sufficiently large C_ϵ, such that all terms within the brackets in (162) are dominated by the term $\widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}}$, which is positive due to the fact that $\mathbf{C}_i \succ 0$ from Theorem 5. This completes the proof. □

Now, we prove Theorem 6. Let $r_{\mathrm{T}} = \sqrt{1/\mathrm{T}}$ and $\widetilde{\boldsymbol{u}} = (u_0, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_V)^\top \in \mathbb{R}^V$. Denote by $\ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)$ the objective function in (52), expressed as:

$$\ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) = \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i) + \sum_{j \in \mathcal{V} \setminus \{i\}} w_{j,i,\mathrm{T}} |\beta_{j,i}|. \tag{163}$$

Similar to the proof of Lemma B.18, it suffices to show that for any given $\epsilon > 0$, there exists a sufficiently large constant $C_\epsilon \in (0, \infty)$ such that, for large T:

$$\mathrm{P}\Big(\inf_{\|\widetilde{\boldsymbol{u}}\| = C_\epsilon} \ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^* + r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}) > \ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) \Big) \ge 1 - \epsilon.$$

Starting from (163), we obtain:

$$\begin{aligned} &\ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^* + r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}) - \ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*) \\ &\quad \ge \{\mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^* + r_{\mathrm{T}}\, \widetilde{\boldsymbol{u}}) - \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)\} \\ &\qquad + \sum_{j \in \mathrm{Pa}^*(i)} w_{j,i,\mathrm{T}} \cdot (|\beta_{j,i}^* + r_{\mathrm{T}} u_j| - |\beta_{j,i}^*|) \\ &\quad \equiv I_1 + I_2. \end{aligned}$$

For the term I_1, using (162), we derive:

$$I_1 = \frac{1}{\mathrm{T}} \{O_{\mathrm{P}}(1) \|\widetilde{\boldsymbol{u}}\| + \widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}} + o_{\mathrm{P}}(1) \|\widetilde{\boldsymbol{u}}\|^2 + o_{\mathrm{P}}(1) \|\widetilde{\boldsymbol{u}}\|^3\}.$$

For the term I_2, applying the triangle inequality and condition (55), we have:

$$\begin{aligned} |I_2| &\le \sum_{j \in \mathrm{Pa}^*(i)} w_{j,i,\mathrm{T}}\, r_{\mathrm{T}}\, |u_j| \\ &\le r_{\mathrm{T}}\, \|\widetilde{\boldsymbol{u}}\|_1 \max_{j \in \mathrm{Pa}^*(i)} w_{j,i,\mathrm{T}} = O_{\mathrm{P}}(1/\mathrm{T})\, \|\widetilde{\boldsymbol{u}}\|_1, \end{aligned}$$

which is dominated by $\widetilde{\boldsymbol{u}}^\top \mathbf{C}_i \widetilde{\boldsymbol{u}}/\mathrm{T}$ for a sufficiently large C_ϵ. Hence, we conclude that I_2 is dominated by I_1. The remaining proof is the same as that of Lemma B.18. □

N. Proof of Theorem 7

For a $\sqrt{1/\mathrm{T}}$-consistent estimator $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ of $\widetilde{\boldsymbol{\beta}}_i^*$, for any $\epsilon > 0$, there exists a constant C_ϵ such that for sufficiently large T,

$$\mathrm{P}(\|\widehat{\widetilde{\boldsymbol{\beta}}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon) > 1 - \epsilon. \tag{164}$$

Let $r_{\mathrm{T}} = \sqrt{1/\mathrm{T}}$. Recall condition A5 implies that $\widetilde{\boldsymbol{x}}_i(t)$ is bounded above. This, together with (151), yields that there exists a constant $c \in (0, \infty)$ such that

$$\sup_{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon} \Big| \frac{\partial^2 \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i} \partial \beta_{0;i}} \Big| \le c, \text{ for any } j \in \mathcal{V} \setminus \{i\},$$
$$\sup_{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon} \Big| \frac{\partial^2 \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i} \partial \beta_{k,i}} \Big| \le c, \text{ for any } j, k \in \mathcal{V} \setminus \{i\}.$$

Combining this with Taylor's expansion, (150), and (155), for $j \in \mathcal{V} \setminus \{i\}$, we have:

$$\begin{aligned} &\sup_{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon} \Big| \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} \Big| \\ &\le \Big| \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)}{\partial \beta_{j,i}} \Big| \\ &\qquad + \sup_{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon} \Bigg[\Big\{ \Big| \frac{\partial^2 \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i} \partial \beta_{0;i}} \Big| \\ &\qquad\qquad + \sum_{k \in \mathcal{V} \setminus \{i\}} \Big| \frac{\partial^2 \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i} \partial \beta_{k,i}} \Big| \Big\} \cdot \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \Bigg] \\ &\le \Big| \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i^*)}{\partial \beta_{j,i}} \Big| + V \cdot c\, r_{\mathrm{T}}\, C_\epsilon \\ &= O_{\mathrm{P}}(\sqrt{1/\mathrm{T}}) + O(r_{\mathrm{T}}) = O_{\mathrm{P}}(\sqrt{1/\mathrm{T}}). \end{aligned} \tag{165}$$

Consider $\widetilde{\boldsymbol{\beta}}_i$ in the ball $\{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}}\, C_\epsilon\}$. For $j \in \mathcal{V} \setminus \{\mathrm{Pa}^*(i) \cup i\}$, if $\beta_{j,i} > 0$, then (57) and (165) imply:

$$\begin{aligned} \frac{\partial \ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} &= \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} + w_{j,i,\mathrm{T}} \operatorname{sign}(\beta_{j,i}) \\ &\ge - \sup_{\widetilde{\boldsymbol{\beta}}_i : \|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \le r_{\mathrm{T}} C_\epsilon} \Big| \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} \Big| \\ &\qquad + \min_{j \in \mathcal{V} \setminus \{\mathrm{Pa}^*(i) \cup i\}} w_{j,i,\mathrm{T}} \\ &> 0, \end{aligned} \tag{166}$$

with probability tending to 1 as $\mathrm{T} \to \infty$. Likewise, if $\beta_{j,i} < 0$, then:

$$\frac{\partial \ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} = \frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}} + w_{j,i,\mathrm{T}} \operatorname{sign}(\beta_{j,i})$$

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

$$\leq \sup_{\widetilde{\boldsymbol{\beta}}_i:\|\widetilde{\boldsymbol{\beta}}_i-\widetilde{\boldsymbol{\beta}}_i^*\|\leq r_{\mathrm{T}}C_\epsilon} \left|\frac{\partial \mathcal{L}_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)}{\partial \beta_{j,i}}\right| - \min_{j\in\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}} w_{j,i,\mathrm{T}} < 0, \quad (167)$$

with probability tending to 1 as $\mathrm{T}\to\infty$.

By (166) and (167), the following argument holds with probability tending to 1 as $\mathrm{T}\to\infty$: For all $\widetilde{\boldsymbol{\beta}}_i$ with $\|\widetilde{\boldsymbol{\beta}}_i - \widetilde{\boldsymbol{\beta}}_i^*\| \leq r_{\mathrm{T}}C_\epsilon$ and all $j\in\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}$, $\partial\ell_{i,\mathrm{T}}(\widetilde{\boldsymbol{\beta}}_i)/\partial\beta_{j,i}$ has the same sign as $\beta_{j,i}$. Together with (164) and the first-order condition of $\widehat{\widetilde{\boldsymbol{\beta}}}_i$, it follows that

$$\mathrm{P}\big(\widehat{\beta}_{j,i}=0,\ \text{for all } j\in\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}\big)\geq 1-2\epsilon \quad (168)$$

holds for sufficiently large T. Since ϵ is arbitrary, letting $\epsilon\to 0$ in (168) yields that

$$\mathrm{P}\big(\widehat{\beta}_{j,i}=0,\ \text{for all } j\in\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}\big)\to 1, \quad (169)$$

as $\mathrm{T}\to\infty$. Note that the vector $\widehat{\boldsymbol{\beta}}_i^{(\mathrm{II})}$ collects all the components in $\widehat{\widetilde{\boldsymbol{\beta}}}_i$ whose indices belong to the set $\mathcal{V}\setminus\{\mathrm{Pa}^*(i)\cup i\}$. Hence, (169) implies that $\mathrm{P}(\widehat{\boldsymbol{\beta}}_i^{(\mathrm{II})}=\mathbf{0})\to 1$, as $\mathrm{T}\to\infty$. This completes the proof. □

O. Proof of Corollary 1

Recall that the true edge set is non-empty ($\mathcal{E}^*\neq\varnothing$), implied by condition A8. To prove Corollary 1, it suffices to show that for each pair of distinct nodes $(j,i)\in\mathcal{V}\times\mathcal{V}$,

$$\begin{cases}\mathrm{P}((j,i)\in\widehat{\mathcal{E}}_+)\to 1, & \text{if } (j,i)\in\mathcal{E}_+^*,\\ \mathrm{P}((j,i)\in\widehat{\mathcal{E}}_-)\to 1, & \text{if } (j,i)\in\mathcal{E}_-^*,\\ \mathrm{P}((j,i)\notin\widehat{\mathcal{E}})\to 1, & \text{if } (j,i)\notin\mathcal{E}^*,\end{cases} \quad \text{as } \mathrm{T}\to\infty.$$

If $(j,i)\in\mathcal{E}_+^*$, then (14) implies that $\beta_{j,i}^*>0$. By Theorem 6, we have $\widehat{\beta}_{j,i}\xrightarrow{\mathrm{P}}\beta_{j,i}^*>0$. Thus, $\mathrm{P}((j,i)\in\widehat{\mathcal{E}}_+)=\mathrm{P}(\widehat{\beta}_{j,i}>0)\to 1$, as $\mathrm{T}\to\infty$.

Similarly, if $(j,i)\in\mathcal{E}_-^*$, then we have $\mathrm{P}((j,i)\in\widehat{\mathcal{E}}_-)=\mathrm{P}(\widehat{\beta}_{j,i}<0)\to 1$, as $\mathrm{T}\to\infty$.

If $(j,i)\notin\mathcal{E}^*$, then (13) implies that $\beta_{j,i}^*=0$. By (58) in Theorem 7, we obtain $\mathrm{P}((j,i)\notin\widehat{\mathcal{E}})=\mathrm{P}(\widehat{\beta}_{j,i}=0)\to 1$, as $\mathrm{T}\to\infty$. This completes the proof. □

ACKNOWLEDGMENT

The authors thank the Associate Editor and three reviewers for insightful comments that significantly improved the article's presentation.

REFERENCES

[1] D. H. Perkel, G. L. Gerstein, and G. P. Moore, "Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains," *Biophysical J.*, vol. 7, no. 4, pp. 419–440, 1967.

[2] G. L. Gerstein and D. H. Perkel, "Simultaneously recorded trains of action potentials: Analysis and functional interpretation," *Science*, vol. 164, no. 3881, pp. 828–830, May 1969.

[3] D. R. Brillinger and A. E. P. Villa, "Examples of the investigation of neural information processing by point process analysis," in *Advanced Methods of Physiological System Modeling*, vol. 3. Boston, MA, USA: Springer, 1994, pp. 111–127.

[4] K. D. Harris, J. Csicsvari, H. Hirase, G. Dragoi, and G. Buzsáki, "Organization of cell assemblies in the hippocampus," *Nature*, vol. 424, no. 6948, pp. 552–556, Jul. 2003.

[5] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: State-of-the-art and future challenges," *Nat. Neurosci.*, vol. 7, no. 5, pp. 456–461, 2004, doi: 10.1038/NN1228.

[6] O. O. Aalen, "Dynamic modelling and causality," *Scandin. Actuarial J.*, vol. 1987, nos. 3–4, pp. 177–190, Jul. 1987.

[7] V. Didelez, "Graphical models for marked point processes based on local independence," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 70, no. 1, pp. 245–264, Feb. 2008.

[8] T. Dean and K. Kanazawa, "A model for reasoning about persistence and causation," *Comput. Intell.*, vol. 5, no. 2, pp. 142–150, Feb. 1989. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.1989.tb00324.x

[9] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*. Berkeley, CA, USA: Univ. of California, 2002.

[10] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," *J. Neurophysiol.*, vol. 93, no. 2, pp. 1074–1089, Feb. 2005.

[11] C. Zhang, Y. Chai, X. Guo, M. Gao, D. Devilbiss, and Z. Zhang, "Statistical learning of neuronal functional connectivity," *Technometrics*, vol. 58, no. 3, pp. 350–359, Jul. 2016.

[12] M. Zhao et al., "An L_1–regularized logistic model for detecting short-term neuronal interactions," *J. Comput. Neurosci.*, vol. 32, no. 3, pp. 479–497, Jun. 2012.

[13] M. Lukasik, P. K. Srijith, T. Cohn, and K. Bontcheva, "Modeling tweet arrival times using log-Gaussian cox processes," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 250–255.

[14] O. Nicolis, L. M. R. Quezada, and G. Ibacache-Pulgar, "Temporal cox process with folded normal intensity," *Axioms*, vol. 11, no. 10, p. 513, Sep. 2022.

[15] S. Rajaram, T. Graepel, and R. Herbrich, "Poisson-networks: A model for structured point processes," in *Proc. 10th Intl. Workshop AI Stat.*, 2005, pp. 277–284.

[16] S. Chen, D. Witten, and A. Shojaie, "Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process," *Electron. J. Statist.*, vol. 11, no. 1, p. 1207, Jan. 2017.

[17] P. Embrechts and M. Kirchner, "Hawkes graphs," *Theory Probab. Appl.*, vol. 62, no. 1, pp. 132–156, Jan. 2018.

[18] H. Xu, M. Farajtabar, and H. Zha, "Learning Granger causality for Hawkes processes," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1717–1726.

[19] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard, "Lasso and probabilistic inequalities for multivariate point processes," *Bernoulli*, vol. 21, no. 1, pp. 83–143, 2015.

[20] X. Tang and L. Li, "Multivariate temporal point process regression," *J. Amer. Stat. Assoc.*, vol. 118, no. 542, pp. 830–845, Apr. 2023.

[21] W. Dempsey, B. Oselio, and A. Hero, "Hierarchical network models for exchangeable structured interaction processes," *J. Amer. Stat. Assoc.*, vol. 117, no. 540, pp. 2056–2073, Oct. 2022.

[22] C. T. Butts, "4. A relational event framework for social action," *Sociol. Methodol.*, vol. 38, no. 1, pp. 155–200, Aug. 2008.

[23] P. O. Perry and P. J. Wolfe, "Point process modelling for directed interaction networks," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 75, no. 5, pp. 821–849, Nov. 2013.

[24] B. Yuan, H. Li, A. L. Bertozzi, P. J. Brantingham, and M. A. Porter, "Multivariate spatiotemporal Hawkes processes and network reconstruction," *SIAM J. Math. Data Sci.*, vol. 1, no. 2, pp. 356–382, Jan. 2019.

[25] M. Costa, C. Graham, L. Marsalle, and V. C. Tran, "Renewal in Hawkes processes with self-excitation and inhibition," *Adv. Appl. Probab.*, vol. 52, no. 3, pp. 879–915, Sep. 2020.

[26] V. V. Kalashnikov, *Mathematical Methods in Queuing Theory*, vol. 271. Dordrecht, The Netherlands: Springer, 2013.

[27] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London, U.K.: Springer, 2012.

[28] R. Azaïs, J.-B. Bardet, A. Génadot, N. Krell, and P.-A. Zitt, "Piecewise deterministic Markov process—Recent results," *ESAIM Proc.*, vol. 44, pp. 276–290, Jan. 2014.

[29] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. New York, NY, USA: Springer, 2003.

[30] I. Rubin, "Regular point processes and their detection," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 5, pp. 547–557, Sep. 1972.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.

[31] A. Bonnet, M. Martinez Herrera, and M. Sangnier, "Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity," *Statist. Comput.*, vol. 33, no. 4, p. 91, Aug. 2023.
[32] S. Wei, Y. Xie, C. S. Josef, and R. Kamaleswaran, "Granger causal chain discovery for sepsis-associated derangements via continuous-time Hawkes processes," in *Proc. 29th ACM SIGKDD Conf. Knowl. Disc. Data Min.*, 2023, pp. 2536–2546.
[33] P. Brémaud and L. Massoulié, "Stability of nonlinear Hawkes processes," *Ann. Probab.*, vol. 24, no. 3, pp. 1563–1588, Jul. 1996.
[34] S. M. Ross, *Stochastic Processes*. Hoboken, NJ, USA: Wiley, 1995.
[35] P. K. Andersen and N. Keiding, "Multi-state models for event history analysis," *Stat. Methods Med. Res.*, vol. 11, no. 2, pp. 91–115, Apr. 2002.
[36] M. F. Neuts, "A versatile Markovian point process," *J. Appl. Probab.*, vol. 16, no. 4, pp. 764–779, Dec. 1979.
[37] F.-B. Dolivo, *Counting Processes and Integrated Conditional Rates: A Martingale Approach With Application to Detection*. Ann Arbor, MI, USA: Univ. of Michigan, 1974.
[38] T. Fischer, "On simple representations of stopping times and stopping time sigma-algebras," *Statist. Probab. Lett.*, vol. 83, no. 1, pp. 345–349, Jan. 2013.
[39] A. G. Hawkes and D. Oakes, "A cluster process representation of a self-exciting process," *J. Appl. Probab.*, vol. 11, no. 3, pp. 493–503, Sep. 1974.
[40] P. Reynaud-Bouret and S. Schbath, "Adaptive estimation for Hawkes processes; Application to genome analysis," *Ann. Statist.*, vol. 38, no. 5, pp. 2781–2822, Oct. 2010.
[41] R. E. Kass, U. T. Eden, and E. N. Brown, *Analysis of Neural Data*, vol. 491. New York, NY, USA: Springer, 2014.
[42] G. Nieuwenhuis, "Asymptotic mean stationarity and absolute continuity of point process distributions," *Bernoulli*, vol. 19, no. 5A, pp. 1612–1636, Nov. 2013.
[43] E. Bacry and J.-F. Muzy, "First- and second-order statistics characterization of Hawkes processes and non-parametric estimation," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2184–2202, Apr. 2016.
[44] M. Krumin, I. Reutsky, and S. Shoham, "Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input," *Frontiers Comput. Neurosci.*, vol. 4, p. 147, Nov. 2010.
[45] H. Zou, "The adaptive lasso and its Oracle properties," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006.
[46] A. W. Van der Vaart, *Asymptotic Statistics*, vol. 3. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[47] R. Nishii, "Asymptotic properties of criteria for selection of variables in multiple regression," *Ann. Statist.*, vol. 12, no. 2, pp. 758–765, Jun. 1984.
[48] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, Dec. 2007, doi: 10.1214/07-AOAS131.
[49] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.
[50] X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing, "DAGs with NO TEARS: Continuous optimization for structure learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 9492–9503.
[51] C. Zhang, Y. Jiang, and Y. Chai, "Penalized Bregman divergence for large-dimensional regression and classification," *Biometrika*, vol. 97, no. 3, pp. 551–566, Sep. 2010.

Muhong Gao received the B.S. degree in statistics from Peking University, China, in 2015, and the Ph.D. degree in statistics from the University of Wisconsin–Madison, USA, in 2022. He is currently a Post-Doctoral Scholar with the Academy of Mathematics and System Science, Chinese Academy of Sciences. His research interests include statistical learning theory, signal processing, and applied probability.

Chunming Zhang received the Ph.D. degree in statistics from the University of North Carolina at Chapel Hill. She is a Professor with the Department of Statistics, University of Wisconsin–Madison. Her work spans multiple hypotheses testing, large-scale simultaneous inference, dimension reduction, high-dimensional inference, non-parametric and semi-parametric modeling and inference, functional and longitudinal data analysis, and robust statistics. Her research interests include statistical methods in computational neuroscience, biostatistics, and financial econometrics; and the analysis of neuroimaging, spatial, and temporal data. She is an Elected Fellow of the Institute of Mathematical Statistics and the American Statistical Association. She received the 2024 IMS Medallion Award and Lecture.

Jie Zhou received the Ph.D. degree in statistics from the Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He is a Professor with the School of Mathematics, Capital Normal University. His research interests include the theory and application of statistical modeling and inference of complex data, including survival data, recurrent event data, and longitudinal data. In recent years, he has also been focusing on big data and deep learning.

Authorized licensed use limited to: University of Wisconsin. Downloaded on July 16,2024 at 19:30:33 UTC from IEEE Xplore. Restrictions apply.