Paper Sharing

Structural Hawkes Processes for Learning Causal Structure from Discrete-Time Event Sequences

Lecturer：Yuxin Wu

2023.8.9

Introduction

- Many data are recorded in the form of events
  - system logs
  - social network interactions
  - shopping behaviors
  - browsing behaviors

⋆ Learning causal structure among event types on multi-type event sequences is an important and challenging task

Introduction

- Existing methods, such as the multivariate Hawkes processes based methods, mostly boil down to learning the so-called Granger causality

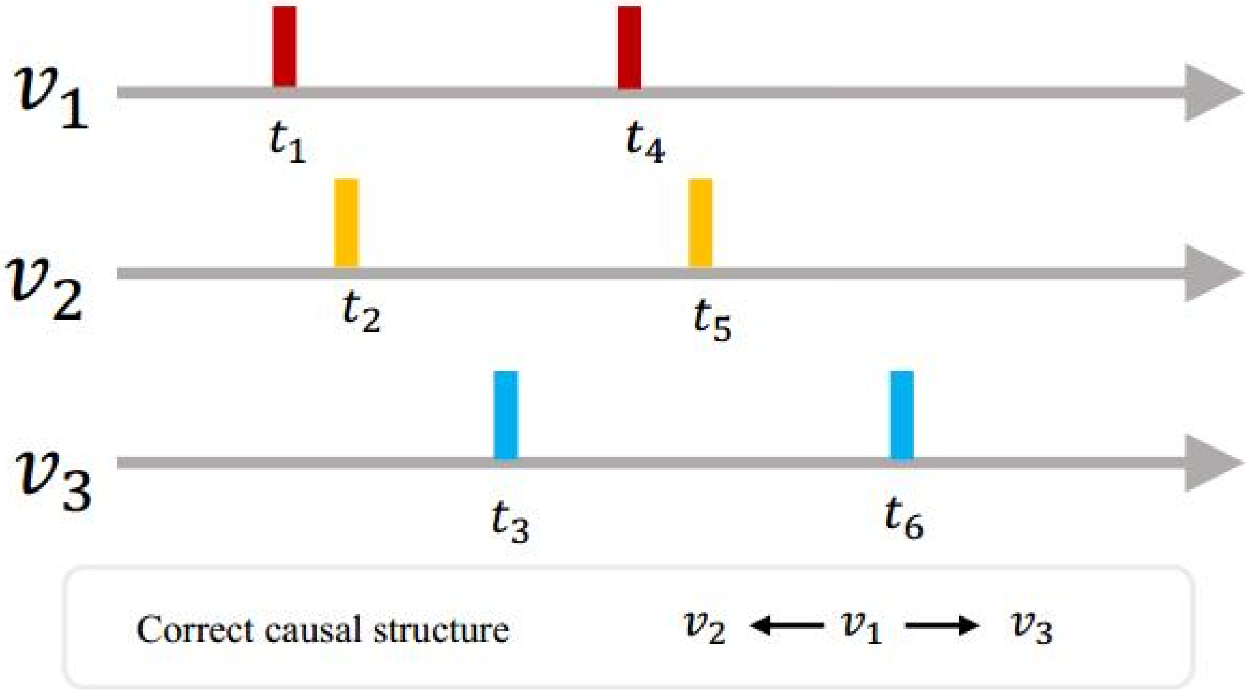
- a implicit assume: temporal precedence assumption

——all events are recorded instantaneously and accurately such that the cause event happens strictly prior to its effect event

Introduction

- a implicit assume: temporal precedence assumption

——all events are recorded instantaneously and accurately such that the cause event happens strictly prior to its effect event



Introduction

- Limited recording capabilities and storage capacities

↓

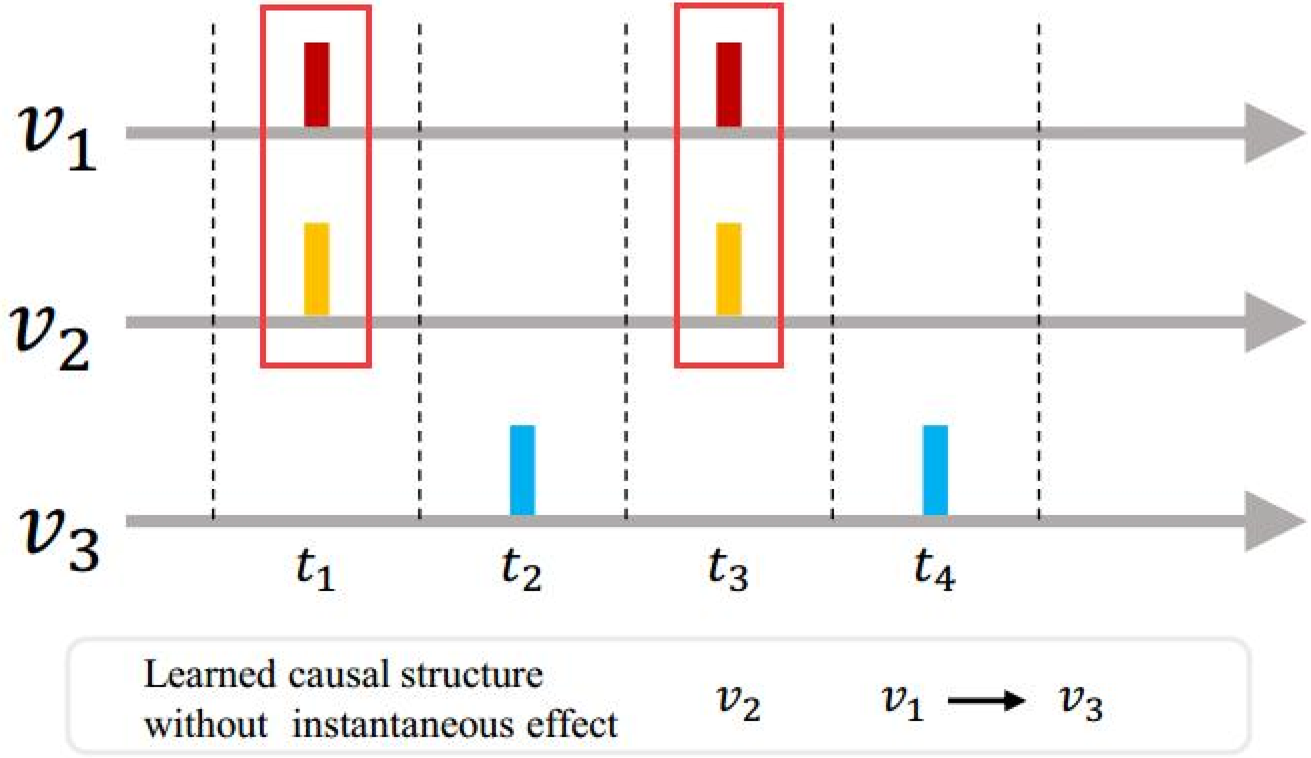
- Retaining event’s occurred times with high-resolution is expensive or practically impossible

↓

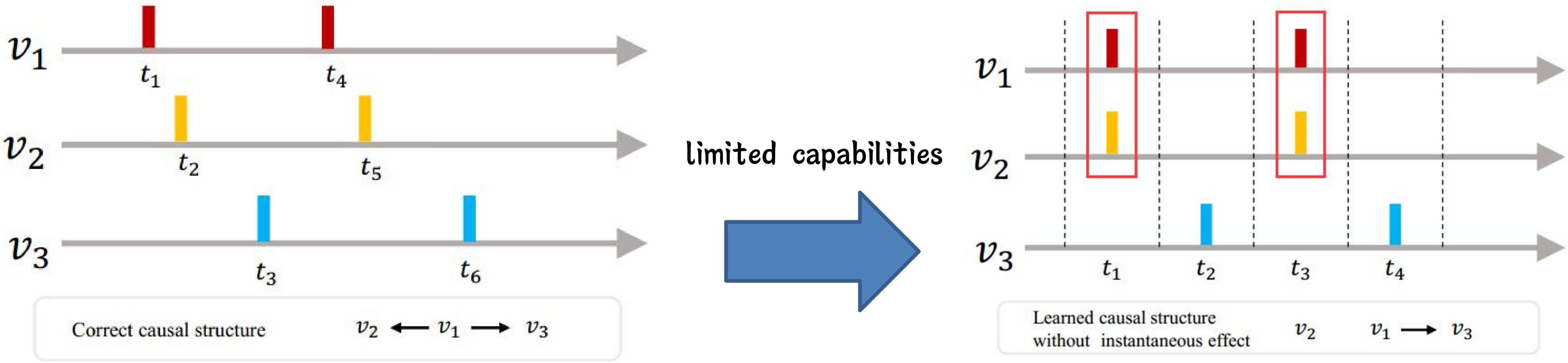
- In many real-world applications, and we usually only can access the corresponding discrete-time event sequences.

Introduction

- In many real-world applications, and we usually only can access the corresponding discrete-time event sequences.



Introduction



Typical event sequence in continues-time

Event sequence in discrete-time

Granger-based methods will fail to identify $v_1 \rightarrow v_2$ in discrete-time

Two questions

- In this paper, we aim to answer the following two questions:

(1) How to design and learn a Hawkes process that leverages the instantaneous effect in discrete time?

(2) Can we identify the causal relationship in event sequences under the existence of the instantaneous effect?

Design of Structural Point Processes

Structural counting processes

A structural counting process is a multivariate counting process $\boldsymbol{N}^{(\Delta)}$ in discrete-time with the conditional intensity of $N_v^{(\Delta)}$ for each $v \in \boldsymbol{V}$ satisfying:

$$\lambda_v(k\Delta)\Delta = \mathbb{E}\big[X_{v,k} | \underbrace{\mathcal{F}_{(k-1)\Delta}}_{\text{past}} \cup \underbrace{\mathcal{F}_{k\Delta}^{-v}}_{\text{presence}}\big]$$

where $\mathcal{F}_{(k-1)\Delta} = \bigcup_{0 \le s \le k-1, v \in \mathbf{V}} \mathcal{F}_{s\Delta}^{v}$ *is the filtration with discrete-time in the past and* $\mathcal{F}_{k\Delta}^{-v} \coloneqq \{\mathcal{F}_{k\Delta}^{v\prime} | v' \in \mathbf{V} \setminus v\}$ *is the filtration that except for type-*v *event.*

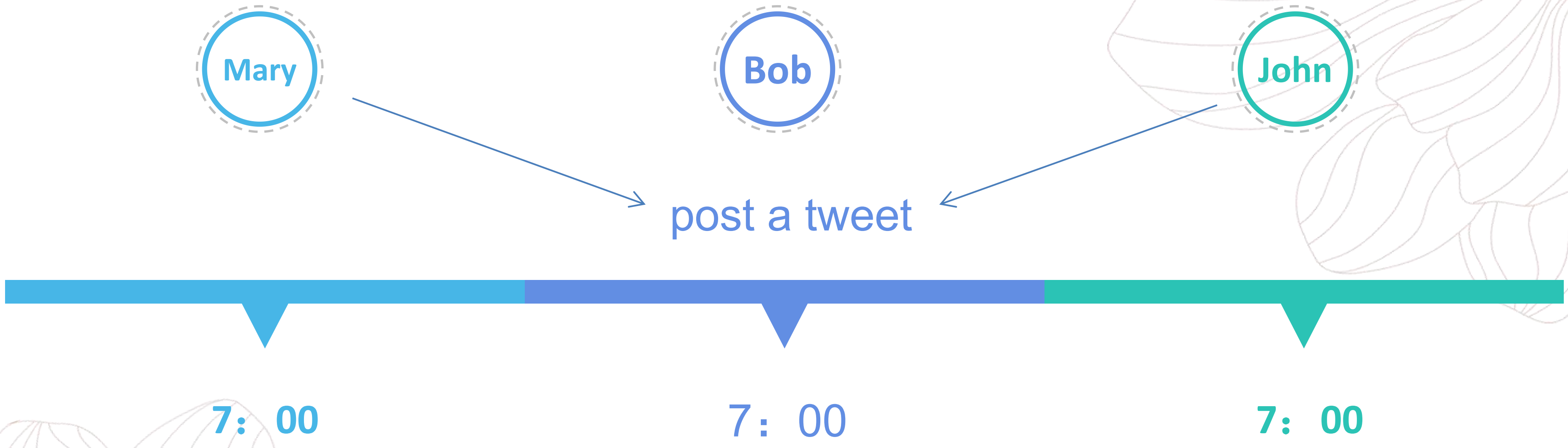
Structural Hawkes processes

A structural Hawkes process is a structural counting process such that for each $v \in \mathbf{V}$, the intensity of $N_v^{(\Delta)}$ can be written as:

$$\lambda_v(k\Delta) = \mu_v + \sum_{v\prime \in \mathbf{V}} \sum_{i=1}^{k} \phi_{v\prime,v}\big((k-i)\Delta\big) X_{v,i},$$

where $\phi_{v,v}(0) \equiv 0$ ensures the exclusive of type-v event at time $k\Delta$, *and* $X_{v,i} \coloneqq N_v(i\Delta) - N_v((i-1)\Delta)$ *denotes the number of events occurs in time* $[i\Delta, (i-1)\Delta]$.

Design of Structural Point Processes



- the similiar application: virus transmition

——people in close contact may experience symptoms on the same day, rather than in strict order

Learning of Structural Point Processes

Estimation using Minorization-Maximization Algorithm

By applying Jensen's inequality and obtain the following lower bound: $Q(\Theta|\Theta^{(j)})$

$$= \sum_{v\in \mathbf{V}} \sum_{k=1}^{K} \Bigg[-\Big(\mu_v + \sum_{v'\in \mathbf{V}} \sum_{i=1}^{k} \phi_{v',v}\left((k-i)\Delta\right) X_{v',i} \Big)\Delta$$

$$+ X_{v,k} \left(q_{v,k}^{\mu} \log\left(\frac{\mu_v}{q_{v,k}^{\mu}} \right) \right.$$

$$\left. + \sum_{v'\in \mathbf{V}} \sum_{i=1}^{k} q_{v,k}^{\alpha}(v',i) \log\left(\frac{\phi_{v',v}((k-i)\Delta) X_{v',i}}{q_{v,k}^{\alpha}(v',i)} \right) \right) \Bigg]$$

where $q_{v,k}^{\mu} = \frac{\mu_v^{(j)}}{\lambda_v^{(j)}(k\Delta)}$ and $q_{v,k}^{\alpha}(v',i) = \frac{\phi_{v\prime,v}^{(j)}((k-i)\Delta) X_{v\prime,i}}{\lambda_v^{(j)}(k\Delta)}$. By setting $\frac{\partial Q(\Theta|\Theta^{(j)})}{\partial \mu_v} = 0, \frac{\partial Q(\Theta|\Theta^{(j)})}{\partial \alpha_{v\prime,v}} = 0$, we have:

$$\mu_v^{(j+1)} = \frac{\sum_{k=1}^{K} X_{v,k}\, q_{v,k}^{\mu}}{K\Delta}$$

$$\alpha_{v\prime,v}^{(j+1)} = \begin{cases} \dfrac{\sum_{k=1}^{K}\sum_{i=1}^{k} q_{v,k}^{\alpha}(v',i) X_{v,k}}{\sum_{k=1}^{K}\sum_{i=1}^{k} \kappa\left((k-i)\Delta\right) X_{v\prime,i}\Delta} & v' \neq v \\[2ex] \dfrac{\sum_{k=1}^{K}\sum_{i=1}^{k-1} q_{v,k}^{\alpha}(v',i) X_{v,k}}{\sum_{k=1}^{K}\sum_{i=1}^{k-1} \kappa\left((k-i)\Delta\right) X_{v\prime,i}\Delta} & v' = v \end{cases}$$

Causal discovery using Hill-climb based algorithm

Algorithm 1 Learning causal structure using SHP

Input: Data set $\mathbf{X}$

Output: G^*, Θ^*

```
G' ← empty graph, L*_p ← −∞
while L*_p(G*, Θ*; X) < L'_p(G', Θ'; X) do
    G*, Θ* ← G', Θ' with largest L'_p(G', Θ'; X)
    for every G' ∈ V(G*) do
        Update Θ' via iteration in Eq. (7)
        Record score L'_p(G', Θ'; X)
    end for
end while
return G*, Θ*
```

7

Two questions

- In this paper, we aim to answer the following two questions:

(1) How to design and learn a Hawkes process that leverages the instantaneous effect in discrete time?

(2) Can we identify the causal relationship in event sequences under the existence of the instantaneous effect?

Identifiability

Definition 3 (INAR(∞)). *For $\theta_k \geq 0, k \in \mathbb{N}_0$, let $\epsilon_t \overset{i.i.d.}{\sim} \text{Pois}(\theta_0), t \in \mathbb{N}$, and $\xi_i^{(t,k)} \sim \text{Pois}(\theta_k)$. An Integer-valued autoregressive time series of infinite order (INAR(∞)) process $X_t, t \in \mathbb{N}$ is defined by*

$$X_t = \sum_{k=1}^{\infty} \theta_k \circ X_{t-k} + \epsilon_t$$

where $\circ$ is a reproduction operator given by $\theta_k \circ X_{t-k} \equiv \sum_{i=1}^{X_{t-k}} \xi_i^{(t,k)}$ with $\xi_i^{(t,k)}$ be a sequence of i.i.d. non-negative integer-valued random variables that depends on the reproduction coefficients θ_k, $\{\epsilon_t\}_{t \in \mathbb{N}}$ is an i.i.d. integer-valued immigration sequence that are independent of $\{\xi_i^{(t,k)}\}$, and X_{t-k} is independent of ϵ_t for all k.

Theorem 1 (Kirchner (2016)). *Let N be a Hawkes process with immigration intensity μ and let $\phi: \mathbb{R} \to \mathbb{R}_0^+$ be a reproduction intensity that is piecewise continuous with $\phi(t) = 0, t \leq 0$ and $\int \phi(t)dt < 1$. For $\Delta \in (0, \delta)$, let $\left(X_t^{(\Delta)}\right)$ be an INAR(∞) sequence with immigration parameter $\Delta\mu$ and reproduction coefficients $\Delta\phi(k\Delta), k \in \mathbb{N}$. From the sequences $\left(X_t^{(\Delta)}\right)_{\Delta \in (0,\delta)}$, we define a family of point processes by*

$$N^{(\Delta)}(A) := \sum_{k: k\Delta \in A} X_k^{(\Delta)}, \quad A \in \mathcal{B}, \Delta \in (0, \delta),$$

where $\mathcal{B} := \mathcal{B}(\mathbb{R})$ is the Borel set in $\mathbb{R}$. Then, we have that

$$N^{(\Delta)} \xrightarrow{\text{w}} N \quad \text{for } \Delta \to 0.$$

9

Identifiability

Definition 4 (Instantaneous causal structure in structural Hawkes process). *Let* $\epsilon_{v,t} \overset{i.i.d.}{\sim} \text{Pois}(\mu_v)$*, and* $\xi_i^{(v',v)} \sim \text{Pois}(\alpha_{v',v})$*. The instantaneous causal structure in the structural Hawkes process consists of a set of equations of the form* $X_{v,t} = \sum_{v' \in \mathbf{V}} \alpha_{v',v} \circ X_{v',t} + \epsilon_v, \quad v \in \mathbf{V}$*, where* $\alpha_{v',v} > 0$ *for*

Theorem 2. *Let* $X \to Y$ *be the correct causal direction that follows*

$$Y \quad = \textstyle\sum_{i=1}^{X} \xi_i + \epsilon, \qquad X, \xi_i, \text{ and } \epsilon \text{ are independent},$$

where $\xi_i \sim \text{Pois}(\alpha_{X,Y})$, $\epsilon \sim \text{Pois}(\mu_Y)$, $X \sim \text{Pois}(\mu_X)$*. Then, there does not exist a backward model that admits the following equation:*

$$X \quad = \textstyle\sum_{i=1}^{Y} \hat{\xi}_i + \hat{\epsilon}, \qquad Y, \hat{\xi}_i, \text{ and } \hat{\epsilon} \text{ are independent},$$

where $\hat{\xi}_i \sim \text{Pois}(\hat{\alpha}_{Y,X})$, $\hat{\epsilon} \sim \text{Pois}(\hat{\mu}_X)$, $Y \sim \text{Pois}(\hat{\mu}_Y)$

Theorem 3. *With the causal faithfulness assumption and causal sufficiency assumption, the multivariate instantaneous causal structure is identifiable.*

Experiments: Synthetic Data

- **Description**: the author conduct six different control experiments for SHP

- **Result**: In general, our proposed SHP method outperforms all the baseline methods in all six control experiments.

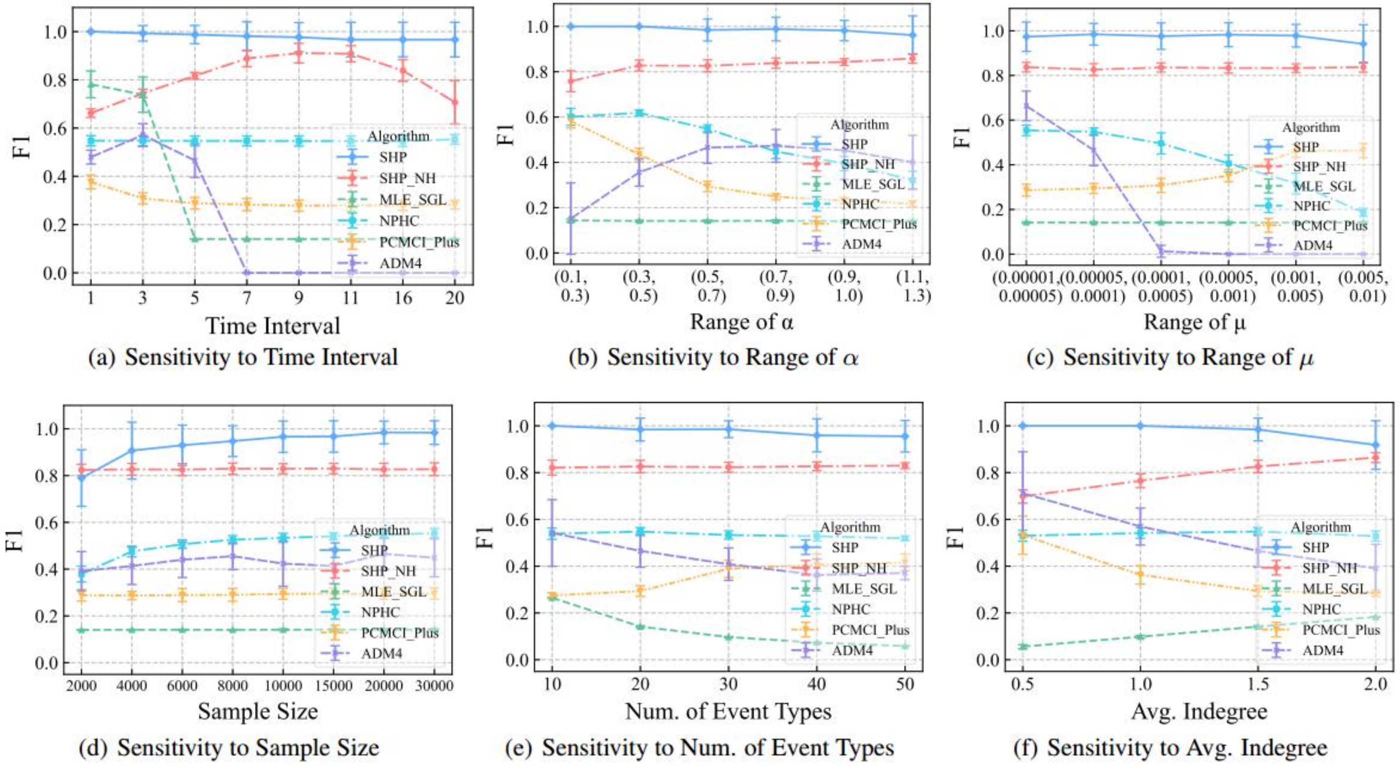


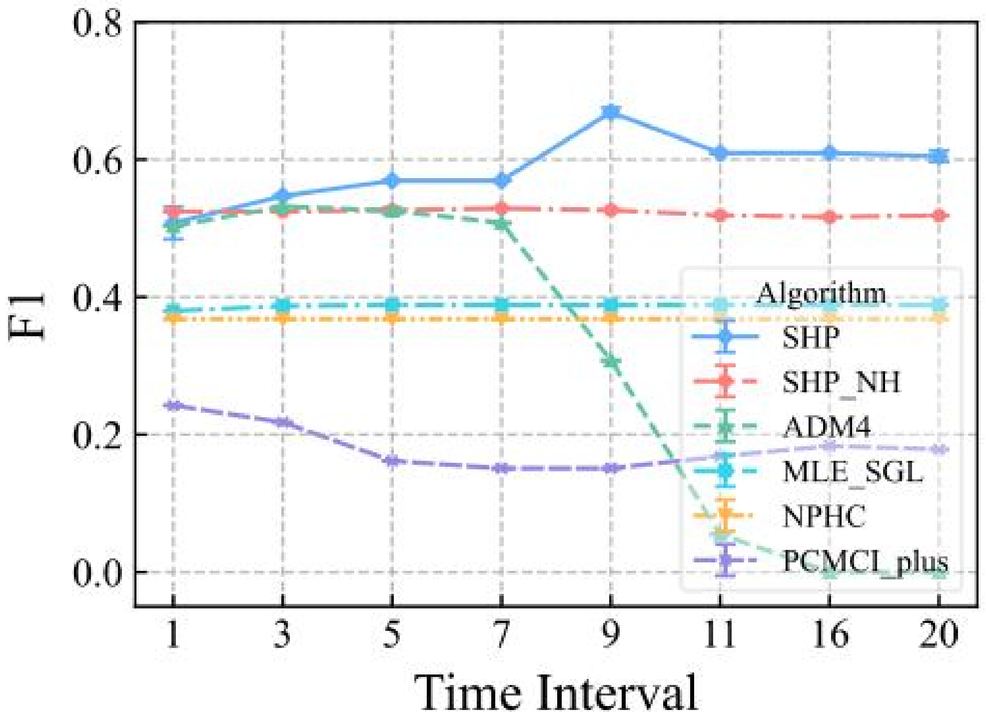
Figure 3: F1 in the Sensitivity Experiments

Experiments: Real world Data

- **Description:** the dataset records eight months of alarms that occurred in a real metropolitan cellular network.

- **Goal:** find the causal structure among eighteen alarms.

- **Result:** It verifies the effectiveness of SHP in capturing the instantaneous effect.



Conclusion

Hawkes process + Instantaneous effect = Structural Hawkes Processes

Thank you for listening

主講人：吳雨欣

2023.8.9