

Mining periodic patterns and cascading bursts phenomenon in individual e-mail communication

Chenlong Li, Zhanjie Song, Wenjun Wang & Xu (Sunny) Wang

To cite this article: Chenlong Li, Zhanjie Song, Wenjun Wang & Xu (Sunny) Wang (2019): Mining periodic patterns and cascading bursts phenomenon in individual e-mail communication, Journal of Applied Statistics, DOI: [10.1080/02664763.2019.1605340](https://doi.org/10.1080/02664763.2019.1605340)

To link to this article: <https://doi.org/10.1080/02664763.2019.1605340>



Published online: 21 Apr 2019.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)



Mining periodic patterns and cascading bursts phenomenon in individual e-mail communication

Chenlong Li^{a,d}, Zhanjie Song^{a,b}, Wenjun Wang^c and Xu (Sunny) Wang^d

^aSchool of Mathematics, Tianjin University, Tianjin, People's Republic of China; ^bVisual Pattern Analysis Research Lab, Tianjin University, Tianjin, People's Republic of China; ^cSchool of Computer Science and Technology, Tianjin University, Tianjin, People's Republic of China; ^dDepartment of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada

ABSTRACT

Quantitative understanding of human activity is very important as many social and economic trends are driven by human actions. We propose a novel stochastic process, the Multi-state Markov Cascading Non-homogeneous Poisson Process (M2CNPP), to analyze human e-mail communication involving both periodic patterns and bursts phenomenon. The model parameters are estimated using the Generalized Expectation Maximization (GEM) algorithm while the hidden states are treated as missing values. The empirical results demonstrate that the proposed model adequately captures the major temporal cascading features as well as the periodic patterns in e-mail communication.

ARTICLE HISTORY

Received 12 July 2018
Accepted 31 March 2019

KEYWORDS

Individual e-mail communication; human dynamic modeling; Markov cascading non-homogeneous poisson process

1. Introduction

Human beings participate in all kinds of daily social activities and exhibit certain patterns in their behavior. These patterns are by no means new – they drove human behavior for centuries dominating everything from wars to Einstein's correspondences [7]. It has become well-known from exploring large data sets describing human actions that everything human beings do they do in bursts – brief periods of intensive activity followed by long periods of nothingness. Recent advances in the study of human dynamics mainly focus on verifying the existence of the bursts phenomenon and uncovering the underlying mechanism in various human actions, such as human mobility and communication [1,6,7,13,16,27,37,39].

Scale-free random walks known as Lévy flights have been proposed to describe the bursts phenomenon in human mobility [13]. Recent research, however, has showed that the dispersal of human mobility is actually slower than that described by Lévy flights. Human mobility exhibits bursts phenomenon, i.e. it concentrates in a small region during a long time period with short trajectories [6]. There is a high probability that an individual

CONTACT Chenlong Li  lichenlong@tju.edu.cn  School of Mathematics, Tianjin University, 300072 Tianjin, People's Republic of China Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada

likely returns to a few highly frequently visited locations [13]. These features have been considered in recent research of human dynamics modeling [3,14].

E-mail records have been analyzed to uncover human behavior in recent publications [5,17,38] and can be represented as the realization of a point process in time [1,9,22]. The prototype of point processes is the Poisson process, which is characterized by the fact that the intervals, T 's, between events are exponentially distributed. It has been shown, however, that certain heavy-tailed power law distributions can be used to better approximate the inter-event time for e-mail communication [1,27] by modeling highly clustered inter-event time at the tails.

The research of point processes has been growing rapidly in the last decade and in particular for modeling various bursts phenomena [8]. For instance, the Markov Modulated Poisson Process (MMPP) [11] and the self-exciting point process [15] were used in a broad range of research fields, such as stars [29], biotic populations [28], air pollution [32], crime and security [25,26,35], transportation [34], marine mammal abundance [18], web traffic [36], and the occurrence of a rare DNA motif [31]. Some of the early applications for telecommunication used the cluster point process [2] and the MMPP model [33]. Recently, the self-exciting process was used to model e-mail networks infer leadership [12].

The most related work is the Cascading Non-homogeneous Poisson Process (referred to as CNPP₁) proposed to model the cascading behavior as well as periodic cycles (such as circadian and weekly cycles) of human activities [22]. In the CNPP₁ model, the primary behavior is described by an extremely low rate non-homogeneous Poisson process, while the secondary cascading behavior is described by a high rate homogeneous Poisson process. This approach, unfortunately, has intensive computation cost. A simplified version of the CNPP₁ model (referred to as CNPP₂) was proposed [20,21] to describe the cascading behaviors using a double-chain hidden Markov model, which reduces the computation burden.

In this paper, we expand the idea of the CNPP₂ model from the two-state hidden Markov chain to the multi-state hidden Markov. The new model is called Multi-state Markov Cascading Non-homogeneous Poisson Process (M2CNPP) model, which focuses on the overall behavior of e-mail communication without incorporating the detailed information from senders. In [20], one can derive a geometric distribution for the cascading pattern based on the two-state hidden Markov model. However, the proposed multi-state hidden Markov model can describe more complex structures of bursts phenomenon and better approximate realistic cascading patterns (see Figures 1 and 2). We additionally show that the maximum likelihood estimates of the intensity parameters of the M2CNPP model are the same as those from the CNPP₁ model using full likelihood function [22]. To demonstrate our approach, we apply the proposed model to the same simulated data as well as the university e-mail data set used in [22].

The paper is organized as follows. Section 2 introduces the data used in this paper and some challenges of modeling these data. Section 3 describes the basic CNPP model with its inter-event time distribution and parameter estimation. The proposed M2CNPP model and the Generalized Expectation-Maximization (GEM) parameter estimation algorithm are described in Section 4. The parameter estimates using GEM are also derived explicitly in the same section. In Section 5, the simulated data and the university e-mail data are used to demonstrate the proposed approach. Finally, Section 6 summarizes the proposed model with a discussion and future research directions.

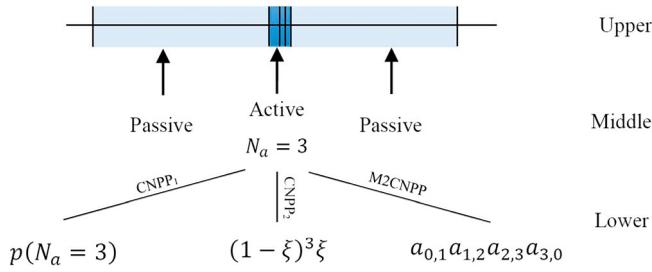


Figure 1. Example of human communication. Upper: One synthetic time series with $N_a = 3$. Middle: The states of the three mutually exclusive intervals. Lower: The probability structure of the three CNPP models for event $N_a = 3$. Lower left: the non-parametric cascading structure of $CNPP_1$. Lower middle: the two-state Markov cascading structure of $CNPP_2$. Lower right: the proposed multi-state Markov cascading structure of $M2CNPP$.

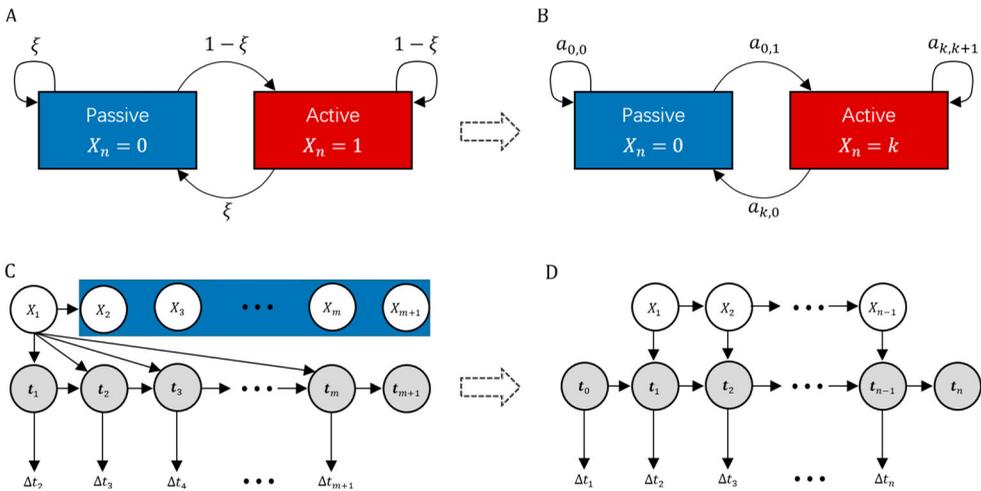


Figure 2. A ($CNPP_2$): Human activity patterns are characterized by a first-order Markov process with the transition rates $1 - \xi$ and ξ between an active state ($X_n = 1$) and a passive state ($X_n = 0$) [20]. C ($CNPP_1$): the structure of $CNPP_1$ includes $N_a = m$ cascading events within an active session, where m could be $0, 1, 2, \dots$; the highlighted part indicates these cascading events are governed by the state of X_1 . B & D ($M2CNPP$): The structure of the proposed $M2CNPP$ is a multi-stage double-chain Markov model where the time between two consecutive events Δt_n is described by a Poisson process with rate λ_a or $\lambda(t)$ depending on the hidden states. The white circles represent the missing information, while the gray circles the observed data.

2. Data, empirical patterns and modeling challenges

The data we studied were extracted from the log files of one of the main e-mail servers at one of the European universities, named Université de Genève. The data consist of more than 2×10^6 e-mail messages sent during a period of 83 days and connect about 10,000 users. The content of the messages was never accessible, and the only information taken from the log files was the ‘to,’ ‘from,’ and ‘time’ fields, and aliases were resolved. The time stamps have a precision of 1 second. After cleaning and tidying, 394 accounts were left (for more details, see the Supporting Information in [22]).

The empirical evidence has showed that the inter-event time for e-mail communication might be better approximated by a heavy-tailed distribution instead of the traditional Poisson point process. The heavy tail means the inter-event time is highly clustered at tails. It is well known that e-mail exchange follows a power law distribution [1]. However, recently some researches have indicated that e-mail exchange may not follow power law distributions at all [7,10,30,37]. Further research showed that cascades and periodic patterns exist simultaneously in e-mail communication [12,21,22]. That is, individuals are much more likely to continue writing e-mails once they have written one e-mail, in order to use their time more efficiently [21]. For example, a person, named Li, works a ‘nine to five’ job. Li arrives at the company at 9:00 a.m. and checks his e-mails. He reviews incoming e-mails and decides to reply 3 important e-mails. After sending these e-mails, Li begins to work. At 10:30 a.m., Li has important information to be sent to somebody. He sends the important information using e-mail immediately, and checks his incoming e-mails and decides to send 4 more e-mails. Li checks his e-mails before lunch. Li performs the similar routine in the afternoon. Li will not send any e-mails at night until next morning or at weekends until next Monday. The cascades with periodic patterns make time series more complex and nonstationary, and exist in many other human activities, such as individual telephone calls, running errands and mobility. Our goal is to build a simple but effective model to account for the cascades and periodic patterns simultaneously in e-mail communication. The model allows us to understand human behaviors.

3. Cascading non-homogeneous poisson process (CNPP)

3.1. Fundamental concepts

A CNPP model consists of a primary process and a secondary process. The primary process is a non-homogeneous Poisson process with a rate function $\lambda(t)$. Each event in the primary process triggers a secondary process, also called an *active session*, i.e. ‘cascades of activity’, which is modeled as a homogeneous Poisson process with a rate λ_a . During each active session there occur N_a additional events, which can be viewed as randomly drawn from some probability distribution $p(N_a)$. The intervals between any pair of consecutive active sessions are the *passive sessions*. For instance, in our toy example, Li may send a sequence of e-mails separated by a few of minutes or even hours. There exist two states in his e-mail communication. The first one is a primary process, which accounts for his sleep and work routine. The second one is a secondary process, or called *active sessions*. That is, once Li sends an e-mail, he would focus on his e-mail communication, in order to use his time more efficiently.

The intensity function $\lambda(t)$ for the primary process is periodic in time, i.e. $\lambda(t) = \lambda(t + W)$, where W is a period, and often assumed to be one week. More explicitly, we have

$$\lambda(t) = N_w p_d(t) p_w(t), \quad (1)$$

where the constant N_w denotes the average number of active intervals per week, while the rate functions $p_d(t)$ and $p_w(t)$ are the step functions for every hour and every day, respectively [22].

As noted in Section 1, the two models proposed in [22] and [20] are referred to as CNPP₁ and CNPP₂, respectively. The CNPP₁ model consists of three separate but related components, as illustrated in Figure 1. The distribution of the number of events between starting and ending time of an active session is described by the distribution $p(N_a)$, which is independent of any other active sessions. A random variable X with values in $\{\mathbf{0}, \mathbf{1}\}$ specifies the state of the next e-mail sending event. More precisely, after the k th event, $X_k = \mathbf{0}$ implies that the $k+1$ th event follows a non-homogeneous Poisson process with intensity function $\lambda(t)$, while $X_k = \mathbf{1}$ implies that the $k+1$ th event follows a homogeneous Poisson process with intensity λ_a . The sequence of states defines the underlying active interval configuration C , which can be represented by the sequence $J = \{0 = j_0 < j_1 < j_2 < \dots < j_m \leq n - 1\}$, where the state is $\mathbf{0}$ after the k th event iff $k \in J$. The sequence J represents not only if the k th event occurs on the primary process, but also counts the number of secondary events in each active session.

For the CNPP₂ model, the secondary events occurring in an active session are assumed to be independent and the change of states is determined by the transition probability ξ (Figure 2(a)). The lower middle plot of Figure 1 shows the probability that three events occur in an active session, which can be calculated as $p(N_a = 3) = (1 - \xi)^3 \xi$, and thus follows a geometric distribution. The CNPP₂ model can be viewed as a Double-Chain Hidden Markov model (Figure 2(d)).

3.2. Parameter estimation

Let $t_0 = 0$, and assume t_1, \dots, t_n be the observed time points where t_k is the occurring time of the k th event. Let $\mathbf{t}_1^n = (t_1, \dots, t_n)$ represent the set of time points. The value of X_k represents the state in the time interval $(t_k, t_{k+1}]$. Then the full likelihood functions of models CNPP₁ and CNPP₂ are given below, respectively:

$$\mathcal{L}_1 = P\{\Phi_1, \mathbb{P}; C, \mathbf{t}_1^n\} = \prod_{1 \leq i \leq m} p(N_a = j_i - j_{i-1} - 1) \prod_{i \in J} f(t_{i+1}) \prod_{\substack{\bar{i} \in J \\ 0 < i \leq n-1}} g(t_{i+1}), \quad (2)$$

$$\mathcal{L}_2 = P\{\Phi_2; C, \mathbf{t}_1^n\} = \xi^m (1 - \xi)^{n-m} \prod_{i \in J} f(t_{i+1}) \prod_{\substack{\bar{i} \in J \\ 0 < i \leq n-1}} g(t_{i+1}), \quad (3)$$

where Φ_1 and Φ_2 are the parameter sets; \mathbb{P} represents the distribution of N_a ; $f(t_k)$ represents the probability density function of an event from the non-homogeneous Poisson process occurring at t_k and no event occurring in $(t_{k-1}, t_k]$; while $g(t_k)$ represents the probability density function of an event from the homogeneous Poisson process occurring at t_k and no event occurring in $(t_{k-1}, t_k]$. Then by taking logarithm and expectation on both sides of (2) and (3), we get:

$$Q_1(\Phi_1, \mathbb{P}; \mathbf{t}_1^n) = \mathbb{E}_C[\log \mathcal{L}_1 | \mathbf{t}_1^n; \Phi_1, \mathbb{P}]; \quad (4)$$

$$Q_2(\Phi_2; \mathbf{t}_1^n) = \mathbb{E}_C[\log \mathcal{L}_2 | \mathbf{t}_1^n; \Phi_2]. \quad (5)$$

For the CNPP₁ model, one can estimate the parameter set Φ_1 and the distribution \mathbb{P} given a sequence of observations by minimizing $Q_1(\Phi_1, \mathbb{P}; \mathbf{t}_1^n)$. Thus, the Expectation Maximization (EM) algorithm can be used to simultaneously estimate Φ_1 and \mathbb{P} . However,

the EM algorithm needs to consider all the 2^{n-1} possible arrangements of the sequence J . Alternatively, one can estimate the parameter set Φ_1 , the distribution \mathbb{P} , and the configuration C given a sequence of observations by minimizing the area test statistic S (see Section 4.4.2). The method of simulated annealing inference can also be used [22]. For the CNPP₂ model, the EM algorithm can be used to estimate the parameters quickly and efficiently by minimizing $Q_2(\Phi_2; \mathbf{t}_1^n)$ [20].

4. Multi-state Markov cascading non-homogeneous Poisson process (M2CNPP)

In this section, we introduce our proposed model, the multi-state Markov cascading non-homogeneous Poisson process (M2CNPP). The fundamental concepts of the M2CNPP model are explained in Section 4.1. Then the parameter estimation using the Generalized Expectation and Maximization (GEM) algorithm is given in Section 4.3. Finally, the model selection and the performance evaluation criteria are introduced in Section 4.4.

4.1. The basic concepts of M2CNPP

There present habitual patterns in human activities such as completing the jobs with top priorities first. These patterns are formed over time in order to achieve efficiency. In e-mail communication, the responding time to e-mail directly reflects our habitual behavior. This behavior appears to grow intensively at the beginning, and then fades away. However, for the CNPP₂ model, because of the time-invariant feature, the parameter ξ is unchanged and remains constant. This assumption is not realistic. Therefore, a new multi-stage Markov cascading non-homogenous Poisson process (M2CNPP) model is proposed to count for more complicated patterns in human e-mail communication.

Let $\{\mathbf{0}, \mathbf{1}, \dots, \mathbf{r}\}$ represent the hidden states of an *active session*. We assume that at state $k \in \{\mathbf{0}, \mathbf{1}, \dots, \mathbf{r}\}$, an individual moves to the passive state $\mathbf{0}$ with the transition probability $a_{k,0}$ or to the next active state $k+1$ with the transition probability $a_{k,k+1}$. All other transition probabilities are assumed zero. With this assumption, the independence property is retained and the M2CNPP model can also be viewed as a less constrained double-chain hidden Markov model. The M2CNPP model is visualized in the bottom right graph of Figures 1 and 2(b).

Denote $S(X) = \{\mathbf{0}, \mathbf{1}, \dots, \mathbf{r}\}$ as the hidden state space. The transition probability matrix of the hidden states is written as $A = \{a_{i,j}\}$, $i, j \in S(X)$. More specifically, we have the form specified as Equation (6).

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & 0 & 0 & \cdots & 0 \\ a_{1,0} & 0 & a_{1,2} & 0 & \cdots & 0 \\ a_{2,0} & 0 & 0 & a_{2,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{\mathbf{r}-1,0} & 0 & 0 & \cdots & 0 & a_{\mathbf{r}-1,\mathbf{r}} \\ a_{\mathbf{r},0} & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad (6)$$

where $a_{k,0} + a_{k,k+1} = 1$, $0 \leq k \leq r - 1$ and $a_{r,0} = 1$. The probability distribution of the first hidden state is assumed as $\pi = \{a_{0,0}, a_{0,1}, 0, \dots, 0\}$. The rate of the M2CNPP model at state $\mathbf{0}$ is defined as $\lambda(t)$ (Equation (1)); otherwise it is a constant.

Remark 4.1: A hidden variable is introduced to describe whether an e-mail belongs to an *active session* or not, rather than a sequence of e-mails belong to an *active session*. For the former case, there are only n hidden variables to estimate, but the latter case has 2^n hidden variables to estimate. As a summary, the proposed M2CNPP has the following unique features:

- (a) The secondary process follows a multi-stage Markov homogeneous Poisson process, which accounts for the cascading patterns;
- (b) The hidden states in the secondary process are treated as missing values, and the GEM algorithm is used to estimate parameters.

4.2. Simple examples

In this section, we provide two simple examples to illustrate the proposed M2CNPP model. Let

$$\lambda(t) = 0.01 \cdot \mathbf{1}_{t \in \{\text{Sun., Sat.}\}} + 0.3 \cdot \mathbf{1}_{t \in \{\text{Mon., \dots, Fri.}\}},$$

where $\mathbf{1}_{t \in \{\cdot\}}$ is an indicator function. The function $\lambda(t)$ is the intensity function of the primary process, which is a non-homogeneous Poisson process. Furthermore, we define $\lambda_a = 10$ be the intensity rate of the secondary process, which is an homogeneous Poisson process. The hidden state space $S(X)$ has three states $\{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$. In particular, $X_i = \mathbf{0}$ implies that the $i+1$ th event follows a non-homogeneous Poisson process with intensity function $\lambda(t)$; $X_i = \mathbf{1}$ or $X_i = \mathbf{2}$ implies that the $i+1$ th event follows the homogeneous Poisson process with intensity λ_a . The sequence of states defines the underlying active/passive status of next e-mail, and Poisson processes determine the time of event occurrence. Furthermore, we let

$$A = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.8 & 0 & 0.2 \\ 1 & 0 & 0 \end{bmatrix}$$

be the transition probability matrix of the hidden states. Given an initial time t_0 on Monday, and the initial state X_0 is $\mathbf{0}$, an e-mail is sent at random time point t_1 where t_1 has probability density function $\lambda(t_1) \exp\{-\int_{t_0}^{t_1} \lambda(t) dt\}$. We consider the following two cases.

Case 1. Consider state transfer random variables $X_1 = \mathbf{1}$, $X_2 = \mathbf{2}$ and $X_3 = \mathbf{0}$. $X_1 = \mathbf{1}$ indicates the state at t_1 is $\mathbf{1}$, i.e. an active session, with transition probability $a_{0,1} = 0.3$. The next e-mail is sent at random time point t_2 with conditional probability density function $10 \exp\{-10(t_2 - t_1)\}$ given time point t_1 . The second state is $X_2 = \mathbf{2}$, with transition probability $a_{1,2} = 0.2$. The third e-mail is sent at random time point t_3 with conditional probability density function $10 \exp\{-10(t_3 - t_2)\}$ given time point sequence t_1, t_2 . The third state is $X_3 = \mathbf{0}$, i.e. the active session ends, with transition probability $a_{2,0} = 1$. The fourth e-mail is sent at random time

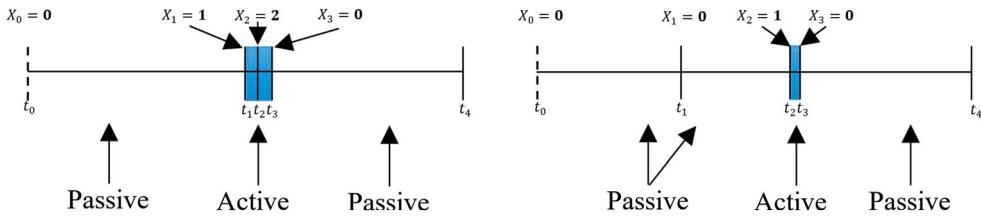


Figure 3. Example of M2CNPP model. Left: Case 1; Right: Case 2.

point t_4 with conditional probability density function $\lambda(t_4) \exp\{-\int_{t_3}^{t_4} \lambda(t) dt\}$ given time point sequence t_1, t_2, t_3 . The complete likelihood function of the time point sequence $\{t_1, t_2, t_3, t_4\}$ is

$$\begin{aligned} \mathcal{L} &= P(\mathbf{t}_1^4, X_0 = \mathbf{0}, X_1 = \mathbf{1}, X_2 = \mathbf{2}, X_3 = \mathbf{0}) \\ &= \lambda(t_1) \exp\left\{-\int_{t_0}^{t_1} \lambda(t) dt\right\} \cdot 0.3 \cdot 10 \exp\{-10(t_2 - t_1)\} \\ &\quad \cdot 0.2 \cdot 10 \exp\{-10(t_3 - t_2)\} \cdot 1 \cdot \lambda(t_4) \exp\left\{-\int_{t_3}^{t_4} \lambda(t) dt\right\}. \end{aligned}$$

Case 2. Consider state transfer random variables $X_1 = \mathbf{0}$, $X_2 = \mathbf{1}$ and $X_3 = \mathbf{0}$. The first state is $X_1 = \mathbf{0}$, with transition probability $a_{0,0} = 0.7$. The next e-mail is sent at random time point t_2 with conditional probability density function $\lambda(t_2) \exp\{-\int_{t_1}^{t_2} \lambda(t) dt\}$ given time point t_1 . The second state is $X_2 = \mathbf{1}$, with transition probability $a_{0,1} = 0.3$. The third e-mail is sent at random time point t_3 with conditional probability density function $10 \exp\{-10(t_3 - t_2)\}$ given time point sequence t_1, t_2 . The third state is $X_3 = \mathbf{0}$, with transition probability $a_{1,0} = 0.8$. The fourth e-mail is sent at random time point t_4 with conditional probability density function $\lambda(t_4) \exp\{-\int_{t_3}^{t_4} \lambda(t) dt\}$ given time point sequence t_1, t_2, t_3 . For this case, the complete likelihood function of the time point sequence $\{t_1, t_2, t_3, t_4\}$ is expressed as

$$\begin{aligned} \mathcal{L} &= P(\mathbf{t}_1^4, X_0 = \mathbf{0}, X_1 = \mathbf{0}, X_2 = \mathbf{1}, X_3 = \mathbf{0}) \\ &= \lambda(t_1) \exp\left\{-\int_{t_0}^{t_1} \lambda(t) dt\right\} \cdot 0.7 \cdot \lambda(t_2) \exp\left\{-\int_{t_1}^{t_2} \lambda(t) dt\right\} \\ &\quad \cdot 0.3 \cdot 10 \exp\{-10(t_3 - t_2)\} \cdot 0.8 \cdot \lambda(t_4) \exp\left\{-\int_{t_3}^{t_4} \lambda(t) dt\right\}. \end{aligned}$$

Figure 3 demonstrates the two cases. These examples are extended in Section 5.1 for evaluating the proposed M2CNPP model.

4.3. The generalized expectation maximization algorithm (GEM)

It is impossible to directly estimate the parameter set $\Phi = \{\lambda_a, \lambda(t) : \lambda(t) > 0 \text{ for } t, t \in \mathbb{R}_+\}$ and the transition probability matrix A based on a sequence of observed time points,

since the states of sessions (i.e. either passive or active) are missing. Therefore, we introduce the GEM Algorithm for the purpose of parameter estimation.

Let $I = \{i_1, \dots, i_{n-1}\}$ represent the realization of the hidden states $\{X_1, \dots, X_{n-1}\}$. The complete likelihood function of an observed time point sequence can be written as follows:

$$\mathcal{L} = P(\mathbf{t}_1^n, I; \Phi, A) = f(t_1) a_{0,i_1} b_{i_1}(t_2) a_{i_1,i_2} b_{i_2}(t_3) \cdots a_{i_{n-2},i_{n-1}} b_{i_{n-1}}(t_n), \tag{7}$$

where

$$b_i(\cdot) = \begin{cases} f(\cdot), & i = 0, \\ g(\cdot), & i = 1, \dots, r. \end{cases}$$

$f(\cdot)$ and $g(\cdot)$ are defined as Equations (A1) and (A2) in Appendix 1, respectively.

Let $\lambda(t) = \lambda_{i,j} = \lambda_i^h \lambda_j^d, i = 1, \dots, 24, j = 1, \dots, 7$ where λ_i^h and λ_j^d are the hourly and daily rates, respectively. Given the values of $\Phi^{(l)}$ and $A^{(l)}$ at the l th iteration, the expectation of the log-likelihood function is written as follows:

$$\begin{aligned} Q(\Phi, A; \Phi^{(l)}, A^{(l)}) &= \mathbb{E}_I[\log \mathcal{L} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)}] \\ &= \sum_I \log a_{0,i_1} P(I | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)}) \\ &\quad + \sum_{k=1}^{n-2} \sum_I \log a_{i_k, i_{k+1}} P(I | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)}) + \log f(t_1) \\ &\quad + \sum_{k=1}^{n-1} \sum_I \log b_{i_k}(t_{k+1}) P(I | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)}). \end{aligned} \tag{8}$$

Note that the optimal solutions of the intensity rates λ_i^h and $\lambda_j^d, i = 1, \dots, 24, j = 1, \dots, 7$ in maximizing function (8) are not unique as $\{\lambda_i^h, \lambda_j^d\}$ and $\{(1/k)\lambda_i^h, k\lambda_j^d\}$ for $k \in \mathbb{R}_+$ give the same value. In order to avoid this situation, we normalize the results, that is, let $\lambda_{i,j} = N_w \lambda_i^h \lambda_j^d$ where $\sum_{j=1}^7 \lambda_j^d = 1$ and $\sum_{i=1}^{24} \lambda_i^h = 1$. Then the set of parameters is $\Phi = \{N_w, \lambda_a, \lambda_i^h, \lambda_j^d, i = 1, \dots, 24, j = 1, \dots, 7 : \sum_{j=1}^7 \lambda_j^d = 1, \sum_{i=1}^{24} \lambda_i^h = 1\}$. The normalized λ_j^d is the intensity for a day of a week, and the normalized λ_i^h for an hour of a day and N_w can be seen as the average number of active intervals per period. Then the maximum likelihood estimates of the parameters can be estimated by GEM [23,24] as outlined in Algorithm 1.

The property of convergence of Algorithm 1 is shown in Theorem 4.1 (see Appendix 1 for the detailed proof).

Theorem 4.1: *Algorithm 1 is a GEM algorithm, that is,*

$$Q(\Phi^{(l+1)}, A^{(l+1)}; \Phi^{(l)}, A^{(l)}) \geq Q(\Phi^{(l)}, A^{(l)}; \Phi^{(l)}, A^{(l)}), \tag{9}$$

where $A^{(l)}$ and $\Phi^{(l)} = \left\{ \widehat{N}_w^{(l)}, \widehat{\lambda}_a^{(l)}, \widehat{\lambda}_i^h^{(l)}, \widehat{\lambda}_j^d^{(l)}, i = 1, \dots, 24, j = 1, \dots, 7 \right\}$ are the estimated parameter values at the l th iteration.

Algorithm 1 GEM-based Maximum Likelihood Estimation for the M2CNPP model

Input: Model $\mathcal{M}(\theta)$, $\theta \in \Theta$; Randomly initialize $\theta^{(0)} = \{\Phi^{(0)}, A^{(0)}\}$; Convergence criterion tol ; The total number of iteration N

Output: The Maximum Likelihood estimate $\hat{\theta}$ for θ

– Set $l = 0$

while $\frac{|Q^{(l+1)} - Q^{(l)}|}{|Q^{(l)}|} \leq tol$ or $l \leq N$ **do**

 % *E-step: Calculate the Expectation*

 – Using Forward-Backward procedure [4] to compute $P(X_k = \mathbf{0}, X_{k+1} = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$, $P(X_k = \mathbf{0} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$, $P(X_k = i, X_{k+1} = j | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$ and $P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$, $i, j \in S(X)$, $k = 0, 1, \dots, n-2$,

 % *M-step: Maximize the Expectation* $Q^{(l+1)}(\Phi, A; \Phi^{(l)}, A^{(l)})$

$$\hat{a}_{\mathbf{0},i}^{(l+1)} \leftarrow \frac{\sum_{k=0}^{n-2} P(X_k = \mathbf{0}, X_{k+1} = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=0}^{n-2} P(X_k = \mathbf{0} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}, i = \mathbf{0}, \mathbf{1},$$

$$\hat{a}_{i,j}^{(l+1)} \leftarrow \frac{\sum_{k=1}^{n-2} P(X_k = i, X_{k+1} = j | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=1}^{n-2} P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}, j \in \{\mathbf{0}, i+1\}, i = \mathbf{1}, \dots, r-1,$$

$$\hat{\lambda}_j^d{}^{(l+1)} \leftarrow \frac{\tilde{\lambda}_j^d{}^{(l+1)}}{\sum_{j=1}^7 \tilde{\lambda}_j^d{}^{(l+1)}}, j = 1, \dots, 7,$$

$$\hat{\lambda}_i^h{}^{(l+1)} \leftarrow \frac{\tilde{\lambda}_i^h{}^{(l+1)}}{\sum_{i=1}^{24} \tilde{\lambda}_i^h{}^{(l+1)}}, i = 1, \dots, 24,$$

$$\hat{N}_w^{(l+1)} \leftarrow (\sum_{j=1}^7 \tilde{\lambda}_j^d{}^{(l+1)}) (\sum_{i=1}^{24} \tilde{\lambda}_i^h{}^{(l+1)}),$$

$$\hat{\lambda}_a^{(l+1)} \leftarrow \frac{\sum_{k=1}^{n-1} \sum_{i=1}^r P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=1}^{n-1} [(t_{k+1} - t_k) \sum_{i=1}^r P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})]},$$

where $\tilde{\lambda}_j^d{}^{(l+1)} = \frac{\sum_{i=1}^{24} d_{ij}^{(l)}}{\sum_{i=1}^{24} \tilde{\lambda}_i^h{}^{(l)} c_{ij}^{(l)}}$, $j = 1, \dots, 7$, $\tilde{\lambda}_i^h{}^{(l+1)} = \frac{\sum_{j=1}^7 d_{ij}^{(l)}}{\sum_{j=1}^7 \tilde{\lambda}_j^d{}^{(l+1)} c_{ij}^{(l)}}$, $i = 1, \dots, 24$,

$$c_{i,j}^{(l)} = \sum_k (t_{k+1} - t_k)_{i,j} P(X_k = \mathbf{0} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)}) \quad \text{and} \quad d_{i,j}^{(l)} = \sum_{k \in K_1} P(X_k = \mathbf{0} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$$

– Set $l = l + 1$

end while

4.4. Model selection and performance evaluation

4.4.1. Parameter selection criteria

The number of states \mathbf{r} can be pre-defined or can be estimated using model selection criteria [33]. In this paper, two model evaluation criteria are compared based on how accurately the models identify the true number of states \mathbf{r} for a simulated data set. The two criteria are Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC). Then the criterion with the best performance will be chosen for the real data. The two criteria are defined as Equations (10) and (11), respectively.

$$\text{BIC} = -2LL + p \log(n), \quad (10)$$

$$\text{AIC} = -2LL + 2p, \quad (11)$$

where LL is the log-likelihood function, p is the number of independent parameters and n the number of observations. Specifically, we generate many synthetic time series from the M2CNPP model, each of which has a fixed number of events with r cascading states and a set of fixed parameters. The procedure of generating simulated data of the M2CNPP model is described in Algorithm 2. This algorithm is a modified version of the classic thinning algorithm [19]. The algorithm works as follows: (1) Generate $t_1^* < t_2^* < \dots < t_{N_T}^*$ from $(0, T]$ to form a homogeneous Poisson process with constant intensity λ^* where $\lambda^* \geq \sup_{t \in (0, T]} \lambda(t)$; (2) Remove the time points t_j^* with probability $1 - \lambda_{t_j^*}/\lambda^*$; (3) The remaining points form a non-homogeneous point process with intensity function $\lambda(t)$.

Algorithm 2 Simulation of the M2CNPP Model

Input: Primary process rate $\hat{\lambda}(t)$; Secondary process rate $\hat{\lambda}_a$; Transition probability matrix \hat{A} ; Data size n
Output: The M2CNPP points t_1, t_2, \dots, t_n
 – Set $\lambda^* = \sup_t \hat{\lambda}(t)$, $s_i = 0$ and $t_j = 0$ where $i = j = 0$
while $j < N$ **do**
 –① Set $s_i = 0$ where $i = 0$
 –② Generate a homogeneous Poisson process with intensity λ^* on $(0, +\infty)$; Denote the first point by Δs , set $i = i + 1$ and $s_i = t_j + s_{i-1} + \Delta s$
 –③ Generate a uniform random number u_0 on $[0, 1]$
 –④ If $u_0 \leq \frac{\lambda(s_i)}{\lambda^*}$, set $j = j + 1$ and $t_j = s_i$; Else, return to ②
 –⑤ Generate r uniform random numbers u^k ($k = 1, \dots, r$) on $[0, 1]$; Find the first k s.t. $u^k < \hat{a}_{m-1,0}$ ($m = 1, \dots, r + 1$) and set $N_a = k - 1$
 –⑥ If $N_a > 0$, generate a homogeneous Poisson process with intensity $\hat{\lambda}_a$ on $(0, +\infty)$, and denote Δt_l ($l = 1, \dots, \min\{N_a, N - j\}$) as the time interval, then set $t_{j+l} = t_{j+l-1} + \Delta t_l$, $l = 1, \dots, \min\{N_a, N - j\}$; Else, return to ①
 –⑦ Collect $\{t_j\}$ and set $j = j + \min\{N_a, n - j\}$
end while

4.4.2. Model performance evaluation with the simulated data

[22] quantified the agreement between the model $\mathcal{M}(\theta)$ with parameters θ and data \mathcal{D} by measuring the area statistic S between the empirical cumulative distribution function $P_{\mathcal{D}}(u)$ and the model cumulative distribution function $P_{\mathcal{M}}(u; \theta)$ presented as follows:

$$S = \int |P_{\mathcal{D}}(u) - P_{\mathcal{M}}(u; \theta)| du. \tag{12}$$

Rather than using Equation (12) directly to calculate the area statistic S , we decide to compute p -values of the agreement using a Monte Carlo procedure: (1) Estimate the parameters θ and denote as $\hat{\theta}$; (2) Calculate the area test statistic S between $P_{\mathcal{M}}(u; \hat{\theta})$ and $P_{\mathcal{D}}(u)$; (3) Generate M synthetic datasets \mathcal{D}_s ($s = 1, \dots, M$) from the model $\mathcal{M}(\hat{\theta})$ using the estimates $\hat{\theta}$. The simulation procedure results in an assemble of M test statistics S_s , $s = 1, \dots, M$. Finally, a two-tailed p -value with a precision of $1/M$, i.e. $P_r(|S_s - < S_s >| > |S - < S_s >|)$ is calculated. The procedure is described in Algorithm 3.

For the proposed M2CNPP model, $M = 400$ Monte Carlo synthetic datasets are generated. The models with p -values less than 5% are rejected. This method indicates that a

model is rejected if the predictions are inconsistent with the empirical observations \mathcal{D} at a conservative 5% significance level.

Algorithm 3 Monte Carlo Hypothesis Testing

Input: Dataset \mathcal{D} ; Model $\mathcal{M}(\theta)$, $\theta \in \Theta$; Number M

Output: Two-tailed p -value $Pr(|S_s - < S_s >| > |S - < S >|)$

Initialization:

– Estimate $\hat{\theta}$ using GEM

– Calculate the area test statistic S between model $\mathcal{M}(\hat{\theta})$ and data \mathcal{D}

for $s = 1$ to M **do**

– Generate a synthetic dataset \mathcal{D}_s from model $\mathcal{M}(\hat{\theta})$

– Estimate $\hat{\theta}_s$ using data \mathcal{D}_s

– Calculate the area test statistic S_s between model $\mathcal{M}(\hat{\theta}_s)$ and data \mathcal{D}_s

– Collect $\{S_s\}$

end for

– Calculate a two-tailed p -value $Pr(|S_s - < S_s >| > |S - < S >|)$

5. Results

Both simulated and real data are used to evaluate the proposed M2CNPP model. For the simulated data, the performance of the two model evaluation criteria is compared based on how accurately the models identify the true total number of the states \mathbf{r} . Then the better criterion is chosen and used for parameter estimation for the real data.

5.1. Simulation

Firstly, we verify that the GEM procedure provides valid parameter estimates. Four hundred synthetic time series are generated, each of which has 400 events with 4 cascading states (i.e. $\mathbf{r} = 4$). The values of the intensity parameters $\lambda(t)$ and λ_a are chosen based on the following assumptions: (1) people usually do not work at night; (2) people only work from 9: 00 a.m. to 6: 00 p.m.; (3) people check e-mails from 5: 00 p.m. to 6: 00 p.m before going home. Table 1 lists the parameter values, and Equation (13) provides the transition matrix A for model simulation.

$$A = \begin{bmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 & 0 \\ 0.1 & 0 & 0 & 0.9 & 0 \\ 0.7 & 0 & 0 & 0 & 0.3 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (13)$$

5.1.1. Estimating \mathbf{r}

Two criteria (i.e. BIC and AIC) are used to estimate the hidden states \mathbf{r} . Figure 4 displays the results based on the simulated data. For AIC, there are 82 out of 400 trials that fail to estimate the correct number of the hidden states \mathbf{r} . However, for BIC, the number of trials that fail to estimate the correct \mathbf{r} decreases to 53. For both AIC and BIC, most trials estimate $\mathbf{r} = 3$. The reason why \mathbf{r} is more frequently estimated as 3 by two criteria is because we set

Table 1. The values $\lambda_i^h, \lambda_j^d, \lambda_a$ and N_w used for simulation.

λ_i^h	hour	value	λ_i^h	hour	value	λ_j^d	day	value
λ_1^h	0 ~ 1	0	λ_{13}^h	12 ~ 13	0.06	λ_1^d	Mon.	0
λ_2^h	1 ~ 2	0	λ_{14}^h	13 ~ 14	0.058	λ_2^d	Tue.	0.15
λ_3^h	2 ~ 3	0	λ_{15}^h	14 ~ 15	0.1	λ_3^d	Wed.	0.2
λ_4^h	3 ~ 4	0	λ_{16}^h	15 ~ 16	0.08	λ_4^d	Thu.	0.2
λ_5^h	4 ~ 5	0	λ_{17}^h	16 ~ 17	0.05	λ_5^d	Fri.	0.15
λ_6^h	5 ~ 6	0	λ_{18}^h	17 ~ 18	0.15	λ_6^d	Sat.	0.1
λ_7^h	6 ~ 7	0	λ_{19}^h	18 ~ 19	0.05	λ_7^d	Sun.	0.2
λ_8^h	7 ~ 8	0	λ_{20}^h	19 ~ 20	0.04			
λ_9^h	8 ~ 9	0.025	λ_{21}^h	20 ~ 21	0.002			
λ_{10}^h	9 ~ 10	0.1	λ_{22}^h	21 ~ 22	0.007			
λ_{11}^h	10 ~ 11	0.15	λ_{23}^h	22 ~ 23	0.004	λ_a		20
λ_{12}^h	11 ~ 12	0.16	λ_{24}^h	23 ~ 24	0	N_w		20

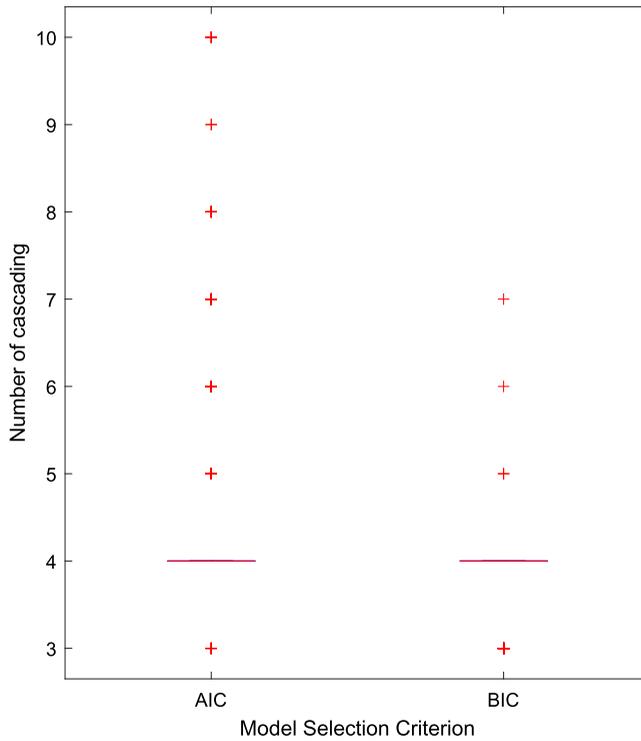


Figure 4. Estimating \mathbf{r} with AIC and BIC.

up $a_{3,0} = 0.7$, a considerably large transition probability. Both AIC and BIC try to select simple models.

After \mathbf{r} is chosen, the estimates of other parameters can be easily obtained. The biases and standard deviations of $\widehat{A} - A, \widehat{\lambda}_i^h - \lambda_i^h (i = 1, \dots, 24), \widehat{\lambda}_j^d - \lambda_j^d (j = 1, \dots, 7), \widehat{\lambda}_a - \lambda_a$ and $\widehat{N}_w - N_w$ are shown in Matrix (14) and Table 2 (standard deviation (sd)

Table 2. The estimates of $\widehat{\lambda}_i^h, \widehat{\lambda}_j^d, \widehat{\lambda}_a$ and \widehat{N}_w with standard deviations in parentheses.

$\widehat{\lambda}_i^h - \lambda_i^h$	mean $\times 10^{-4}$ (sd $\times 10^{-2}$)	$\widehat{\lambda}_i^h - \lambda_i^h$	mean $\times 10^{-4}$ (sd $\times 10^{-2}$)	$\widehat{\lambda}_j^d - \lambda_j^d$	mean $\times 10^{-4}$ (sd $\times 10^{-2}$)
$\widehat{\lambda}_1^h - \lambda_1^h$	0 (0)	$\widehat{\lambda}_{13}^h - \lambda_{13}^h$	-7.21 (1.57)	$\widehat{\lambda}_1^d - \lambda_1^d$	0 (0)
$\widehat{\lambda}_2^h - \lambda_2^h$	0 (0)	$\widehat{\lambda}_{14}^h - \lambda_{14}^h$	-1.77 (1.52)	$\widehat{\lambda}_2^d - \lambda_2^d$	-6.05 (2.16)
$\widehat{\lambda}_3^h - \lambda_3^h$	0 (0)	$\widehat{\lambda}_{15}^h - \lambda_{15}^h$	-5.98 (2.02)	$\widehat{\lambda}_3^d - \lambda_3^d$	-10.30 (2.47)
$\widehat{\lambda}_4^h - \lambda_4^h$	0 (0)	$\widehat{\lambda}_{16}^h - \lambda_{16}^h$	-12.00 (1.70)	$\widehat{\lambda}_4^d - \lambda_4^d$	17.20 (2.39)
$\widehat{\lambda}_5^h - \lambda_5^h$	0 (0)	$\widehat{\lambda}_{17}^h - \lambda_{17}^h$	0.41 (1.33)	$\widehat{\lambda}_5^d - \lambda_5^d$	12.00 (2.12)
$\widehat{\lambda}_6^h - \lambda_6^h$	0 (0)	$\widehat{\lambda}_{18}^h - \lambda_{18}^h$	-0.92 (2.35)	$\widehat{\lambda}_6^d - \lambda_6^d$	10.10 (1.75)
$\widehat{\lambda}_7^h - \lambda_7^h$	0 (0)	$\widehat{\lambda}_{19}^h - \lambda_{19}^h$	5.79 (1.38)	$\widehat{\lambda}_7^d - \lambda_7^d$	1.02 (2.54)
$\widehat{\lambda}_8^h - \lambda_8^h$	0 (0)	$\widehat{\lambda}_{20}^h - \lambda_{20}^h$	0.78 (0.38)		
$\widehat{\lambda}_9^h - \lambda_9^h$	2.78 (0.98)	$\widehat{\lambda}_{21}^h - \lambda_{21}^h$	0.92 (0.27)		
$\widehat{\lambda}_{10}^h - \lambda_{10}^h$	11.80 (1.81)	$\widehat{\lambda}_{22}^h - \lambda_{22}^h$	1.81 (0.52)		
$\widehat{\lambda}_{11}^h - \lambda_{11}^h$	14.20 (2.23)	$\widehat{\lambda}_{23}^h - \lambda_{23}^h$	0.96 (0.39)	$\widehat{\lambda}_a - \lambda_a$	9.42(130.77)
$\widehat{\lambda}_{12}^h - \lambda_{12}^h$	0.25 (2.38)	$\widehat{\lambda}_{24}^h - \lambda_{24}^h$	5.14×10^{-6} (7.92×10^{-6})	$\widehat{N}_w - N_w$	-2.5×10^3 (213.54)

in parentheses).

$$\widehat{A} - A : 10^{-2} \times \begin{bmatrix} -4.68(3.46) & 4.22(3.46) & 0 & 0 & 0 \\ 9.81(8.77) & 0 & -10.29(8.72) & 0 & 0 \\ 7.06(7.35) & 0 & 0 & -6.9(7.14) & 0 \\ -2.17(9.54) & 0 & 0 & 0 & 3.36(9.33) \\ 0(0) & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{14}$$

Matrix (14) and Table 2 show that the biases are very close to zero for most parameters. Even for the parameters with large biases, the true values of the parameters actually fall in the 95% confidence interval of the estimates. This results suggest that these parameter estimates are asymptotically unbiased. Furthermore, we also confirm empirically that parameter estimates are insensitive to the choice of initial values. The mean and the standard deviation of iterations are 6.46 and 0.474, respectively. This indicates that the algorithm converges to the true values very quickly.

5.2. Results from real data

In this section, we implement the proposed M2CNPP model to the university e-mail data described in Section 2. The data have been studied in [9] and [22]. Figure 5 displays the e-mail communication patterns of 4 users with an increasing order of e-mail usage. Figure 5(a,b) indicates that the active intervals much more likely happen during weekdays rather than weekends and during daytime rather than nighttime. These observations are similar to those reported in the previous research [22]. In fact, the parameter estimates of the M2CNPP proposed in this paper and CNPP₁ proposed in [22] are consistent, see Appendix 1.

Figure 6 compares the prediction of the proposed M2CNPP model with the empirical cumulative distribution of inter-event times $P(\tau)$ for the 4 users from Figure 5. In these empirical inter-event time distributions of e-mail communication, the heavy tail exists because of the prolonged periods of inactivity and short-term bursts of active intervals. The method of Monte Carlo hypothesis testing is used to assess the significance of the agreement between the prediction of our model and the data. Except for the user with id 2881, the results presented from the remaining three users have high p -values and are

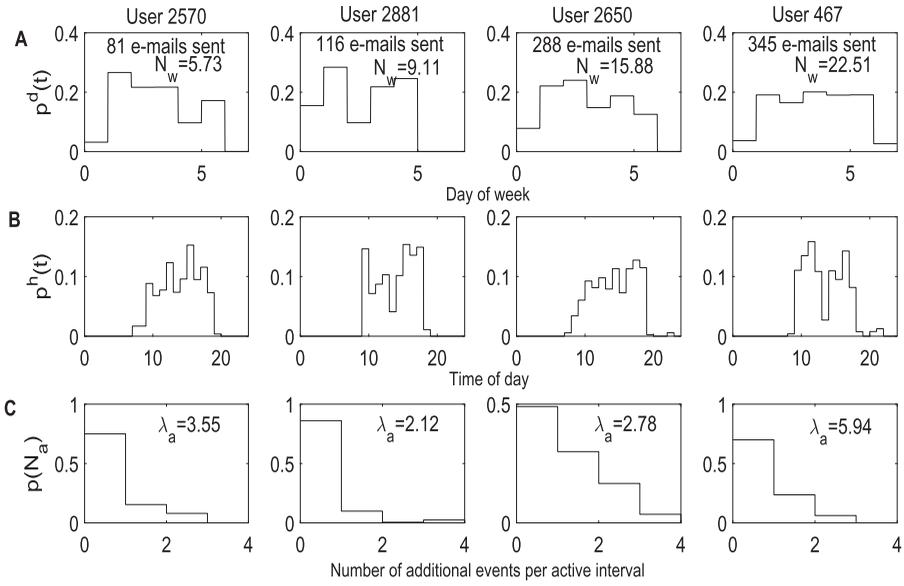


Figure 5. E-mail communication patterns of four users in an increasing order of e-mail usage. These cases exemplify the e-mail usage patterns that are typically observed in the e-mail communication. (A) Distribution of active sessions occurring within a week; (B) Distribution of active sessions occurring within a day; (C) Transition probability.

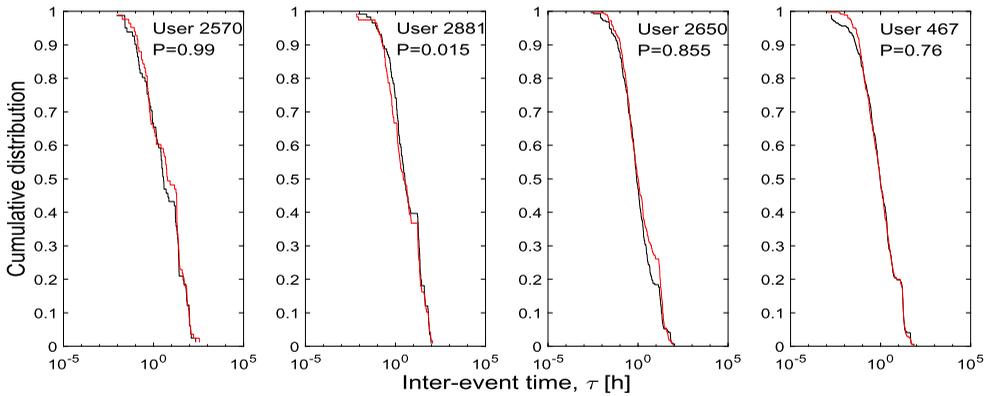


Figure 6. Comparison of the prediction of the proposed M2CNPP model (red line) with the empirical cumulative distribution of inter-event times $P(\tau)$ (black line) for the 4 users in Figure 5.

consistent with the empirical cumulative distribution of inter-event times of e-mail communication. Figure 7 displays the conditional probability density $p(R|\tau)$ with log-residual $R = \ln(p_{\mathcal{M}}(\tau|\hat{\theta})/p(\tau))$, where $p_{\mathcal{M}}(\tau|\hat{\theta})$ is the inter-event time distribution of the best-fit M2CNPP model $\mathcal{M}(\hat{\theta})$, and $p(\tau)$ is the empirical inter-event time distribution obtained for all 394 users under consideration. From Figure 7, it can be seen that our model does not display any large systematic deviations. We also discover there are no systematic deviations between the model prediction and the data at the tail of the inter-event time distribution where the power-law scaling approximately holds. The cascading pattern distributions are

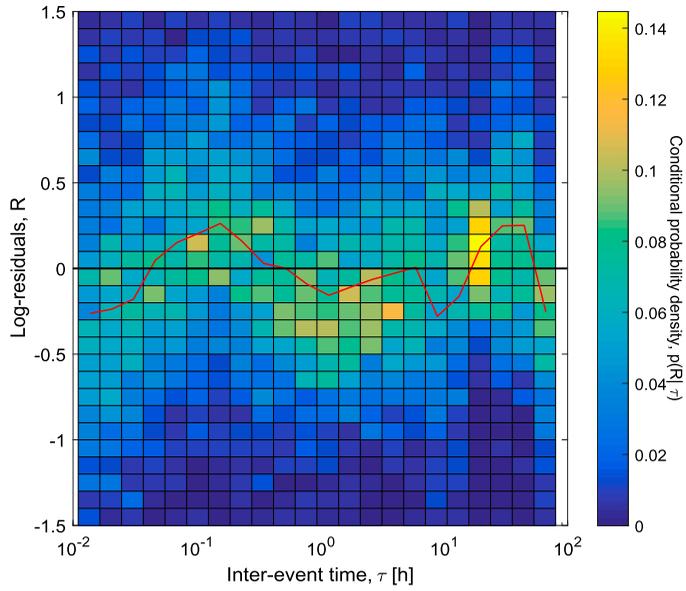


Figure 7. Conditional probability density $p(R|\tau)$ with log-residual $R = \ln(p_{\mathcal{M}}(\tau|\hat{\theta})/p(\tau))$, where $p_{\mathcal{M}}(\tau|\hat{\theta})$ is the inter-event time distribution of the best-fit M2CNPP model $\mathcal{M}(\hat{\theta})$ and $p(\tau)$ is the empirical inter-event time distribution obtained for all 394 users under consideration. The average log-residual at each inter-event time is represented by the dashed line.

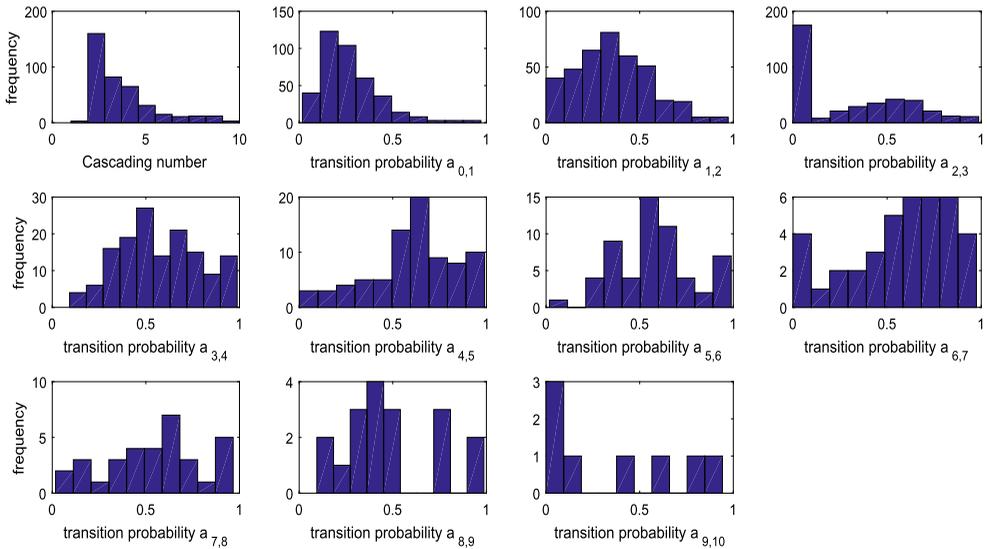


Figure 8. The cascading pattern distributions. The frequency of transition probabilities equaling zero is removed to clearly display the distribution.

shown in Figure 8, from which we can see that most people successively send less than 4 e-mails, and there are about half of the people who successively send 2 e-mails with the sending probability (transition probability $a_{1,2}$) less than 0.5.

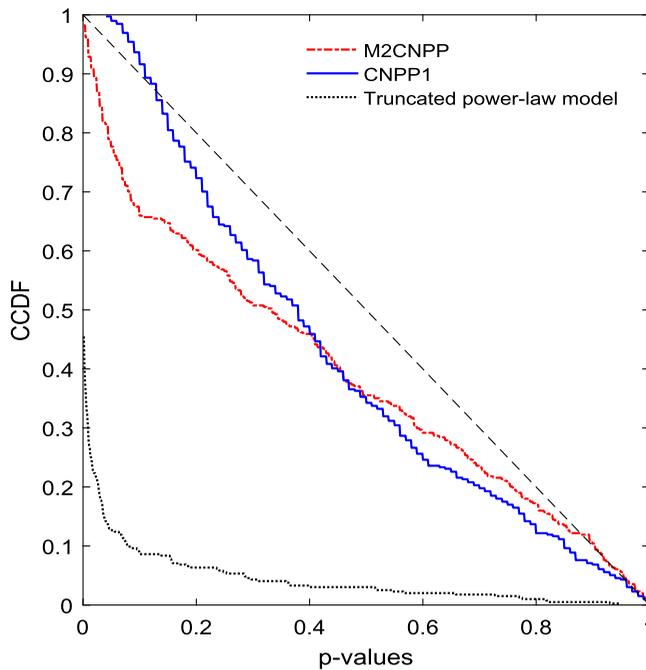


Figure 9. Model comparison: the complementary cumulative distribution function (CCDF) vs. p -values for all users.

5.3. Comparison

In this section, we compare the proposed M2CNPP model with the CNPP₁ model and a truncated power-law model using Algorithm 3. Figure 9 describes the complementary cumulative distribution function (CCDF) (i.e. one minus CDF) of p -values for all users. Note that if the data were actually generated by one of the models compared, we would expect to see a uniform distribution of p -values (dashed line). From Figure 9, it can be seen that our model is approximately consistent with the CNPP₁ model and the empirical data. At the 5% significance level, the proposed M2CNPP model fails to model the periodic patterns and bursts phenomenon for 83 users. In comparison, the CNPP₁ model failed to describe 1 user, whereas the truncated power-law model failed for 344 users [22]. We will provide an explanation of the discrepancy between the number of users failed to be modeled by the M2CNPP and the CNPP₁ models in the following paragraphs.

(1) Parameter estimation. One reason why there is a big discrepancy between the number of users that were failed to be modeled by our proposed model and the CNPP₁ model is because the parameter estimation method used in [22] tends to result in larger p -values than those of our estimation method. In fact, the parameters for our model are estimated by Maximum Likelihood Estimation (MLE), whereas the CNPP₁ model used Least Area Estimation (LAE), which minimizes the area test statistic S . Because the results of the Monte Carlo hypothesis testing procedure rely on S . The LAE tends to result in a small value of S with a larger p -value. Consequently, the LAE is overfitting, which can be directly verified by the fact that the CNPP₁ model failed to describe only 1 user (User 2241). To further investigate the difference of the estimated parameters by the M2CNPP and the CNPP₁

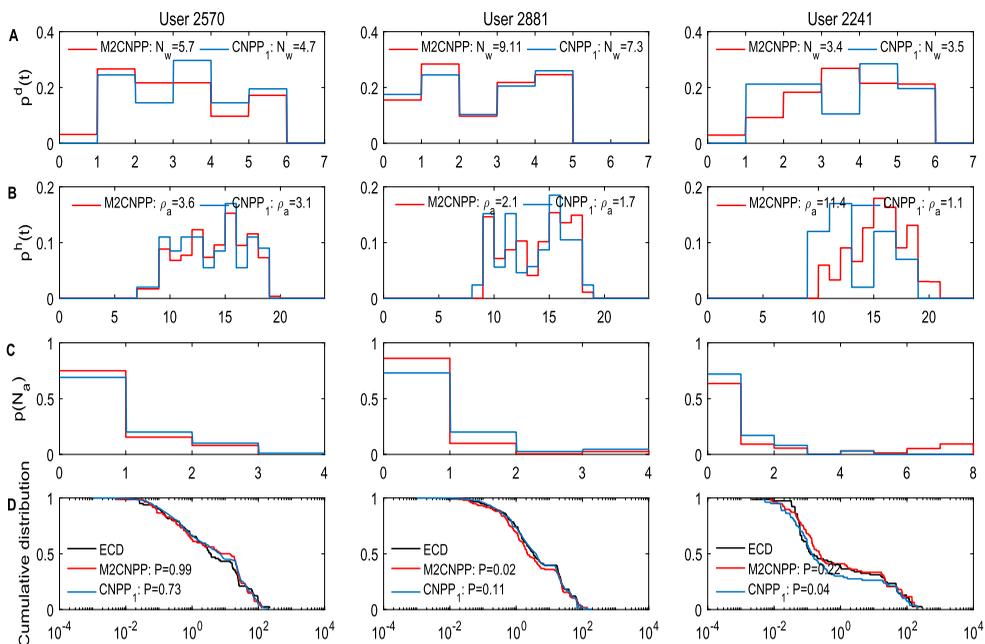


Figure 10. Comparison of parameter estimation (red line: the M2CNPP model; blue line: the CNPP₁ model; black line: the empirical cumulative distribution (ECD) of inter-event times). (A) Distribution of active sessions occurring within a week; (B) Distribution of active sessions occurring within a day; (C) Transition probability; (D) ECD and cumulative distributions of the M2CNPP and the CNPP₁ models.

models, we plot Figure 10. In Figure 10, the estimated parameters of three users (User 2570, User 2881 and User 2241) by LAE and MLE are displayed. The red line is the estimated parameters of the M2CNPP model, the blue line is the estimated parameters of the CNPP₁ model and the black line is the empirical cumulative distribution (ECD) of inter-event times. User 2570 was successfully modeled by both the M2CNPP and the CNPP₁ models. User 2881 was successfully modeled by only CNPP₁. User 2241 was successfully modeled by only M2CNPP. Although the ECDs of the M2CNPP and the CNPP₁ models have only small difference, the estimated parameters by the CNPP₁ and M2CNPP models vary a lot. Thus, the results of the Monte Carlo hypothesis testing strongly rely on the accuracy of the parameter estimation.

(2) Pattern diversity. The other reason is because the CNPP₁ model is more general than the M2CNPP model, i.e. the CNPP₁ model does not have too many constraints. Although we try to model the general pattern of individual e-mail communication, we also want to identify those different behaviors. Our M2CNPP model is able to capture the e-mail communication patterns, i.e. the periodic cascading patterns and the heavy tail property, of 79% individuals. This indicates the rest of 21% individuals have slightly different patterns. While the CNPP₁ model is too general to distinguish the slightly different e-mail communication patterns. Further the M2CNPP model spends much less computation time than the CNPP₁ model.

To improve our M2CNPP model, we will investigate more generalized forms of $\lambda(t)$. In this paper, we set $\lambda(t)$ to be a stepwise function, then the estimated parameters are

influenced by the division of the time window, and there are too many parameters to be estimated. To solve these problems, a general intensity function $\lambda(t)$ should be considered, such as kernel density functions, which were used in a self-exciting process proposed by [12].

6. Conclusions and future research

In this paper, we propose an extension of the CNPP₁ and the CNPP₂ models in order to describe more complex periodic patterns and cascades phenomenon of human communication as well as efficiently compute the parameter estimates. Our motivation is to leverage the rapidly increasing volume of e-mail communication to characterize individual behaviors. We apply the new model, M2CNPP to simulated data and a university e-mail data set. Our results clearly demonstrate that the proposed M2CNPP model can accurately describe both the periodic cascading patterns and the heavy tail property observed in e-mail communication. Additionally, we use BIC to reduce the possibility of overfitting, whereas in [22], the area statistic S in Equation 12 is directly minimized using numerical methods without considering the overfitting problem. Our model does not display any large systematic deviations from data (Figure 7). In addition to providing a better explanation of patterns in e-mail communication, this model can be easily adapted to describing other human actions such as individual mobility patterns [14]. Furthermore, our method uses a non-parametric method rather than specifying a deterministic distribution structure to mining the cascading behavior.

Future research will involve the generalization of $\lambda(t)$ including many different forms. As in [22], we assume our estimate $\hat{\lambda}(t)$ be a stepwise function, then the estimated parameters are influenced by the division of time window. To solve the problem, a general intensity function $\lambda(t)$ should be considered, such as kernel density functions.

Acknowledgments

The authors thank Drs G. Kossinets and J.P. Eckmann for providing the data used in this research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is supported by National Natural Science Foundation of China [grant no. 91746107]; the State Scholarship Fund of China Scholarship Council (CSC); and National Science and Engineering Research Council of Canada.

References

- [1] A.L. Barabási, *The origins of bursts and heavy tails in human dynamics*, Nature 435 (2005), pp. 207–211.
- [2] M.S. Bartlett, *The spectral analysis of point processes*, J. Roy. Stat. Soc. 25 (1963), pp. 264–296.
- [3] V. Belik, T. Geisel, and D. Brockmann, *Natural human mobility patterns and spatial spread of infectious diseases*, Phys. Rev. X. 1 (2011), pp. 3103–3106.

- [4] A. Berchtold, *The double chain Markov model*, Commun. Stat-Theor. M. 28 (1999), pp. 2569–2589.
- [5] K. Bhattacharya and K. Kaski, *Social Physics: Uncovering Human Behaviour from Communication*, arXiv preprint (2018) Available at arXiv:1804.04907.
- [6] D. Brockmann, L. Hufnagel, and T. Geisel, *The scaling laws of human travel*, Nature 439 (2006), pp. 462–465.
- [7] A. Clauset, C.R. Shalizi, and M.E.J. Newman, *Power-law distributions in empirical data*, Siam Review 51 (2009), pp. 661–703.
- [8] D.R. Cox, *Some statistical methods connected with series of events*, J. Roy. Stat. Soc. B. 17 (1955), pp. 129–164.
- [9] J.P. Eckmann, E. Moses, D. Sergi, and L.P. Kadanoff, *Entropy of dialogues creates coherent structures in e-mail traffic*, P. Natl. A. Sci. 101 (2004), pp. 14333–14337.
- [10] A.M. Edwards, R.A. Phillips, and N.W. Watkins, *Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer*, Nature 449 (2007), pp. 1044–1048.
- [11] W. Fischer, and K. Meier-Hellstern, *The Markov-modulated Poisson process (MMPP) cookbook*, Performance Evaluation 18 (1993), pp. 149–171.
- [12] E.W. Fox, M.B. Short, F.P. Schoenberg, K.D. Coronges, and A.L. Bertozzi, *Modeling e-mail networks and inferring leadership using self-exciting point processes*, J. Am. Stat. Assoc. 111 (2016), pp. 564–584.
- [13] M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabási, *Understanding individual human mobility patterns*, Nature 453 (2008), pp. 779–782.
- [14] X.P. Han, X.W. Wang, X.Y. Yan, and B.H. Wang, *Cascading walks model for human mobility patterns*, PLoS ONE. 10 (2013), pp. 1–19.
- [15] A.G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*, Biometrika 58 (1971), pp. 83–90.
- [16] M. Karsai, K. Kaski, A.L. Barabási, and J. Kertész, *Universal features of correlated bursty behaviour*, Sci. Rep. 2 (2012), pp. 539–539.
- [17] M. Kivela and M.A. Porter, *Estimating interevent time distributions from finite observation periods in communication networks*, Phys Rev E 92 (2015), pp. 1–10.
- [18] R. Langrock, D.L. Borchers, and H.J. Skaug, *Markov-modulated nonhomogeneous poisson processes for modeling detections in surveys of marine mammal abundance*, J. Am. Stat. Assoc. 108 (2013), pp. 840–851.
- [19] P.A. Lewis and G.S. Shedler, *Simulation of nonhomogeneous Poisson processes by thinning*, Nav. Res. Log. (NRL) 26 (1979), pp. 403–413.
- [20] R.D. Malmgren, J.M. Hofman, L.A.N. Amaral, and D.J. Watts, *Characterizing individual communication patterns*, in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, Association for Computing Machinery, 2009, pp. 607–615.
- [21] R.D. Malmgren, D.B. Stouffer, A.S. Campanharo, and L.A.N. Amaral, *On universality in human correspondence activity*, Science 325 (2009), pp. 1696–1700.
- [22] R.D. Malmgren, D.B. Stouffer, A.E. Motter, and L.A. Amaral, *A Poissonian explanation for heavy tails in e-mail communication*, P. Nat. A. Sci. 105 (2008), pp. 18153–18158.
- [23] G.J. McLachlan and T. Krishnan, *The EM algorithm and extensions*, Vol. 382, John Wiley and Sons, New Jersey, 2007.
- [24] X.L. Meng and D.B. Rubin, *Recent extensions of the EM algorithm (with discussion)*, Bayesian stat. 4 (1992), pp. 307–320.
- [25] G. Mohler, *Modeling and estimation of multi-source clustering in crime and security data*, Ann. Appl. Stat. 7 (2013), pp. 1525–1539.
- [26] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, and G.E. Tita, *Self-exciting point process modeling of crime*, J. Am. Stat. Assoc. 106 (2011), pp. 100–108.
- [27] M.E.J. Newman, *Power Laws, Pareto Distributions and Zipf’s Law*, Contemp. Phys. 46 (2006), pp. 323–351.
- [28] J. Neyman and E.L. Scott, *Statistical Approach to Problems of Cosmology*, J. Roy. Stat. Soc. B. 20 (1958), pp. 1–43.

- [29] J. Neyman, E.L. Scott, and C.D. Shane, *Statistics of Images of Galaxies with Particular Reference to Clustering*, in Berkeley Symposium on Mathematical Statistics and Probability. 1958, pp. 75–106
- [30] R. Perline, *Strong, weak and false inverse power laws*, Stat. Sci. 20 (2005), pp. 68–88.
- [31] P. Reynaud-Bouret and S. Schbath, *Adaptive estimation for hawkes processes; application to genome analysis*, Ann. Stat. 38 (2010), pp. 2781–2822.
- [32] E.R. Rodrigues and J.A. Achcar, *Applications of discrete-time Markov chains and Poisson processes to air pollution modeling and studies*, Springer, New York, 2013.
- [33] T. Rydén, *An EM algorithm for estimation in Markov-modulated Poisson processes*, Comput. Stat. Data Anal. 21 (1996), pp. 431–447.
- [34] P. Salvador, R. Valadas, and A. Pacheco, *Multiscale fitting procedure using Markov modulated Poisson processes*, Telecommun. Syst. 23 (2003), pp. 123–148.
- [35] S.L. Scott, *Bayesian Analysis of a Two-State Markov Modulated Poisson Process*, J. Comput. Graph. Stat. 8 (1999), pp. 662–670.
- [36] S.L. Scott and P. Smyth, *The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data*, Bayesian Stat. 7 (2003), pp. 671–680.
- [37] D.B. Stouffer, R.D. Malmgren, and L.A.N. Amaral, *Log-normal statistics in e-mail communication patterns*, (2006) Available at arXiv:physics/060527
- [38] S.F. Wang, X. Feng, Y. Wu, and J.H. Xiao, *Double dynamic scaling in human communication dynamics*, Phys. A Stat. Mech. Appl. 473 (2017), pp. 313–318.
- [39] W.J. Wang, L. Pan, N. Yuan, S. Zhang, and D. Liu, *A comparative analysis of intra-city human mobility by taxi*, Phys. A Stat. Mech. Appl. 420 (2015), pp. 134–147.

Appendices

Appendix 1. Proof of Theorem 4.1

For easy reference, a set of properties about a point process are described below.

Let $\{N^p(t)\}$ be the primary non-homogeneous Poisson process, and let $\{N^c(t)\}$ be the cascading homogeneous Poisson process. Define $N^p_{t,t+h} = N^p(t+h) - N^p(t)$, $\Lambda(t) = \int_0^t \lambda(x) dx$ and $\Lambda_{t,t+h} = \Lambda(t+h) - \Lambda(t)$, one gets

$$P\{N^p_{t,t+h} = n\} = \frac{\Lambda^n_{t,t+h}}{n!} \exp\{-\Lambda_{t,t+h}\}.$$

For the non-homogeneous Poisson counting process, the probability density function of an event occurring at t_k and no event occurring in $(t_{k-1}, t_k]$ is

$$f(t_k) \triangleq \lim_{\Delta t \rightarrow 0} \frac{P\{N^p_{t_k-\Delta t, t_k} = 1, N^p_{t_{k-1}, t_k-\Delta t} = 0\}}{\Delta t} = \lambda(t_k) \exp\{-\Lambda_{t_{k-1}, t_k}\}, \quad (A1)$$

$t_{k-1} \leq t_k$, otherwise, $f(t_k) = 0$. For the homogeneous Poisson counting process, the probability density function of an event occurring at t_k and no event occurring in $(t_{k-1}, t_k]$ is

$$g(t_k) \triangleq \lim_{\Delta t \rightarrow 0} \frac{P\{N^c_{t_k-\Delta t, t_k} = 1, N^c_{t_{k-1}, t_k-\Delta t} = 0\}}{\Delta t} = \lambda_a \exp\{-\lambda_a(t_k - t_{k-1})\} \quad (A2)$$

when $t_{k-1} \leq t_k$, otherwise, $g(t_k) = 0$.

Since the M2CNPP model has both discrete and continuous random variables involved, the likelihood function can be written as in Equation A3,

$$\begin{aligned} \mathcal{L} &= P\{X_j = \mathbf{0}, j \in J = \{j_1, \dots, j_m\} \subset \{1, \dots, n-1\}, \mathbf{t}_i^n\} \\ &\triangleq \lim_{\Delta t \rightarrow 0} \frac{P\{X_j = \mathbf{0}, j \in J, t_i - \Delta t \leq \mathbf{t}_i \leq t_i, 1 \leq i \leq n\}}{\Delta t^n}. \end{aligned} \quad (A3)$$

Then one can obtain Equation 7.

Proof: The conditional expectation of log-likelihood function with the current state variables $\{X_1, \dots, X_{n-1}\}$ can be calculated as follows:

$$\begin{aligned}
Q(\Phi, A; \Phi^{(l)}, A^{(l)}) &= \sum_{i=0}^1 \log a_{0,i} P(i_1 = i | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{k=1}^{n-2} \sum_{i=0}^{\mathbf{r}} \sum_{\substack{j \in \{0, i+1\} \\ j \leq \mathbf{r}}} \log a_{i,j} P(X_k = i, X_{k+1} = j | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \log f(t_1) + \sum_{k=1}^{n-1} \sum_{i=0}^{\mathbf{r}} \log b_i(t_{k+1}) P(X_k = i | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&= \sum_{i=0}^1 \log a_{0,i} \sum_{k=0}^{n-2} P(X_k = \mathbf{0}, X_{k+1} = i | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{k=1}^{n-2} \sum_{i=1}^{\mathbf{r}-1} \sum_{j \in \{0, i+1\}} \log a_{i,j} P(X_k = i, X_{k+1} = j | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{k=0}^{n-1} \log f(t_{k+1}) P(X_k = \mathbf{0} | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{k=1}^{n-1} \sum_{i=1}^{\mathbf{r}} \log g(t_{k+1}) P(X_k = i | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&= \sum_{i=0}^1 \log a_{0,i} \sum_{k=0}^{n-2} P(X_k = \mathbf{0}, X_{k+1} = i | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{k=1}^{n-2} \sum_{i=1}^{\mathbf{r}-1} \sum_{j \in \{0, i+1\}} \log a_{i,j} P(X_k = i, X_{k+1} = j | t_1^n; \Phi^{(l)}, A^{(l)}) \\
&\quad + \sum_{i=1}^{24} \sum_{j=1}^7 \left[\sum_{k \in K_1} P(X_k = \mathbf{0} | t_1^n; \Phi^{(l)}, A^{(l)}) \log \lambda_{i,j} \right. \\
&\quad \left. - \lambda_{i,j} \left(\sum_{k=0}^{n-1} (t_{k+1} - t_k)_{i,j} P(X_k = \mathbf{0} | t_1^n; \Phi^{(l)}, A^{(l)}) \right) \right] \\
&\quad + \left[\sum_{k=1}^{n-1} \sum_{i=1}^{\mathbf{r}} P(X_k = i | t_1^n; \Phi^{(l)}, A^{(l)}) \right] \log \lambda_a \\
&\quad - \lambda_a \sum_{k=1}^{n-1} \left[(t_{k+1} - t_k) \sum_{i=1}^{\mathbf{r}} P(X_k = i | t_1^n; \Phi^{(l)}, A^{(l)}) \right]
\end{aligned}$$

where $P(i_0 = 0) = P(X_0 = \mathbf{0}) = 1$; $\log a_{\mathbf{r},0} = 0$; $K_1 = \{k : [t_{k+1}/24]_7 = j, [t_{k+1}]_{24} = i, 0 \leq k \leq n-1\}$, $[\cdot]$ is the function rounding up to the nearest integer; $(t_{k+1} - t_k)_{i,j}$ is the time interval $(t_k, t_{k+1}]$ corresponding to the intensity rate $\lambda_{i,j}$; $a|_b = a \bmod b$ represents arithmetic modulus b .

Under the constraint $a_{k,0} + a_{k,k+1} = 1$, $\mathbf{0} \leq k \leq \mathbf{r} - 1$ and using Lagrangian multiplier method, one gets the estimated parameters $a_{k,0}, a_{k,k+1}$, $\mathbf{0} \leq k \leq \mathbf{r} - 1$ by letting the corresponding partial

derivatives equal zero. That is,

$$\widehat{a}_{\mathbf{0},i}^{(l+1)} = \frac{\sum_{k=0}^{n-2} P(X_k = \mathbf{0}, X_{k+1} = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=0}^{n-2} P(X_k = \mathbf{0} | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}, \quad i = \mathbf{0}, \mathbf{1},$$

$$\widehat{a}_{i,j}^{(l+1)} = \frac{\sum_{k=1}^{n-2} P(X_k = i, X_{k+1} = j | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=1}^{n-2} P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}, \quad j \in \{\mathbf{0}, i + 1\}, \quad i = \mathbf{1}, \dots, \mathbf{r} - 1,$$

The parameter λ_a can also be estimated by letting the corresponding partial derivative equal zero, which is

$$\widehat{\lambda}_a^{(l+1)} = \frac{\sum_{k=1}^{n-1} \sum_{i=1}^r P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})}{\sum_{k=1}^{n-1} [(t_{k+1} - t_k) \sum_{i=1}^r P(X_k = i | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})]}.$$

For estimating $N_w, \lambda_i^h, \lambda_j^d, i = 1, \dots, 24, j = 1, \dots, 7$, we first fix $N_w = \widehat{N}_w^{(l)}$ and $\lambda_i^h = \widehat{\lambda}_i^{h(l)}$. Under the constraint $\sum_{j=1}^7 \lambda_j^d = 1$ and using Lagrangian multiplier method, one gets the estimated parameters $\lambda_j^d, j = 1, \dots, 7$ by letting the corresponding partial derivatives equal zero. That is,

$$\widehat{\lambda}_j^{d(l+1)} = \frac{\widetilde{\lambda}_j^{d(l+1)}}{\sum_{j=1}^7 \widetilde{\lambda}_j^{d(l+1)}}, \quad j = 1, \dots, 7,$$

where

$$\widetilde{\lambda}_j^{d(l+1)} = \frac{\sum_{i=1}^{24} d_{ij}^{(l)}}{\sum_{i=1}^{24} \widehat{\lambda}_i^{h(l)} c_{ij}^{(l)}}, \quad j = 1, \dots, 7,$$

$c_{ij} = \sum_k (\overline{t_{k+1} - t_k})_{ij} P(X_k = 0 | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$ and $d_{ij} = \sum_{k \in K_1} P(X_k = 0 | \mathbf{t}_1^n; \Phi^{(l)}, A^{(l)})$. We then fix $N_w = \widehat{N}_w^{(l)}$ and $\lambda_j^h = \widehat{\lambda}_j^{h(l+1)}$. Under the constraint $\sum_{i=1}^{24} \lambda_i^d = 1$ and using Lagrangian multiplier method, one gets the estimated parameters $\lambda_i^d, i = 1, \dots, 24$ by letting the corresponding partial derivatives equal zero. That is,

$$\widehat{\lambda}_i^{h(l+1)} = \frac{\widetilde{\lambda}_i^{h(l+1)}}{\sum_{i=1}^{24} \widetilde{\lambda}_i^{h(l+1)}}, \quad i = 1, \dots, 24,$$

where

$$\widetilde{\lambda}_i^{h(l+1)} = \frac{\sum_{j=1}^7 d_{ij}^{(l)}}{\sum_{j=1}^7 \widehat{\lambda}_j^{d(l+1)} c_{ij}^{(l)}}, \quad i = 1, \dots, 24.$$

Finally, let $\lambda_j^h = \widehat{\lambda}_j^{h(l+1)}$ and $\lambda_i^d = \widehat{\lambda}_i^{d(l+1)}$, we have the estimated parameter N_w , which is

$$\widehat{N}_w^{(l+1)} = \left(\sum_{j=1}^7 \widetilde{\lambda}_j^{d(l+1)} \right) \left(\sum_{i=1}^{24} \widetilde{\lambda}_i^{h(l+1)} \right).$$

Let $\Phi_1^{(l+1)} = \left\{ \widehat{N}_w^{(l)}, \widehat{\lambda}_a^{(l+1)}, \widehat{\lambda}_i^{(l)}, \widehat{\lambda}_j^{(l+1)}, i = 1, \dots, 24, j = 1, \dots, 7 \right\}$ and $\Phi_2^{(l+1)} = \left\{ \widehat{N}_w^{(l)}, \widehat{\lambda}_a^{(l+1)}, \widehat{\lambda}_i^{(l+1)}, \widehat{\lambda}_j^{(l+1)}, i = 1, \dots, 24, j = 1, \dots, 7 \right\}$, one has

$$\begin{aligned} Q(\Phi^{(l)}, A^{(l)}; \Phi^{(l)}, A^{(l)}) &\leq Q(\Phi_1^{(l+1)}, A^{(l+1)}; \Phi^{(l)}, A^{(l)}) \\ &\leq Q(\Phi_2^{(l+1)}, A^{(l+1)}; \Phi^{(l)}, A^{(l)}) \\ &\leq Q(\Phi^{(l+1)}, A^{(l+1)}; \Phi^{(l)}, A^{(l)}). \end{aligned}$$

This completes the proof. ■

Remark A.1: In application, the basic time unit is an hour. The beginning of the day, regardless of what day of the week it is, of the first time in the data is determined as $t_0 = 0$. $\lambda_{1,1}$ represents the intensity rate parameter of the first hour in the day that contains the first time in the data. Then the time can be adjusted to the standard one after the parameters are estimated. Note that the probability distribution of the first hidden state needs not to be $\pi = \{a_{0,0}, a_{0,1}, 0, \dots, 0\}$ in practice, however, with this assumption, it is easier to derive the parameter estimates. Therefore in this paper, we assume that the initial status is $\mathbf{0}$.

Remark A.2: For efficiently calculating the above estimates, one can refer to [4].

Appendix 2. The relationship between CNPP₁ and the proposed M2CNPP

In [22], the proposed cascading structure of CNPP₁ requires intensive computation. To handle this problem, we propose our multi-state Markov CNPP. Although the assumption of the Markov property for cascading process is stronger than the assumed cascading structure in the CNPP₁ model, we can show that the maximum likelihood estimates of the intensity parameters of the M2CNPP model are the same as those estimated in the CNPP₁ model with the full likelihood function in the following.

In fact, since $\sum_{k=0}^r P(N_a = k) = 1$, one gets the unique feasible solution that matches the transition probability A to \mathbb{P} , which is the distribution law of N_a in the CNPP₁ model by solving the system of equations (Equation (A4)).

$$\begin{aligned} P(N_a = 0) &= a_{0,0}, \\ P(N_a = k) &= \prod_{i=1}^k a_{i-1,i} a_{k,0}, \quad k = 1, 2, \dots, r. \end{aligned} \tag{A4}$$

The transition probabilities are expressed in Equation (A5).

$$\begin{aligned} a_{0,0} &= P(N_a = 0), \\ a_{0,1} &= 1 - P(N_a = 0), \\ a_{k,0} &= \frac{P(N_a = k)}{1 - \sum_{i=0}^{k-1} P(N_a = i)}, \quad k = 1, 2, \dots, r, \\ a_{k,k+1} &= \frac{1 - \sum_{i=0}^k P(N_a = i)}{1 - \sum_{i=0}^{k-1} P(N_a = i)}, \quad k = 1, 2, \dots, r-1. \end{aligned} \tag{A5}$$

Suppose A^* , Φ^* and I^* are the unique optimal solutions of the likelihood functions \mathcal{L} . Then the \mathbb{P} that satisfies Equations (A4) with Φ^* and I^* are the unique optimal solutions of the likelihood functions \mathcal{L}_1 . Otherwise, by using Equations (A5), one can always obtain another set of solutions, which is different from (A^*, Φ^*, I^*) and can achieve a higher likelihood value.