9-2024

Unraveling the dynamics of stable and curious audiences in web systems

Rodrigo ALVES

Antoine LEDENT
Singapore Management University, aledent@smu.edu.sg

Renato ASSUNÇÃO

Pedro VAZ-DE-MELO

Marius KLOFT

Citation

ALVES, Rodrigo; LEDENT, Antoine; ASSUNÇÃO, Renato; VAZ-DE-MELO, Pedro; and KLOFT, Marius. Unraveling the dynamics of stable and curious audiences in web systems. (2024). *WWW '24: Proceedings of the ACM Web Conference 2024, Singapore, May 13-17*. 2464-2475.
Available at: https://ink.library.smu.edu.sg/sis_research/9304

Unraveling the Dynamics of Stable and Curious Audiences in Web Systems

Rodrigo Alves
rodrigo.alves@fit.cvut.cz
Czech Technical University in Prague
Prague, Czech Republic

Antoine Ledent*
aledent@smu.edu.sg
Singapore Management University
Singapore, Singapore

Renato Assunção
assuncao@dcc.ufmg.br
Federal University of Minas Gerais
Belo Horizonte, Brazil

Pedro Vaz-De-Melo
olmo@dcc.ufmg.br
Federal University of Minas Gerais
Belo Horizonte, Brazil

Marius Kloft
kloft@cs.uni-kl.de
RPTU Kaiserslautern-Landau
Kaiserslautern, Germany

ABSTRACT

We propose the Burst-Induced Poisson Process (BPoP), a model designed to analyze time series data such as feeds or search queries. BPoP can distinguish between the slowly-varying regular activity of a stable audience and the bursty activity of a curious audience, often seen in viral threads. Our model consists of two hidden, interacting processes: a self-feeding process (SFP) that generates bursty behavior related to viral threads, and a non-homogeneous Poisson process (NHPP) with step function intensity that is influenced by the bursts from the SFP. The NHPP models the normal background behavior, driven solely by the overall popularity of the topic among the stable audience. Through extensive empirical work, we have demonstrated that our model fits and characterizes a large number of real datasets more effectively than state-of-the-art models. Most importantly, BPoP can quantify the stable audience of media channels over time, serving as a valuable indicator of their popularity.

CCS CONCEPTS

• **Information systems → Web mining**; **Web applications**; • **Computing methodologies → Machine learning approaches**.

KEYWORDS

Temporal Dynamics of Web Systems; Time-series; EM-algorithm

ACM Reference Format:
Rodrigo Alves, Antoine Ledent, Renato Assunção, Pedro Vaz-De-Melo, and Marius Kloft. 2024. Unraveling the Dynamics of Stable and Curious Audiences in Web Systems. In *Proceedings of the ACM Web Conference 2024 (WWW '24), May 13–17, 2024, Singapore, Singapore*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3589334.3645473

*The first two authors, Rodrigo Alves and Antoine Ledent, contributed equally to this research.

WWW '24, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0171-9/24/05…$15.00
https://doi.org/10.1145/3589334.3645473

1 INTRODUCTION

Why do content creators on YouTube frequently request their subscribers to enable notifications? This practice stems from their awareness that not all subscribers are regular viewers of their channels [30], despite the argument that the primary determinant of sustained interest lies in the number of subscriptions a channel garners [14, 48]. In fact, the popularity dynamics of online items can be explained by several endogenous and exogenous factors [13, 54], which include the quality of the content [25, 50], their metadata [18], their age [5], the recommendation algorithm and its rank on keyword-based queries [59], promotions [38], and social media effects [7, 50].

Accurately predicting the enduring appeal of online content remains a notably challenging task due to the distinct patterns exhibited in the popularity dynamics of online items [12, 26, 46]. These dynamics often involve one or more peaks of *popularity bursts* that intermingle with the *regular and stable audience* of the content. Existing literature mainly focuses on supervised and feature-based approaches for predicting long-term popularity [17, 25, 26, 38, 51], while some models overlook the bursty nature of popularity dynamics [1, 27, 41]. Moreover, these approaches aim to forecast an item's overall popularity, taking into account the influence of exogenous factors such as media exposure and virality on social media.

Unlike previous studies, this paper introduces an approach that (1) distinguishes between *popularity bursts* and the *consistent and stable audience* of content, and (2) investigates the underlying dynamics of these audiences. We specifically focus on differentiating two audience types: the curious, attracted by external and viral events such as gossip [2, 8, 38], and the stable audience, representing stable viewership. Empirical evidence reveals that content such as keyword-discovered videos [5], popular TV episodes, and music videos [50] maintains steady popularity over time, dominated by the stable audience. However, news, sports, and movie content often undergo rapid popularity surges followed by quick declines, mostly due to temporally limited events (e.g. breaking news). In these cases, the curious tends to prevail over the stable audience.

The primary challenge in distinguishing between these audience types arises from the lack of individual labels that distinguish stable and curious viewers. Instead, we typically only have the total number of viewers for an observed random series of events (RSE), which is a combination of both hidden processes associated with the two

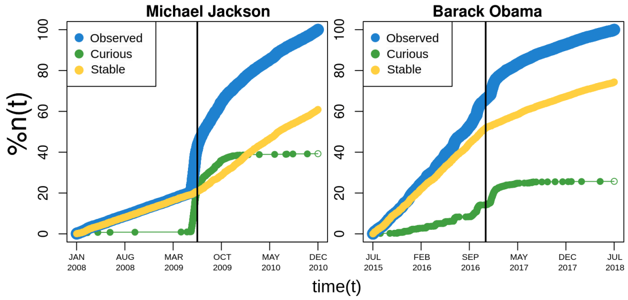

Figure 1: Two example time series that motivate the BPoP model. Left: Michael Jackson (Jan/01/2008 to Dec/31/2010). Right: Barack Obama (Jul 20 2015 Jul 19 2018). The data is taken from Google Trend (country: USA; search engine: Youtube).

types of audiences. Disentangling these two audience types poses difficulties, particularly because during viral events, curious users tend to dominate the channel's activity, leading to a burst in the overall RSE [6, 47, 53]. Such bursts can become so prominent that they fully obscure the presence of the stable audience during these events. To accurately identify and quantify the stable audience, it becomes essential to pinpoint the timestamps associated with these bursts. However, this task is challenging as the activity from the stable audience remains unlabelled and gets mixed with the bursts.

Another significant challenge arises from the potential for the stable audience to modify its typical behavior in response to bursts or external events [16, 23, 33, 39, 49, 55, 56]. For instance, the unexpected death of Michael Jackson in June 2009 triggered a surge in media and web activity, leading to increased music sales, video views, and reactions in posts. During this period, both existing and new fans engaged with Jackson's work, transforming him into an enduring musical icon. This behavior is illustrated in the left-hand side of Figure 1, where the blue line represents the cumulative web activity associated with Michael Jackson's RSE. Initially, it displayed a relatively constant growth rate until his death (vertical black line), followed by a sudden spike in interest. Over time, the blue line returned to a consistent growth rate. We will illustrate how to distinguish between two types of web activities during such events: (1) regular stable audience activity (yellow line) and (2) activity driven by unexpected events generated by the curious audience (green line). Notably, the yellow curve changed its slope after Jackson's unexpected death, marking a significant transition event that not only led to a short-term burst of activity but also *consistently altered* the stable audience [39, 56]. The revival of his songs, tribute notes, and the younger generation's discovery of Jackson's work contributed to a sustained increase in interaction. Conversely, the end of Barack Obama's presidential term (right-hand side of Figure 1) resulted in reduced political activity and a decline in mentions.

Point processes form a statistical framework to learn and infer about RSEs [9, 40]. There are two contrasting approaches in this domain. One focuses on self-exciting point processes, which model correlations between past and future events [8, 11, 44]. On the other hand, the homogeneous Poisson process and its variants have also been deemed suitable [8, 20, 21, 28]. This divergence has led to extensive research examining the diversity of human actions. For example, studies have found that Twitter (X) hashtag activity can be continuous, periodic, or concentrated around a single peak [22]. Similarly, research on YouTube videos has shown that the current tweeting rate and tweet volume since a video's upload are crucial parameters for identifying its virality or popularity [43]. Moreover, the popularity of YouTube videos can undergo multiple phases of growth and decline, likely influenced by various background random processes superimposed on bursty behavior [55].

Therefore, in theory, point processes could be used to solve the problem of estimating the stable audience of online items, but existing models are not appropriate for this particular setting: they focus on different aspects of RSE characterization, and do not provide methods to identify and measure burst-induced **changes** in *background* popularity. While Poisson processes (PPs) [21, 28] can easily estimate the stable audience when **all** incoming events arrive at a fixed and predictable rate, they fail to mimic the bursts of events seen in real data. On the other hand, self-exciting processes, such as Hawkes and Wold processes, are able to capture the correlations between consecutive events that generate bursts of activity, but existing approaches do not model the *time-varying* nature of the stable audience [1, 2, 24, 29, 39, 45, 52].

To address these concerns, we propose the Burst-induced Poisson Process (**BPoP**) model, which is able to flexibly incorporate dependencies between the two hidden and underlying point processes involving the stable audience and the curious audience. We show that **BPoP** mimics the bursts of events seen in real data and is also able to efficiently capture the time-varying background rates that realistically represent the stable audience. Our main contributions are: **(a) A New Model**, namely **BPoP**, which is able to disentangle the slowly-varying regular activity of the stable audience from the curious activity occurring in bursts. This model does not depend on hard-to-get external information but uses only random series of events (RSEs) (Section 2); **(b) An EM algorithm** to cope with our intensity function's complex dependence on the history of the process (Section 3); **(c) Novel findings** describing and quantifying the stable audience for eleven real world data containing more than a hundred thousand RSEs (Section 4). **(d)** Extensive empirical investigations have demonstrated that **BPoP** consistently outperforms alternative models in fitting both real and synthetic data (Section 5).

2 THE BPOP MODEL

Formal construction: filtrations are formal constructions in probability theory required for the formal description of time-dependent processes. Consider a general continuous-time Markov process adapted to the filtration $(\mathcal{H}_t)_{t\in\mathbb{R}^+}$: $\mathcal{H}_t$ represents the information that is realised at time t. Let $N(a,b)$ be the random number of events in $(a,b]$. The conditional intensity rate function characterizes the distribution and is given by $\lambda(t|\mathcal{H}_t) = \lim_{h\to 0}\mathbb{E}\left(N(t,t+h)|\mathcal{H}_t\right)/h$.

Figure 2 shows the main idea of **BPoP**. We observe the point process timestamps $0 < t_1 < t_2 < \ldots$ of events up to a time t (depicted as blue dots in the fourth row). We assume that these events are a mixture of events coming from the stable audience and the curious audience, which are two dependent point processes, represented as yellow and green dots in the first and third rows, respectively. On the one hand, we model the curious audience

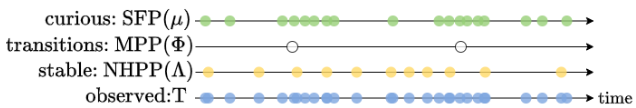

Figure 2: The Burst-induced Poisson Process (BPoP) model. The curious (SFP) and stable audience (NHPP) labels, as well as the transitions (events of MPP), are not observed.

generating the occasional bursts as a *Self-Feeding process* (SFP) [1, 44, 45]. SFPs are simple self-exciting processes, and their intensity function is given by $\lambda_s(t|\mathcal{H}_t) = 1/(\mu/e + \Delta t_i)$ (where $\Delta t_i = t_i - t_{i-1}$, $t_i = \max_k\{t_k : t_k \leq t\}$ and $\mu > 0$). Typical SFP instances alternate between bursts and calm periods, making them an ideal model for capturing bursts. On the other hand, the events associated with the stable audience are modelled by a second process, the classical non-homogeneous Poisson process (NHPP), shown in the third row. A third underlying meta Poisson process (MPP) controls the times when the stable audience (or NHPP) transitions occur. These transitions are shown as white dots in the second row in Figure 2. The intensity of the meta Poisson process generating the transitions is proportional to the intensity of the SFP process, which acts as a soft proxy for "whether a burst is currently occurring".

Remark: The main difficulty with this model is that we only observe the blue dots in the fourth row. The labels associated with each event (the green and yellow colors) and the transitions (the white dots) are not directly observed.

Therefore, our **BPoP** model involves the combination of an SFP, representing the curious, and a NHPP, representing the stable audience, which interact with each other. At time t, the history of the process is composed of the observed event timestamps $\{t_1, t_2, \ldots\} < t$, unobserved labels $\{z_1, z_2, \ldots\}$ as well as unobserved MPP events $\{\varphi_1, \varphi_2, \ldots\}$, which represent the transitions. We use the convention that $z_i = 0$ if $t_i \in$ NHPP and $z_i = 1$ if $t_i \in$ SFP. Thus, **BPoP** is governed by the following three intensity functions: (1) the SFP intensity $\lambda_s(t) = 1/[(g(t) - g(g(t))) + \mu/e]$ where $g(u) = [\max(t_i : z_i = 1 \quad \wedge \quad t_i < u)]_+$ denotes the last SFP event before t, with the convention that $g(t) = 0$ if $\nexists i : t_i \leq t \wedge z_i = 1$ and $\mu > 0$ is the SFP parameter; (2) $\lambda_\Phi(t) = c\lambda_s(t)$, where $c \in [0, 1]$ is a parameter that controls the NHPP transition sensitivity; and (3) $\lambda_p(t) = \lambda_{m_t}$, where $\Lambda = \{\lambda_0, \lambda_1, \lambda_2, \cdots\}$ is an infinite set of positive numbers (parameters) and, $m_t = \sum_{j=1}^{n_\phi} 1_{\varphi_j < t}$. Similarly we can define $\lambda_s^+(t) = \lim_{\delta\to 0} \lambda_s(t + \delta)$ (resp. $\lambda_\phi^+(t)$ and $\lambda_p^+(t)$) as the intensity of the SFP (resp. MPP and NHPP) immediately after t. Thus, to generate the next time stamp, we first generate three exponential variables E_s, E_ϕ and E_p with intensities $\lambda_s^+(t)$, $\lambda_\phi^+(t)$ and $\lambda_p^+(t)$ respectively. Then, the next event will take place at $t + E$, where E=min(E_s, E_ϕ, E_p), and it will belong to the SFP (resp. NPHH, MPP) component if $E = E_s$ (resp. E_ϕ, E_p). Likewise, we continue generating the rest of the process from time $t + E$.

Considering the described generative model we aim to infer the parameters of **BPoP**. When performing inference, c is set as a hyperparameter[1] and the parameters Λ, μ are determined via maximum likelihood. To optimize the likelihood, we will use the EM algorithm, relying on Gibbs sampling in the E-step. However, the EM algorithm in the case of point processes requires great care, since the events are not independent data and the usual derivations are not appropriate.

3 FITTING USING EM

Our optimization approach to fitting the model relies on the EM algorithm [35, 36]. The EM algorithm represents a broad class of alternating optimization methods used to estimate the maximum likelihood estimate of parameters θ in statistical models involving unobserved latent variables Z. The strategy consists of two steps: (1) the E-step estimates the conditional distribution of the latent variables (given the observations) based on the current estimate of the parameters; (2) the M-step computes the maximum likelihood estimate of the parameters based on the current estimate of the latent distribution. These two steps are performed alternately until convergence. In our specific case, the latent variables are the labels z_i (SFP/curious versus Poisson/stable) of the observed timestamps and the transitions m_i, while the parameters are the λ's and the μ. In this context, since the conditional distribution of the labels (given the observations and the parameters) is not analytically tractable, we estimate it using Gibbs sampling [15].

In this section, we compute the likelihood for our model, the marginal conditional probabilities required for Gibbs simulation, and present the overall details of our approach. To do that, we must first introduce some notation. Let $T = \{t_1, t_2, \cdots, t_n\}$ be the observed event timestamps from the mixture of the SFP (curious) and the NHPP (stable audience). Also, let $N(t) = \sum_{i=1}^{n} 1_{t_i \leq t}$ be a function that computes the cumulative number of events up to time t. The number of transitions that occurred before t_i is given by $M := \{m_1, m_2, \cdots, m_n\}$, where $m_i = \sum_{j=1}^{n_\phi} 1_{\varphi_j < t_i}$, and the set of NHPP transitions between t_{i-1} and t_{i+1} is given by $\Phi_i := \{\varphi_j | t_{i-1} < \varphi_j < t_{i+1}\}$. Finally, we set $\theta = Z \cup \Phi \cup M$ (the latent variables), $\theta_{-i} = \theta \setminus \theta_i$ where $\theta_i := (\{z_i, m_i\} \cup \Phi_i)$.

E-Step: Here we will explain how to use *Gibbs sampling* to draw a set of latent variables Z, M and Φ (collectively referred to as θ) from the conditional distribution given a fixed set of parameters μ and Λ. *Gibbs sampling* is a general statistical method which allows one to draw samples from complicated high-dimensional distributions. To draw a sample from a distribution p on $\mathbb{R}^d$, we start with an arbitrary vector $x \in \mathbb{R}^d$, and proceed to iteratively replace each coordinate x_i ($i \leq d$) by a sample from the *conditional distribution* of x_i given the current values of the other coordinates x_j ($j \neq i$). In many situations, the conditional distribution is easier to compute than the multivariate probability density function (PDF) due to the intractability of the calculation of the multivariate normalization constant. It is known that under mild conditions, *after convergence*, Gibbs sampling leads to a sample from the original multivariate distribution p [15]. In our model, the distribution to estimate is the latent *joint* distribution of the labels z_i and transitions m_i. The distribution is proportional to the corresponding likelihood, but the normalization constant is intractable. On the other hand, the

[1] Indeed, one cannot simply optimize over it since larger c allows for far more transitions and makes the model prone to overfitting

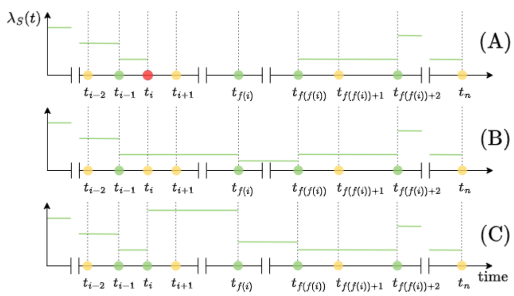

Figure 3: A: Start: behaviour of SFP intensity function given that label of t_i is unknown (red dot). Green dots refer to SFP events while yellow ones are NHPP events; B: Behaviour of SFP intensity function given that $t_i \in$ NHPP; C: Behaviour of SFP intensity function given that $t_i \in$ SFP.

conditional marginal distributions are easy to compute, making Gibbs sampling a practical solution.

We start with an initialized value for θ, and we perform a large number N_{Gibbs} of updates on its components. At each update step, we pick $i \leq n$ and update the value of the component θ_i according to the conditional distribution of θ_i given the current value of θ_{-i} (and, as always, the value of T). After a large number of iterations, this procedure yields a sample whose distribution is approximately that of a sample of θ given T only. Indeed, the distribution in question is the only stationary distribution of the Markov chain corresponding to the updates, as long as the chain is aperiodic and irreducible[2]. To perform this procedure, we need to compute the conditional probability $\mathbb{P}(\theta_i|T, \theta_{-i})$ for any i, θ_i, θ_{-i}.

We will do that in two steps: first, we compute the conditional probability $\mathbb{P}(z_i, m_i|T, \theta_{-i})$, and second, we compute the conditional probability density function of Φ_i given θ_{-i}, T and m_i, z_i. Those conditional probabilities and densities are proportional to the corresponding likelihoods. Note that we have the following expression for the likelihood of our model $\mathcal{L}(\theta) =$

$$\prod_{i=1}^{n} \lambda_s(t_i)^{z_i} \lambda_p(t_i)^{1-z_i} \prod_{j=1}^{n_\phi} \lambda_\phi(\varphi_j) \times e^{-\int_0^{t_n} \lambda_s(t)+\lambda_\phi(t)+\lambda_p(t)dt}. \quad (1)$$

This naturally factorizes as $\mathcal{L}(\theta) = \mathcal{L}_s(\theta)\mathcal{L}_\phi(\theta)\mathcal{L}_p(\theta)$, where $\mathcal{L}_s(\theta)$ $\mathcal{L}_\phi(\theta)$ and $\mathcal{L}_p(\theta)$ are, respectively, the components of the likelihood function (evaluated at θ) corresponding to the SFP, MPP and NHP components: $\mathcal{L}_s(\theta) = \prod_{i=1}^n \lambda_s(t_i)^{z_i} e^{-\int_0^{t_n} \lambda_s(t)dt}$. $\mathcal{L}_\phi(\theta)$ and $\mathcal{L}_p(\theta)$ are defined similarly.

A key observation now is that the factors in (1) corresponding to the intervals $(0, t_{i-1}]$ and $(t_{f(f(i))}, t_n]$ do not depend on the values $\{z, m\}$, where $f(u) = \text{argmin}_j\{t_j | t_u < t_j \wedge z_j = 1\}$ denotes the index of the next SFP event after t_u. Indeed, whether t_i is an SFP or Poisson event only influences the SFP intensity of the next two SFP events. Therefore, we can write equivalently:

$$\mathbb{P}(z_i = z, m_i = m|T, \theta_{-i}) \propto \mathcal{L}_s^i(\Omega_{z,m})\mathcal{L}_\phi^i(\Omega_{z,m})\mathcal{L}_p^i(\Omega_{z,m}), \quad (2)$$

[2]Those properties follow from the fact that the conditional distributions considered all have full support, as can be the seen below.

where $\mathcal{L}_s^i(\Omega_{z,m})$ (resp. $\mathcal{L}_\phi^i(\Omega_{z,m})$ and $\mathcal{L}_p^i(\Omega_{z,m})$) corresponds to the component of the likelihood corresponding the SFP (resp. MPP, NHPP) and to the interval $[t_{i-1}, t_{f(f(i))})$ and $\Omega_{z,m} := \theta_{-i} \cup \{z, m\}$. Thus $\theta = \Omega_{z,m} \cup \Phi_i$ and we have

$$\begin{aligned} \mathbb{P}(z_i = z, m_i = m|T, \theta_{-i}) &\propto \int_{\Phi_i \in F_{\Omega_{z,m}}} \mathcal{L}(\Omega_{z,m} \cup \Phi_i) d\Phi_i \\ &\propto \mathcal{L}_s^i(\Omega_{z,m})\mathcal{L}_\phi^i(\Omega_{z,m}) \int_{\Phi_i \in F_{\Omega_{z,m}}} \mathcal{L}_p^i(\Omega_{z,m} \cup \Phi_i) d\Phi_i, \end{aligned} \quad (3)$$

where we can write $\mathcal{L}_s^i(\Omega_{z,m})$ for $\mathcal{L}_s^i(\Omega_{z,m} \cup \Phi_i)$ for any Φ_i since $\mathcal{L}_s^i$ doesn't depend on Φ_i and $F_{\Omega_{z,m}}$ is the set of Φ_is compatible with the values of T, M when m_i is set to the index m: for instance, if $m_{i-1} = m_{i+1}$, then there are no transitions in the interval $[t_{i-1}, t_{i+1})$, so $F_{\Omega_{z,m}} = \{\emptyset\}$. On the other hand, if $m_{i+1} - m_{i-1} = 1$ and $m = m_{i+1}$, $F_{\Omega_{z,m}}$ is the interval $[t_{i-1}, t_i)$.

To develop an intuition of how the label of t_i affects $\mathcal{L}_s^i(\Omega_{z,m})$, we will demonstrate how to compute the intensities in the interval of interest. A similar explanation can be reproduced for $\mathcal{L}_\phi^i(\Omega_{z,m})$ and $\mathcal{L}_p^i(\Omega_{z,m})$. Consider Figure 3. By definition, given the parameter μ, the SFP intensity depends solely on the two last SFP events. Therefore the computation of $\lambda_s(t)$ before t_i only depends on the labels of the events before t_i and, therefore, as they are known, such labels are not influenced by whether t_i is a Poisson or a SFP event (note the green lines before t_i, Figure 3-A). Similarly, after $t_{f(f(i))}$ all the labels of the events are known and $\lambda_s(t)$ can be directly computed (note the green lines after $t_{f(f(i))}$, Figure 3-A). However, the label of event t_i impacts the value of $\lambda_s(t)$ between t_i and $t_{f(f(i))}$. In Figure 3-B, we consider the case where $t_i \in$ Poisson. In this case, the only SFP shift will happen at $t_{f(i)}$, which is (by definition) the first SFP event after t_i. Observe that the intensity between t_i and $t_{f(i)}$ remains the same as before t_i, with the next change occurring at $t_{f(i)}$. On the other hand, if $t_i \in$ SFP (see Figure 3-C), two shifts happen in the interval. The first one immediately after t_i and the second one after $t_{f(i)}$.

Therefore, $\mathcal{L}_s^i(\Omega_{z,m})$ and $\mathcal{L}_\phi^i(\Omega_{z,m})$ can be computed directly. The integral $\mathcal{L}_p^i(\Omega_{z,m}) := \int_{\Phi_i \in F_{\Omega_{z,m}}} \mathcal{L}_p^i(\Omega_{z,m} \cup \Phi_i) d\Phi_i$ can be expressed as

$$\begin{aligned} &C \prod_{j: t_j \in [t_{i-1}, t_{f(f(i))})} \left(\lambda_p(t_j|\Omega_{z,m})\right)^{1-z_j} \\ &\times \mathcal{I}(\Omega_{z,m}, t_{i-1}, t_i, m_{i-1}, m) \times \mathcal{I}(\Omega_{z,m}, t_i, t_{i+1}, m, m_{i+1}). \end{aligned} \quad (4)$$

The constant C is independent of m, z, defined as

$$C = e^{t_{i-1}\lambda_p(t_{t-i})} e^{-t_{i+1}\lambda_p(t_{t+i})} e^{-\int_{t_{i+1}}^{t_{f(f(i))}} \lambda_p(t)dt},$$

and

$$\mathcal{I}(\Omega, t_b, t_e, m_b, m_e) = \int_{\mathcal{T}} e^{\sum_{j=m_b}^{m_e} \left(\lambda_{(j+1)} - \lambda_{(j)}\right) x_{(j-m_b+1)}} dx,$$

where $\mathcal{T} = \{x = (x_1, \ldots, x_{m_e - m_b}) | t_b \leq x_1 \leq x_2 \leq \cdots \leq x_{m_e - m_b} \leq t_e\}$. The integrand in the definition of $\mathcal{I}$ is proportional to the likelihood of observing no Poisson event between t_e and t_b assuming x_i is the $m_b + i$th transition for all $1 \leq i \leq m_e - m_b$. The integrals in the above equations can be computed using the strategy described in the Appendix of this paper. This concludes the explanation of the computation of $\mathbb{P}(z_i = z, m_i = m|T, \theta_{-i})$. Since z_i, m_i are discrete random variables, it is then straightforward to sample from the corresponding distribution.

To complete our description of the Gibbs update which yields θ_i, we must describe how to draw Φ_i from its conditional distribution assuming $\Omega_{z,m}$ is given. Note that given $\Omega_{z,m}$, Φ_i^- and Φ_i^+ are independent, where $\Phi_i^- := \{\varphi_j | t_{i-1} < \varphi_j < t_i\}$ and $\Phi_i^+ := \{\varphi_j | t_i < \varphi_j < t_{i+1}\}$. The probability density function of Φ_i^- (resp. Φ_i^+) is proportional to $\mathcal{I}(\Omega_{z,m}, t_{i-1}, t_i, m_{i-1}, m)$ (resp. $\mathcal{I}(\Omega_{z,m}, t_i, t_{i+1}, m_i, m_{i+1})$). To generate a sample from the distribution of Φ_i^- in practice (Φ_i^+ is completely analogous), we make the following observations. For any interval $[a, b]$, let $N_{a,b} = \#(j : \phi_j \in [a, b])$ and $f_1 = (t_{i-1} + t_i)/2$. We have that the joint distribution of $(N_{t_{i-1},f_1}, N_{f_1,t_i})$ evaluated at (n_1, n_2) (with $n_1 + n_2 = m_{i+1} - m_{i-1}$) is proportional to $\mathcal{I}(\Omega_{z,m}, t_{i-1}, f_1, m_{i-1}, m_{i-1}+n_1)\mathcal{I}(\Omega_{z,m}, f_1, t_i, m_{i-1}+n_1, m_i)$. Thus, a sample can be drawn from it. We can continue to split the interval $[t_{i-1}, t_i)$ iteratively, choosing at each step how many φ_js are on each side of each subinterval by drawing from the relevant discrete distributions. This can be done until only one φ_i is in each interval, and its precise position can then be determined by a draw from its now one-dimensional probability distribution. This concludes the generation procedure for the E step.

M-Step: Now, we will elucidate the process of maximizing the log-likelihood, which corresponds to the current estimate of the conditional distribution of θ, over the parameter set $\{\mu, \Lambda\}$. The procedure described in the **E-Step** section allows us to draw N_θ samples $\{\theta_1, \theta_2, \cdots, \theta_{N_\theta}\}$ from the conditional distribution of θ given the current estimate of $\{\mu, \Lambda\}$. We then update μ via the formula $\hat{\mu} = \left(\sum_{j=1}^{N_\theta} \operatorname{argmin}_\mu \log(\mathcal{L}_s(\theta_j))\right)/N_\theta$, where the likelihood minimization steps are performed via binary search. Note that μ is an easy parameter to estimate as it affects the whole interval, thus $\operatorname{argmin}_\mu \mathcal{L}_s(\theta_j)$ is a good estimate even for a single value of j.

Regarding the set of parameters Λ, a key observation is that $\mathcal{L}_s(\theta)$ and $\mathcal{L}_\phi(\theta)$ are independent of Λ, allowing us to maximize over $\mathcal{L}_s(\theta)$ alone. Let $U_j(\theta) = \sum_j 1_{t_j \in [\bar{\varphi}_j, \bar{\varphi}_{j+1}) \wedge z_j = 0}$. The part of $\log(\mathcal{L}_s(\theta))$ which depends on λ_j is $-(\bar{\varphi}_{j+1} - \bar{\varphi}_j) + U_j(\theta)\log(\lambda_j) - \log(U_j(\theta))$. Averaging over all values of θ and optimizing over λ_j, we immediately obtain the following formula for the λs:

$$\hat{\lambda}_i = \frac{\sum_{j=1}^{N_\theta} U_i(\theta_j)}{\sum_{j=1}^{N_\theta} (\bar{\varphi}(\theta_j)_{i+1} - \bar{\varphi}(\theta_j)_i)}, \tag{5}$$

i.e. we are treating the observations of the Poisson events on the intervals corresponding to λ_j as if they came from a fixed homogeneous Poisson process. This is valid since the value of Λ only influences the Poisson likelihood component.

4 EXPERIMENTS

In this section, we present our experiments with real-world RSEs collected from various web systems. Additionally, we conduct synthetic data experiments (cf. Appendix) to validate our model under different ground truth scenarios and evaluate the effectiveness of our EM algorithm's derivation in recovering the model's underlying parameters. We showcase **BPoP**'s utility on 11 real-world datasets involving RSEs from diverse web systems across various domains. For more details regarding the datasets, we refer to the appendix.

Table 1 shows the total number of RSEs, the average number of events, as well as the average of the indices absolute stability κ and relative stability $\tilde{\kappa}$ (defined further in this section) for each dataset considering the entire observed time interval of each time series. Out of the total population, 91% of individuals exhibited a value of $0.05 < P_{NHPP} < 0.95$: this suggests a combination of stable behavior and a curious audience. Among those, 21% have at least one transition, i.e., their stable audience changed during the analyzed period, an assumption that motivated the conception of **BPoP**. The fifth column ($|\Phi|$) in Table 1 shows the the average number of transitions for each dataset among the individuals that had transitions. In total, we analyzed more than 78 million events.

Disentangling Stable and Ephemeral Audiences: We demonstrate the effectiveness of our disentangling method using data from the #ACL Twitter hashtag during the *Austin City Limits* (ACL) festival in 2009. ACL is an annual three-day music and art festival held in Austin, Texas, USA, attracting over 130 bands, with around 65,000 daily attendees. In 2009, ACL promoted the "The Sound and the Jury competition (SJC)," a virtual band contest offering a festival slot to the winner. Our model, relying solely on event timestamps related to the Twitter hashtag #ACL (Figure 5, top left), effectively separates the series of events into stable audience (NHPP) and curious (SFP) components. Notably, in this context, the audience is not tied to the festival itself but rather to the related **hashtag** in the period *before, during*, **and** *after the festival.* The *stable audience* (of the hashtag) comprises dedicated music fans who regularly engage with the Twitter feed, while the *curious audience* consists of individuals intrigued by the festival. The first component (Figure 5, bottom left) maintains a constant event arrival rate between transitions $\varphi_1, \varphi_2, \ldots$ (indicated by vertical lines). This rate $\lambda_p(t)$ signifies the stable audience rate at a given time t. In our model, bursts (Figure 5, top right) are associated with the intensity of the SFP, $\lambda_s(t)$ (Figure 5, bottom right), representing the curious audience. Importantly, we acknowledge that the rate $\lambda_p(t)$ is not stationary, as bursts (modeled by the SFP component) are triggered by external or internal events related to the topic. These factors not only generate short bursts of intense activity but also lead to enduring changes in the topic's discussion dynamics. Examples of such incidents could include retweets by prominent celebrities or the passing of influential figures related to the topic, both of which can alter the composition and behavior of the stable audience.

These facts can be observed in the behavior of the hashtag #ACL and its disentangled representation provided by **BPoP**. In the pre-SJC period, the arrival rate of the events was very low: at this time, only hard-core fans were actively posting tweets, namely the stable audience. This period of calm was disrupted by the SJC campaign period. The SJC campaign (highlighted in purple) was a short period of time during which the bands involved in the contest together with their fans (the curious) posted a large number of tweets asking users to vote for them. In addition to the short burst of tweets asking for votes, this effect altered the topic dynamic: now, it was not only the hard-core fans of the festival, but also the bands' supporters which were active in the social network. This effect was to continue until the end of the first round, and the announcement of the TOP-20 bands, which would be selected to compete in the next stage. After this announcement, since the number of bands in the contest has decreased, we expect a decrease in the number of supporters, resulting in fewer hashtag users and tweets. From **BPoP**, we indeed observe a transition at this event and a decreased stable audience (NHPP) rate afterwards. The new rate remained constant until the ACL event itself. Finally, a huge burst of events occurred during the

Table 1: Fitting and characterization of the datasets

	#RSE	n	κ	$\tilde{\kappa}$	$\lvert\Phi\rvert$	R^2_{BPoP}	R^2_{BP}	R^2_{H}	R^2_{BH}
AskMe	490	133	0.36	0.45	1.03	**0.8514 ± 0.11**	0.6204 ± 0.10	0.3211 ± 0.15	0.2967 ± 0.2057
Digg	974	122	0.31	0.45	1.74	**0.8944 ± 0.10**	0.7231 ± 0.09	0.4824 ± 0.17	0.5142 ± 0.1786
Enron	147	1589	0.59	0.49	2.21	**0.9449 ± 0.07**	0.9178 ± 0.06	0.5954 ± 0.25	0.8686 ± 0.1459
GitHub(U)	40385	675	0.76	0.50	1.40	0.9565 ± 0.03	0.9526 ± 0.04	0.8876 ± 0.10	**0.9657 ± 0.0524**
GitHub(P)	35085	696	0.77	0.50	1.39	**0.9570 ± 0.03**	0.9523 ± 0.04	0.8853 ± 0.10	0.9503 ± 0.0449
G. Trends	579	2975	0.66	0.48	2.09	**0.9632 ± 0.07**	0.9596 ± 0.01	0.7973 ± 0.29	0.8148 ± 0.2128
MetaFilter	8249	172	0.42	0.43	1.23	**0.8931 ± 0.09**	0.7123 ± 0.11	0.3906 ± 0.18	0.4624 ± 0.2020
MetaTalk	2465	203	0.43	0.49	1.26	**0.9176 ± 0.08**	0.7921 ± 0.11	0.4535 ± 0.20	0.6093 ± 0.2145
Twitter(X)	18888	1142	0.71	0.50	1.60	**0.9469 ± 0.05**	0.8882 ± 0.12	0.8098 ± 0.23	0.8355 ± 0.2312
Yelp	1931	128	0.22	0.38	1.34	0.9193 ± 0.13	0.9411 ± 0.08	0.8402 ± 0.14	**0.9470 ± 0.0422**
YouTube	250	3241	0.59	0.49	1.90	**0.9720 ± 0.02**	0.9696 ± 0.01	0.7010 ± 0.18	0.7623 ± 0.1661

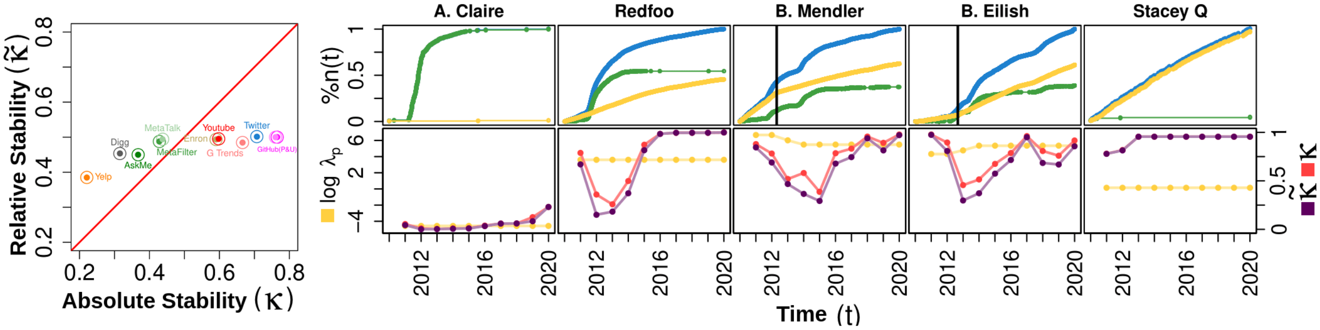

Figure 4: Left: Scatterplot of the average $\tilde{\kappa}$ versus the average κ for each of the real datasets. Right: Each column of the plots corresponds to the Google Trend time series associated with one of 5 artists. The top row shows the raw counts. The bottom plots show the yearly average of $\log(\lambda_p(t))$ (yellow line), κ (magenta line), and $\tilde{\kappa}$ (red line). Vertical black lines mean NHPP transitions.

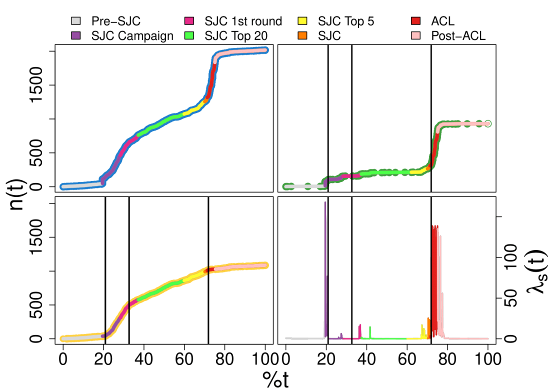

Figure 5: Analysis of the #ACL Twitter hashtag associated with the 2009 *Austin City Limits* (ACL) festival. Top left: cumulative events; Top right: curious (burst) component; Bottom left: stable audience (NHPP) component; Bottom right: intensity of the SFP.

festival, which was correctly modeled by our model as mostly part of the SFP process, which represents the curious. Our model also detects a transition at this event, and the rate goes back to pre-SJC levels afterwards. After the festival ends, the audience is once again composed of the stable audience only.

Our model's core hypothesis is that short bursts of activity are likely to simultaneously change the stable audience constitution because they share a common cause: topic related unusual incidents. We model this by making the transition occurrence rate equal to $c\lambda_s(t)$ or proportional to the arrival rate of atypical (SFP) events ($\lambda_s(t)$), where c is small. SFP events appear also during non-bursty periods. **BPoP** allows for the possibility of transitions occurring also during calm periods, not only during bursts, and our algorithm is able to detect such transitions.

Absolute and Relative Stability: After learning the component intensities $\lambda_s(t)$ and $\lambda_p(t)$ we can contrast their absolute and relative influence on the observed events. We define two indexes, both in the interval $[0, 1]$:

$$\kappa = \int_a^b \frac{\lambda_p(t)(b-a)^{-1}}{\lambda_p(t)+\lambda_s(t)} dt \text{ and } \widetilde{\kappa} = \frac{\int_a^b \lambda_p(t)dt}{\int_a^b (\lambda_p(t)+\lambda_s(t))dt}. \quad (6)$$

The absolute stability κ tells us the importance of the stable audience averaged over the time interval $[a, b]$, whilst the relative stability $\widetilde{\kappa}$ describes the proportion of the activity in the interval $[a, b]$ that is assigned to the stable audience. The plot on the left hand side of Figure 4 shows the average $\tilde{\kappa}$ versus the average κ for

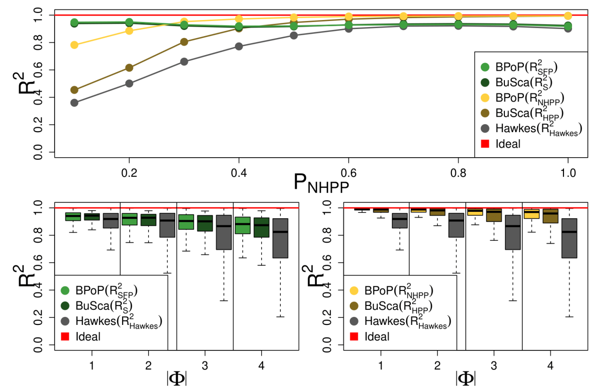

Figure 6: Top: fitting performance (determination coefficient) as a function of the proportion $|PP|/n$ of Poisson events. Bottom: fitting performance as a function of the number of transitions.

each of the real datasets. Consider the two pairs associated with GitHub: $\kappa \approx 0.8$ but $\tilde{\kappa} \approx 0.5$. This shows that while the stable audience (NHPP component) dominates most of the time, only half of the activities are carried out by them, i.e., the other half comes from the curious.

The plots in the right hand side of Figure 4 show the parameters λ_p, κ and $\widetilde{\kappa}$ calculated in each year separately. We demonstrate the versatility of our method in capturing a wide range of time series dynamics, including those with and without popularity transitions, as illustrated in the top graphs of Figure 4. For example, the time series for (1) Claire reveals a lack of a stable audience, with activity largely driven by transient curiosity. The middle graphs all highlight a mix of audiences, showcasing (2) Redfoo (without transitions), (3) Mendler (experiencing a single transition), and (4) Ellish (also with one transition). In contrast, (5) Stacey Q's time series depicts a consistently stable audience devoid of abrupt fluctuations. The analysis of popularity shifts is depicted in the bottom graphs of Figure 4 (indicated by yellow lines). Notably, the popularity levels for (1), (2), and (4) remain unchanged throughout the observed period. Conversely, (3) exhibits a decline in popularity around 2012, while (4) sees an uptick in stable popularity around the same timeframe. Furthermore, the BPoP metric provides a quantitative comparison between stable popularity and curiosity-driven behaviors, as analyzed in the red and purple lines of the bottom graphs in Figure 4.

5 GOODNESS OF FIT

Baselines: BPoP is a model that relies only on the observed event timestamps. We compare our model with three similar models.

Hawkes processes [2, 24, 29, 39, 52] are a class of self-exciting processes which are widely used for modeling web communications. The Hawkes process model assumes that any event increases the probability of additional events. Its conditional intensity is $\lambda(t|\mathcal{H}_t) = \lambda + \sum_{t_i<t} K(t-t_i)$, where $K(x) > 0$ is the kernel function, which satisfies $\int_0^\infty K(x)dx < 1$, to ensure stationarity.

BuSca [1]: similarly to our model **BPoP**, BuSca is a mixture process involving a homogeneous Poisson process and a self-exciting process. The conditional intensity of the BuSca model is given by $\lambda(t|\mathcal{H}_t) = \lambda + \frac{1}{\Delta_t+\mu/e}$, where $\lambda \geq 0$ and $\mu > 0$ are constants and Δ_t is the last SFP interval before t. However, in contrast to our model, BuSca assumes that the stable audience (Poisson component) remains constant and continues indefinitely.

BuSca-Hawkes: this model is similar to BuSca. However, the bursty component is a Hawkes process instead of an SFP. This can be considered a particular case of the method described in [32].

Metrics: To assess the Hawkes Process method's performance we used the random time change theorem to transform a HP into a unit rate Poisson process (see [9]). After the transformation we computed the determination coefficient R^2_H corresponding to the linear regression problem predicting the cumulative number of events $N(t)$ for all ts in the transformed process. Similarly, for BuSca, we computed R^2_S (SFP) and R^2_{HPP} (homogeneous Poisson process) for the disentangled processes (see [1]) and computed the final coefficient $R^2_{BP} = (R^2_S + R^2_{HPP})/2$. For the BuSca-Hawkes process, we compute the combined metric $R^2_{BH} = (R^2_S + R^2_H)/2$, where R^2_S and R^2_H reflect the contributions of the respective underlying processes.

To check the goodness of fit of **BPoP**, we first output the $\{\hat{Z}, \hat{\Phi}\} \subset \theta_i$, where $i = \text{argmax}_j\{\mathcal{L}(\theta_j)\ ;\ 1 \leq j \leq N_\theta\}$, which allows us to disentangle the NHPP from the SFP. For the SFP fitting, we took the inter-event times sample and built the empirical cumulative distribution function $\mathbb{F}(t)$ leading to the odds-ratio function $OR(t) = \mathbb{F}(t)/(1-\mathbb{F}(t))$. Then, we computed the R^2_{SFP} coefficient of the linear regression problem predicting the cumulative number of events $N(t)$ versus the $OR(t)$ (see [44]).

The computation of R^2_{NHPP} requires some explanation. Let $\bar{\Phi}$ be the list of the elements of $\{0\cup\Phi\cup t_n\}$ ordered from smallest to largest, so that $\bar{\varphi}_0 = 0$, $\bar{\varphi}_{m_t+1} = t_n$. For all $i \in \{0, 1, \cdots, |\bar{\Phi}|-1\}$, we construct the set $T_i = \{t_j|\bar{\varphi}_i < t_j < \bar{\varphi}_{i+1}\}$ and then estimate $R_{NHPP}{}^2_i$ as the determination coefficient corresponding to the linear regression problem predicting $N(t)$ from t on the interval $[\bar{\varphi}_i, \bar{\varphi}_{i+1})$ with the datapoints obtained from T_i. Finally, we compute R^2_{NHPP} as the weighted average of the $R_{NHPP}{}^2_i$s. The weight is a multiple of the respective interval. More formally, $R^2_{NHPP} = \sum_i(\bar{\varphi}_{i+1} - \bar{\varphi}_i)R_{NHPP}{}^2_i/t_n$. Finally, we compute $R^2_{BPoP} = (R^2_{SFP} + R^2_{NHPP})/2$.

All R^2 coefficients vary between 0 (*worst case*) and 1 (*best case*). Table 1 shows the goodness-of-fit statistics (average and standard deviation) for **BPoP** and for the baselines, grouped by dataset. **BPoP** surpasses the Hawkes process method in all datasets considered. It also consistently outperforms BuSca (in 9 out of 11 datasets). Indeed, the high concentration of R^2_{BPoP} (as R^2_{SFP} and R^2_{NHPP}, in Figure 6) close to the maximum value of 1 shows that our model can accurately fit the time series considered, as well as disentangle the mixed process into its two hidden components (NHPP and SFP).

Results: Figure 6 shows how our disentangled models behave under different regimes. To construct the top graph, we computed P_{NHPP} and rounded the value considering the range $\{0, 0.1, 0.2, \cdots, 1\}$. The extremes correspond, respectively, to a pure Poisson processes and a pure SFP. We can observe that our model improves significantly when the mixture is dominated by the bursty behavior. With respect to the number of transitions $|\Phi|$, the boxplots on the bottom show that **BPoP** outperforms the baselines across the whole spectrum, with the performance increasing with the number of transitions.

6 RELATED WORK

Human Communication Dynamics: Characterizing the dynamics of human communication on the web has significant implications for various applications, including trend detection, clustering, anomaly detection, and popularity prediction [11, 29, 52, 57]. This research is inspired by a substantial body of work focused on predicting the popularity of online content, such as YouTube videos, hashtags, and forum posts [2, 16, 33, 38, 49, 55]. The primary goal is to estimate the total number of events associated with a given item. At first, Crane and Sornette posited two primary mechanisms for the occurrence of viewing activity: random occurrences influenced by external factors such as featuring, or internally driven through sharing [8]. However, recent studies have identified additional factors influencing popularity, such as content quality [25, 50], item metadata [18], item age [5], recommendation algorithms, ranking in keyword-based queries [59], and social network effects [7, 50].

More recently, Yu et al. [55] have introduced a phase representation model for online videos, extending Crane and Sornette's endogenous growth and exogenous shock model. They discovered that online videos undergo multiple stages of popularity fluctuations over several months. This finding received further support from Rizoiu et al. [38], who identified a strong correlation between external promotion and online video popularity. Their research highlighted the substantial impact of external attention and promotion on video popularity. Additionally, Gleeson et al. [16] emphasized the significance of recent popularity over cumulative popularity in the adoption of Facebook apps.

RSEs Modeling: Human activity on the web displays a broad spectrum of unpredictability, ranging from complete randomness [8, 20, 21, 28] to high correlation and burstiness [3, 19, 29, 45, 52]. These diverse patterns have prompted the adoption of point process stochastic models, which provide statistical frameworks for comprehending sequences of random events [9, 40]. In principle, these models can be employed to estimate the audience size (*fanbase*) of online items. However, existing models are not well-suited for this specific task. Poisson processes (PPs) [21, 28] are suitable when events arrive regularly at a fixed rate, allowing for stable audience estimation. While a significant portion of online items can be accurately described by such a simple model [7, 8, 12, 27], PPs have limitations. For example, Malmgren et al. [28] introduced a non-homogeneous Poisson process model that accounts for circadian cycles with varying rate $\lambda(t)$, but it lacks the self-exciting property. This means that the probability of observing an event at a small time interval $[t, t + \Delta t)$ does not depend on previous events within that interval. This limitation hampers PPs from effectively capturing event bursts observed in real-world data.

On the other hand, while self-exciting processes effectively capture correlations between consecutive events responsible for activity bursts in real data, existing methods often overlook the time-varying nature of the fanbase. Hawkes processes [2, 24, 29, 39, 52], one of the most widely used models, maintain a constant baseline rate and incorporate event history via a conditional intensity formula: $\lambda(t|\mathcal{H}_t) = \lambda + \sum_{t_i<t} K(t - t_i)$, where $K(x) > 0$ is typically a decreasing exponential kernel. Hawkes processes promote bursty behavior and fall into the category of pure self-exciting models with a constant background intensity. They offer an alternative to pure Poisson processes, with nuanced mathematical properties and applications. However, they cannot capture changes in background intensity, a feature addressed in our work.

The recent literature reflects a growing interest in exploring alternatives to the widely-used Hawkes-based processes for modeling self-exciting point process data. For instance, Etesami et al. [10] and Trouleau et al. [42] have applied variational inference algorithms to fit Bayesian models for multivariate self-feeding processes, enabling the analysis of real-world communication dynamics. Moreover, Noorbakhsh and Rodriguez [37] have introduced a novel class of Gumbel-max point processes specifically designed to address causal issues in point process modeling. Lastly, [34, 58] study state-dependent Hawkes processes, where point processes interact based on a latent state. Similarly to our work, they involve a latent state, which in our corresponds to the current metapoisson intensity. However, in their case, the type of the event (which random point process it is drawn from) is known for each event time.

Our results consistently make a strong case for the adoption of self-feeding processes as a compelling alternative to widely-used Hawkes-based models. Previous research, including [1, 4, 31, 45], strongly supports this approach. Its appeal lies in its simplicity and its ability to accurately capture the short-memory and power-law behavior common in real-world data. Self-feeding processes produce point patterns characterized by bursts of intense activity followed by periods of low activity, aligning well with real-world observations. By applying our model to another real case, we aim to demonstrate its effectiveness as a competitive alternative for modeling self-exciting point processes. Finally, Alves et al. [1] employed a Wold process to model social media events effectively. However, this model, with a constant background rate, presents significant training challenges due to multiple approximations required for EM algorithm expectations. Our novel model addresses these limitations by accurately mimicking event bursts while efficiently capturing time-varying background rates, providing a more realistic representation of our fanbase dynamics.

7 CONCLUSION

In this article, we presented Burst-induced Poisson Process (**BPoP**), a model that separates stable and curious media audiences. **BPoP** combines an SFP for viral thread bursts (representing the curious audience) with a non-homogeneous Poisson process for regular user behavior (the stable audience). These components interact, and we develop a tailored EM algorithm to address this complexity. Our model excels in identifying audience dynamics in both synthetic and real data.

ACKNOWLEDGEMENTS

This work is supported by the authors' individual grants from FAPEMIG and CNPq. Part of this work was conducted within the DFG research unit FOR 5359 on Deep Learning on Sparse Chemical Process Data (KL 2698/6-1 and KL 2698/7-1). Alves acknowledges Recombee for supporting his research. Kloft acknowledges support by the Carl-Zeiss Foundation, the DFG awards KL 2698/2-1 and KL 2698/5-1, and the BMBF awards 03|B0770E and 01|S21010C.

REFERENCES

[1] Rodrigo Augusto da Silva Alves, Renato Martins Assuncao, and Pedro Olmo Stancioli Vaz de Melo. 2016. Burstiness scale: A parsimonious model for characterizing random series of events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1405–1414.

[2] Peng Bao. 2016. Modeling and predicting popularity dynamics via an influence-based self-excited Hawkes process. In *International Conference on Information and Knowledge Management, Proceedings*. https://doi.org/10.1145/2983323.2983868

[3] Albert-László Barabási. 2005. The origin of bursts and heavy tails in human dynamics. *Nature* 435, 7039 (may 2005), 207–211. https://doi.org/10.1038/nature03459

[4] Guilherme Borges, Flavio Figueiredo, Renato M Assunçao, and Pedro OS Vaz-de Melo. 2020. Networked Point Process Models Under the Lens of Scrutiny. In *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*.

[5] Youmna Borghol, Siddharth Mitra, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. 2011. Characterizing and modelling popularity of user-generated videos. *Performance Evaluation* 68, 11 (2011), 1037–1055.

[6] Cody Buntain and Jimmy Lin. 2016. Burst Detection in Social Media Streams for Tracking Interest Profiles in Real Time. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 777–780. https://doi.org/10.1145/2911451.2914733

[7] Xu Cheng, Cameron Dale, and Jiangchuan Liu. 2008. Statistics and social network of youtube videos. In *2008 16th Interntional Workshop on Quality of Service*. IEEE, 229–238.

[8] R. Crane and D. Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences* 105, 41 (oct 2008), 15649–15653. https://doi.org/10.1073/pnas.0803685105

[9] Daryl J Daley and David Vere-Jones. 2007. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

[10] Jalal Etesami, William Trouleau, Negar Kiyavash, Matthias Grossglauser, and Patrick Thiran. 2021. A variational inference approach to learning multivariate wold processes. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2044–2052.

[11] Alceu Ferraz Costa, Yuto Yamaguchi, Agma Juci Machado Traina, Caetano Traina, and Christos Faloutsos. 2015. RSC: Mining and Modeling Temporal Activity in Social Media. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. ACM Press, New York, New York, USA, 269–278. https://doi.org/10.1145/2783258.2783294

[12] Flavio Figueiredo, Fabricio Benevenuto, and Jussara M Almeida. 2011. The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of the fourth ACM international conference on Web search and data mining*. 745–754.

[13] Kazuki Fujita, Alexey Medvedev, Shinsuke Koyama, Renaud Lambiotte, and Shigeru Shinomoto. 2018. Identifying exogenous and endogenous activity in social media. *Physical Review E* 98, 5 (2018), 052304.

[14] Florencia García-Rapp. 2017. Popularity markers on YouTube's attention economy: the case of Bubzbeauty. *Celebrity Studies* (2017).

[15] Alan E Gelfand. 2000. Gibbs sampling. *Journal of the American statistical Association* 95, 452 (2000), 1300–1304.

[16] J. P. Gleeson, D. Cellai, J.-P. Onnela, M. A. Porter, and F. Reed-Tsochas. 2014. A simple generative model of collective online behavior. *Proceedings of the National Academy of Sciences* 111, 29 (jul 2014), 10411–10415. https://doi.org/10.1073/pnas.1313895111

[17] Masoud Hassanpour, Seyed Amir Hoseinitabatabaei, Payam Barnaghi, and Rahim Tafazolli. 2020. Improving the accuracy of the video popularity prediction models through user grouping and video popularity classification. *ACM Transactions on the Web (TWEB)* 14, 1 (2020), 1–28.

[18] William Hoiles, Anup Aprem, and Vikram Krishnamurthy. 2017. Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data. *IEEE Transactions on Knowledge and Data Engineering* 29, 7 (2017), 1426–1437.

[19] Hao Jiang and Constantinos Dovrolis. 2005. Why is the Internet Traffic Bursty in Short Time Scales?. In *Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'05)*. 241–252.

[20] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido. 2004. A nonstationary poisson view of internet traffic. In *IEEE INFOCOM 2004*, Vol. 3. IEEE, 1558–1569. https://doi.org/10.1109/INFCOM.2004.1354569

[21] Jon Kleinberg. 2002. Bursty and hierarchical structure in streams. In *Proceedings of the eighth ACM SIGKDD (KDD '02)*. ACM, New York, NY, USA, 91–101. https://doi.org/10.1145/775047.775061

[22] Janette Lehmann, Bruno Gonçalves, José J Ramasco, and Ciro Cattuto. 2012. Dynamical classes of collective attention in twitter. In *Proceedings of the 21st international conference on World Wide Web*. 251–260.

[23] Haoyang Li, Peng Cui, Chengxi Zang, Tianyang Zhang, Wenwu Zhu, and Yishi Lin. 2019. Fates of Microscopic Social Ecosystems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, NY, USA, 668–676. https://doi.org/10.1145/3292500.3330827

[24] Sha Li, Xiaofeng Gao, Weiming Bao, and Guihai Chen. 2017. FM-Hawkes. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 1119–1128. https://doi.org/10.1145/3132847.3132883

[25] Dongliang Liao, Jin Xu, Gongfu Li, Weijie Huang, Weiqing Liu, and Jing Li. 2019. Popularity prediction on online articles with deep fusion of temporal process and content features. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 200–207.

[26] Changsha Ma, Zhisheng Yan, and Chang Wen Chen. 2017. Larm: A lifetime aware regression model for predicting youtube video popularity. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 467–476.

[27] Xinyu Ma. 2022. Online video popularity modeling research. In *International Conference on Applied Statistics, Computational Mathematics, and Software Engineering (ASCMSE 2022)*, Vol. 12345. SPIE, 375–381.

[28] R. Dean Malmgren, Jake M. Hofman, Luis A.N. Amaral, and Duncan J. Watts. 2009. Characterizing individual communication patterns. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*. ACM Press, New York, New York, USA, 607. https://doi.org/10.1145/1557019.1557088

[29] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, and Christos Faloutsos. 2012. Rise and fall patterns of information diffusion. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*. ACM Press, New York, New York, USA, 6. https://doi.org/10.1145/2339530.2339537

[30] Ega Maulana, Usep Suhud, Juhasdi Susono, Agung Dharmawan Buchdadi, and K Amiruddin. 2020. Measuring YouTube Channel Subscriber Loyalty. *The International Journal of Social Sciences World (TIJOSSW)* 2, 2 (2020), 32–39.

[31] Pedro OS Vaz De Melo, Christos Faloutsos, Renato Assunção, Rodrigo Alves, and Antonio AF Loureiro. 2015. Universal and distinct properties of communication dynamics: how to generate realistic inter-event times. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 3 (2015), 1–31.

[32] Xenia Miscouridou, Samir Bhatt, George Mohler, Seth Flaxman, and Swapnil Mishra. 2023. Cox-Hawkes: doubly stochastic spatiotemporal Poisson processes. *Transactions on Machine Learning Research* (2023). https://openreview.net/forum?id=xzCDD9i4IZ

[33] Swapnil Mishra. 2019. Bridging models for popularity prediction on social media. In *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/3289600.3291598

[34] Maxime Morariu-Patrichi and Mikko S Pakkanen. 2022. State-dependent Hawkes processes and their application to limit order book modelling. *Quantitative Finance* 22, 3 (2022), 563–583.

[35] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.

[36] Shu Kay Ng, Thriyambakam Krishnan, and Geoffrey J McLachlan. 2012. The EM algorithm. In *Handbook of computational statistics*. Springer, 139–172.

[37] Kimia Noorbakhsh and Manuel Rodriguez. 2022. Counterfactual temporal point processes. *Advances in Neural Information Processing Systems* 35 (2022), 24810–24823.

[38] Marian-Andrei Rizoiu, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. 2017. Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*. 735–744.

[39] Tiago Santos, Simon Walk, Roman Kern, Markus Strohmaier, and Denis Helic. 2019. Self- and Cross-Excitation in Stack Exchange Question & Answer Communities. In *The World Wide Web Conference*. ACM, New York, NY, USA, 1634–1645. https://doi.org/10.1145/3308558.3313440

[40] Donald L Snyder and Michael I Miller. 2012. *Random point processes in time and space*. Springer Science & Business Media.

[41] Zhiyi Tan, Yanfeng Wang, Ya Zhang, and Jun Zhou. 2016. A novel time series approach for predicting the long-term popularity of online videos. *IEEE Transactions on Broadcasting* 62, 2 (2016), 436–445.

[42] William Trouleau. 2021. *Learning Self-Exciting Temporal Point Processes Under Noisy Observations*. Ph.D. Dissertation. École Polytechnique Fédérale de Lausanne - EPFL. https://infoscience.epfl.ch/record/288643

[43] David Vallet, Shlomo Berkovsky, Sebastien Ardon, Anirban Mahanti, and Mohamed Ali Kafaar. 2015. Characterizing and predicting viral-and-popular video content. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 1591–1600.

[44] Pedro Olmo Stancioli Vaz de Melo, Christos Faloutsos, Renato Assunção, Rodrigo Alves, and Antonio A.F. Loureiro. 2015. Universal and Distinct Properties of Communication Dynamics: How to Generate Realistic Inter-event Times. *ACM Transactions on Knowledge Discovery in Data* (2015).

[45] Pedro Olmo S Vaz de Melo, Christos Faloutsos, Renato Assunção, and Antonio Loureiro. 2013. The self-feeding process: a unifying model for communication dynamics in the web. In *Proceedings of the 22nd international conference on World Wide Web*. 1319–1330.

[46] Aining Wang, Chen Zhang, and Yuedong Xu. 2016. A first view on mobile video popularity as time series. In *Proceedings of the 8th ACM International Workshop on Hot Topics in Planet-scale mObile computing and online Social neTworking*. 7–12.

[47] Senzhang Wang, Zhao Yan, Xia Hu, Philip S Yu, and Zhoujun Li. 2015. Burst Time Prediction in Cascades. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9338

[48] Mirjam Wattenhofer, Roger Wattenhofer, and Zack Zhu. 2012. The YouTube social network. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 6. 354–361.

[49] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. 2016. Time Matters. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, New York, NY, USA, 1336–1344. https://doi.org/10.1145/2964284.2964335

[50] Jiqiang Wu, Yipeng Zhou, Dah Ming Chiu, and Zirong Zhu. 2016. Modeling dynamics of online video popularity. *IEEE Transactions on Multimedia* 18, 9 (2016), 1882–1895.

[51] Qitian Wu, Chaoqi Yang, Hengrui Zhang, Xiaofeng Gao, Paul Weng, and Guihai Chen. 2018. Adversarial Training Model Unifying Feature Driven and Point Process Perspectives for Event Popularity Prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, USA, 517–526. https://doi.org/10.1145/3269206.3271714

[52] Shuang-hong Yang and Hongyuan Zha. 2013. Mixture of Mutually Exciting Processes for Viral Diffusion. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, Sanjoy Dasgupta and David Mcallester (Eds.), Vol. 28. JMLR Workshop and Conference Proceedings, 1–9. http://jmlr.csail.mit.edu/proceedings/papers/v28/yang13a.pdf

[53] Junjie Yao, Bin Cui, Yuxin Huang, and Xin Jin. 2010. Temporal and social context based burst detection from folksonomies. In *Proceedings of the National Conference on Artificial Intelligence*.

[54] Hongzhi Yin, Bin Cui, Hua Lu, Yuxin Huang, and Junjie Yao. 2013. A unified model for stable and temporal topic detection from social media data. In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*. IEEE, 661–672.

[55] Honglin Yu, Lexing Xie, and Scott Sanner. 2015. The Lifecyle of a Youtube Video: Phases, Content and Popularity. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10537

[56] Chengxi Zang, Peng Cui, Christos Faloutsos, and Wenwu Zhu. 2017. Long Short Memory Process. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, 565–574. https://doi.org/10.1145/3097983.3098055

[57] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1513–1522.

[58] Feng Zhou, Quyu Kong, Yixuan Zhang, Cheng Feng, and Jun Zhu. 2021. Nonlinear Hawkes processes in time-varying system. *arXiv preprint arXiv:2106.04844* (2021).

[59] Renjie Zhou, Samamon Khemmarat, Lixin Gao, Jian Wan, and Jilin Zhang. 2016. How YouTube videos are discovered and its impact on video views. *Multimedia Tools and Applications* 75 (2016), 6035–6058.

APPENDIX

A DATASETS' DESCRIPTION

Below, we provide a description of the real-world datasets used in our study:

- **AskMe**, **MetaFilter** and **MetaTalk**[3]: the time-series are from topics of an online discussion forum and the events consist of comments timestamps.
- **Digg**: each time-series corresponds to a different news post in the website and the events are the *diggs* (a "digg" is a concept akin to a Facebook "like") given to the respective post.
- **Enron**[4]: a sequence of events is associated with an e-mail account and the events are the incoming and outgoing messages timestamps.
- **Github** (Collected by the authors): we split this dataset into two parts: **Github** (**U**sers) and **Github** (**P**rojects). In the first one, the events are activities of a user (in different projects). In the second one, the events correspond to different activities by the user on the corresponding project.
- **Google Trends** (Collected by the authors): time series of YouTube views. We consider only USA users. Each topic is related to famous people such as singers and politicians.
- **Twitter(X)**: the event timestamps correspond to tweets featuring a given hashtag. (Collected by the authors)
- **Yelp**: the dataset consists of timestamps of user ratings for several restaurants.
- **Youtube**: each time series correspond to a YouTube video and the events are the timestamps of users' comments.

The GitHub dataset in this study was collected by the authors and it includes user and repository events, with each entry being a User ID, Project ID, and Timestamp triplet, indicating user actions like "Fork" or "Commit". We then aggregate time series data for users (GitHub U) and projects (GitHub P). For Google Trends data, collected via a Python API, preprocessing was needed as timestamps are not provided. The data consist of normalized search volumes on a 0 to 100 scale, collected in batches with a single anchor in each for rescaling purposes. Post-collection, each data point has a 120-number sequence indicating relative search volumes. The data transformation involved generating random event series from these numbers, e.g., from a series $\{12, 7, 50, \dots, 43\} \in \mathbb{R}^{120}$, sampling uniformly 12 points from $(0, 1]$, 7 from $(1, 2]$, and 43 from $(119, 120]$.

B SYNTHETIC DATA EXPERIMENTS

BPoP is a *generative model* that combines an SFP to represent the curious audience and an NHPP to represent the stable audience, both interacting with each other. While the complete generation procedure is originally detailed in the main paper's model description, we're providing a concise summary in this appendix for ease of reference.

Generation procedure: consider that, at time t, the history of the process is composed of the observed event timestamps $\{t_1, t_2, \dots\} < t$, unobserved labels $\{z_1, z_2, \dots\}$ as well as unobserved MPP events $\{\varphi_1, \varphi_2, \dots\}$, which represent the transitions. We use the convention

[3]Avaiable in: http://stuff.metafilter.com/infodump/

[4]Available at: https://www.cs.cmu.edu/~./enron/

that $z_i = 0$ if $t_i \in$ NHPP and $z_i = 1$ if $t_i \in$ SFP. Define the following three intensity functions: (1) the SFP intensity $\lambda_s(t) = 1/[(t_{g(t)} - t_{g(g(t))}) + \mu/e]$ where $g(u) = [\max(t_i : z_i = 1 \quad \wedge \quad t_i < u)]_+$ denotes the last SFP event before t, with the convention that $g(t) = 0$ if $\nexists i : t_i \leq t \wedge z_i = 1$ and $\mu > 0$ is the SFP parameter; (2) $\lambda_\Phi(t) = c\lambda_s(t)$, where $c \in [0, 1]$ is a parameter that controls the NHPP transition sensitivity; and (3) $\lambda_p(t) = \lambda_{m_t}$, where $\Lambda = \{\lambda_0, \lambda_1, \lambda_2, \cdots\}$ is an infinite set of positive numbers (parameters) and, $m_t = \sum_{j=1}^{n_\phi} 1_{\varphi_j < t}$. Similarly we can define $\lambda_s^+(t) = \lim_{\delta \to 0} \lambda_s(t + \delta)$ (resp. $\lambda_\phi^+(t)$ and $\lambda_p^+(t)$) as the intensity of the SFP (resp. MPP and NHPP) immediately after t. Thus, to generate the next time stamp, we first generate three exponential variables E_s, E_ϕ and E_p with intensities $\lambda_s^+(t)$, $\lambda_\phi^+(t)$ and $\lambda_p^+(t)$ respectively. Then, the next event will take place at $t + E$, where E=min(E_s, E_ϕ, E_p), and it will belong to the SFP (resp. NPHH, MPP) component if $E = E_s$ (resp. E_ϕ, E_p). Likewise, we continue generating the rest of the process from time $t + E$.

Experimental setup: Let $P_{NHPP} = \sum_{i=1}^{n}(1 - z_i)/n$ be the proportion of observed NHPP events. We chose sets of μ, Λ and c corresponding to estimated values of $(\mathbb{E}(n), \mathbb{E}(|\Phi|), \mathbb{E}(P_{NHPP}))$ in the set $\{500, 750, 1000\} \times \{0, 1, 2\} \times \{0, 0.25, 0.5, 0.75, 1\}$. In the case where transitions are present, we considered only the cases in which the expected number of events of both processes (SFP and NHPP) is greater than 0. For example, the tuple $(500, 1, 0)$ was not considered since we would expect only one transition and zero NHPP events. Concerning the parameter Λ we selected values such that $\forall i \min(\lambda_i, \lambda_{i+1})/\max(\lambda_i, \lambda_{i+1}) = 1/3$.

For each tuple $(\mathbb{E}(n), \mathbb{E}(|\Phi|), \mathbb{E}(P_{NHPP}))$, we conducted 50 simulations (we use the generation procedure explained above) and assessed our methods via two metrics. To assess our method's ability to recover the ground truth model accurately given the observations, we aggregate the relative difference between the total intensities corresponding to our recovered labels and the ground truth labels:

$$\delta(\theta, \hat{\theta}) = \int_0^{t_n} \frac{|(\lambda_s(t|\theta) + \lambda_p(t|\theta)) - (\lambda_s(t|\hat{\theta}) + \lambda_p(t|\hat{\theta}))|}{\lambda_s(t|\theta) + \lambda_p(t|\theta)} dt. \tag{7}$$

The reason we do this instead of simply counting the proportion of correct labels is as follows. Correctly classifying the timestamps is both more difficult and less interesting inside a burst compared to calm periods. Furthermore, the parametrization of the model could present some redundancy, in which case very different parameter combinations could correspond to similar point processes. On the other hand, $\delta(\theta, \hat{\theta})$ is far less sensitive to the label assignments in a short bursty period, but a small value of $\delta(\theta, \hat{\theta})$ still indicates excellent performance. Indeed, it shows that the model accurately represents the position of the set of observations in the probability space and would perform well at predicting the positions of further observations if they had been left unobserved.

The second metric simply aims at verifying convergence. We evaluate the log-likelihood $\log(\mathcal{L}(\hat{\theta}))$ at our model parameters (and at one high-likelihood draw of the conditional labels), and compare it with the log-likelihood evaluated with the ground truth parameters and labels $\log(\mathcal{L}(\theta))$.

We report the results of our experiments evaluated with both metrics in Figure 7. The box plot shows that our method has a strong ability to recover the underlying components (SFP and NHPP) based only on the observed timestamps. Larger values of the number of events (n) correspond to smaller values of $\delta(\theta, \hat{\theta})$. Mixtures with higher P_{NHPP} tend to produce fewer bursts and therefore have a more uniform behavior over the whole observed period. Consistently with this, we observe that larger values of P_{NHPP} correspond to smaller values of $\delta(\theta, \hat{\theta})$. The number of transitions ($|\Phi|$) has a lower impact in comparison to the other parameters', though smaller values of $\delta(\theta, \hat{\theta})$ are associated with fewer transitions. The rightmost graph in Figure 7 shows that $\log(\mathcal{L}(\hat{\theta}))$ is systematically close to $\log(\mathcal{L}(\theta))$, and even surpasses it in more than 80% of the cases. Both metrics' behavior jointly indicate that our method can accurately recover the ground truth based only on the timestamps of the observed mixture process.

C EXPLICIT COMPUTATION OF THE INTEGRAL $\mathcal{I}$

There exists an explicit formula for the multiple integral

$$\mathcal{I}(\Omega, t_b, t_e, m_b, m_e) = \int_{\mathcal{T}} e^{\sum_{j=m_b}^{m_e} \left(\lambda_{(j+1)} - \lambda_{(j)}\right) x_{(j-m_b+1)}} dx \tag{8}$$

where $\mathcal{T} = \{x_1, \ldots, x_{m_e-m_b} : t_b \leq x_1 \leq x_2 \leq \cdots \leq x_{m_e-m_b} \leq t_e\}$. For any N and $a = (a_1, \ldots, a_N)$ define $G_a = G_a^N = \int_{\mathcal{T}} e^{\sum_{i=1}^{N} a_i x_i} dx$ where $\mathcal{T} = \{x_1, \ldots, x_N : T_1 \leq x_1 \leq x_2 \leq \cdots \leq x_{m_e-m_b} \leq T_2\}$ (where we omit the dependence on T_1, T_2 for notational simplicity). Thus $\mathcal{I}(\Omega, t_b, t_e, m_b, m_e) = G^{m_e-m_b}_{\left(\lambda_{(j+1)} - \lambda_{(j)}\right)_{j \leq m_e - m_b}}$ and computing G_a for any choice of a is enough.

First, we observe that we have the following recurrence relation

$$\begin{aligned} G^N_{a_1,\ldots,a_N} &= \int_{T_1}^{T_2} \int_{x_1}^{T_2} \cdots \int_{x_{N-1}}^{T_2} e^{\sum_{i=1}^{N} a_i x_i} dx_N \ldots, dx_2 dx_1 \\ &= \int_{T_1}^{T_2} \int_{x_1}^{T_2} \cdots \int_{x_{N-2}}^{T_2} \left[\frac{e^{T_2 a_N} - e^{a_N x_{N-1}}}{a_N} \right] dx_{N-1} \ldots dx_1 \\ &= \frac{e^{T_2 a_N} G^{N-1}_{a_1,\ldots,a_{N-1}}}{a_N} - \frac{e^{T_1 a_N} G^{N-1}_{a_1,\ldots,a_{N-1}+a_N}}{a_N}. \end{aligned} \tag{9}$$

Based on iteratively applying this recurrence relation, we can get the following formula:

$$\begin{aligned} G^N_{a_1,\ldots,a_N} = e^{T_1(a_1+\ldots+a_N)} \Bigg(&\sum_{\delta \in \{0,1\}^N} (-1)^{\sum_{i=1}^{N} \delta_i} \\ &\times \left[\frac{\prod_{\{i:\delta_i=0\}} e^{(T_2-T_1)(a_i + \sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j} \delta_u)}}{\prod_{i=1}^{N}(a_i + \sum_{j=i+1} a_j \prod_{u=i+1}^{j} \delta_u)} \right] \Bigg). \end{aligned} \tag{10}$$

Whilst the iterative computation in question is reasonably straightforward, we reproduce the details here for the reader's convenience.

Proof of Formula (10). The proof is by induction. Note that for $N = 1$, we have indeed $G^N_{a_1} = \frac{e^{T_2 a_1}}{a_1} - \frac{e^{T_1 a_1}}{a_1}$ as expected. Suppose the result holds for $N - 1$ and let us prove it holds for N.

Note that by the scaling of the formula and the definition of G, it is clear that we can restrict ourselves to the case $T_1 = 0$. Now, by

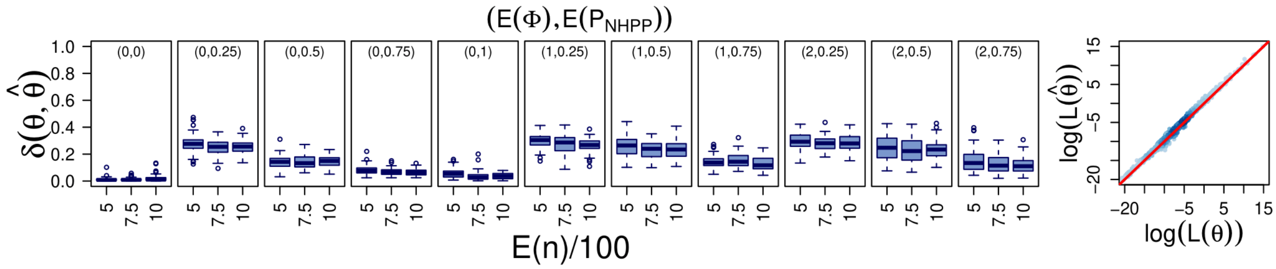

Figure 7: Summary of the results of synthetic data experiments. Left: $\delta(\theta,\hat{\theta})$ distribution grouped by $(\mathbb{E}(|\Phi|),\mathbb{E}(P_{NHPP}))$. The x-axis shows the expected number of observed events. Right: plot of $\log(\mathcal{L}(\theta))\times\log(\mathcal{L}(\hat{\theta}))$ with the $y=x$ line in red.

equation (9), we have $G^N_{a_1,\dots,a_N} =$

$$\begin{aligned}
&\frac{e^{T_2 a_N} G^{N-1}_{a_1,\dots,a_{N-1}}}{a_N} - \frac{G^{N-1}_{a_1,\dots,a_{N-1}+a_N}}{a_N} \qquad (11)\\
&= \frac{e^{T_2 a_N}}{a_N} \sum_{\delta\in\{0,1\}^{N-1}} \Bigg((-1)^{\sum_{i=1}^{N-1}\delta_i} \\
&\quad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg) \\
&\quad - \frac{1}{a_N} \sum_{\delta\in\{0,1\}^{N-1}} \Bigg((-1)^{\sum_{i=1}^{N-1}\delta_i} \\
&\qquad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg),
\end{aligned}$$

where $\tilde{a}_i = a_i$ for $i \le N-2$ and $\tilde{a}_{N-1} = a_{N-1} + a_N$. Now, note that

$$\begin{aligned}
&\sum_{\substack{\delta\in\\ \{0,1\}^{N-1}}} \frac{-(-1)^{\sum_{i=1}^{N-1}\delta_i}}{a_n} \left[\frac{\prod_{i:\delta_i=0} e^{T_2(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)} \right] \qquad (12)\\
&= \frac{1}{a_n} \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}\times\{1\}}} \Bigg((-1)^{\sum_{i=1}^{N}\delta_i} \\
&\qquad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg) \\
&= \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}\times\{1\}}} \Bigg((-1)^{\sum_{i=1}^{N}\delta_i} \\
&\qquad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N}(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg). \qquad (13)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{e^{T_2 a_N}}{a_N} \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}}} (-1)^{\sum_{i=1}^{N-1}\delta_i} \left[\frac{\prod_{\substack{\{i:\\ \delta_i=0\}}} e^{T_2(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \\
&= \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}\times\{0\}}} \frac{(-1)^{\sum_{i=1}^{N}\delta_i}}{a_N} \left[\frac{\prod_{\{i:\delta_i=0\}} e^{T_2(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \\
&= \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}\times\{0\}}} (-1)^{\sum_{i=1}^{N}\delta_i} \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N}(a_i+\sum_{j=i+1}^{N} a_j \prod_{u=i+1}^{j}\delta_u)} \right].
\end{aligned}$$

Plugging Eqs (12) and (**??**) back into Eq. (11), we obtain $G^N_{a_1,\dots,a_N} =$

$$\begin{aligned}
&\frac{e^{T_2 a_N}}{a_N} \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}}} \Bigg((-1)^{\sum_{i=1}^{N-1}\delta_i} \\
&\qquad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(a_i+\sum_{j=i+1}^{N-1} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg) \\
&\quad - \frac{1}{a_N} \sum_{\substack{\delta\in\\ \{0,1\}^{N-1}}} \Bigg((-1)^{\sum_{i=1}^{N-1}\delta_i} \\
&\qquad \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N-1}(\tilde{a}_i+\sum_{j=i+1}^{N-1} \tilde{a}_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg) \\
&= \sum_{\substack{\delta\in\\ \{0,1\}^{N}}} \Bigg((-1)^{\sum_{i=1}^{N}\delta_i} \times \left[\frac{\prod_{i:\delta_i=0} e^{T_2(a_i+\sum_{j=i+1} a_j \prod_{u=i+1}^{j}\delta_u)}}{\prod_{i=1}^{N}(a_i+\sum_{j=i+1} a_j \prod_{u=i+1}^{j}\delta_u)} \right] \Bigg).
\end{aligned}$$

□