Check for updates

Received: 1 November 2023 | Revised: 30 December 2023 | Accepted: 18 January 2024

DOI: 10.1002/sta4.660

ORIGINAL ARTICLE

WILEY

Graph-based mutually exciting point processes for modelling event times in docked bike-sharing systems

Francesco Sanna Passino | Yining Che | Carlos Cardoso Correia Perello

Department of Mathematics, Imperial College London, London, UK

Correspondence
Francesco Sanna Passino, Department of Mathematics, Imperial College London, 180 Queen's Gate, SW7 2AZ, London, UK.
Email: f.sannapassino@imperial.ac.uk

Abstract

This paper introduces graph-based mutually exciting processes (GB-MEP) to model event times in network point processes, focusing on an application to docked bike-sharing systems. GB-MEP incorporates known relationships between nodes in a graph within the intensity function of a node-based multivariate Hawkes process. This approach reduces the number of parameters to a quantity proportional to the number of nodes in the network, resulting in significant advantages for computational scalability when compared with traditional methods. The model is applied on event data observed on the Santander Cycles network in central London, demonstrating that exploiting network-wide information related to geographical location of the stations is beneficial to improve the performance of node-based models for applications in bike-sharing systems. The proposed GB-MEP framework is more generally applicable to any network point process where a distance function between nodes is available, demonstrating wider applicability.

KEYWORDS
bike-sharing systems, dynamic networks, Hawkes processes, mutually exciting processes

1 | INTRODUCTION

In recent years, a number of convenient and sustainable alternatives for short-distance travel have emerged in world cities, with the objective of reducing traffic congestion and promoting healthier and more sustainable urban environments. Bike-sharing systems (BSSs) represent one such solution, which have become increasingly popular especially after the end of restrictions to movement due to the COVID-19 pandemic (Heydari et al., 2021). Two types of BSSs are usually offered in cities: traditional docked systems, which require users to pick up and return bikes at designated stations, and dockless systems, which utilise GPS technology, allowing users to locate and unlock bikes using smartphone applications. From a statistical perspective, the literature about BSSs is mostly focused on two problems: rebalancing, consisting in the redistribution of bikes across stations and locations according to the predicted demand (see, e.g., the survey of Vallez et al., 2021), and demand forecasting (e.g. Rennie et al., 2023). In general, modelling docked BSSs is still an active research problem, extensively studied in the literature (see, e.g., Choi et al., 2022; Deng et al., 2023; Busatto & Stingo, 2023; Paul et al., 2023).

Motivated by the application to docked BSSs, this work proposes a statistical network-informed model for event times, generically called *graph-based mutually exciting process* (GB-MEP), which could be used to predict demand across stations. The proposed methodology is tested on publicly available data from the Santanter Cycles bike-sharing network in central London. In this work, journeys observed on BSSs are interpreted

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC-BY) licence to any *Author Accepted Manuscript* version arising.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Authors. *Stat* published by John Wiley & Sons Ltd.

as a realisation of a network point process, with events having a corresponding duration. In general, demand for bicycles in docked BSSs has often been modelled as a multivariate point process in the literature (see, e.g., Gervini & Khanal, 2018; Okawa et al., 2019; Foschi, 2023), but network information, expressing the interaction between different stations, is often not explicitly incorporated within the existing models. This work proposes a parsimonious representation for the temporal point processes observed at each docking station, directly utilising the information provided by the distance between stations in the network for imposing restrictions on the model parameters.

In particular, this work considers a particular class of point processes, called mutually exciting processes, with a focus on Hawkes processes (see, for example, Hawkes, 1971; Laub et al., 2022). Hawkes processes have been extensively used in the literature for modelling a number of different phenomena, such as earthquakes (Ogata, 1988), crime (Mohler et al., 2011), social media (Zhao et al., 2015; Rizoiu et al., 2017), emails (Fox et al., 2016; Miscouridou et al., 2018; Sanna Passino & Heard, 2023) and computer networks (Price-Williams & Heard, 2020; Shlomovich et al., 2022; Sanna Passino & Heard, 2023). Hawkes processes appear to be a particularly suitable model for a BSS, since they capture periods of increased demand in a data-driven way, avoiding specific parametric assumptions on the process, such as periodic components or weather conditions. This property has been proved particularly powerful in cyber-security, where self-exciting models largely outperform alternatives that model periodic components parametrically (Price-Williams & Heard, 2020).

The remainder of this article is structured as follows: Section 2 gives background on point processes on graphs and Hawkes processes, describing related literature and challenges of existing approaches when applied to network point processes. Next, Section 3 introduces GB-MEPs, the main contribution of this work, discussing the method, parameter estimation and model evaluation procedures. The proposed methodology is then applied on publicly available data from the Santander Cycles BSS in central London (Section 4), followed by a conclusion and discussion in Section 5.

2 | BACKGROUND

Consider a network $\mathbb{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, \ldots, M\}, M \in \mathbb{N}$, is the node set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, such that $(x, y) \in \mathcal{E}$ if nodes x and y are connected. Within the context of BSSs, nodes represent docking stations, and an edge between two stations can be drawn if at least one trip between the stations is observed within an observation period. It is further assumed that event data with a duration are observed on the network, corresponding to a sequence of quadruples $(x_k, y_k, t_k, t'_k), k = 1, 2, \ldots, N, N \in \mathbb{N}$, where $(x_k, y_k) \in \mathcal{E}$ is an edge, $t_k \in \mathbb{R}_+$ is the start time of the event, and $t'_k > t_k$ is the end time of the event. This corresponds to observing events with duration $\delta_k = t'_k - t_k$. In practice, events are ordered by end time, such that $t'_1 < \ldots < t'_N$. In BSSs, the quadruple (x_k, y_k, t_k, t'_k) denotes a bike journey between stations x_k and y_k, started at time t_k and ended at time t'_k. From the observed sequence of quadruples, two left-continuous counting processes $N_i : \mathbb{R}_+ \to \mathbb{N}_0$ and $N'_i : \mathbb{R}_+ \to \mathbb{N}_0$ could be associated with each node:

$$N_i(t) = \sum_{k \geq 1} \mathbb{1}_{[0,t)}(t_k) \mathbb{1}_{\{i\}}(x_k), N'_i(t) = \sum_{k \geq 1} \mathbb{1}_{[0,t)}(t'_k) \mathbb{1}_{\{i\}}(y_k), \tag{1}$$

where $\mathbb{1}_{\cdot}(\cdot)$ denotes the indicator function: $\mathbb{1}_S(x) = 1$ if $x \in S$, and 0 otherwise. In the BSS application, the process $N_i(t)$ counts the number of journeys started from station i before time t, whereas the process $N'_i(t)$ counts the number of journeys ended at station i before time t. For simplicity, event times for each node are often denoted with the notation $t_{i,k}, i \in \mathcal{V}, k \in \mathbb{N}$, representing the start time of the k-th event started from node i and $t'_{i,k}$, corresponding to the ending time of the k-th event ended on node i. This notation can be used to obtain a simpler representation for $N_i(t)$ and $N'_i(t)$:

$$N_i(t) = \sum_{k \geq 1} \mathbb{1}_{[0,t)}(t_{i,k}), N'_i(t) = \sum_{k \geq 1} \mathbb{1}_{[0,t)}(t'_{i,k}). \tag{2}$$

In general, a counting process could be characterised by its conditional intensity function (CIF). The CIF $\lambda_i(t)$ for $N_i(t)$ is defined as follows:

$$\lambda_i(t) = \lim_{dt \to 0} \frac{\mathbb{E}[N_i(t + dt) - N_i(t) | \mathcal{H}(t)]}{dt}, \tag{3}$$

where $\mathcal{H}(t)$ is the history of the process on the entire network up to time t.

The counting processes $\{N_i(t)\}_{i \in \mathcal{V}}$ and $\{N'_i(t)\}_{i \in \mathcal{V}}$ could be considered as realisations of multivariate temporal point processes. For this class of processes, a popular model in the literature is the multivariate Hawkes process (Hawkes, 1971) with scaled exponential kernel (see, e.g., Laub et al., 2022), which models the CIF $\lambda_i(t)$ on each node as a function of all start times of events observed in the entire network before time t:

$$\lambda_i(t) = \lambda_i + \sum_{j=1}^{M} \sum_{t_{j,k} < t} \alpha_{ij} \exp\{-\beta_{ij}(t - t_{j,k})\}, \tag{4}$$

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\lambda_i \in \mathbb{R}+$ is a baseline intensity, $\alpha_{ij} \in \mathbb{R}_+$ is the jump in the CIF of node i caused by an event on node j and $\beta_{ij} \in \mathbb{R}_+$, with $\alpha_{ij} < \beta_{ij}$, is the corresponding decay rate. The end times of events could also be considered in the intensity in Equation (4), obtaining the following self-and-mutually exciting Hawkes process CIF (see, e.g., Laub et al., 2022):

$$\lambda_i(t) = \lambda_i + \sum_{j=1}^{M} \left[\sum_{t_{j,k} < t} \alpha_{ij} \exp\{-\beta_{ij}(t - t_{j,k})\} + \sum_{t'_{j,h} < t} \alpha'_{ij} \exp\{-\beta'_{ij}(t - t'_{j,h})\} \right], \tag{5}$$

where $\alpha'_{ij}, \beta'_{ij} \in \mathbb{R}_+, \alpha'_{ij} < \beta'_{ij}$. The models in Equations 4 and (5) have a number of issues when applied to temporal point processes on networks: First, the number of parameters is $\mathcal{O}(M^2)$ across the entire network, implying that it grows quadratically with the number of nodes, which is unfeasible for estimation in practice, especially when the network becomes large; second, the network structure is not explicitly incorporated within the intensity function in Equations 4 and (5); third, information from additional processes, such as end times of events, is not included within the CIF in Equation (4).

In the application to BSSs, these issues appear to be particularly relevant: The number of docking station could be very large (approximately 800 in the Santander Cycles BSS in central London), making estimation of the parameters α_{ij} and β_{ij} a high-dimensional problem. Using the example of $M = 800$, the total number of parameters required for model in Equation (4) for the entire network would be $M + 2M^2 = 1{,}280{,}800$. This number is more than the average number of bike journeys per month in the busiest year in the Santander Cycles BSS (over 950,000 hires per month in 2022; see Transport for London, 2022). Additionally, another drawback of the multivariate Hawkes process models in Equations (4) and (5) is that information about *known*, *pre-existing* network structure is not incorporated within the CIF: for modelling network point processes in real-world applications, information about the nodes in the BSS could be very important. For example, in a BSS, stations which are geographically close are expected to have similar characteristics compared with stations which are further apart, suggesting that geographical information about nodes should be taken into account within the model.

It must be remarked that the model in Equation (4) has been often used successfully in the literature for modelling multivariate point processes, including spatio-temporal processes (see, e.g., the survey of Reinhart, 2018). In particular, Hawkes processes are in general suitable to *discover* existing network structure and causal relationships between nodes (see, e.g., Linderman & Adams, 2014; Etesami et al., 2016; Xu et al., 2016; Eichler et al., 2017; Yuan et al., 2019), but known characteristics about the network and nodes are not explicitly incorporated within the model. Using the approach of Linderman and Adams (2014), the network adjacency matrix, if known, could be used to set the coefficients α_{ij} to zero where $(i,j) \notin \mathcal{E}$, but in general, the use of node covariates and known network information has been overlooked in the literature. Also, it must be remarked that Hawkes process models on networks are usually applied at two different levels of resolution: node-based (see, e.g., Fox et al., 2016; Price-Williams & Heard, 2020), or edge-based (see, e.g., Blundell et al., 2012; Perry & Wolfe, 2013; Miscouridou et al., 2018; Huang et al., 2022; Sanna Passino & Heard, 2023).

This work mostly concerns with node-based models, proposing an extension of the standard multivariate Hawkes process, exploiting existing network structure between the nodes to reduce the number of parameters to estimate to $\mathcal{O}(M)$, providing advantages for scalability to large networks. Additionally, information from end times is added to the CIF, providing a more complete modelling framework for event data on networks with an associated event duration.

3 | GRAPH-BASED MUTUALLY EXCITING PROCESSES

The main objective of this work is to propose a GB-MEP for the node-specific CIFs $\lambda_i(t)$ for the temporal point processes of start times of connection events on a graph, applied to journeys observed on BSSs. The GB-MEP intensity function is assumed to take the following generic form:

$$\lambda_i(t) = \lambda_i + \sum_{j=1}^{M} \left\{ \sum_{k=1}^{N_j(t)} \omega_i\{\gamma(i,j), t - t_{j,k}\} + \sum_{h=1}^{N'_j(t)} \omega'_i\{\gamma(i,j), t - t'_{j,h}\} \right\}, \tag{6}$$

where $\lambda_i \in \mathbb{R}_+$ is a baseline rate, $\omega_i, \omega'_i : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ are spatio-temporal excitation kernel functions and $\gamma : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_+$ is a distance function between nodes. The distance function γ quantifies how close in space two nodes are. For example, in BSSs, the distance function could be taken to be the haversine metric, which provides an *'as-the-crow-flies'* measure of distance between the geographic location of two nodes. Given the latitude and longitude (φ_x, ℓ_x) and (φ_y, ℓ_y) of the bike stations x and y (in radians), the haversine distance between the stations is as follows:

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\gamma(x,y) = 2\rho \arcsin\left\{\sqrt{\sin^2\left(\frac{\varphi_y - \varphi_x}{2}\right) + \cos(\varphi_x)\cos(\varphi_y)\sin^2\left(\frac{\ell_y - \ell_x}{2}\right)}\right\}, \tag{7}$$

where ρ is the Earth radius. Note that, for more general point processes on networks where information about the node locations is not present, the distance $\gamma(x,y)$ between two nodes $x,y \in \mathcal{V}$ could depend on graph-related statistics, such as the length of the shortest path between the nodes or the number of common neighbours. In this case, γ takes the interpretation of a *dissimilarity metric*.

In this work, it is further assumed that the spatial and temporal components of the excitation kernels ω_i are separable, with the temporal element of the kernel taking a scaled exponential form. This results in the following excitation kernel:

$$\omega_i(\gamma^*, t) = \kappa_i(\gamma^*) \cdot \alpha_i \exp(-\beta_i t), \tag{8}$$

where $\alpha_i, \beta_i \in \mathbb{R}_+, \alpha_i < \beta_i$, and γ^* correspond to a realised value of the distance function γ. The function $\kappa_i : \mathbb{R}_+ \to (0,1]$ rescales the node-specific temporal mutual excitation given by $\alpha_i \exp(-\beta_i t)$, using the distance γ between the nodes. In the application to BSSs, the excitation effect of an event occurring on a node with small distance from station i is likely to have a larger effect on the intensity $\lambda_i(t)$ than an event occurring on a node located further away. Therefore, κ_i could be set to a monotonically decreasing function such as $\kappa(0) = 1$ and $\lim_{\gamma^* \to \infty} \kappa_i(\gamma^*) = 0$. A possible choice, used in this work, is the truncated exponential function $\kappa_i(\gamma^*) = \exp(-\theta_i \gamma^*) \mathbb{1}_{[0,\varepsilon]}(\gamma^*)$, where $\theta_i \in \mathbb{R}_+$ is a decay parameter and $\varepsilon \in \mathbb{R}_+$ is a *fixed constant* defining the radius of neighbouring nodes considered to have nonzero effect on the CIF $\lambda_i(t)$. Note that setting $\varepsilon \to 0$ would result in $\kappa_i(\gamma^*) = \mathbb{1}_{\{0\}}(\gamma^*)$, which would only give weight to events involving station i, removing the spatial mutual excitation component and making the process exclusively self-exciting. A similar spatio-temporal excitation kernel could also be postulated for $\omega_i'(\gamma^*, t)$:

$$\omega_i'(\gamma^*, t) = \exp(-\theta_i' \gamma^*) \mathbb{1}_{[0,\varepsilon]}\{\gamma^*\} \cdot \alpha_i' \exp(-\beta_i' t), \tag{9}$$

with $\alpha_i', \beta_i', \theta_i' \in \mathbb{R}_+$, and $\alpha_i' < \beta_i'$. Overall, this gives the following extended form for the proposed GB-MEP CIF in Equation (6) for applications to BSSs:

$$\lambda_i(t) = \lambda_i + \sum_{j=1}^{M} \left\{ \sum_{k=1}^{N_j(t)} \exp\{-\theta_i \gamma(i,j)\} \mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\} \cdot \alpha_i \exp\{-\beta_i (t - t_{j,k})\} + \sum_{h=1}^{N_j'(t)} \exp\{-\theta_i' \gamma(i,j)\} \mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\} \cdot \alpha_i' \exp\{-\beta_i' (t - t_{j,h}')\} \right\}. \tag{10}$$

An example of the intensity functions arising under a GB-MEP process is given in Figure 1.

3.1 | Parameter estimation

The model parameters $\boldsymbol{\phi}_i = (\lambda_i, \alpha_i, \beta_i, \theta_i, \alpha_i', \beta_i', \theta_i')$ for each node-based process could be learned via maximum likelihood estimation, by maximising the node-specific log-likelihood function separately for each station. The log-likelihood function for a temporal point process takes the following form (see, e.g., Daley & Vere-Jones, 2002):

$$\ell_i(\boldsymbol{\phi}_i) = \sum_{k=1}^{N_i(T)} \log[\lambda_i(t_{i,k})] - \Lambda_i(T), \tag{11}$$

where $\Lambda_i : \mathbb{R}_+ \to \mathbb{R}_+$ is the compensator, defined as

$$\Lambda_i(t) = \int_0^t \lambda_i(s) \mathrm{d}s. \tag{12}$$

For the CIF in Equation (10), the compensator takes the following form:

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

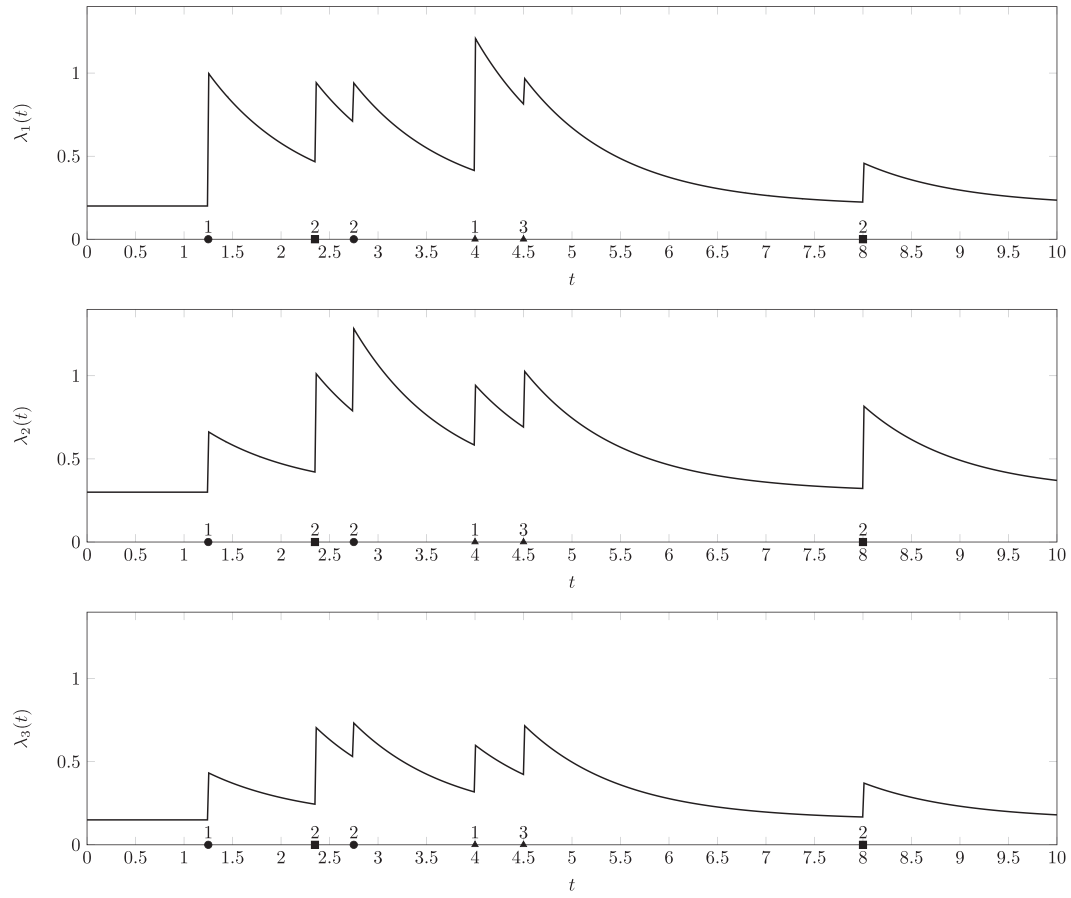


FIGURE 1 Example of conditional intensity functions (CIFs) for a graph-based mutually exciting process (GB-MEP) with $\mathcal{V}=\{1,2,3\}$ and three realised events expressed as quadruples $(x,y,t,t'), x,y\in\mathcal{V}, t,t'\in\mathbb{R}_+, t<t'$. Events and their symbols: $(1,2,1.25,2.75)$ correspond to ●; $(1,3,4,4.5)$ to ▲ and $(2,2,2.35,8)$ to ■. Parameters:
$\lambda_1=0.2,\lambda_2=0.3,\lambda_3=0.15,\alpha_1=0.8,\alpha_2=\alpha_3=0.6,\beta_1=\beta_2=\beta_3=1,\theta_1=\theta_2=\theta_3=1,\alpha'_1=\alpha'_2=0.5,\alpha'_3=0.3,\beta'_1=\beta'_2=\beta'_3=1,\theta'_1=\theta'_2=\theta'_3=1.5$.
Node distances: $\gamma(1,2)=\gamma(2,1)=0.5,\gamma(1,3)=\gamma(3,1)=0.75,\gamma(2,3)=\gamma(3,2)=0.25,\gamma(1,1)=\gamma(2,2)=\gamma(3,3)=0$

$$\Lambda_i(t)=\lambda_i t-\sum_{j=1}^{M}\left\{\frac{\exp\{-\theta_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}\cdot\alpha_i}{\beta_i}\sum_{k=1}^{N_j(t)}[\exp\{-\beta_i(t-t_{j,k})\}-1]-\frac{\exp\{-\theta'_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}\cdot\alpha'_i}{\beta'_i}\sum_{h=1}^{N'_i(t)}[\exp\{-\beta'_i(t-t'_{j,h})\}-1]\right\}. \tag{13}$$

Note that, because of the double summations in the CIF in Equation (10), the computational complexity for calculating the term $\sum_{k=1}^{N_i(T)}\log[\lambda_i(t_{i,k})]$ in the log-likelihood in Equation (11) is $\mathcal{O}\left\{N_i(T)\sum_{j=1}^{M}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}[N_j(T)+N'_j(T)]\right\}$, proportional to the sum of the number of events occurring on each node, multiplied by the total number of events started from the station under consideration. In practice, this makes model fitting particularly challenging and almost unfeasible for large datasets.

This issue is largely alleviated by the scaled exponential excitation kernel, which admits a computationally cheaper form of the log-likelihood, written in recursive form using the following expression for the CIF in Equation (10):

$$\lambda_i(t_{i,k})=\lambda_i+\sum_{j=1}^{M}\left[\exp\{-\theta_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}\cdot\alpha_i A_{ij}(k)+\exp\{-\theta'_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}\cdot\alpha'_i A'_{ij}(k)\right]. \tag{14}$$

In the above expression, $A_{ij}(k)$ and $A'_{ij}(k)$ are defined as follows:

$$A_{ij}(k)=\sum_{l=1}^{N_j(t_{i,k})}\exp\{-\beta_i(t_{i,k}-t_{j,l})\}, A'_{ij}(k)=\sum_{l=1}^{N'_j(t_{i,k})}\exp\{-\beta'_i(t_{i,k}-t'_{j,l})\}. \tag{15}$$

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The initial values $A_{ij}(1)$ and $A'_{ij}(1)$ are simply as follows:

$$A_{ij}(1) = \sum_{l=1}^{N_j(t_{i,1})} \exp\{-\beta_i(t_{i,1} - t_{j,l})\}, A'_{ij}(1) = \sum_{l=1}^{N'_j(t_{i,1})} \exp\{-\beta'_i(t_{i,1} - t'_{j,l})\}. \tag{16}$$

The recursive form of the log-likelihood stems from the following proposition.

Proposition 1. The expressions for $A_{ij}(k)$ and $A'_{ij}(k)$ in Equation (15) can be rewritten recursively as follows:

$$A_{ij}(k) = \exp\{-\beta_i(t_{i,k} - t_{i,k-1})\} \cdot A_{ij}(k-1) + \sum_{l=N_j(t_{i,k-1})+1}^{N_j(t_{i,k})} \exp\{-\beta_i(t_{i,k} - t_{j,l})\}, A'_{ij}(k) = \exp\{-\beta'_i(t_{i,k} - t_{i,k-1})\} \cdot A'_{ij}(k-1) + \sum_{l=N'_j(t_{i,k-1})+1}^{N'_j(t_{i,k})} \exp\{-\beta'_i(t_{i,k} - t'_{j,l})\}. \tag{17}$$

Proof. The result is proved for $A_{ij}(k)$ here; the proof for $A'_{ij}(k)$ is analogous. The summation for $A_{ij}(k)$ in Equation (15) could be broken up at $N_j(t_{i,k-1})$ as follows:

$$A_{ij}(k) = \sum_{l=1}^{N_j(t_{i,k-1})} \exp\{-\beta_i(t_{i,k} - t_{j,l})\} + \sum_{l=N_j(t_{i,k-1})+1}^{N_j(t_{i,k})} \exp\{-\beta_i(t_{i,k} - t_{j,l})\}. \tag{18}$$

Premultiplying the first element of Equation (18) by $\exp\{-\beta_i(t_{i,k-1}) - \beta_i(t_{i,k-1})\}$ gives the following:

$$\sum_{l=1}^{N_j(t_{i,k-1})} \exp\{-\beta_i(t_{i,k} - t_{j,l})\} = \exp\{-\beta_i(t_{i,k-1}) - \beta_i(t_{i,k-1})\} \sum_{l=1}^{N_j(t_{i,k-1})} \exp\{-\beta_i(t_{i,k} - t_{j,l})\} = \exp\{-\beta_i(t_{i,k}) - \beta_i(t_{i,k-1})\} \sum_{l=1}^{N_j(t_{i,k-1})} \exp\{-\beta_i(t_{i,k-1} - t_{j,l})\}. \tag{19}$$

By the definition in Equation (15), $\sum_{l=1}^{N_j(t_{i,k-1})} \exp\{-\beta_i(t_{i,k-1} - t_{j,l})\} = A_{ij}(k-1)$, which gives the result in Equation (17). □ ■

It follows that writing $B_{ij}(k) = \exp\{-\theta_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\} \cdot \alpha_i A_{ij}(k) + \exp\{-\theta'_i\gamma(i,j)\}\mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\} \cdot \alpha'_i A'_{ij}(k)$; Proposition 1 gives a recursive form for the log-likelihood in Equation (11):

$$\ell_i(\boldsymbol{\phi}_i) = \sum_{k=1}^{N_i(T)} \log\left\{\lambda_i + \sum_{j=1}^{M} B_{ij}(k)\right\} - \Lambda_i(T). \tag{20}$$

Under this recursive representation, $\ell_i(\boldsymbol{\phi}_i)$ can be evaluated in $\mathcal{O}\{N_i(T)M_i^\varepsilon\}$, where $M_i^\varepsilon = \sum_{j=1}^{M} \mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\}$ is the number of nodes within distance ε of node i. Note that this assumes that the update of $B_{ij}(k)$ given $B_{ij}(k-1)$ is $\mathcal{O}(M_i^\varepsilon)$.

The node-specific log-likelihoods $\ell_i(\boldsymbol{\phi}_i), i = 1,\ldots,M$ can be optimised independently and in parallel via standard statistical optimisation techniques. For example, the node-based expectation–maximisation (EM) algorithm described in Fox et al. (2016) can be adapted to GB-MEP utilising the same latent variables and reparametrisation discussed in Fox et al. (2016) and Sanna Passino and Heard (2023). The main drawback of optimising the log-likelihood via the EM algorithm consists in its computation cost: The algorithm augments the parameter space with latent variables used to reconstruct the branching structure associated with the process (see, e.g., Section 3.3 in Laub et al. Laub et al., 2022). Therefore, it is computationally more convenient to use gradient-based methods for optimisation, leveraging the recursive structure of the log-likelihood stemming from Proposition 1. In particular, this work uses the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm for optimisation (see, e.g., Zhu et al., 1997), a quasi-Newton method which approximates the inverse of the Hessian matrix using gradients. Using the form of the log-likelihood in Equation (20), recursive expressions for the gradients of $\ell_i(\boldsymbol{\phi}_i)$ can be obtained with respect to all parameters in $\boldsymbol{\phi}_i$, implying that estimation via L-BFGS is computationally feasible even for large datasets.

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2 | Model evaluation

The quality of the model fit could be assessed by evaluating the distribution of p-values calculated on a training and a test set observation periods, using the same methodology discussed in Sanna Passino and Heard (2023). By the time rescaling theorem (see, e.g., Brown et al., 2002), the transformed event times $\Lambda_i(t_{i,k}), k = 1,2,\ldots$ are event times of a homogeneous Poisson process with unit rate, under the null hypothesis of correct specification of the conditional intensity and its parameters. Therefore, under the same null hypothesis, the upper tail p-values calculated from the observed events should follow a uniform distribution in $(0,1)$. Since the interarrival times of a unit rate Poisson process are exponentially distributed with unit rate, the corresponding upper tail p-values for the observed events are as follows:

$$p_{i,k} = \exp\left\{-\int_{t_{i,k-1}}^{t_{i,k}} \lambda_i(s)\mathrm{d}s\right\} = \exp\{-\Lambda_i(t_{i,k}) + \Lambda_i(t_{i,k-1})\}, i = 1,\ldots,M, k = 1,\ldots,N_i(T), \tag{21}$$

assuming $\Lambda_i(t_{i,0}) = 0$ for all $i = 1,\ldots,M$. Using the recursive expressions in Proposition 3.1 and the counting processes in Equation (2), the expression in Equation (21) reduces to the following:

$$\begin{aligned} p_{i,k} = \exp\Bigg\{ & -\lambda_i(t_{i,k} - t_{i,k-1}) + \sum_{j=1}^{M} \frac{\exp\{-\theta_i\gamma(i,j)\}\mathbb{1}_{[0,\epsilon]}\{\gamma(i,j)\}\cdot\alpha_i}{\beta_i}[A_{ij}(k) - A_{ij}(k-1) - N_j(t_{i,k}) + N_j(t_{i,k-1})] \\ & + \sum_{j=1}^{M} \frac{\exp\{-\theta_i'\gamma(i,j)\}\mathbb{1}_{[0,\epsilon]}\{\gamma(i,j)\}\cdot\alpha_i'}{\beta_i'}[A'_{ij}(k) - A'_{ij}(k-1) - N'_j(t_{i,k}) + N'_j(t_{i,k-1})]\Bigg\}. \end{aligned} \tag{22}$$

Considering an observation period $[0,T]$, the model parameters could first be estimated from a testing period $[0,T^*], T^* < T$, using the procedure described in Section 3.1. The p-values can then be calculated directly by plugging in the estimates into Equation (21). The estimates could also be used to evaluate the performance of the model in a testing period $(T^*,T]$, by calculating the p-values using the same methodology. In both cases, under a correct model specification and parameter estimation, the distributions of the p-value on the training and test set events should be approximately uniform in $(0,1)$. The quality of the fit to the uniform distribution could be evaluated using the Kolmogorov–Smirnov (KS) statistic, denoted here KS_i.

For the sequence of node-specific p-values $p_{i,k}, k = 1,\ldots,N_i(T)$ on the training set, KS_i is defined as follows:

$$\mathrm{KS}_i = \sup_{p \in [0,1]} \left| p - \frac{1}{N_i(T)} \sum_{k=1}^{N_i(T)} \mathbb{1}_{[0,p)}(p_{i,k}) \right|. \tag{23}$$

4 | APPLICATION TO THE SANTANDER CYCLES BIKE-SHARING SYSTEM

The proposed models were tested on data from the Santander Cycles bike-sharing network in central London. Transport for London publicly released data about journeys on the Santander Cycles network, publicly available online (https://cycling.data.tfl.gov.uk/) (powered by TfL). Each record includes information about the start and end times, IDs of source and destination stations, journey duration, a bike ID number and an ID number for the rental. For the analysis in this section, 32 weeks of data were considered, starting from 2 March 2022 until 9 October 2022. A total of 4, 098, 756 journeys recorded between 2 March 2022 and 21 June 2022 (approximately 16 weeks) were used to construct the training set, whereas the remaining 3, 995, 591 journeys occurred between 22 June 2022 and 9 October 2022 (approximately 16 weeks) formed the test set. A total of $M = 787$ unique active docking stations were observed in the data under consideration. Distances between docking stations were calculated using the haversine distance formula in Equation (7), with the value $\rho = 6365.079$km (corresponding to the Earth radius at the latitude of London, with no altitude). A more precise estimate of the cycling distance between two nodes could be obtained via calls to the TfL Journey Planner API or any other tool providing calculations of cycling distances between locations. The main practical limitation for this approach is that the calculation of the full distance matrix would require $\binom{M}{2}$ API requests, which normally largely exceeds the number of permitted requests under a basic plan.

The proposed GB-MEP model in Equation (3) has been compared with a number of alternatives:

- A standard Poisson process model for each station (Poisson): $\lambda_i(t) = \lambda_i$,

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- A mutually exciting model for each station (MEP): $\lambda_i(t) = \lambda_i + \sum_{h=1}^{N_i'(t)} \alpha_i' \exp\{-\beta_i'(t - t_{i,h}')\}$,
- A self-exciting Hawkes process model for each station (SEP): $\lambda_i(t) = \lambda_i + \sum_{k=1}^{N_i(t)} \alpha_i \exp\{-\beta_i'(t - t_{i,k})\}$,
- A self-and-mutually exciting process for each station (SMEP): $\lambda_i(t) = \lambda_i + \sum_{k=1}^{N_i(t)} \alpha_i \exp\{-\beta_i'(t - t_{i,k})\} + \sum_{h=1}^{N_i'(t)} \alpha_i' \exp\{-\beta_i'(t - t_{i,h}')\}$,
- A spatial mutually exciting process for each station (SpMEP): $\lambda_i(t) = \lambda_i + \sum_{j=1}^{M} \sum_{k=1}^{N_j(t)} \exp\{-\theta_i \gamma(i,j)\} \mathbb{1}_{[0,\varepsilon]}\{\gamma(i,j)\} \cdot \alpha_i \exp\{-\beta_i(t - t_{j,k})\}$.

Note that all models above can be written as special cases of GB-MEP for specific choices of the functions κ and κ', *cf.* Equation (8). The value of ε in SpMEP and GB-MEP has been set to $\varepsilon = 0.5$. For some stations, $M_i^\varepsilon < 3$ for $\varepsilon = 0.5$; in these cases, ε was set to the minimum value such that $M_i^\varepsilon \geq 3$. All model parameters for GB-MEP and its competitors were estimated via maximum likelihood, using numerical optimisation via L-BFGS. Note that the Poisson model has a closed form solution: $\hat{\lambda}_i = N_i(T^*)/T^*$ for a training period $[0, T^*)$. The parameters of SEP and MEP were initialised to $(\lambda_i, \alpha_i, \beta_i) = (e^{-4}, e^{-4}, 2e^{-4})$, and the corresponding results were used to initialise SMEP. In turn, the results obtained from SMEP were used to initialise SpMEP and GB-MEP.

The results are reported in Table 1 and Figures 2 and 3. Table 1 shows the KS score for the distribution of *p*-values across *all M stations*, calculated against the uniform distribution. Figure 2 reports uniform quantile–quantile (Q–Q) plots for all stations individually and for all stations combined. Figure 3 contains the boxplots of KS scores for all six models considered in this section. The results show that GB-MEP has the best performance across all models, both on the training and test set, demonstrating that taking into account the geographical location of nodes within the network is important to improve the goodness-of-fit of statistical models to real-world network point process data. GB-MEP is the only model which achieves a median KS score below 0.05 across stations, both in the training and test sets (cf. Figure 3). Additionally, Table 1 shows remarkable agreement between the results on the training and test sets, which suggests that the models are not overfitting the data.

For most stations, the two best performing models are SMEP and GB-MEP, as observed in Table 1 and Figures 2 and 3. GB-MEP has a better performance than SMEP on approximately 73.88% of stations in the test set. Figure 4 analyses more closely the differences between the results of GB-MEP and SMEP, showing all docking stations in London, coloured by the difference in KS scores on the test set, and a histogram of differences between KS scores for SMEP and GB-MEP on the test set.

The results demonstrate that GB-MEP is an appropriate choice for modelling BSSs. The spatial component in GB-MEP implies that events are not only influenced by past occurrences, but they also depend on the spatial proximity to other locations. This means that high demand for bikes

TABLE 1 Kolmogorov-Smirnov scores for the distribution of *p*-values for all stations on training and test sets for six different models.

	Poisson	MEP	SEP	SMEP	SpMEP	GB-MEP
Training set	0.2186	0.1610	0.0751	0.0326	0.0734	0.0246
Test set	0.2217	0.1584	0.0752	0.0350	0.0738	0.0264

Abbreviations: GB-MEP, graph-based mutually exciting process; MEP, mutually exciting process; SEP, self-exciting process; SMEP, self-and-mutually exciting process; SpMEP, spatial mutually exciting process.

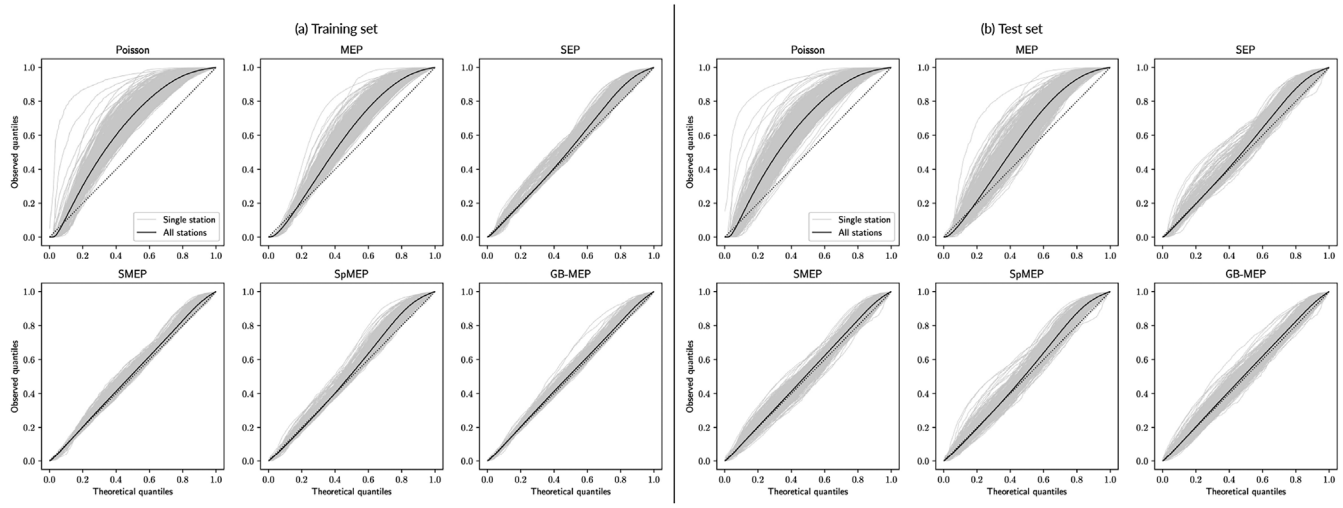


FIGURE 2 Uniform quantile–quantile (Q–Q) plots of *p*-values, calculated separately for each station on the training and test sets for six different models.

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

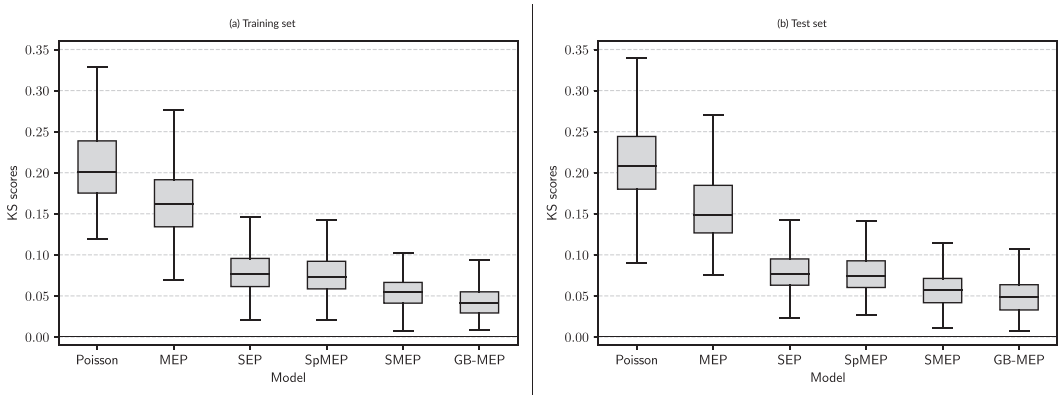


FIGURE 3 Boxplots of Kolmogorov–Smirnov (KS) scores across all station, calculated on the training and test sets for six different models.

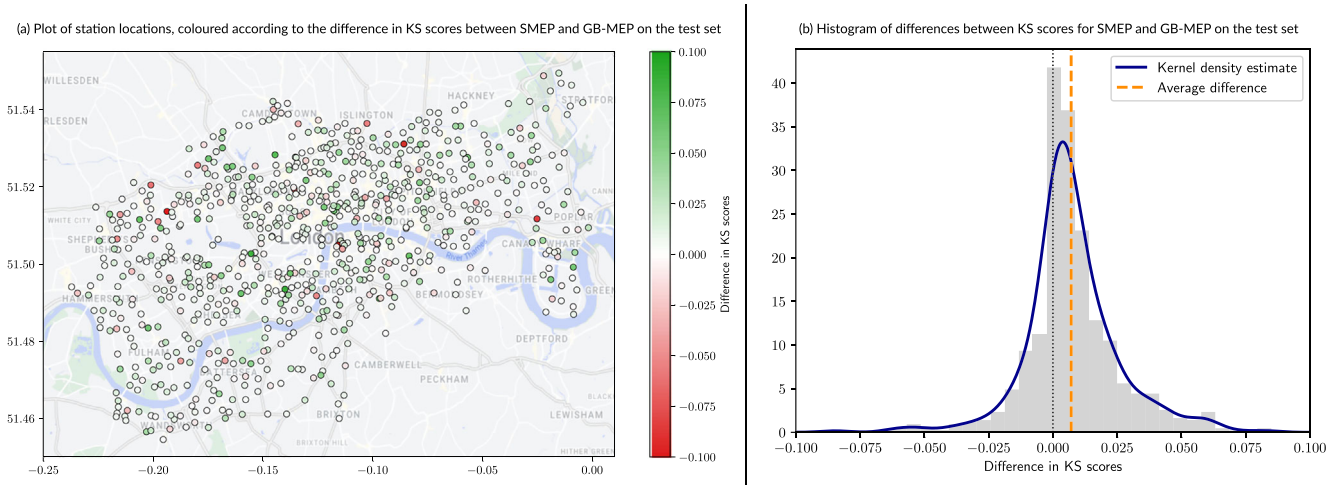


FIGURE 4 Station locations and histogram with information about differences in Kolmogorov–Smirnov (KS) scores between self-and-mutually exciting process (SMEP) and graph-based mutually exciting process (GB-MEP) on the test set.

in one station can influence neighbouring stations. For example, all docking stations around railway stations in central London periodically witness increased demand from commuters seeking transportation to their workplaces during morning rush hour. As one moves further away from the railway stations, the demand for bikes diminishes, mainly due to decreased proximity and preferences of commuters. This behaviour is easily captured by GB-MEP. In addition, GB-MEP can also capture increased demand due to nonperiodic events. For example, consider a sporting event or a concert in central London, which often result in concentrated demands for transportation. As fans leave the stadium or concert hall, there would likely be a surge in demand for bikes at the docking stations close to the event location. Nearby stations, especially those within a reasonable walking distance from the stadium or concert hall, would likely experience increased demand as well. The spatial component in GB-MEP explicitly accounts for the spatial proximity of bike stations, resulting in a particularly suitable model for situations when one station experiences a high demand for bikes due to a specific event or circumstance, triggering a chain reaction affecting nearby stations as people move around the area.

The results also show an effect of the end times on the intensity of start times of journeys, which may seem counterintuitive, but it is consistent with some characteristics of the Santander Cycles BSS: For example, if a station is empty, a user returning a bike to that station may imply that there is a higher chance of a journey starting immediately from that station, with a different user picking up the newly available bike. Also, until September 2022, users were charged £2 for unlimited journeys with Santander Cycles in the subsequent 24 hour. For journeys longer than 30 minutes, users are additionally charged. Therefore, many users typically returned bikes before the 30 minutes mark, and they would pick up another bike immediately for a new journey, meaning that returning a bike to a station could increase the likelihood of a new journey starting from the same station.

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | CONCLUSION

This article discussed a general strategy to incorporate measures of distance between nodes on a graph within the context of multivariate Hawkes processes on graphs, proposing GB-MEP, a graph-based mutually exciting point process. Scalable inference procedures were discussed, using a recursive form of the likelihood function under a scaled exponential form of the excitation kernel. The model was fitted on a large dataset of journeys within the Santander Cycles BSS, with over 4 million events in the training set, and close to $M = 800$ nodes. The model performance was evaluated on close to 4 million events comprising the test set, demonstrating superior performance of GB-MEP over alternative approaches.

A possible limitation of the proposed framework for the application to BSSs is that a bound on the capacity of each station is not included within the model: Each station has a limited number of docks, which could be used to correct the CIF when a station is at capacity. Also, additional distance functions could be considered for the excitation component. This work is framed within the context of docked BSSs, but it could be more generally applied to mutually exciting processes on any type of network where a distance metric between nodes is available. If an adjacency matrix is observed, and not a distance matrix, measures of similarity between nodes could be used for γ within the GB-MEP framework (e.g. the length of the shortest path between nodes).

A possible extension of the version of GB-MEP presented in this work would be to consider different excitation kernels, such as power laws (see, e.g., Laub et al., 2022). Furthermore, removing the separability assumption on the excitation kernel could also prove interesting, as it would allow to capture more complex patterns arising from the interaction between the spatial and temporal components. Finally, adding a probabilistic mixture model component in the bike-sharing application could further improve the model performance, reflecting the fact that bikes could be used for different purposes, such as leisure or commuting.

AUTHOR CONTRIBUTIONS

Francesco Sanna Passino: Methodology; software; formal analysis; writing—original draft. **Yining Che**: Methodology; software; formal analysis; writing—review and editing. **Carlos Cardoso Correia Perello**: Methodology; software; writing—review and editing.

CODE

A *Python* library with code and examples to reproduce all results in this article can be found in the GitHub repository fraspass/gb-mep. The README file in the repository provides detailed instructions on how to download the Santander Cycles data, preprocess them and run the experiments.

ORCID

Francesco Sanna Passino https://orcid.org/0000-0002-4571-6681

REFERENCES

Blundell, C., Beck, J., & Heller, K. A. (2012). Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems 25* (Pereira, F., Burges, C. J. C., Bottou, L., & Weinberger, K. Q., Eds.), Curran Associates, pp. 2600–2608.

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, *14*(2), 325–346.

Busatto, C., & Stingo, F. C. (2023). Inference of multiple high-dimensional networks with the graphical horseshoe prior. arXiv e-prints.

Choi, Y., Cho, H., & Son, H. (2022). Capturing usage patterns in bike sharing system via multilayer network fused Lasso. arXiv e-prints.

Daley, D. J., & Vere-Jones, D. (2002). *An introduction to the theory of point processes—volume I: elementary theory and methods*, Probability and Its Applications: Springer.

Deng, P., Zhao, Y., Liu, J., Jia, X., & Wang, M. (2023). Spatio-temporal neural structural causal models for bike flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(4), 4242–4249.

Eichler, M., Dahlhaus, R., & Dueck, J. (2017). Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, *38*(2), 225–242.

Etesami, J., Kiyavash, N., Zhang, K., & Singhal, K. (2016). Learning network of multivariate Hawkes processes: a time series approach. In *Proceedings of the thirty-second conference on uncertainty in artificial intelligence*, UAI'16, AUAI Press, pp. 162–171.

Foschi, R. (2023). A point processes approach to bicycle sharing systems' design and management. *Socio-Economic Planning Sciences*, *87*, 101608.

Fox, E. W., Short, M. B., Schoenberg, F. P., Coronges, K. D., & Bertozzi, A. L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, *111*(514), 564–584.

Gervini, D., & Khanal, M. (2018). Exploring patterns of demand in bike sharing systems via replicated point process models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *68*(3), 585–602.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.

Heydari, S., Konstantinoudis, G., & Behsoodi, A. W. (2021). Effect of the COVID-19 pandemic on bike-sharing demand and hire time: evidence from Santander Cycles in London. *PLOS ONE*, *16*(12), 1–16.

Huang, Z., Soliman, H., Paul, S., & Xu, K. S. (2022). A mutually exciting latent space Hawkes process model for continuous-time networks. In *Proceedings of the thirty-eighth conference on uncertainty in artificial intelligence* (Cussens, J., & Zhang, K., Eds.), Proceedings of Machine Learning Research, *180*, PMLR, pp. 863–873.

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Laub, P. J., Lee, Y., & Taimre, T. (2022). *The elements of Hawkes processes*: Springer International Publishing.
Linderman, S. W., & Adams, R. P. (2014). Discovering latent network structure in point process data. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, pp. II–1413–II–1421.
Miscouridou, X., Caron, F., & Teh, Y. W. (2018). Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 2349–2358.
Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.
Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27.
Okawa, M., Tanaka, Y., Kurashima, T., Toda, H., & Yamada, T. (2019). Marked temporal point processes for trip demand prediction in bike sharing systems. *IEICE Transactions on Information and Systems*, *E102.D*(9), 1635–1643.
Paul, A., Flynn, K., & Overney, C. (2023). Estimating censored spatial-temporal demand with applications to shared micromobility. arXiv e-prints.
Perry, P. O., & Wolfe, P. J. (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 821–849.
Price-Williams, M., & Heard, N. A. (2020). Nonparametric self-exciting models for computer network traffic. *Statistics and Computing*, *30*(2), 209–220.
Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, *33*(3), 299–318.
Rennie, N., Cleophas, C., Sykulski, A. M., & Dost, F. (2023). Analysing and visualising bike-sharing demand with outliers. *Discover Data*, *1*(1), 1.
Rizoiu, M.-A., Xie, L., Sanner, S., Cebrian, M., Yu, H., & Van Hentenryck, P. (2017). Expecting to be HIP: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pp. 735–744.
Sanna Passino, F., & Heard, N. A. (2023). Mutually exciting point process graphs for modelling dynamic networks. *Journal of Computational and Graphical Statistics*, *32*(1), 116–130.
Shlomovich, L., Cohen, E. A. K., Adams, N., & Patel, L. (2022). Parameter estimation of binned Hawkes processes. *Journal of Computational and Graphical Statistics*, *31*(4), 990–1000.
Transport for London (2022). London's iconic Santander Cycles hire scheme sees yet another record-breaking year in 2022. https://bit.ly/3PPUwbg
Vallez, C. M., Castro, M., & Contreras, D. (2021). Challenges and opportunities in dock-based bike-sharing rebalancing: a systematic review. *Sustainability*, *13*(4).
Xu, H., Farajtabar, M., & Zha, H. (2016). Learning Granger causality for Hawkes processes. In *Proceedings of The 33rd International Conference on Machine Learning* (Balcan, M. F., & Weinberger, K. Q., Eds.), Proceedings of Machine Learning Research, *48*, PMLR, New York, New York, USA, pp. 1717–1726.
Yuan, B., Li, H., Bertozzi, A. L., Brantingham, P. J., & Porter, M. A. (2019). Multivariate spatiotemporal Hawkes processes and network reconstruction. *SIAM Journal on Mathematics of Data Science*, *1*(2), 356–382.
Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A., & Leskovec, J. (2015). SEISMIC: a self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1513–1522.
Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, *23*(4), 550–560.

AUTHOR BIOGRAPHIES

Francesco Sanna Passino is a Lecturer in Statistics in the Department of Mathematics at Imperial College London. His research interests are broadly based on statistical models for dynamic networks, statistical analysis of event-time data, and latent factor models, mostly from a Bayesian perspective.

Yining Che holds a MSci Mathematics from Imperial College London. She completed part of this work for her capstone research project.

Carlos Cardoso Correia Perello holds a BSc Mathematics from Imperial College London, and he is currently a student at the Mathematical Institute at the University of Oxford.

How to cite this article: Sanna Passino, F., Che, Y., & Cardoso Correia Perello, C. (2024). Graph-based mutually exciting point processes for modelling event times in docked bike-sharing systems. *Stat*, *13*(1), e660. https://doi.org/10.1002/sta4.660

20491573, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sta4.660 by Imperial College London, Wiley Online Library on [13/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License