SPATIOTEMPORAL HAWKES PROCESSES WITH A GRAPHON-INDUCED CONNECTIVITY STRUCTURE

BY JUSTIN BAARS[1,a], ROGER J. A. LAEVEN[1,b] AND MICHEL MANDJES[1,2,c]

[1]*University of Amsterdam,* [a]*J.R.Baars@uva.nl;* [b]*R.J.A.Laeven@uva.nl*

[2]*Leiden University,* [c]*M.R.H.Mandjes@math.leidenuniv.nl*

We introduce a spatiotemporal self-exciting point process $(N_t(x))$, boundedly finite both over time $[0,\infty)$ and space $\mathscr{X}$, with excitation structure determined by a graphon W on $\mathscr{X}^2$. This *graphon Hawkes process* generalizes both the multivariate Hawkes process and the Hawkes process on a countable network, and despite being infinite-dimensional, it is surprisingly tractable. After proving existence, uniqueness and stability results, we show, both in the *annealed* and in the *quenched* case, that for compact, Euclidean $\mathscr{X} \subset \mathbb{R}^m$, any graphon Hawkes process can be obtained as the suitable limit of d-dimensional Hawkes processes $\tilde{N}^d$, as $d \to \infty$. Furthermore, in the stable regime, we establish an FLLN and an FCLT for our infinite-dimensional process on compact $\mathscr{X} \subset \mathbb{R}^m$, while in the unstable regime we prove divergence of $N_T(\mathscr{X})/T$, as $T \to \infty$. Finally, we exploit a cluster representation to derive fixed-point equations for the Laplace functional of N, for which we set up a recursive approximation procedure. We apply these results to show that, starting with multivariate Hawkes processes $\tilde{N}^d_t$ converging to stable graphon Hawkes processes, the limits $d \to \infty$ and $t \to \infty$ commute.

1. Introduction. In a wide variety of applied fields, self-exciting point processes are used to model natural, economic and social phenomena that interact locally with their own history, in such a way that past events can trigger events in the future. Besides modeling self-exciting behavior in the temporal dimension, nowadays spatiotemporal models are often used, where the excitation occurs both in time and in space. This is usually done by introducing a mutually exciting, or multivariate Hawkes, process on d coordinates, interpreting the coordinates as locations in space. The basic self- and mutually exciting point process was introduced in 1971, see [30, 31].

Such multivariate Hawkes processes are employed to capture excitation behavior over time and space in, e.g.: seismology ([18, 33, 53]), where earthquakes tend to trigger sequences of aftershocks both at the same location and at nearby or adjacent fault lines; neuroscience ([58]), where high-dimensional processes can be used to assess neuronal spike trains; epidemiology ([15]), where outbreaks of infectious diseases spread contagiously both over time and over different geographical locations; finance ([3, 6, 44]), where financial shocks and transactions, such as stock trades or order arrivals in electronic markets, cluster over time and across markets; social network analysis ([22, 59]), for the prediction of online user activity, or to explain the contagious nature of information diffusion and the spread of online content in social media platforms; and criminology ([54]), where crimes like burglaries or gang-related activities tend to cluster in time and across neighborhoods.

In these modeling approaches, the respective space is typically divided into d locations, and one measures the number of events $N_t = (N_t(1), \dots, N_t(d))$ that have occurred by time t at those d locations. By not distinguishing between more granular locations *within* each

MSC2020 subject classifications: Primary 60G55, 60F15; secondary 05C80.

Keywords and phrases: Spatiotemporal point processes, Hawkes processes, Limit theorems, Random graphs, Transform analysis.

of those d locations, one essentially divides the space into d approximating, 'homogeneous' locations or 'subpopulations', possibly neglecting important information. The main advantage of assuming this homogeneity within subspaces is that the probabilistic behavior of the corresponding (finite-dimensional) multivariate Hawkes process is well understood; see e.g., [12, 48, 63, 64]. Furthermore, it is highly tractable: moments can readily be determined for Markovian models ([17, 20, 41]); transforms can be approximated iteratively ([38]); heavy-tailed and heavy-traffic asymptotics have been identified ([38, 41]); a broad range of scaling limits have been studied ([5, 32, 35, 36]); and recently results on the distribution of the cluster duration of the Hawkes process have become available along with the combinatoric structure inherent to its branching representation ([19]). Moreover, a substantial amount of research has been done on statistical inference, both for the parametric case ([52]) and, more recently, for the nonparametric case ([11, 29, 40]).

From this well-understood realm, the focus within the applied probability literature has gradually shifted towards understanding Hawkes processes in high-dimensional settings; see [26, 48] for early research in this direction. More recently, Hawkes processes on large, *countable* networks have been defined in the influential paper [21], where the connectivity structure is provided by an infinite directed graph. In such a model, each node still represents a location or subpopulation. When a spatial structure is absent, and a complete graph on d nodes is imposed for the interaction structure, the large population limit yields an inhomogeneous Poisson process with a *deterministic* intensity function solving a certain convolution equation. This model is still quite tractable: its large-time behavior can be described in the mean-field setting and in a nearest-neighbor version of the model exact asymptotics have been provided ([21]); and recently, a large deviations principle for the mean-field limit has been established ([27]).

The idea of [21] has been extended in [2], where the mean-field prelimit becomes a model on d nodes sampled from a Euclidean space $I = [0,1]$ or $I = \mathbb{R}^d$, and where the connectivity between those nodes is modeled by a graphon $W : I^2 \to \mathbb{R}_+$. In contrast to [21], the coordinates in [2] are sampled from an *uncountable* space. In particular, the mean-field limit is an uncountable collection of inhomogeneous Poisson processes, whose intensity functions satisfy a convolution equation. Interestingly, in this setting, the large-time behavior can still be analyzed, and is related to the spectral radius of an operator corresponding to this convolution equation ([2]).

In different subfields of applied probability, there has been a rise in research on describing, or rather approximating, behavior in an infinite network with a graphon-induced ([47]) connectivity structure. Typically, finite-dimensional models with a graph-induced connectivity structure approximate infinite-dimensional models with a graphon-induced connectivity structure, as the dimension grows large. Such approximations have been investigated for macro-level, deterministic SEIR models ([50]), for micro-level, stochastic SIR models ([55]), for linear threshold models ([23]), for stochastic n-player games ([4, 13, 14]), and for stochastic n-particle systems ([7, 8]). In the case of Hawkes processes, spatiotemporal and ETAS models have been studied in several applied fields, including seismology and statistics ([25, 34, 43, 57]), but to the best of our knowledge, there is no account in the literature of a Hawkes process with a graphon-induced connectivity structure. Perhaps the closest to such a model are d-dimensional ($d \in \mathbb{N}$) Hawkes processes with connectivity between nodes sampled from a graphon ([2, 62]), but those mutually exciting processes themselves are *finite-dimensional*.

In the present work, we introduce, and formally establish, a spatiotemporal self-exciting point process $N_t(x)$ on $[0,\infty) \times \mathscr{X}$, where $(\mathscr{X}, \mathscr{A}, \mu)$ is some σ-finite measure space, that is locally finite in both space and time, with cross-excitation determined by a graphon W :

$\mathscr{X}^2 \to \mathbb{R}_+$. We define our process through a *conditional intensity density* $\lambda_t(x)$, which is such that the simple point process N satisfies

$$\mathbb{P}(N_{t+\mathrm{d}t}(x+\mathrm{d}x) - N_t(x) = 1|\mathcal{F}_t) = \lambda_t(x)\,\mathrm{d}t\mathrm{d}x,$$

where $(\mathcal{F}_t)_{t\in\mathbb{R}}$ is the σ-algebra generated by $(N_t)_{t\in\mathbb{R}}$, i.e., $\mathcal{F}_t = \sigma(N_s(A) : s \leq t, A \in \mathscr{A})$. In the linear case, for some baseline intensity λ_∞, a mark stochastic process B and an excitation function h, we set

$$\lambda_t(x) = \lambda_\infty(x) + \sum_{\substack{(s,y,B_{xy})\in N \\ s<t}} B_{xy}(s)W(x,y)h(t-s).$$

In this way, $\lambda_t(x)$ can be interpreted as the *spatial density* of the conditional intensity. In particular, this representation of the process implies that it is locally finite over space, so while being *infinite-dimensional*, it generates a finite number of events over a bounded subset of time and space. We coin this process a *graphon Hawkes process*.

If we partition our space $\mathscr{X}$ into finite measure sets $\mathscr{X}_k$, and take the parameters of the graphon Hawkes process constant on $\mathscr{X}_k$, then the location *within* $\mathscr{X}_k$ becomes irrelevant, and our process reduces to a Hawkes process on a discrete spatial set. In this way, our novel process generalizes both the multivariate Hawkes process and the process studied in [21]. Note that our graphon Hawkes process is fundamentally different from the process studied in [2], which is a Hawkes process on *finitely many* coordinates with a sampled connectivity graph, and its resulting mean-field limit intensity function, though defined on a continuum of space, is *deterministic* and not boundedly finite over space.

We outline our main results and contributions to the literature. After having formally introduced the graphon Hawkes process in Section 2, we prove corresponding existence, uniqueness and stability results in Theorem 1 of Section 3, under a condition on the spectral radius $\rho(T_{\mathrm{hom}})$ of a linear operator T_{hom} related to the excitation part of the conditional density intensity function. This provides general, weak stability conditions. Furthermore, our general infinite-dimensional framework requires functional-analytic arguments in later sections of the paper that would not be required should one work in a finite-dimensional setting. Whereas the results of Section 3 can be proved by suitably leveraging and generalizing existing techniques from [48], the results of Sections 4–6 require genuinely novel techniques and ideas.

In Section 4, Theorems 2 and 3 establish that (under suitable regularity assumptions) the d-variate process $\tilde{N}^d$ obtained by averaging the parameters of a linear graphon Hawkes process N — defined on compact $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^m$ — over the elements of partitions $\mathcal{P}^d$ of $\mathscr{X}$, converges uniformly on compact sets in probability to N, as $d \to \infty$. We prove this both in the *annealed* and in the *quenched* case. This shows that the graphon Hawkes process is a natural continuous-space generalization of the multivariate Hawkes process. In passing, we prove in Lemma 2 uniform boundedness for summations over operator norms of iterates of the kernel operators corresponding to steppings of a graphon (cf. [47], §7.5 and §9.2), a result that may be of interest in its own right, and might be of use when infinite-dimensional systems are approximated by finite-dimensional systems in different subfields of applied and numerical mathematics. In Lemma 3, we prove a version of the Poincaré inequality that might be interesting in its own right as well. Relying on those functional-analytic arguments and on stochastic coupling, we can give insightful proofs for our convergence results in Theorems 2 and 3.

In Section 5, we provide a functional law of large numbers (FLLN) in the stable case $\rho(T_{\mathrm{hom}}) < 1$ in Theorem 4 and we prove divergence of this prelimit in the unstable case $\rho(T_{\mathrm{hom}}) > 1$ in Theorem 5; hence, we provide a dichotomy in large-time behavior between those two cases. In the stable regime, we also describe a functional central limit theorem

(FCLT) in Theorem 6. In proving our FLLN and FCLT, we show how to essentially reduce to a finite-dimensional setting, allowing for relatively simple proofs of functional limit theorems for infinite-dimensional systems. Given the difficulty of proving such results, our approach may be of interest for other infinite-dimensional models as well.

Finally, we describe in Section 6 the probabilistic behavior of the graphon Hawkes process by characterizing its Laplace functional through fixed-point equations, which provides an iterative approximation procedure in the transform domain, non-trivially extending results from [38, 41] to our infinite-dimensional setting. As an application, we show that, starting with multivariate Hawkes processes $\tilde{N}^d_t$ converging to stable graphon Hawkes processes, we can interchange the limits $d \to \infty$ and $t \to \infty$.

By not discretizing space, we obtain a natural model for phenomena occurring in time and space that is more general, and possibly more realistic, than the classical multivariate Hawkes process, especially in applications where heterogeneous behavior over space is expected. The inaccuracy originating from discretizing space can be reduced by making the approximating, 'homogeneous' sublocations smaller, i.e., by working in a higher-dimensional setting. In the present work, this is taken a step further by working on a continuous spatial set. In applications where excitation or infectivity depends continuously on the distance between two locations or on other geographical characteristics, a continuous-space model provides a clear advantage over finite-dimensional models. For example, the likelihood of an aftershock in location x following an earthquake with epicenter y naturally depends continuously on the distance between x and y, and may also depend on geographical characteristics. Similarly, the likelihood of infection for COVID-19 depends on the distance between an individual and a patient, and on characteristics such as the amount and type of restrictive measures, which vary over space. Moreover, when applying Hawkes processes to crime modeling ([54]), the likelihood of a crime in location y following a crime in x may depend on the distance between x and y, as well as on sociological factors, such as the level of law enforcement, of social control, or of education. Therefore, the model developed in the present work may be of importance to applied fields where multivariate Hawkes processes are used and spatial aspects are relevant.

Conventions. Throughout this paper, $\mathbb{N} := \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Furthermore, the positive half-line is denoted by $\mathbb{R}_+ := [0, \infty)$. In general, the temporal variable is denoted by t, whereas the spatial variable is denoted by x.

When writing 'a multivariate Hawkes process', we mean a Hawkes process on $d \in \mathbb{N}$ coordinates; in particular, a multivariate Hawkes process is finite-dimensional, and we do not refer to Hawkes processes on countable networks or to graphon Hawkes processes by this term.

Appendix. The proofs that are not given in the main text and some supplementary results are collected in four appendices.

2. Definitions of the graphon Hawkes process. In this section, we formally define graphon Hawkes processes, i.e., spatiotemporal self-exciting point processes on an uncountable spatial set, with a connectivity structure defined through a graphon W. We provide two constructions: one through a *conditional intensity density*, and one using a *cluster process representation*, i.e., by describing the dynamics as a spatial Poisson branching process. The conditional intensity density representation allows for a form of nonlinearity, whereas the cluster representation does not. In the linear case the two representations are equivalent. In Section 3, we mainly use the conditional intensity density representation; in Section 6, we mainly use the cluster representation; in Sections 4 and 5, we use both.

We start by defining a spatiotemporal point process on $[0,\infty)\times\mathscr{X}$ through a conditional intensity density, for which we need a formal definition. Let $L^p_{\mathrm{loc}}(\mathscr{X})$ be the set of measurable functions f on a σ-finite measure space $(\mathscr{X},\mathscr{A},\mu)$ such that $\int_A f\,\mathrm{d}\mu<\infty$ for all $A\in\mathscr{A}$ of finite measure.

DEFINITION 1. *Let $(\mathscr{X},\mathscr{A},\mu)$ be a σ-finite measure space. Let $(N_t)_{t\ge 0}$ be a spatiotemporal point process, where to each event at time t, a spatial coordinate $x\in\mathscr{X}$ is attached. Let $(\mathcal{F}_t)_{t\ge0}$ be the natural filtration generated by N, i.e., $\mathcal{F}_t:=\sigma(N_s(A):s\le t, A\in\mathscr{A})$. Then any $(\mathcal{F}_t)_{t\ge0}$-predictable process $(t,x)\mapsto\lambda_t(x)$ is called a* conditional intensity density*, if: $\lambda_t(\cdot)\in L^1_{\mathrm{loc}}$ for all $t\ge0$; and for all $A\in\mathscr{A}$ with $\mu(A)<\infty$, N satisfies, as $\Delta t\downarrow 0$,*

$$\mathbb{P}(N_{t+\Delta t}(A)-N_t(A)=0|\mathcal{F}_t)=1-\int_A\lambda_t(x)\,\mathrm{d}\mu(x)\Delta t+o(\Delta t);$$

$$\mathbb{P}(N_{t+\Delta t}(A)-N_t(A)=1|\mathcal{F}_t)=\int_A\lambda_t(x)\,\mathrm{d}\mu(x)\Delta t+o(\Delta t).$$

Note that the probabilistic structure of N is determined uniquely by the conditional intensity density $\lambda_t(x)$. This can be seen by treating spatial coordinates as marks, and applying [18], Proposition 7.3.IV. Also, it follows from [18], Definition 7.3.II, that any *regular* spatiotemporal point process (see [18], Proposition 7.3.I) admits a conditional intensity density. Furthermore, note that when there is an event in a finite measure set $A\in\mathscr{A}$ at time t, the conditional intensity density can be rescaled to a probability density function determining the spatial coordinate: the spatial coordinate x follows the law $\mathbb{P}(x\in\mathrm{d}x)=\lambda_t(\mathrm{d}x)/\lambda_t(A)$.

Before we define a graphon Hawkes process through its conditional intensity representation, we need some regularity assumptions.

ASSUMPTION 1. *Let $(\mathscr{X},\mathscr{A}=\mathcal{B}(\mathscr{X}),\mu)$ be a topological σ-finite measure space. Let $\lambda_\infty\in L^1_{\mathrm{loc}}(\mathscr{X})$, $h\in L^1(\mathbb{R}_+)$, let $W:\mathscr{X}^2\to\mathbb{R}_+$ be a measurable (di)graphon, i.e., a (possibly non-symmetric) measurable function $(x,y)\mapsto W(x,y)$, and let, for each fixed $s\in\mathbb{R}$, $(x,y)\mapsto B_{xy}(s,\omega)$ be a stochastic process in the variables x and y on $\mathscr{X}^2$, defined on a probability space $(\Omega_B,\mathcal{F}_B,\mathbb{P}_B)$, that is separable w.r.t. the class $\mathcal{U}$ of open subsets of $\mathscr{X}^2$ and is jointly measurable in (x,y,ω) with respect to $\mathcal{B}(\mathscr{X}^2)\otimes\mathcal{F}_B$. Consider a collection of i.i.d. copies of such processes. Also assume that there are $p,q\in[1,\infty]$ with $p^{-1}+q^{-1}=1$ such that $W(\cdot,y)\in L^p_+(\mathscr{X})$ and $\mathbb{E}[B_{\cdot y}]\in L^q_+(\mathscr{X})$ a.s., for μ-a.e. $y\in\mathscr{X}$.*

The definition of *separability* of a stochastic process can be found in [51], §III.4.

DEFINITION 2 (Conditional intensity density representation). *Grant Assumption 1. Furthermore, assume that we are given some history $\mathcal{F}_0$ on $(-\infty,0]$ of a B-marked spatiotemporal point process N with $\mathscr{X}$-valued locations, meaning that for each event at time s, $s\le0$, a mark $(x,y)\mapsto B_{xy}(s)$ is attached, independently of everything else. Let $\mathscr{X}\times[0,\infty)\to[0,\infty):(x,y)\mapsto f_x(y)$ be a measurable function, and assume that $f_x(\cdot)$ is c_x-Lipschitz, for each $x\in\mathscr{X}$, where $x\mapsto c_x$ is assumed to be measurable. We say that N is a* graphon Hawkes process *on $\mathbb{R}_+\times\mathscr{X}$ when N is generated on $\mathbb{R}_+$ by the conditional intensity density*

$$\lambda_t(x)=f_x\Bigg(\lambda_\infty(x)+\sum_{\substack{(s,y,B_{xy})\in N\\ s<t}}B_{xy}(s)W(x,y)h(t-s)\Bigg), \tag{1}$$

where the summation is over all events of N at times s strictly before t, at location y and with mark $B_{xy}(s)\overset{\mathcal{D}}{=}B_{xy}$ satisfying Assumption 1.

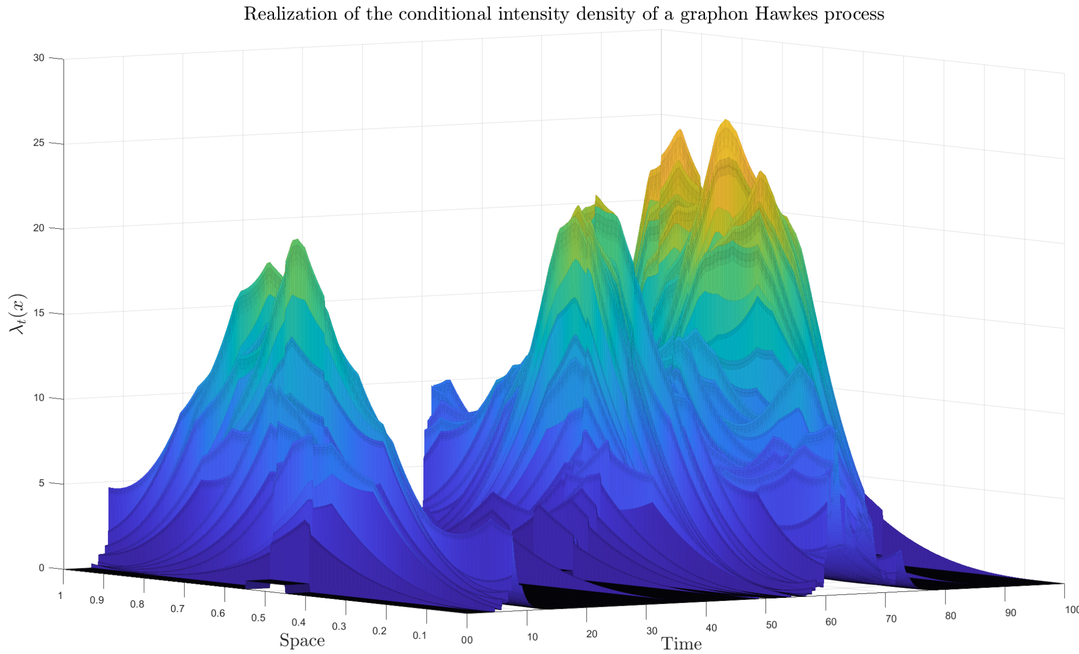

Fig 1: A realization of $\lambda_t(x)$ for a linear, unmarked model on $\mathscr{X} = [0,1] \ni x$ and $[0,100] \ni t$. We use $\lambda_\infty(x) = \mathbf{1}\{x \in [.45,.55]\}$ and a graphon $W(x,y) = 30(1/2 - d(x,y))^3$, where $d(x,y) = \min\{|x-y|, 1-|x-y|\}$ describes the distance on the unit circle.

We illustrate the dynamics of the graphon Hawkes process in Figure 1, where a realization of the conditional intensity density for a linear, unmarked process on $\mathscr{X} = [0,1]$ is shown. Here, we use a graphon W with values $W(x,y)$ depending negatively on the distance between x and y on the unit circle.

In this paper, we mainly focus on *linear* processes for which f_x is the identity mapping, for all $x \in \mathscr{X}$. We assume throughout that we work in this setting, unless explicitly stated otherwise.

REMARK 1. *When we take $\mathscr{X}$ compact, e.g., $\mathscr{X} = [0,1]$, and parameters λ_∞, B, W that are piecewise constant on the elements of some partition $\mathcal{P}^d$ of $[0,1]$ into d disjoint measurable sets $[0,1] = \bigcup_{i=1}^d \mathscr{X}_i$, then the locations within the sets $\mathscr{X}_i$ are irrelevant, and the process is essentially a d-variate Hawkes process. We employ this idea in Section 4.*

Similarly, when $\mathscr{X}$ is unbounded, σ-finite and partitioned into finite measure sets $(\mathscr{X}_k)_{k\in\mathbb{N}}$, then a process with parameters constant on each $\mathscr{X}_k$ is essentially a Hawkes process on a countable network; cf. [21].

As is the case for multivariate Hawkes processes, a linear conditional intensity density specification allows us to give a definition of the graphon Hawkes process in terms of a Poisson branching process. In the linear case, events (t,x) arrive over time and space at rate

$$\lambda_t(x) = \lambda_\infty(x) + \sum_{\substack{(s,y,B_{xy}) \in N \\ s<t}} B_{xy}(s) W(x,y) h(t-s). \tag{2}$$

With this linear rate specification, one can say that there are two types of events: those generated by the *baseline intensity* λ_∞, which are coined *immigrant events*, and those generated by the second term, which consists of intensity increases caused by previous events; therefore, those events are coined *offspring events*. Note that the immigrant events create clusters that are i.i.d., modulo the time shift corresponding to the immigrant arrival times. Furthermore, each event creates offspring according to the same iterative procedure, with different clusters being independent of each other, implying that there is a branching structure, exhibiting self-similarity. This structure is made explicit by the following cluster representation, which is fully exploited in Section 6, when we derive fixed-point equations in the transform domain.

DEFINITION 3 (Cluster representation). *Grant Assumption 1. Consider the time interval $[0,T]$, where $T \in [0,\infty]$. We generate the events of the graphon Hawkes process according to the following procedure.*

(i) *Generate immigrant arrival times according to a time-homogeneous Poisson process $I(\cdot)$ on $[0,T]\times\mathscr{X}$ with intensity $\lambda_\infty(x)$, and write $(T_1^{(0)},x_1^{(0)}),\dots,(T_{I(T)}^{(0)},x_{I(T)}^{(0)})$ for the pairs of event times and locations strictly before time T. Sample the corresponding random mark functions $B_{\cdot x_i^{(0)}}(T_i^{(0)})$ according to the law of B.*

(ii) *Set $n=0$. For each arrival with characteristics $(T_i^{(n)},x_i^{(n)},B_{\cdot x_i^{(n)}}(T_i^{(n)}))$, sample next-generation offspring events $(T_1^{(n+1)},x_1^{(n+1)}),\dots,(T_{K_i^{(n)}(T)}^{(n+1)},x_{K_i^{(n)}(T)}^{(n+1)})$ strictly before time T according to a time-inhomogeneous Poisson process $K_i^{(n)}$ on $[0,T]\times\mathscr{X}$ having intensity at time $t\geq T_i^{(n)}$ equal to $B_{xx_i^{(n)}}(T_i^{(n)})W(x,x_i^{(n)})h(t-T_i^{(n)})$. Sample the corresponding random mark functions $B_{\cdot x_i^{(n+1)}}(T_i^{(n+1)})$ according to the law of B.*

(iii) *Iterate for $n\in\mathbb{N}$, obtaining for finite $t\in[0,T]$ the event sequence*

$$E_n(t) := \left\{ \left(T_i^{(n)},x_i^{(n)},B_{\cdot x_i^{(n)}}\left(T_i^{(n)}\right)\right) : T_i^{(n)} \leq t \right\}, \quad n\in\mathbb{N}_0.$$

Then the process $N(\cdot)$ given by $N(t)=\bigcup_{n\in\mathbb{N}_0} E_n(t)$ constitutes a graphon Hawkes process.

As for multivariate Hawkes processes, with linear models one can decompose the conditional intensity density into a baseline intensity, first-generation offspring intensity, and so on, implying that particles arrive at the same rates according to Definitions 2 and 3, given that their spatial coordinates coincide. It is then intuitively clear that the intensity-based and the cluster representation-based definitions are equivalent for linear processes on compact $\mathscr{X}$ starting on an empty history, though we need to check that the different procedures used to sample the spatial coordinates are equivalent. This is the content of the next elementary lemma. Part (i), corresponding to the cluster representation definition, describes a sampling method where the intensity is decomposed, whereas part (ii), corresponding to the conditional intensity density definition, does not use such a decomposition. The proof of Lemma 1 can be found in Appendix A.

LEMMA 1. *Grant Assumption 1. Suppose that we sample spatial coordinates using an intensity function $\lambda(x)=\lambda^1(x)+\lambda^2(x)$. Then the following two procedures are equivalent.*

(i) *First, sample whether λ^i caused the event, where λ^i gets probability $\frac{\|\lambda^i(\cdot)\|_{L^1(\mathscr{X})}}{\|\lambda^1(\cdot)+\lambda^2(\cdot)\|_{L^1(\mathscr{X})}}$. Then, sample the location according to the law having Lebesgue density $\frac{\lambda^i(\cdot)}{\|\lambda^i(\cdot)\|_{L^1(\mathscr{X})}}$.*

(ii) *Sample the location according to the law having Lebesgue density $\frac{\lambda(\cdot)}{\|\lambda(\cdot)\|_{L^1(\mathscr{X})}}$.*

For most results in this paper (except for the existence, uniqueness and stability results of Section 3), we assume compactness of our space $\mathscr{X}$ and uniform boundedness of the parameters of our model, as follows.

ASSUMPTION 2. *It holds that $W(x,y)$, $\mathbb{E}[B_{xy}]$ are bounded by constants $C_W, C_B \geq 0$, respectively, uniformly over $x,y \in \mathscr{X}$. Furthermore, the Lipschitz constants c_x are bounded by C_{Lip}, uniformly over $x \in \mathscr{X}$, and $\lambda_\infty(\cdot) \in L^\infty_{\mathrm{loc}}(\mathscr{X})$.*

3. Existence, uniqueness and stability. In this section, we consider the general nonlinear model and the corresponding conditional intensity density representation given by Definition 2, where $(\mathscr{X}, \mathcal{B}(\mathscr{X}), \mu)$ is assumed to be some (topological) σ-finite measure space. We still need to prove that this is a 'good' definition, in the sense that there exists a unique process N satisfying the dynamics given by (1), and that such a solution is stable in distribution, meaning that there exists a unique stationary distribution, to which any 'sufficiently regular' solution converges weakly, as $t \to \infty$. We consider the model under relatively weak stability conditions, reminiscent of those given in [12], Theorem 7, for multivariate Hawkes processes. The proofs of this section leverage methodology from [48] and can be found in Appendix A. In Appendix A, we also show how to apply [48], Theorem 2, more directly under the stronger stability conditions considered there. When one considers the case of compact, Euclidean $\mathscr{X}$, e.g., $\mathscr{X} = [0,1]$, the proofs and methodology become somewhat simpler.

In this section, we assume that the spatial coordinate takes values in some σ-finite measure space $(\mathscr{X}, \mathscr{A}, \mu)$, meaning that there exists a countable, measurable partition $\mathscr{X} = \bigsqcup_{i\in\mathbb{N}} \mathscr{X}_i$ into sets of finite measure, i.e., $\mu(\mathscr{X}_i) < \infty$, for all $i \in \mathbb{N}$. We refer to an element $\mathscr{X}_i$ of this partition as a *site*. Of course, $\mathbb{N}$ may be replaced by any other countable set, e.g., $\mathbb{Z}^m$. Note that when $\mathscr{X} = [0,1]$, there is no need to introduce multiple sites.

In the finite-dimensional case, existence, uniqueness and stability of Hawkes processes is related to the spectral radius of a matrix consisting of magnitudes of excitation, see [12], Theorem 7. In the current infinite-dimensional setting, we are interested in the linear transformation

$$T_{\mathrm{hom}} : L^1(\mathscr{X},\mu) \to L^1(\mathscr{X},\mu) : f(\cdot) \mapsto \|h\|_{L^1(\mathbb{R}_+)} c_\cdot \int_{\mathscr{X}} \mathbb{E}[B_{\cdot y}] W(\cdot, y) f(y)\, \mathrm{d}\mu(y), \tag{3}$$

which, heuristically, maps 'a particle in y', i.e., the Dirac delta function δ_y, into its expected (first-generation) offspring size density throughout the space $\mathscr{X}$. Recall that c_x is the Lipschitz constant of f_x; see Definition 2. In the linear case where $f_x \equiv 1$ and $c_x \equiv 1$, the operator T_{hom} corresponds to the homogeneous part of the conditional intensity density, whence the subscript. We work in the regime where $\rho(T_{\mathrm{hom}})$, the spectral radius of T_{hom}, is smaller than 1. Here,

$$\rho(T_{\mathrm{hom}}) := \sup\{|\alpha| : \alpha \in \sigma(T_{\mathrm{hom}})\},$$
$$\sigma(T_{\mathrm{hom}}) := \{\alpha \in \mathbb{C} : T_{\mathrm{hom}} - \alpha I \in B(L^1(\mathscr{X},\mu)) \text{ is not invertible}\},$$

with $I : L^1(\mathscr{X},\mu) \to L^1(\mathscr{X},\mu)$ the identity mapping, and with $B(\mathscr{Z})$ the set of bounded linear operators on a Banach space $\mathscr{Z}$. The spectral radius is related to the operator norm through Gelfand's formula, see [16], Proposition VII.3.8:

$$\rho(T_{\mathrm{hom}}) = \lim_{n\to\infty} \|T^n_{\mathrm{hom}}\|^{1/n} = \inf_{n\in\mathbb{N}} \|T^n_{\mathrm{hom}}\|^{1/n}. \tag{4}$$

We now describe a two-step construction of (possibly nonlinear) graphon Hawkes processes. In the first step, for each $i \in \mathbb{N}$, we define a Poisson random measure $\bar{N}^i$ corresponding to site $\mathscr{X}_i$, which is used to construct events in $\mathscr{X}_i$. Next, for an arrival in $\mathscr{X}_i$, we use uniform marks to set up an acceptance-rejection method for determining the spatial coordinate $x \in \mathscr{X}_i$. To this end, for $i \in \mathbb{N}$, let $(L^i, \mathcal{L}^i, \mathcal{Q}^i)$ be a factor space of two spaces $(L^i_j, \mathcal{L}^i_j, \mathcal{Q}^i_j)$, $j = 1,2$, the first space modeling the mark stochastic processes B^i_{xy} ($x \in \mathscr{X}, y \in \mathscr{X}_i$), and the second one supporting i.i.d. sequences of couples of independent

$\mathrm{Uni}(\mathscr{X}_i, \mu_i)$ and $\mathrm{Uni}(0,1)$ marks; denote such a sequence by $U^i = (U^i_{1,k}, U^i_{2,k})_{k\in\mathbb{N}}$. Here, the $\mathrm{Uni}(\mathscr{X}_i, \mu_i)$-distribution is defined by $\mathbb{P}(U^i_{1,k} \in \mathrm{d}x) = \mu(\mathrm{d}x)/\mu(\mathscr{X}_i)$. For later use, for $j = 1, 2$, let $(L_j, \mathcal{L}_j, \mathcal{Q}_j)$ be the product space $\prod_{i\in\mathbb{N}}(L^i_j, \mathcal{L}^i_j, \mathcal{Q}^i_j)$, and let $(L, \mathcal{L}, \mathcal{Q})$ be the product of $(L_1, \mathcal{L}_1, \mathcal{Q}_1)$ and $(L_2, \mathcal{L}_2, \mathcal{Q}_2)$.

Our acceptance-rejection method is as follows. When there is an arrival at time t in site $\mathscr{X}_i$, we have $\lambda_t(x) \in L^\infty(\mathscr{X}_i)$ a.s., since T_{hom} is a bounded operator under Assumption 1 (use Hölder's inequality). In step k, $U^i_{1,k}$ is used as a proposal. We accept the proposal with probability proportional to $\lambda_t(U^i_{1,k})$. This means that we accept the proposal if and only if

$$\frac{\lambda_t(U^i_{1,k})}{\|\lambda_t(\cdot)\|_{L^\infty(\mathscr{X}_i)}} \geq U^i_{2,k}.$$

If this condition is not satisfied, we reject, and try again for $k+1$. Now, given an arrival in site $\mathscr{X}_i$ at time t, the location within this site is determined by the realization of U^i, since $\lambda_t(\cdot)$ is known. We refer to this location as $X^i_t(U_t)$, with the understanding that U_t is the realization of U^i at time t.

REMARK 2. *In case $\mathscr{X} = [0,1]$ and μ is the Lebesgue measure on $[0,1]$, we need only a single $U \sim \mathrm{Uni}(0,1)$ random variable. Indeed, given an event at time t, the location can be determined by*

$$\mathbb{P}(X_t \in A) = \frac{\int_A \lambda_t(x)\,\mathrm{d}x}{\|\lambda_t(\cdot)\|_{L^1[0,1]}},$$

for $A \in \mathcal{B}[0,1]$, meaning that the location is given by

$$X_t(U) = \inf\left\{ z \in [0,1] : \frac{\int_0^z \lambda_t(x)\,\mathrm{d}x}{\int_0^1 \lambda_t(x)\,\mathrm{d}x} \geq U \right\}. \tag{5}$$

We are now equipped to define the graphon Hawkes process through the Poisson random measures $\bar{N}^i$ on $\mathbb{R} \times L^i \times \mathbb{R}_+$ having intensity $\mathrm{d}t \times \mathcal{Q}^i(\mathrm{d}z) \times \mathrm{d}s$. Let $S_t N_-$ be the history of N at time t; here, N_- denotes the history at time 0, and S_t is the left-shift operator w.r.t. the first variable. Let $N^i(\mathrm{d}t \times \mathrm{d}z)$ be the process counting events in site i, with $z \in L$. Given an initial condition $S_0 N_-$, we assume the following dynamics on $\mathbb{R}_+$:

$$\begin{aligned} &N^i(\mathrm{d}t \times \mathrm{d}z) = \bar{N}^i(\mathrm{d}t \times \mathrm{d}z \times [0, \psi(S_t N_-, i)]);\\ &\psi(S_t N_-, i) = \Lambda^i_t = \int_{\mathscr{X}_i} \lambda_t(x)\,\mathrm{d}\mu(x) = \|\lambda_t(\cdot)\|_{L^1(\mathscr{X}_i,\mu)};\\ &\lambda_t(x) = f_x\left(\lambda_\infty(x) + \sum_{j\in\mathbb{N}} \sum_{\substack{(s, B_{xX^j_s(U_s)}, U_s) \in N^j \\ s<t}} B_{xX^j_s(U_s)} W(x, X^j_s(U_s)) h(t-s)\right). \end{aligned} \tag{6}$$

For the definition of strongly regular point processes, which we use in the next result, we refer to [48], Definition 2.

THEOREM 1. *Grant Assumption 1. Consider a graphon Hawkes process N on a σ-finite measure space $(\mathscr{X}, \mathscr{A}, \mu)$ having (possibly nonlinear) conditional intensity density specification*

$$\lambda_t(x) = f_x\left(\lambda_\infty(x) + \int_{(-\infty,t)} \int_{\mathscr{X}} B_{xy}(s) W(x,y) h(t-s)\,\mathrm{d}N_s(y)\,\mathrm{d}\sigma(s)\right), \tag{7}$$

where σ denotes the counting measure on $\mathscr{X}$. Assume that we have a partition of $\mathscr{X}$ into finite, non-null measure sets $\bigsqcup_{i\in\mathbb{N}} \mathscr{X}_i$ that is such that the following conditions hold:

$$C := \sup_{t\geq 0, i\in\mathbb{N}} \|f.(\lambda_\infty(\cdot) + \eta(t,\cdot))\|_{L^1(\mathscr{X}_i,\mu)} < \infty; \tag{8}$$

$$\rho(T_{\text{hom}}) < 1; \tag{9}$$

where

$$\eta(t,x) := \sum_{\substack{(s,y,B_{xy})\in N \\ s<0}} B_{xy} W(x,y) h(t-s). \tag{10}$$

Then there exists a unique strongly regular solution N to (6) such that

$$\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\psi(S_t N_-, i) < \infty.$$

REMARK 3. *In the linear case, (8) is equivalent to the following two conditions.*

$$\alpha := \sup_{i\in\mathbb{N}} \int_{\mathscr{X}_i} \lambda_\infty(x)\, \mathrm{d}\mu(x) < \infty; \tag{11}$$

$$\sup_{t>0, i\in\mathbb{N}} \int_{\mathscr{X}_i} \eta(t,x)\, \mathrm{d}\mu(x) < \infty. \tag{12}$$

Besides the proof of Theorem 1, we provide in Appendix A some results that more directly follow the work of [48]. First, in Proposition 2, we give a stability result, which is [48], Theorem 4, specified to the current setting; there we also use [48], Remarks 3 and 4. Next, in Proposition 3, we apply [48], Theorem 2, directly to prove existence and uniqueness under a stability condition related to the operator norm of T_{hom}, which we calculate to be

$$\|T_{\text{hom}}\| = \|h\|_{L^1(\mathbb{R}_+)} \sup_{y\in\mathscr{X}} \int_{\mathscr{X}} c_x \mathbb{E}[B_{xy}] W(x,y)\, \mathrm{d}\mu(x). \tag{13}$$

REMARK 4. *When we work with a linear model on a finite measure space $\mathscr{X}$, the parameter ρ from equation (74) is given by $\rho = \|T_{\text{hom}}\|$. In this case, condition (74) implies $\rho(T_{\text{hom}}) < 1$ (cf. (9)) by Gelfand's formula.*

4. Convergence of mutually exciting processes to graphon Hawkes processes. In this section, we show that the *linear* graphon Hawkes process on a compact, convex, Euclidean spatial set $\mathscr{X} \subset \mathbb{R}^m$, which is an infinite-dimensional model, occurs as the limit of suitably chosen multivariate Hawkes processes on d coordinates, as $d \to \infty$. More specifically, we show that, under some regularity assumptions, partitioning $\mathscr{X}$ and averaging the parameters of the graphon Hawkes process over the sets of this partition yields a piecewise constant graphon Hawkes process — i.e., a multivariate Hawkes process, see Remark 1 — that is close to the original graphon Hawkes process, for large d. This averaging procedure can be followed for infinite Euclidean spaces $\mathscr{X} = \mathbb{R}^m$ as well, by partitioning the space into hyperrectangles of uniformly bounded size. The proofs of the present section, however, do not carry over to this setting, due to the possibility that the excitation of the prelimit 'comes from infinity', meaning that an event may be caused by excitation stemming from an event that is arbitrarily far away.

We start by introducing some definitions and assumptions (Section 4.1), the prelimit model $(\tilde{N}^d)_{d\in\mathbb{N}}$ (Section 4.2), and establish some required technical results (Section 4.3). Then we consider the *annealed* case (Section 4.4), where the prelimit is a multivariate Hawkes process with a complete weighted connectivity graph. Next, we treat the *quenched* case (Section 4.5), where we sample edges from this complete connectivity graph, yielding a Hawkes process on a finite network with an unweighted, directed connectivity graph as prelimit. Note that for such a sampling procedure, we need W to map into $[0,1]$, which can be accomplished by a rescaling such as $\breve{W}(x,y)\breve{B}_{xy} = (W(x,y)/C_W)\,(C_W B_{xy})$, using Assumption 2.

After providing definitions, the prelimit model and some technical preliminaries on $[0,T]$, we work from Section 4.4 onwards on the finite time interval $[0,1]$. By obvious modifications, everything caries over to bounded time intervals $[0,T]$.

4.1. *Definitions and assumptions.* To prove convergence, we use the notion of *uniform convergence on compacts in probability* (ucp convergence) and we first need to define a metric on the space of marked spatiotemporal point processes. Next, we provide some assumptions.

DEFINITION 4. *Let $\mathscr{X} \subset \mathbb{R}^m$ be a convex, m-dimensional manifold. Let $(N^k)_{k\in\mathbb{N}}$ be a sequence of spatiotemporal point processes on $[0,T]\times\mathscr{X}$, marked by i.i.d. realizations of a stochastic process B on $\mathscr{X}^2$ satisfying Assumption 1. Furthermore, let N be another such process. Write $\mathscr{Z}$ for the space of realizations of such processes. Let $\mathfrak{d}(\cdot,\cdot)$ be a metric on $\mathscr{Z}$. We say that N^k converges uniformly on compact sets in probability to N, as $k\to\infty$, if for every nondegenerate compact, convex m-dimensional manifold $\mathscr{A}\subset\mathscr{X}$ and for every $\delta>0$,*

$$\mathbb{P}\left(\sup_{\substack{\mathscr{A}'\subset\mathscr{A}\\ \mathscr{A}' \text{ compact, convex}}} \mathfrak{d}\left(N^k_{\mathscr{A}'}, N_{\mathscr{A}'}\right) > \delta\right) \to 0, \qquad \text{as } k\to\infty, \tag{14}$$

in which case we write $N^k \xrightarrow{\text{ucp}} N$. Here, $N_{\mathscr{Y}}$ denotes N restricted to the spatial set $\mathscr{Y}$.

We aim to define a metric with the property that two point processes are 'far apart' whenever there is an event of one process that does not coincide with an event of the other; and such that two simultaneous events are 'close' if and only if their spatial coordinates and their marks are close. This motivates the following metric.

DEFINITION 5. *For some $T\geq 0$, let N,M be two spatiotemporal point processes on $[0,T]\times\mathscr{X}$, marked by i.i.d. realizations of a stochastic process B on $\mathscr{X}^2$ satisfying Assumption 1. Write $\mathcal{T}_L := \{t\in\mathbb{R}_+ : \exists y, B \text{ s.t. } (t,y,B)\in L\}$ for the event times of $L=N,M$. Then we define a metric on $\mathscr{Z}$ (see Definition 4) by*

$$\begin{aligned}\mathfrak{d}(N,M) := &\sum_{\substack{t\in\mathcal{T}_N\cap\mathcal{T}_M\\ (t,y^N,B^N)\in N\\ (t,y^M,B^M)\in M}} \left(\|y^N-y^M\|_{\mathbb{R}^m} + \|B^N_{\cdot y^N} - B^M_{\cdot y^M}\|_{L^1(\mathscr{X})}\right)\\ &+ \sum_{\substack{t\in\mathcal{T}_N\triangle\mathcal{T}_M\\ (t,y,B)\in\mathcal{T}_N\cup\mathcal{T}_M}} \left(1+\|y\|_{\mathbb{R}^m} + \|B_{\cdot y}\|_{L^1(\mathscr{X})}\right).\end{aligned} \tag{15}$$

REMARK 5. *The choice of the Euclidean norm on $\mathbb{R}^m$ is unimportant, since all norms on a finite-dimensional vector space are equivalent. Also, we could just as well work with the $L^p(\mathscr{X})$-norm, $p\in(1,\infty)$, by stating an analog of Lemma 3 below, using a Poincaré inequality in $L^p(\mathscr{X})$ and assumptions on the p-variation of B.*

ASSUMPTION 3. *Let $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}] := \prod_{i=1}^{m}[a_i, b_i] \subset \mathbb{R}^m$ be some nondegenerate hyperrectangle in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \mathrm{Leb}^m)$, where Leb^m is the Lebesgue measure on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Assume that we are given a sequence of partitions $(\mathcal{P}^d)_{d\in\mathbb{N}}$ of $\mathscr{X}$ into d hyperrectangles $(\mathscr{X}_n^d)_{n\in[d]}$ with $\mathrm{mesh}(\mathcal{P}^d) \to 0$ monotonically as $d \to \infty$, where $\mathrm{mesh}(\mathcal{P}^d) := \max_{n\in[d]} \mathrm{diam}(\mathscr{X}_n^d)$.*

Next, we need some regularity assumptions on the parameters of our graphon Hawkes process in terms of their *total variation*; cf. [45].

DEFINITION 6. *Let $\Omega \subset \mathbb{R}^m$ be an open region, and let $f \in \mathcal{L}^1_{\mathrm{loc}}(\Omega)$. We define the total variation of f in Ω by*

$$\mathrm{Var}(f, \Omega) := \sup\left\{\sum_{i=1}^{m}\int_\Omega \frac{\partial \Phi_i}{\partial x_i} f \, \mathrm{d}x : \Phi \in C_c^1(\Omega, \mathbb{R}^m), \|\Phi\|_{L^\infty(\Omega)} \le 1\right\}, \tag{16}$$

where $C_c^1(\Omega, \mathbb{R}^m)$ is the set of continuously differentiable functions $\Phi : \Omega \to \mathbb{R}^m$. When working on equivalence classes, i.e., for $f \in L^1_{\mathrm{loc}}(\Omega)$, we set

$$\|f\|_{\mathrm{TV}(\Omega)} := \mathrm{Var}(f, \Omega) := \inf_{\substack{g\in\mathcal{L}^1_{\mathrm{loc}}(\mathbb{R}^m) \\ g=f \text{ a.e.}}} \mathrm{Var}(g, \Omega). \tag{17}$$

Write $\mathrm{BV}(\Omega)$ for the set of (equivalence classes of) integrable functions $f \in L^1_{\mathrm{loc}}(\Omega')$ defined on a subset $\Omega' \supset \Omega$ of $\mathbb{R}^m$ such that $f|_\Omega$ has finite total variation.

REMARK 6. *For univariate functions $g : \mathbb{R} \to \mathbb{R}$, (16) is equivalent to*

$$\mathrm{Var}(g, (a, b)) = \sup\left\{\sum_{i=1}^{n} |g(x_{i+1}) - g(x_i)| : n \in \mathbb{N}, a = x_0 < x_1 < \cdots < x_n = b\right\}. \tag{18}$$

ASSUMPTION 4. *Assume that N is a linear graphon Hawkes process defined on $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}]$, satisfying Assumption 3. Suppose that $\|\lambda_\infty\|_{\mathrm{TV}(\mathscr{X})} < \infty$. Suppose that $B_{\cdot y}$ is of bounded variation a.s. for a.a. $y \in \mathscr{X}$, with $\sup_{y\in\mathscr{X}} \mathbb{E}\|B_{\cdot y}\|_{\mathrm{TV}(\mathscr{X})} < \infty$. Furthermore, suppose that $W(\cdot, y)$ is of bounded variation for all $y \in \mathbb{R}$, with $\sup_{y\in\mathscr{X}} \|W(\cdot, y)\|_{\mathrm{TV}(\mathscr{X})} < \infty$.*

Finally, in Proposition 1 below, we require continuity of W, as follows.

ASSUMPTION 5. *Suppose that $W : \mathscr{X}^2 \to \mathbb{R}_+$ is a.e. continuous.*

4.2. *Prelimit model.* We consider a linear graphon Hawkes process N, defined through its conditional intensity density (2), with corresponding integral operator

$$T_{\mathrm{hom}} : L^1(\mathscr{X}) \to L^1(\mathscr{X}) : f(\cdot) \mapsto \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} \mathbb{E}[B_{\cdot y}] W(\cdot, y) f(y) \, \mathrm{d}y. \tag{19}$$

We define the intended prelimit $(\tilde{N}^d)_{d\in\mathbb{N}}$ by specifying conditional intensity densities. Using the partition $\mathcal{P}^d$ from Assumption 3, we partition the compact hyperrectangle $\mathscr{X}$ into further hyperrectangles $(\mathscr{X}_n^d)_{n\in[d]}$. We define a prelimit model $\tilde{N}^d$ as the linear model having conditional intensity density

$$\tilde{\lambda}_t^d(x) = \tilde{\lambda}_\infty^d(x) + \sum_{\substack{(s,y,\tilde{B}^d_{xy})\in\tilde{N}^d \\ s<t}} \tilde{B}^d_{xy}(s) \tilde{W}^d(x, y) h(t - s), \tag{20}$$

where the parameters are obtained by averaging over the elements $\mathscr{X}_n^d$ of $\mathcal{P}^d$:

$$\tilde{\lambda}_\infty^d(x) = \sum_{i=1}^d \frac{\mathbf{1}_{\mathscr{X}_i^d}(x)}{\mathrm{Leb}^m(\mathscr{X}_i^d)} \int_{\mathscr{X}_i^d} \lambda_\infty(y')\,\mathrm{d}y'; \tag{21}$$

$$\tilde{B}_{xy}^d(s) = \sum_{i=1}^d \sum_{j=1}^d \frac{\mathbf{1}_{\mathscr{X}_i^d}(x)\mathbf{1}_{\mathscr{X}_j^d}(y)}{\mathrm{Leb}^m(\mathscr{X}_i^d)\mathrm{Leb}^m(\mathscr{X}_j^d)} \int_{\mathscr{X}_i^d \times \mathscr{X}_j^d} B_{x'y'}(s)\,\mathrm{d}x'\,\mathrm{d}y'; \tag{22}$$

$$\tilde{W}^d(x,y) = \sum_{i=1}^d \sum_{j=1}^d \frac{\mathbf{1}_{\mathscr{X}_i^d}(x)\mathbf{1}_{\mathscr{X}_j^d}(y)}{\mathrm{Leb}^m(\mathscr{X}_i^d)\mathrm{Leb}^m(\mathscr{X}_j^d)} \int_{\mathscr{X}_i^d \times \mathscr{X}_j^d} W(x',y')\,\mathrm{d}x'\,\mathrm{d}y'. \tag{23}$$

Note that when we we average out the parameters of a graphon Hawkes process over the sets $\mathscr{X}_n^d$ as in (21)–(23), the location *within* $\mathscr{X}_n^d$ of an event in $\mathscr{X}_n^d$ does not influence the dynamics of the process $\tilde{N}^d$. Since we have finitely many coordinates, the resulting process can be interpreted as a Hawkes process on a discrete network with nodes $n \in [d]$ corresponding to sites $\mathscr{X}_n^d$. This multivariate Hawkes process $N = (N_j^d)_{j\in[d]}$ on finitely many coordinates can be specified through its conditional intensity. Indeed, for each $i \in [d]$, we have a conditional intensity in coordinate i given by

$$\Lambda_{t,i}^d = \lambda_{\infty,i}^d + \sum_{j=1}^d \sum_{\substack{s\in N_j^d \\ s<t}} B_{ij}^d(s) W_{ij}^d h(t-s), \tag{24}$$

where, cf. (21)–(23),

$$\lambda_{\infty,i}^d = \int_{\mathscr{X}_i^d} \lambda_\infty(x)\,\mathrm{d}x; \tag{25}$$

$$B_{ij}^d(s) = \frac{1}{\mathrm{Leb}^m(\mathscr{X}_j^d)} \int_{\mathscr{X}_i^d \times \mathscr{X}_j^d} B_{xy}(s)\,\mathrm{d}x\,\mathrm{d}y; \tag{26}$$

$$W_{ij}^d = \frac{1}{\mathrm{Leb}^m(\mathscr{X}_i^d)\mathrm{Leb}^m(\mathscr{X}_j^d)} \int_{\mathscr{X}_i^d \times \mathscr{X}_j^d} W(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{27}$$

Note how (24)–(27) describe the same model as (21)–(23), since integrating a constant density over site i is identical to multiplying by $\mathrm{Leb}^m(\mathscr{X}_i^d)$. The connectivity structure of the model on $[d]$ is described by a complete weighted graph with weights $(W_{ij}^d)_{i,j\in[d]}$.

We refer to the model with conditional intensity density given by (20) as the *annealed case*. In the *quenched case*, the prelimit uses a simple graph, with edges sampled with probabilities equal to the weights W_{ij}^d. For this to make sense, we need $C_W \leq 1$. As observed earlier, this can be accomplished by writing $W(x,y)B_{xy} = (W(x,y)/C_W)(C_W B_{xy})$, assuming $C_W < \infty$. More specifically, in the quenched case, we connect coordinates $i, j \in [d]$ in the prelimit model $\bar{N}^d$ if and only if $Z_{ij}^d = 1$, where $Z_{ij}^d \sim \mathrm{Bernoulli}(W_{ij}^d)$. Then, for each $i \in [d]$, we have a conditional intensity in coordinate i given by

$$\Lambda_{t,i}^d = \lambda_{\infty,i}^d + \sum_{j=1}^d \sum_{\substack{s\in N_j^d \\ s<t}} B_{ij}^d(s) Z_{ij}^d h(t-s). \tag{28}$$

This is equivalent to a graphon Hawkes process $\bar{N}^d$ having conditional intensity density

$$\bar{\lambda}_t^d(x) = \tilde{\lambda}_\infty^d(x) + \sum_{\substack{(s,y,\tilde{B}_{xy}^d)\in\bar{N}^d \\ s<t}} \tilde{B}_{xy}^d(s)\tilde{Z}^d(x,y)h(t-s), \tag{29}$$

where

$$\tilde{Z}^d(x,y) = \sum_{i=1}^{d}\sum_{j=1}^{d} \mathbf{1}_{\mathscr{X}_i^d}(x)\mathbf{1}_{\mathscr{X}_j^d}(y) Z_{ij}^d. \tag{30}$$

We call $\tilde{N}^d$ on the weighted graph $\tilde{W}^d$ the *annealed* model because its environment can be seen as the integrated version of that of the *quenched* model $\bar{N}^d$ on a $\tilde{W}^d$-random graph.

One has to be careful to prevent the multivariate Hawkes process induced by the graphon Hawkes process, as defined with the aid of equations (21)–(23), to be explosive. It is easy to see that (11) still holds for the model with averaged parameters, but (9) is more difficult to verify. Suppose that (9) holds for the original model, but that with high probability, B is large where W is small, and *vice versa*. It may occur that after averaging, the resulting multivariate process is not stable any more. However, it can be argued that this problem does not arise for d sufficiently large whenever $\mathrm{mesh}(\mathcal{P}^d) \to 0$, as $d \to \infty$. This is shown in the next subsection.

4.3. *Technical results.* In the proofs of the convergence results in the next subsections, we need some technical results. These results are collected in Lemmas 2 and 3 below, which might be of independent interest. While new to the literature, they are primarily technical, hence we postpone their proofs to Appendix B. Lemma 2 deals with the uniform boundedness of certain operator norms. Specifically, it demonstrates that if $\rho(T_{\mathrm{hom}}) < 1$, then for d sufficiently large, and hence $\mathrm{mesh}(\mathcal{P}^d)$ sufficiently small, the approximating processes $\tilde{N}^d$ are stable, and in particular have finite expected total offspring, for every immigrant.

LEMMA 2. *Grant Assumptions 1, 2, 3 and 4. Let N be a stable linear graphon Hawkes process satisfying the conditions of Theorem 1, with multivariate approximating processes on weighted graphs $\tilde{N}^d$ as described by* (21)–(23)*, and with corresponding integral operators $\tilde{T}_{\mathrm{hom}}^{(d)}$. Let $\delta := \frac{1}{2}(1-\rho(T_{\mathrm{hom}}))$. Then we can find some $D' \in \mathbb{N}$ such that for $d \geq D'$, $\tilde{N}^d$ satisfies the conditions of Theorem 1, with $\rho(\tilde{T}_{\mathrm{hom}}^{(d)}) \leq \rho(T_{\mathrm{hom}}) + \delta < 1$.*

Furthermore, we can find $D \in \mathbb{N}$ such that the expected cluster size $\mathbb{E}[Z_x^d]$ for $\tilde{N}^d$ of a particle in $x \in \mathscr{X}$ is uniformly bounded over $x \in \mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}]$ and $d \geq D$. Denote this uniform bound by $\mathfrak{K}$.

The next result bounds the L^1-norm between a function and its piecewise constant approximation by the total variation of that function multiplied with the mesh of the corresponding partition. This result, which can be viewed as a suitable version of the Poincaré inequality, may be interesting in its own right, in that it has the potential to be applied more broadly.

LEMMA 3. *Let $\mathscr{X} = \prod_{i=1}^m [a_i, b_i) \subset \mathbb{R}^m$ be some hyperrectangle. Assume that $\mathscr{X}$ shares its boundaries with elements from the partition $\mathcal{P}^d$ of $\mathbb{R}^m$ into $K \in \mathbb{N}$ bounded hyperrectangles $\mathscr{X}_i^d$, $i \in [K]$; if this is not the case, replace $\mathcal{P}^d$ by its coarsest refinement such that the boundaries of $\mathscr{X}$ are also boundaries of elements of $\mathcal{P}^d$. Consider a piecewise constant approximation*

$$f^d|_{\mathscr{X}_n^d} \equiv \frac{1}{\mathrm{Leb}^m(\mathscr{X}_n^d)} \int_{\mathscr{X}_n^d} f(x)\, \mathrm{d}x$$

of $f \in L^1_{\mathrm{loc}}(\mathbb{R}^m)$, which is assumed to be of bounded variation over bounded sets. Then it holds that $f^d \in L^1(\mathscr{X}, \mu)$, $f^d \to f$ a.e. as $\mathrm{mesh}(\mathcal{P}^d) \to 0$, and finally,

$$\|f^d - f\|_{L^1(\mathscr{X},\mu)} \leq \frac{1}{2} \mathrm{Var}(f, \mathscr{X}) \mathrm{mesh}(\mathcal{P}^d). \tag{31}$$

In the case of $\mathscr{X} \subset \mathbb{R}$, we can state a version of Lemma 3 that is slightly more general, and admits an elementary proof. See Lemma 7 in the appendix.

REMARK 7. *Instead of partitioning $\mathbb{R}^m$ into hyperrectangles, as prescribed by Assumption 3, we could take a partition of $\mathbb{R}^m$ into* convex *sets in such a way that the mesh of the partition tends to* 0*, as $d \to \infty$. The particular choice of the tesselation does not alter our analysis. Indeed, in such a scenario, [1], Theorem 3.2, still holds, hence the proof of Lemma 3 can be adapted to different tesselations.*

4.4. *Annealed case.* In this subsection, we prove convergence in the annealed case, where the prelimit $\tilde{N}^d$ is a multivariate Hawkes process on a complete weighted graph $\tilde{W}^d$, with parameters obtained as averages of the parameters of N over the sets of a partition $\mathcal{P}^d$ of $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}]$. As a corollary, we show that any sufficiently regular sequence of mutually exciting point processes with suitably convergent parameters (invoking the cut metric) converges to a graphon Hawkes process. The quenched case is studied in the next subsection.

THEOREM 2. *Grant Assumptions 1, 2, 3 and 4. Let N be a linear graphon Hawkes process on $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^m$ and time interval $[0,1]$, starting on an empty history, satisfying the conditions of Theorem 1, with multivariate approximating processes on weighted graphs $\tilde{N}^d$ as described by (20)–(23), with corresponding integral operators $\tilde{T}^{(d)}_{\text{hom}}$. Using the metric $\mathfrak{d}$ from Definition 5, it holds that $\tilde{N}^d \xrightarrow{\text{ucp}} N$, as $d \to \infty$.*

PROOF. Under Assumptions 1, 2, 3 and 4, as in Lemma 2, for $\delta := \frac{1}{2}(1 - \rho(T_{\text{hom}}))$, let $D \in \mathbb{N}$ be such that: (i) for all $d \geq D$, $\rho(\tilde{T}^{(d)}_{\text{hom}}) \leq \rho(T_{\text{hom}}) + \delta < 1$; and (ii) there exists $\mathfrak{K} < \infty$ bounding the expected cluster size for N and the expected cluster size $\mathbb{E}[Z^d_x]$, uniformly over $x \in \mathscr{X}$, $d \geq D$.

The prelimit $\tilde{N}^d$ is a multivariate Hawkes process, i.e., a graphon Hawkes process with piecewise constant parameters on the sets of the partition $\mathcal{P}^d$. In particular, it is defined on the same space as N. Hence, it is possible to couple those processes by defining them w.r.t. the same Poisson random measure; cf. (6). We assume such coupled sample paths, and we bound the expected discrepancy between N and $\tilde{N}^d$, in the metric $\mathfrak{d}$, for d large. Since $\mathfrak{d}(N_{\mathscr{A}'}, M_{\mathscr{A}'}) \leq \mathfrak{d}(N_{\mathscr{A}}, M_{\mathscr{A}})$ for $\mathscr{A}' \subset \mathscr{A}$ for this metric, it suffices to consider discrepancies on the whole space $[\boldsymbol{a}, \boldsymbol{b}]$. Fix $\delta, \epsilon > 0$. We use the bound on the expected discrepancies to bound the probability of a discrepancy larger than δ by ϵ; more formally, we show that there exists d such that $\mathbb{P}\left(\mathfrak{d}(\tilde{N}^d, N) > \delta\right) < \epsilon$, thus establishing ucp convergence.

The discrepancy between the two coupled processes $\tilde{N}^d$ and N in the metric $\mathfrak{d}$ consists of several parts. For each simultaneous event, we compare the difference in location and the difference in $L^1(\mathscr{X})$-norm between the marks corresponding to those locations. Next, for each non-simultaneous event the contribution to the $\mathfrak{d}$-distance equals unity, plus the distance of the spatial coordinate to the origin, plus the $L^1(\mathscr{X})$-norm of the mark. By [48], Lemma 1, those non-simultaneous events correspond to events of

$$|\tilde{N}^d - N|(\mathrm{d}t \times \mathrm{d}z \times \mathrm{d}x) := \bar{N}(\mathrm{d}t \times \mathrm{d}z \times \mathrm{d}x \times [0, |\tilde{\lambda}^d_t(x) - \lambda_t(x)|]), \tag{32}$$

where $\bar{N}$ is a Poisson random measure on $\mathbb{R} \times L \times \mathscr{X} \times \mathbb{R}_+$ with intensity measure $\mathrm{d}t \times \mathcal{Q}(\mathrm{d}z) \times \mathrm{d}x \times \mathrm{d}s$.

To bound the expected discrepancy in the $\mathfrak{d}$-metric, we consider differences in the parameters that contribute to the discrepancies between the two processes N and $\tilde{N}^d$. A convenient way to do this is by employing the branching structure of the *linear* graphon Hawkes process. Each contribution to the discrepancy in $\mathfrak{d}$-distance is of exactly one of the following types:

(i) The difference $\|\lambda_\infty - \tilde{\lambda}^d_\infty\|_{L^1([\boldsymbol{a},\boldsymbol{b}])}$ in baseline intensities creates a stream of nonsimultaneous events, each of which creates a cluster of events for either N or $\tilde{N}^d$.
(ii) Because of averaging, $B_{\cdot y}$ differs from $\tilde{B}^d_{\cdot y'}$, and $W(\cdot, y)$ differs from $\tilde{W}^d(\cdot, y')$. For each simultaneous event with spatial coordinates y and y' of N and $\tilde{N}^d$, respectively, this creates additional events. Again, each of those additional events generates clusters of events for either N or $\tilde{N}^d$.
(iii) Even when we have a simultaneous event for both processes, the point from $\tilde{N}^d$ is located differently within $\mathscr{X}^d_i$ when compared to the point from N. Because all parameters of $\tilde{N}^d$ are constant on $\mathscr{X}^d_i$, this does not influence the future evolution of the process, although it affects $\mathfrak{d}$. Both the location and the mark may be different, causing discrepancies.

In the following, we formally bound (i)–(iii) separately.

Fix $\epsilon' > 0$. We start by bounding the discrepancy caused by (i), i.e., the events generated by differences in baseline intensities, and offspring thereof. Suppose that $d_1 \geq D$ is sufficiently large such that $\mathrm{mesh}(\mathcal{P}^d) \leq 2\epsilon'/\|\lambda_\infty\|_{\mathrm{TV}(\mathscr{X})}$. Then, $\|\lambda_\infty - \tilde{\lambda}^d_\infty\|_{L^1(\mathscr{X})} \leq \epsilon'$ by Lemma 3. Each of the events caused by the difference in baseline intensities creates a cluster of expected size bounded by $\mathfrak{K}$. Therefore, on the whole space $\mathscr{X}$, the difference in baseline intensities causes a stream of nonsimultaneous events with an expected count over our time frame $[0, 1]$ bounded by $\epsilon'\|\lambda_\infty\|_{\mathrm{TV}(\mathscr{X})}\mathfrak{K}$. The contribution to $\mathfrak{d}$ of each of those events can be bounded by $C_\mathfrak{d} := 1 + |\boldsymbol{a}| \vee |\boldsymbol{b}| + C_B$. Hence, by choosing

$$\epsilon' = \frac{\delta\epsilon}{3\|\lambda_\infty\|_{\mathrm{TV}(\mathscr{X})}\mathfrak{K}C_\mathfrak{d}},$$

we can assure that the discrepancy in the metric $\mathfrak{d}$ caused by (i) can be bounded by $\delta\epsilon/3$.

Next, we bound the discrepancy in $\mathfrak{d}$-distance from (ii), i.e., discrepancies caused by differences between B and $\tilde{B}^d$ and between W and $\tilde{W}^d$, for each simultaneous event of N and $\tilde{N}^d$. To this end, first note that the number of events for N on the time interval $[0, 1]$, and therefore the expected number of simultaneous events of N and $\tilde{N}^d$ on $[0, 1]$, can be bounded by $\alpha\mathfrak{K}$, where $\alpha := \int_\mathscr{X} \lambda_\infty(x)\,\mathrm{d}x$, as in (11). For each of those events, there is a difference in mark and in location between N and $\tilde{N}^d$. By the construction presented in Section 3, we know that the locations of a simultaneous point for N and $\tilde{N}^d$ lie in the same element $\mathscr{X}^d_n$ of the partition $\mathcal{P}^d$. Furthermore, since the parameters of $\tilde{N}^d$ are constant on the sets $\mathscr{X}^d_n$, it follows that the location of an event for $\tilde{N}^d$ *within* $\mathscr{X}^d_n$ does not influence future dynamics, meaning that if we want to bound the expected number of events caused by differences between $B_{\cdot y}$ and $\tilde{B}^d_{\cdot y'}$ and between $W(\cdot, y)$ and $\tilde{W}^d(\cdot, y')$ for a simultaneous event for N and $\tilde{N}^d$, we may in fact assume that $y = y'$. Now for each simultaneous event of N and $\tilde{N}^d$, the difference between mark and graphon parameters and between locations causes an offspring with an expected count bounded by

$$\begin{aligned}
&\sup_{y\in\mathscr{X}} \int_\mathscr{X} \mathbb{E}\left|B_{xy}W(x,y) - \tilde{B}^d_{xy}\tilde{W}^d(x,y)\right| \,\mathrm{d}x \\
&\leq \sup_{y\in\mathscr{X}} \int_\mathscr{X} \mathbb{E}\left|B_{xy} - \tilde{B}^d_{xy}\right| W(x,y)\,\mathrm{d}x + \sup_{y\in\mathscr{X}} \int_\mathscr{X} \mathbb{E}\left[\tilde{B}^d_{xy}\right] \left|W(x,y) - \tilde{W}^d(x,y)\right| \,\mathrm{d}x \\
&\leq \mathrm{mesh}(\mathcal{P}^d)\left(C_W \sup_{y\in\mathscr{X}} \mathbb{E}\|B_{\cdot y}\|_{\mathrm{TV}(\mathscr{X})} + C_B \sup_{y\in\mathscr{X}} \|W(\cdot, y)\|_{\mathrm{TV}(\mathscr{X})}\right)/2,
\end{aligned}$$

using Assumption 2 and Lemma 3. In turn, each of those events generates a cluster of further events with expected size bounded by $\mathfrak{K}$, with each event contributing to $\mathfrak{d}$ at most $C_\mathfrak{d} :=$

$1 + |\boldsymbol{a}| \vee |\boldsymbol{b}| + C_B$. Hence, selecting $d_2 \geq D$ such that

$$\mathrm{mesh}(\mathcal{P}^d) < \frac{2\delta\epsilon}{3\alpha\mathfrak{K}^2 \left(C_W \sup_{y\in\mathscr{X}} \mathbb{E}\|B_{\cdot y}\|_{\mathrm{TV}(\mathscr{X})} + C_B \sup_{y\in\mathscr{X}} \|W(\cdot,y)\|_{\mathrm{TV}(\mathscr{X})}\right) C_{\mathfrak{d}}},$$

guarantees that the expected discrepancy between N and $\tilde{N}^d$ during $[0,1]$ caused by (ii) above is less than $\delta\epsilon/3$.

Finally, we bound the expected discrepancy in $\mathfrak{d}$-distance caused by (iii), i.e., the difference in contributions to $\mathfrak{d}$ for simultaneous events of N and $\tilde{N}^d$ through different locations and marks. Again, the expected number of simultaneous events of N and $\tilde{N}^d$ on $[0,1]$ can be bounded by $\alpha\mathfrak{K}$. For each of those events, the locations are within the same element $\mathscr{X}_n^d$ of the partition $\mathcal{P}^d$, hence the difference in locations contributes at most $\mathrm{mesh}(\mathcal{P}^d)$ to the distance $\mathfrak{d}(N,\tilde{N}^d)$, while by Lemma 3 the difference in marks causes an expected discrepancy bounded by $\mathrm{mesh}(\mathcal{P}^d)\sup_{y\in\mathscr{X}} \mathbb{E}\|B_{\cdot y}\|_{\mathrm{TV}(\mathscr{X})}/2$. Hence, selecting $d_3 \geq D$ such that

$$\mathrm{mesh}(\mathcal{P}^d) < \frac{2\delta\epsilon}{3\alpha\mathfrak{K}(2+\sup_{y\in\mathscr{X}} \mathbb{E}\|B_{\cdot y}\|_{\mathrm{TV}(\mathscr{X})})},$$

guarantees that the expected discrepancy in $\mathfrak{d}$-distance between N and $\tilde{N}^d$ caused by differences in locations and mark sizes is less than $\delta\epsilon/3$.

Combining the bounds for the contributions (i)–(iii), it follows that $\mathbb{E}[\mathfrak{d}(\tilde{N}^d,N)] < \delta\epsilon$ for $d \geq \max\{d_1,d_2,d_3\}$, hence by Markov's inequality, $\mathbb{P}\left(\mathfrak{d}(\tilde{N}^d,N) > \delta\right) < \epsilon$. □

The essential ingredients in the proof of Theorem 2 are the L^1-convergence of the parameters of the prelimit to those of the graphon Hawkes process, and that the prelimit has uniformly bounded cluster sizes. Hence, by imposing appropriate conditions on the parameters of the prelimit, we can state a converse to Theorem 2, which holds for prelimits obtained in different ways than through (20)–(23).

To state such a converse, we want to use the natural notion of graphon convergence, which is convergence in the cut metric, see [47], Chapter 8. With $\mathcal{B}(\mathscr{X})$ the Borel σ-algebra on $\mathscr{X}$, the cut norm of W is defined by

$$\|W\|_{\square} := \sup_{S,T\in\mathcal{B}(\mathscr{X})} \left| \int_{S\times T} W(x,y)\,\mathrm{d}x\,\mathrm{d}y \right|,$$

and the corresponding cut metric is defined by

$$d_{\square}(W_1,W_2) := \|W_1 - W_2\|_{\square}.$$

In applications (e.g., [2, 7, 8]), the cut norm is often connected to more naturally appearing norms, by an application of [47], Lemma 8.11, which states that

$$\sup_{\|g\|_{L^\infty(\mathscr{X})}=1} \int_{\mathscr{X}} \left| \int_{\mathscr{X}} W(x,y) g(y)\,\mathrm{d}y \right| \mathrm{d}x =: \|W\|_{L^\infty(\mathscr{X})\to L^1(\mathscr{X})} \leq 4\|W\|_{\square}.$$

The following corollary establishes that under suitable assumptions, for any sequence of graphs $\breve{W}^d$ on d nodes such that the associated products $\mathbb{E}[\breve{B}^d]\breve{W}^d$ converge in the cut metric to $\mathbb{E}[B]W$, the corresponding d-variate processes N^d converge to the graphon Hawkes process N associated with the graphon W. The proof of the corollary is postponed to Appendix B.

COROLLARY 1. *Let $(N^d)_{d\in\mathbb{N}}$ be a sequence of marked d-variate Hawkes processes. Embed each of the N^d in the space of graphon Hawkes processes by taking a sequence of* *partitions $\mathcal{P}^d$ of $\mathscr{X}$ satisfying Assumption 3, letting coordinate i correspond to $\mathscr{X}_i^d$, and taking the parameters $\breve{\lambda}_\infty^d, \breve{W}^d, \breve{B}^d$ of the graphon Hawkes process $\breve{N}^d$ corresponding to N^d piecewise constant on $\mathscr{X}_i^d$, $\mathscr{X}_i^d \times \mathscr{X}_j^d$, respectively, and such that (24)–(27) hold.*

Suppose that $\mathscr{X}$, $\mathcal{P}^d$, λ_∞, W and B can be chosen such that:

(a) *λ_∞, W, B satisfy Assumptions 1 and 2;*
(b) *$\breve{\lambda}_\infty^d \to \lambda_\infty$ in $L^1(\mathscr{X})$, as $d \to \infty$;*
(c) *$\mathbb{E}[\breve{B}^d]\breve{W}^d \to \mathbb{E}[B]W$ in the cut metric, as $d \to \infty$;*
(d) *$C := \inf_{z\in\mathscr{X}} \int_{\mathscr{X}} B_{xz}W(x,z)\,\mathrm{d}x > 0$, a.s. over the realization of B;*
(e) *The expected cluster size $\mathbb{E}[Z_x^d]$ is uniformly bounded in $x \in \mathscr{X}, d \in \mathbb{N}$.*

Then, using the metric $\mathfrak{d}$ from Definition 5, it holds that $(N^d)_{d\in\mathbb{N}} \xrightarrow{\text{ucp}} N$, as $d \to \infty$.

REMARK 8. *Condition (d) of Corollary 1 requires that the expected offspring resulting from a particle in $z \in \mathscr{X}$ is bounded away from 0, uniformly over $z \in \mathscr{X}$. Condition (e) is satisfied if we assume that there is a $\delta \in (0,1)$ such that*

$$\|h\|_{L^1(\mathbb{R}_+)} \sup_{y\in\mathscr{X}} \int_{\mathscr{X}} \mathbb{E}[\breve{B}_{xy}^d]\breve{W}^d(x,y)\,\mathrm{d}x \le 1-\delta$$

uniformly over $d \in \mathbb{N}$. This implies that the expected cluster size is bounded uniformly by $\sum_{n\ge 0}(1-\delta)^n = 1/\delta$; see also Remark 4.

4.5. *Quenched case.* In this subsection, we consider the quenched case, where the prelimit $\bar{N}^d$ is a multivariate Hawkes process on a looped, directed graph $\mathbb{G}^d$ with edges sampled from $(W_{ij}^d)_{i,j\in[d]}$.

THEOREM 3. *Work in the setting of Theorem 2, with $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}] \subset \mathbb{R}^m$ and time interval $[0,1]$, but different from the annealed process of Theorem 2, consider the conditional intensity density $\bar{\lambda}_t^d(\cdot)$ of the quenched process $\bar{N}^d$ given by*

$$\bar{\lambda}_t^d(x) = \tilde{\lambda}_\infty^d(x) + \sum_{\substack{(s,y,\tilde{B}_{xy}^d)\in\bar{N}^d \\ s<t}} \tilde{B}_{xy}^d(s)\tilde{Z}^d(x,y)h(t-s), \tag{33}$$

with parameters as defined in (21), (22) and (30). Then, using the metric $\mathfrak{d}$ from Definition 5, it holds that $\bar{N}^d \xrightarrow{\text{ucp}} N$, as $d \to \infty$. Here, the uniform convergence in probability *is also w.r.t. the realization of the connectivity graph.*

PROOF. As observed before, we may assume without loss of generality that $C_W \le 1$. Fix $\delta, \epsilon > 0$. Let $\tilde{N}^d$ denote the annealed prelimit on a weighted complete directed graph of Theorem 2, so that $\bar{N}^d$ is the quenched prelimit on a looped directed graph sampled from this complete graph. The idea of this proof is to argue that for large d, with high probability, no set $\mathscr{X}_n^d$ from $\mathcal{P}^d$ contains more than one event from $\tilde{N}^d$, and conditionally on this high probability event, the probabilistic behavior (including stability) of $\bar{N}^d$ is the same as that of $\tilde{N}^d$. Indeed, suppose that there is at most one event in each set $\mathscr{X}_n^d$. Given an arrival in some $\mathscr{X}_j^d \ni y$, first sampling an edge $Z_{ij}^d \sim \text{Bernoulli}(W_{ij}^d)$ to $\mathscr{X}_i^d \ni x$, and in case of an edge (i,j) generating the offspring according to $\tilde{B}_{xy}^d(s)h(\cdot - s)$, is equivalent to thinning the increase of intensity by $\tilde{W}_{ij}^d$, i.e., to generating offspring according to $\tilde{B}_{xy}^d(s)\tilde{W}_{ij}^d h(\cdot - s)$.

To prove ucp convergence, let $\delta \in (0,1)$. For such δ, on the event $\{\mathfrak{d}(\tilde{N}^d, N) \le \delta\}$, the processes $\tilde{N}^d$ and N only have simultaneous events, by definition of $\mathfrak{d}$. Suppose that the partition $\mathcal{P}^d$ has no element $\mathscr{X}_n^d$ in which more than one event for N occurs during $[0,1]$.

Then, conditionally on $\{\mathfrak{d}(\tilde{N}^d, N) \le \delta\}$, the same property holds with N replaced by $\tilde{N}^d$, since a simultaneous event of N and $\tilde{N}^d$ has a location within the same element $\mathscr{X}_n^d$ of the partition $\mathcal{P}^d$. We focus on taking d sufficiently large such that this property holds for $\mathcal{P}^d$ and N.

As in Theorem 2, the expected cluster size of N is bounded by $\mathfrak{K}$, hence N has an expected number of events on $[0,1]$ that is bounded by $\alpha\mathfrak{K} < \infty$. Thus, $N(\omega)$, the realization of N during time interval $[0,1]$, is finite a.s. This implies that $\ell(\omega) := \inf_{\boldsymbol{x},\boldsymbol{y} \in N(\omega)} \|\boldsymbol{x} - \boldsymbol{y}\| > 0$ a.s. For $d \ge 1$ such that $\mathrm{mesh}(\mathcal{P}^d) < \ell(\omega)/2$, all elements of $N(\omega)$ are in different elements $\mathscr{X}_n^d$ from the partition $\mathcal{P}^d$. This property then carries over to $d' \ge d$, since $\mathrm{mesh}(\mathcal{P}^d) \to 0$ monotonically, as $d \to \infty$, by Assumption 3. For $d \in \mathbb{N}$, let

$$A_d := \left\{ \omega \in \Omega : \mathrm{mesh}(\mathcal{P}^d) < \frac{\ell(\omega)}{2} \right\}.$$

Then, $\bigcup_{d \in \mathbb{N}} A_d = \Omega$ (up to a null set), and $A_d \uparrow \Omega$ (up to a null set). Hence, by continuity of probability measures, $\mathbb{P}(A_d) \uparrow 1$ as $d \to \infty$. Select $d_4 \in \mathbb{N}$ such that $\mathbb{P}(A_{d_4}) \ge 1 - \epsilon$; then for $d \ge d_4$ it holds that $\mathbb{P}(A_d) \ge 1 - \epsilon$. In the proof of Theorem 2, we showed that $\mathbb{P}(\mathfrak{d}(\tilde{N}^d, N) \le \delta) \ge 1 - \epsilon$ for $d \ge \max\{d_1, d_2, d_3\}$. Hence, for $d \ge \max\{d_1, d_2, d_3, d_4\}$,

$$\mathbb{P}(\{\mathfrak{d}(\tilde{N}^d, N) \le \delta\} \cap A_d) \ge 1 - 2\epsilon.$$

Since conditionally on A_d, the probabilistic behavior of $\bar{N}^d$ and $\tilde{N}^d$ is identical, we have

$$\mathbb{P}(\mathfrak{d}(\bar{N}^d, N) \le \delta) \ge \mathbb{P}(\{\mathfrak{d}(\bar{N}^d, N) \le \delta\} \cap A_d) = \mathbb{P}(\{\mathfrak{d}(\tilde{N}^d, N) \le \delta\} \cap A_d) \ge 1 - 2\epsilon,$$

which establishes the ucp convergence $\bar{N}^d \xrightarrow{\text{ucp}} N$ as $d \to \infty$. □

We have shown that the graphon Hawkes process can be seen as a continuous-space limit of multivariate Hawkes processes. The next proposition, the proof of which is postponed to Appendix B, shows that the graphon corresponding to N also occurs as the limit of the connectivity graphs corresponding $\bar{N}^d$.

PROPOSITION 1. *Work in the setting of Theorem 3, with* $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}] = \prod_{n=1}^d [a_n, b_n]$, *and grant Assumption 5. Let N be a graphon Hawkes process on W with approximating multivariate Hawkes processes $(\bar{N}^d)_{d \in \mathbb{N}}$ with edges sampled from $(W_{ij}^d)_{i,j \in [d]}$, where the parameters are taken piecewise constant on the intervals of the partition $\mathcal{P}^d$. This partition can be chosen in such a way that $\mathbb{P}$-a.s., the connectivity graph $\mathbb{G}^d$ of $\bar{N}^d$ converges to W in the cut metric, while $\bar{N}^d$ converges ucp to N, as $d \to \infty$.*

REMARK 9. *The construction of this subsection essentially works with digraphons and looped, directed graphs $\mathbb{G}^d$. If we intend to use ‘standard’ graphons and simple graphs, the analysis of this section can be adapted in order to sample edges only for $i < j$.*

5. Large-time behavior. We now specialize to a graphon Hawkes process with spatial coordinates in $(\mathscr{X}, \mathscr{A}, \mu) = ([0,1], \mathcal{B}[0,1], \mathrm{Leb}|_{[0,1]})$. This is for legibility only: as is easily verified, all arguments in this section and in Section 6 only rely on $\mathscr{X}$ being a compact subset of an Euclidean space $\mathbb{R}^m$, and in particular do not require $\mathscr{X}$ to be one-dimensional.

We investigate the large-time behavior of the graphon Hawkes process. First, we prove a functional law of large numbers (FLLN) in the stable case where $\rho(T_{\text{hom}}) < 1$, and show that the corresponding prelimit diverges in the case $\rho(T_{\text{hom}}) > 1$, thus providing a dichotomy between these two cases. Having established the FLLN in the stable case, we next prove a functional central limit theorem (FCLT) for the number of events $(N_t(A))_{t \ge 0}$ for $A \in \mathcal{B}[0,1]$.

We establish those large-time results with the aid of limit theorems for finite-dimensional Hawkes processes.

To obtain explicit results, we work in this section with *unmarked* processes, i.e., we set $B \equiv 1$ in (2). Our approach is to approximate N by multivariate Hawkes processes $\tilde{N}^d$ using Theorem 2, to which we apply explicit limit theorems for the unmarked case from [5]. In principle, it is also possible to consider functional limit theorems for marked Hawkes processes. Such limit theorems, however, often assume univariate marked processes ([32, 37]). If we model multivariate processes as marked univariate processes, we enforce a single common intensity function with i.i.d. coordinates. Other papers — e.g., [24] — work with multivariate processes, but obtain less explicit results.

5.1. *Functional law of large numbers.* Under the condition $\rho(T_{\text{hom}}) < 1$, we expect that $N_T[0,1] = \mathcal{O}(T)$, as $T \to \infty$, by Theorem 1. Therefore, we consider $N_T(A)/T$ for large T and for $A \in \mathcal{B}[0,1]$, with the objective to establish an FLLN for $(N_{Tv}(A)/T)_{v\in[0,1]}$. For the ergodic average of the number of events in any Borel set $A \in \mathcal{B}[0,1]$, we will argue that

$$\left(\frac{N_{Tv}(A)}{T}\right)_{v\in[0,1]} \to (v\bar{\lambda}(A))_{v\in[0,1]},$$

where $\bar{\lambda}(A) := \int_A \bar{\lambda}(x)\,\mathrm{d}x$, with $\bar{\lambda}(x)$ the *expected stationary arrival rate density*, meaning that it solves the Fredholm integral equation of the second kind $\bar{\lambda} = \lambda_\infty + T_{\text{hom}}(\bar{\lambda})$, i.e.,

$$\bar{\lambda}(x) = \lambda_\infty(x) + \|h\|_{L^1(\mathbb{R}_+)} \int_0^1 W(x,y)\bar{\lambda}(y)\,\mathrm{d}y. \tag{34}$$

This is motivated by the following heuristic, supposed to hold for T large and for $A \in \mathcal{B}[0,1]$:

$$\begin{aligned}
\frac{\mathbb{E}[N_{Tv}(A)]}{T} &= \frac{1}{T}\int_A \int_0^{Tv} \mathbb{E}[\lambda_t(x)]\,\mathrm{d}t\,\mathrm{d}x \\
&= \frac{1}{T}\int_A \int_0^{Tv} \left[\lambda_\infty(x) + \int_0^1 \int_0^t W(x,y)h(t-s)\mathbb{E}[\lambda_s(y)]\,\mathrm{d}s\,\mathrm{d}y\right]\mathrm{d}t\,\mathrm{d}x \\
&\approx \frac{1}{T}\int_A \int_0^{Tv} \left[\lambda_\infty(x) + \int_0^1 W(x,y)\bar{\lambda}(y)\,\mathrm{d}y\,\|h\|_{L^1(\mathbb{R}_+)}\right]\mathrm{d}t\,\mathrm{d}x \\
&= \frac{1}{T}\int_A \int_0^{Tv} \bar{\lambda}(x)\,\mathrm{d}t\,\mathrm{d}x = v\bar{\lambda}(A),
\end{aligned}$$

where in stationarity we would expect $\mathbb{E}[\lambda_s(y)]$ to be independent of s, in which case we denote this quantity by $\bar{\lambda}(y)$. In the remainder of this subsection, we make this heuristic exact.

THEOREM 4. *Grant Assumptions 1, 2, 3 and 4. Let $N = (N_t(A))_{t\in\mathbb{R}_+, A\in\mathcal{B}[0,1]}$ be an unmarked (i.e., $B \equiv 1$) linear graphon Hawkes process on $\mathscr{X} = [0,1]$ and time interval $\mathbb{R}_+$, starting on an empty history, and satisfying the conditions of Theorem 1. In particular, suppose $\rho(T_{\text{hom}}) < 1$, where in the present setting*

$$T_{\text{hom}} : L^1[0,1] \to L^1[0,1] : f(\cdot) \mapsto \|h\|_{L^1(\mathbb{R}_+)} \int_0^1 W(\cdot,y)f(y)\,\mathrm{d}y. \tag{35}$$

Let $\bar{\lambda}$ be defined in (34). Then,

$$\sup_{v\in[0,1]} \left|\frac{N_{Tv}(A)}{T} - v\bar{\lambda}(A)\right| \xrightarrow{\mathbb{P}} 0, \quad \textit{as } T \to \infty. \tag{36}$$

PROOF. The proof is structured as follows. We first approximate the graphon Hawkes process by d-dimensional Hawkes processes using Theorem 2; then we use an existing FLLN for this prelimit; and we demonstrate that the limit quantities for graphon Hawkes and multivariate Hawkes processes are close, for large d.

Thus, we first approximate the graphon Hawkes process N by processes $\tilde{N}^d$ on weighted graphs, described by (20)–(23), having constant parameters on each element $\mathscr{X}_n^d$ ($n \in [d]$) of the partition $\mathcal{P}^d$ of $[0,1]$ into d intervals, with $\mathrm{mesh}(\mathcal{P}^d) \to 0$ as $d \to \infty$. Denote the corresponding integral operators by $\tilde{T}_{\text{hom}}^{(d)}$. By Lemma 2 (which requires Assumptions 1, 2, 3 and 4), we can find some $D \in \mathbb{N}$ such that for $d \geq D$, $\tilde{N}^d$ is stable with $\rho(\tilde{T}_{\text{hom}}^{(d)}) \leq \rho(T_{\text{hom}}) + \delta < 1$, where $\delta := \frac{1}{2}(1 - \rho(T_{\text{hom}}))$, and such that the expected cluster size $\mathbb{E}[Z_x^d]$ for $\tilde{N}^d$ of a particle in $x \in [0,1]$ is bounded by $\mathfrak{K} < \infty$, uniformly over x and $d \geq D$. This means that the FLLN [5], Theorem 1, holds for $\tilde{N}^d$, for all $d \geq D$.

For $A \in \mathcal{B}[0,1]$, let $(N_t(A))_{t \geq 0}$ be the simple counting process keeping track of the number of points in A over time. Let $\tilde{K}^d = (\|h\|_{L^1(\mathbb{R}_+)} W_{ij}^d)_{i,j \in [d]}$ (see (27) and [5], Assumption (**A1**)), and let I_d be the $d \times d$ identity matrix. Then it holds that

$$\sup_{v \in [0,1]} \left| \frac{N_{Tv}(A)}{T} - v\bar{\lambda}(A) \right| \leq \sup_{v \in [0,1]} \left| \frac{N_{Tv}(A)}{T} - \frac{\tilde{N}_{Tv}^d(A)}{T} \right| \tag{37}$$

$$+ \sup_{v \in [0,1]} \left| \frac{\tilde{N}_{Tv}^d(A)}{T} - v(I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d(A) \right| \tag{38}$$

$$+ \sup_{v \in [0,1]} \left| v(I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d(A) - v\bar{\lambda}(A) \right|, \tag{39}$$

where $(I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d(A) := \int_A ((I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d)(x)\, \mathrm{d}x$. We aim to prove that the left-hand side of the inequality in the above display goes to 0 in probability as $T \to \infty$, by first selecting $\bar{d}$ large enough such that the right-hand side of (37) and (39) are small for $d \geq \bar{d}$, independently of $T \geq 1$, and next selecting T large enough such that (38) is small.

Fix $\epsilon, \kappa \in (0,1)$. We start by bounding (37). Suppose that $T \geq 1$. Let $\mathfrak{d}(\cdot,\cdot)$ be the metric from Definition 5. If $\mathfrak{d}(N_T, \tilde{N}_T^d) < 2T\epsilon'$, then there are fewer than $2T\epsilon'$ non-simultaneous events on $[0,T]$ for both processes, and this leads to the bound

$$\sup_{v \in [0,1]} \left| \frac{N_{Tv}(A)}{T} - \frac{\tilde{N}_{Tv}^d(A)}{T} \right| < 2\epsilon'.$$

Hence, we focus on establishing that $\mathfrak{d}(N_T, \tilde{N}_T^d) < 2T\epsilon'$. To this end, let M be a process corresponding to N, having the same immigrant streams of events $I_1 < I_2 < \cdots$, but where to each immigrant I_k arriving at time t, at this time t we immediately attach the immigrant's complete offspring $O_k \subset (t, \infty)$. More precisely, $M_T = \bigcup_{I_k \leq T} O_k$. Define $\tilde{M}^d$ similarly through $\tilde{N}^d$. This means that the event times of $M_T, \tilde{M}_T^d$ are discrete subsets of $[0, \infty)$, marked by discrete subsets of (t, ∞) for an immigrant arrival at time t. Informally, M, $\tilde{M}^d$ consist of the immigrant streams of N, $\tilde{N}^d$, with each immigrant replaced by a batch of size equal to the whole cluster corresponding to that immigrant.

Let D be the process defined through $D_T = \bigcup_{I_k \leq T} (O_k \cap (T, \infty))$, where $(I_k)_{k \in \mathbb{N}}$, $(O_k)_{k \in \mathbb{N}}$ are the immigrant times and offspring sets corresponding to N; let D^d be defined similarly through N^d. Then we have, a.s.,

$$\mathfrak{d}\left(N_T, \tilde{N}_T^d\right) \leq \mathfrak{d}\left(N_T, \tilde{N}_T^d\right) + \mathfrak{d}\left(D_T, \tilde{D}_T^d\right) = \mathfrak{d}\left(M_T, \tilde{M}_T^d\right). \tag{40}$$

Furthermore, writing M_A for the process only recording the immigrants arriving in $A \in \mathcal{B}(\mathbb{R}_+)$ — meaning that $M_T \equiv M_{[0,T]}$ —, we easily obtain the a.s. bound

$$\mathfrak{d}\left(M_T, \tilde{M}_T^d\right) \leq \sum_{i=1}^{\lceil T \rceil} \mathfrak{d}\left(M_{[i-1,i)}, \tilde{M}_{[i-1,i)}^d\right). \tag{41}$$

Next, note that clusters arrive independently over time, and that the clusters themselves are i.i.d., modulo the time shift corresponding to the immigrant arrival time. That is, $(M_{[i-1,i)}, \tilde{M}_{[i-1,i)}^d)_{i\in\mathbb{N}}$ represent i.i.d. random sets. Furthermore, by the proof of Theorem 2, we can find $d_1 \geq D$ such that, for $d \geq d_1$, the expectation of $\mathfrak{d}(M_{[i-1,i)}, \tilde{M}_{[i-1,i)}^d)$ can be bounded by ϵ'. Hence, for $d \geq d_1$, $\mathfrak{d}(N_T, \tilde{N}_T^d)$ is bounded in expectation by $\lceil T \rceil \epsilon' \leq 2T\epsilon'$, thus, independently of $T \geq 1$,

$$\sup_{v\in[0,1]} \left| \frac{N_{Tv}(A)}{T} - \frac{\tilde{N}_{Tv}^d(A)}{T} \right|$$

is bounded in expectation by $2\epsilon'$. Choose $\epsilon' = \epsilon\kappa/12$. Markov's inequality implies that with probability of at least $1 - \kappa/2$, (37) is smaller than $\epsilon/3$.

We now bound (39). To this end, note that the solution $\bar{\lambda}$ to a Fredholm integral equation of the second kind depends continuously on the initial conditions, whenever there exists a unique solution; see [28], Corollary 9.3.12. This can be guaranteed under the condition $\rho(T_{\text{hom}}) < 1$, since then $\bar{\lambda}$ as defined in (34) can be expressed by a Liouville-Neumann series:

$$\bar{\lambda} = \sum_{n\geq 0} T_{\text{hom}}^n \lambda_\infty, \tag{42}$$

which is in $L^1[0,1]$ by Gelfand's formula. The continuous dependence of the solution $\bar{\lambda}$ on the initial conditions means that we can find $\eta > 0$, such that

$$\|\lambda_\infty - \tilde{\lambda}_\infty^d\|_{L^1[0,1]} < \eta \quad \text{and} \quad \|h\|_{L^1(\mathbb{R}_+)} \|W - \tilde{W}^d\|_{L^1[0,1]^2} < \eta \tag{43}$$

imply that

$$\|(I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d - \bar{\lambda}\|_{L^1[0,1]} < \frac{\epsilon}{3}.$$

Note that $(I_d - \tilde{K}^d)^{-1} \tilde{\lambda}_\infty^d$ is the unique solution to $f = \tilde{\lambda}_\infty^d + \tilde{T}_{\text{hom}}^d(f)$. By Lemma 3, there exists $d_2 \geq D$ such that (43) holds for all $d \geq d_2$. Hence, for such d, (39) is bounded by $\epsilon/3$.

Finally, given $d := d_1 \vee d_2$, we need to choose T sufficiently large to make (38) small. We achieve this by applying [5], Theorem 1, to a multivariate Hawkes process on $2d$ coordinates corresponding to the sets $\mathscr{X}_n^d \cap A$ and $\mathscr{X}_n^d \cap A^{\mathsf{c}}$, $n \in [d]$, with equal parameters on the coordinates corresponding to $\mathscr{X}_n^d \cap A$ and $\mathscr{X}_n^d \cap A^{\mathsf{c}}$, computed according to (20)–(23). Note that [5], Theorem 1, uses the Euclidean norm on $\mathbb{R}^d$. However, as all norms on a finite-dimensional vector space are equivalent, their result holds for the ℓ^1-norm as well. It follows that (38), which is bounded by its ℓ^1-norm over the $2d$ coordinates, converges to 0 in probability as $T \to \infty$. In other words, we can find $T^* \geq 1$ such that for $T \geq T^*$ it holds that, with probability of at least $1 - \kappa/2$, the expression in (38) is bounded by $\epsilon/3$.

Now we use our choice of d to combine the bounds for (37)–(39). Then, for all $T \geq T^*$,

$$\mathbb{P}\left(\sup_{v\in[0,1]} \left| \frac{N_{Tv}(A)}{T} - v\bar{\lambda}(A) \right| < \epsilon \right) \geq 1 - \kappa. \qquad \square$$

We proceed to the case $\rho(T_{\text{hom}}) > 1$, where we expect $N_T(A)/T \to \infty$ a.s., as $T \to \infty$, for each $A \in \mathcal{B}[0,1]$ of positive Lebesgue measure. We prove this claim formally below. First, however, we provide some heuristics. By Gelfand's formula, we have

$$\rho(T_{\text{hom}}) = \lim_{n\to\infty} \|T_{\text{hom}}^n\|^{1/n} = \inf_{n\in\mathbb{N}} \|T_{\text{hom}}^n\|^{1/n} > 1.$$

This implies (see Proposition 4 in Appendix C) the existence of some $g \in L^1_+[0,1]$ with $\|g\|_{L^1[0,1]} = 1$ such that $\|T_{\text{hom}}^n g\|_{L^1[0,1]} \to \infty$ as $n \to \infty$. Hence, if we can find some (Poisson) stream of particles with random spatial locations having probability density g, then each of those particles has infinite expected offspring, rendering the system unstable. This heuristic, however, while insightful, does not directly lead to a formal proof; see Proposition 5 in Appendix C for details. To prove our divergence claim, we follow a different route, for which we need some assumptions.

ASSUMPTION 6. *Assume that $\alpha = \|\lambda_\infty\|_{L^1[0,1]} > 0$. Furthermore, assume that the time from the birth of a parent to the birth of the child has finite expectation, i.e., $\int_0^\infty th(t)\,\mathrm{d}t < \infty$. Next, let*

$$W^{(k)}(x,y) := \int_0^1 \cdots \int_0^1 W(x,x_1)\cdots W(x_{k-1},y)\,\mathrm{d}x_1\cdots\mathrm{d}x_{k-1}, \tag{44}$$

and assume that there is some $k \in \mathbb{N}$ such that $W^{(k)} > 0$ on all of $[0,1]$, and such that $W^{(k)}$ is bounded away from zero, i.e., that there exists some $\kappa \in (0,1]$ such that $\inf_{x,y\in[0,1]} W^{(k)}(x,y) \geq \kappa$. Finally, assume that W is symmetric, i.e., $W(x,y) \equiv W(y,x)$.

The condition $W^{(k)} > 0$ means that the system is *mixing* after k generations; if $W^{(k)}$ is bounded away from 0, then this mixing is *uniform*. We need a mixing assumption to argue that explosive behavior in one part of the space $\mathscr{X}$ spills over to the rest of the space.

THEOREM 5. *Grant Assumptions 1, 2 and 6. Let N be an unmarked linear graphon Hawkes process on $\mathscr{X} = [0,1]$ and time interval $\mathbb{R}_+$, starting with an empty history. Suppose that we are in the regime $\rho(T_{\text{hom}}) > 1$, with T_{hom} defined in (35). Then, for each $A \in \mathcal{B}[0,1]$ of positive Lebesgue measure, it holds that*

$$\frac{N_T(A)}{T} \to \infty \quad \textit{a.s., as } T \to \infty. \tag{45}$$

PROOF. The proof is structured as follows. We first use the condition $\rho(T_{\text{hom}}) > 1$ to find a function $\phi \in L^1[0,1] \cap L^\infty[0,1]$ such that $\|T_{\text{hom}}^n \phi\|_{L^1[0,1]} \to \infty$. Since this function is *bounded*, the same should hold for the unit profile $\mathbf{1} : x \mapsto 1$. Then we use Assumption 6 to find a stream of immigrant particles that eventually lead to some uniformly distributed particle in $[0,1]$. Since the unit profile has infinite expected offspring, the same holds for our uniformly distributed particle. We conclude by applying the strong LLN.

Since $C_W < \infty$ by Assumption 2, we may apply Luzin's theorem to approximate the measurable function $W \in L^1[0,1]^2$ by continuous functions, and it follows by standard arguments that T_{hom} is a compact operator. Next, since W is symmetric, it follows that when treated as an operator on $L^p[0,1]$, $p \in [1,\infty)$, T_{hom} is self-adjoint; see [16], Example VI.1.6. Hence, $T_{\text{hom}} \in B(L^2[0,1])$ is self-adjoint, where $B(L^p[0,1])$ is the set of bounded linear operators on $L^p[0,1]$. For $T \in B(L^p[0,1])$, its spectrum is defined by

$$\sigma^{L^p}(T) := \{\lambda \in \mathbb{C} : T - \lambda I \in B(L^p[0,1]) \text{ is not invertible}\}.$$

Note that $L^2[0,1] \subset L^1[0,1]$, hence if $T_{\text{hom}} - \lambda I$ is not invertible on $L^1[0,1]$, then it is not invertible on $L^2[0,1]$ either. Thus, $\sigma^{L^1}(T_{\text{hom}}) \subset \sigma^{L^2}(T_{\text{hom}})$. Since the spectrum of a self-adjoint operator on a Hilbert space is a subset of the real line ([16], Proposition VII.6.1), $\sigma(T_{\text{hom}}) := \sigma^{L^1}(T_{\text{hom}}) \subset \mathbb{R}$.

Next, it follows from the compactness of $T_{\text{hom}} \in B(L^1[0,1])$ and [16], Theorem VII.7.1, that T_{hom} has $\rho(T_{\text{hom}})$ as an eigenvalue; denote the corresponding normalized eigenfunction as ϕ. By the Krein-Rutman theorem ([60], Theorem V.6.6) — which applies by symmetry of W and by the assumption that $W^{(k)} > 0$ for some $k \in \mathbb{N}$ — we can even select a normalized eigenfunction satisfying $\phi > 0$ a.e. In other words, ϕ is a probability density on $[0,1]$. Furthermore, it holds that ϕ is essentially bounded, which follows from

$$\begin{aligned}\|\phi\|_{L^\infty[0,1]} &= \frac{1}{\rho(T_{\text{hom}})}\|T_{\text{hom}}\phi\|_{L^\infty[0,1]} = \frac{1}{\rho(T_{\text{hom}})} \sup_{x\in[0,1]} \|h\|_{L^1(\mathbb{R}_+)} \int_0^1 W(x,y)\phi(y)\,\mathrm{d}y \\ &\leq \frac{1}{\rho(T_{\text{hom}})} C_W \|h\|_{L^1(\mathbb{R}_+)} \|\phi\|_{L^1[0,1]}.\end{aligned}$$

Hence, we can bound ϕ by the constant function $\|\phi\|_{L^\infty[0,1]} \times \mathbf{1} : x \mapsto \|\phi\|_{L^\infty[0,1]}$, which is in $L^1[0,1]$. For this unit profile $\mathbf{1} : x \mapsto 1$, we have

$$\|T^n_{\text{hom}}\mathbf{1}\|_{L^1[0,1]} = \frac{\left\|T^n_{\text{hom}}\|\phi\|_{L^\infty[0,1]}\mathbf{1}\right\|_{L^1[0,1]}}{\|\phi\|_{L^\infty[0,1]}} \geq \frac{\|T^n_{\text{hom}}\phi\|_{L^1[0,1]}}{\|\phi\|_{L^\infty[0,1]}} \to \infty, \quad \text{as } n \to \infty.$$

Our goal is to use Assumption 6 to construct a Poisson stream of uniformly distributed particles on $[0,1]$ — i.e., distributed according to the unit profile —, each of which has infinite expected offspring size.

Using Assumption 6, there is a Poisson stream of intensity

$$\theta := \alpha\kappa(1 - \exp(-\|h\|_{L^1(\mathbb{R}_+)}))^k > 0$$

leading after k generations (i.e., eventually) to some *uniformly distributed* particle in $[0,1]$. Indeed, one may bound the stream of events of the original process leading after k generations to some uniformly distributed particle in $[0,1]$ by a coupled stream where a kth-generation event in x originating from an immigrant in y is rejected with probability $1 - \kappa/W^{(k)}(x,y)$. Such a uniformly distributed event within $[0,1]$ has infinite expected offspring; here the expectation is w.r.t. both the location within $[0,1]$ and w.r.t. the offspring realization.

Since the time between the birth of a parent and the birth of a child has the finite expectation $\int_0^\infty th(t)\,\mathrm{d}t < \infty$, we infer that $N_T[0,1]/T \to \infty$ a.s., as $T \to \infty$; this follows by applying the strong LLN to a compound Poisson process with arrival rate θ and infinite expected claim size. By invoking Assumption 6 again, it follows that, for any $A \in \mathcal{B}[0,1]$ of positive Lebesgue measure, we have $N_T(A)/T \to \infty$ a.s., as $T \to \infty$. □

REMARK 10. *Having stated limit theorems both in the subcritical ($\rho(T_{\text{hom}}) < 1$) and supercritical ($\rho(T_{\text{hom}}) > 1$) regime, we ask ourselves what happens in the critical ($\rho(T_{\text{hom}}) = 1$) regime. By the Poisson cluster representation of the graphon Hawkes process, this is closely related to the critical behavior of Galton-Watson processes, i.e., one would expect that in the critical regime $N_T(A)$ is of order T^2. Indeed, asymptotics of this type for* univariate *Hawkes processes can be found in [64], Theorem 21.*

5.2. *Functional central limit theorem.* Having proved the FLLN given in (36), a natural question is whether we can establish an FCLT for the graphon Hawkes process. We continue working in the unmarked setting.

As we have seen before, when we consider a graphon Hawkes process with piecewise constant parameters, treating the intervals of continuity as coordinates, we obtain a multivariate Hawkes process. For those processes, [5], Corollary 1, provides an FCLT. The continuous-space extension of their result, which is our *ansatz* for the FCLT, reads: for every $A \in \mathcal{B}[0,1]$,

(46)
$$\sqrt{T}\left(\frac{N_{Tv}(A)}{T} - v\bar{\lambda}(A)\right)_{v\in[0,1]} \longrightarrow \left(\int_A \left((I - T_{\text{hom}})^{-1}\bar{\lambda}^{1/2}\right)(x)X(v,x)\,\mathrm{d}x\right)_{v\in[0,1]},$$

where $\bar{\lambda} \in L^1[0,1]$ solves (34), $\bar{\lambda}^{1/2}$ is taken componentwise, and $X(t,x) = \int_{(0,t)} W(s,x)\,\mathrm{d}s$, for W a standard L^2 Gaussian white noise on the strip $[0,\infty)\times[0,1]$.

Just as for the FLLN, we consider the number of points $N_{Tv}(A)$ in a single Borel set A. Informally, by a result like (46), we obtain a spatial white noise, temporal Brownian limit process, in the sense of Schwartz distributions.

In proving Theorem 6 below, we use a stronger stability condition than the requirement $\rho(T_{\text{hom}}) < 1$ that we have been using before; see Remark 4.

ASSUMPTION 7. *Let $D(y) := \int_0^1 W(x,y)\,\mathrm{d}x$ be the* outdegree *of $y \in [0,1]$. Suppose that*

$$\|h\|_{L^1(\mathbb{R}_+)} \sup_{y\in[0,1]} D(y) < 1. \tag{47}$$

Equation (47) is equivalent to $\rho < 1$, where ρ is defined in (74) in Appendix A.

THEOREM 6. *Grant Assumptions 1, 2, 3, 4 and 7. Let $(N_t)_{t\geq 0}$ be a stationary (see Proposition 3 and Theorem 2), unmarked linear graphon Hawkes process. Then we have, for $A \in \mathcal{B}[0,1]$, and B a standard Brownian motion,*

$$\sqrt{T}\left(\frac{N_{Tv}(A)}{T} - v\mu_A\right)_{v\in[0,1]} \longrightarrow (\sigma_A B(v))_{v\in[0,1]}, \tag{48}$$

as $T \to \infty$, weakly on $D[0,1]$ equipped with the Skorokhod topology. Here, $\mu_A = \bar{\lambda}(A)$ and $\sigma_A \in \mathbb{R}_+$. Finally, for multivariate Hawkes processes, i.e., graphon Hawkes processes with piecewise constant parameters, σ_A is given by $\int_A \left((I - T_{\text{hom}})^{-1}\bar{\lambda}^{1/2}\right)(x)\,\mathrm{d}x$.

PROOF. First, it follows from the FLLN given in (36), which depends on Assumptions 1, 2, 3 and 4, that if the FCLT (48) holds, then we necessarily have $\mu_A = \bar{\lambda}(A)$.

The proof of (48) employs Poisson embedding as in [63], Theorem 1, and [12], to verify the conditions of [10], Theorem 19.1. To this end, we need to check that

$$\sum_{n\geq 1}\left(\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) - \mu_A | \mathcal{F}_0^{(N)}\right]^2\right]\right)^{1/2} < \infty, \tag{49}$$

where $N_{(t_1,t_2]}(A)$ counts the number of points arriving in A during $(t_1,t_2]$, and where $\mathcal{F}_0^{(N)}$ is the σ-algebra recording the history of the stationary process N before time 0.

The rest of the proof is structured as follows. First, we bound the left-hand side of (49) by the corresponding quantity with $A = [0,1]$. Then, we couple N to a spatiotemporal process M such that: (i) M is essentially a univariate Hawkes process; (ii) M is stationary; and (iii) the conditional variance of $N_{(n,n+1]}[0,1]$ can be bounded by that of $M_{(n,n+1]}[0,1]$. We finish the proof by applying [10], Theorem 19.1, to M.

We start bounding the left-hand side of (49) by the corresponding quantity with $A = [0,1]$:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) - \mu_A | \mathcal{F}_0^{(N)}\right]^2\right] &= \mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) | \mathcal{F}_0^{(N)}\right]^2\right] - \mu_A^2 \\
&\le \mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) | \mathcal{F}_0^{(N)}\right]^2\right] - \mu_A^2 + \mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A^{\mathsf{c}}) | \mathcal{F}_0^{(N)}\right]^2\right] - \mu_{A^{\mathsf{c}}}^2 \\
&\quad + \underbrace{2\left(\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) | \mathcal{F}_0^{(N)}\right]\mathbb{E}\left[N_{(n,n+1]}(A^{\mathsf{c}}) | \mathcal{F}_0^{(N)}\right]\right] - \mu_A \mu_{A^{\mathsf{c}}}\right)}_{\text{spatial covariance } \ge 0 \text{ by nonnegativity of } W \text{ and } h} \\
&= \mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}(A) + N_{(n,n+1]}(A^{\mathsf{c}}) - \mu_A - \mu_{A^{\mathsf{c}}} | \mathcal{F}_0^{(N)}\right]^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}[0,1] - \mu_{[0,1]} | \mathcal{F}_0^{(N)}\right]^2\right]. \qquad (50)
\end{aligned}$$

Next, we bound (50) by the corresponding conditional variance expression for a spatiotemporal point process $M = (M_t)_{t\in\mathbb{R}}$ such that $M[0,1]$ is a stable univariate Hawkes process. Consider the cluster representation for the graphon Hawkes process, Definition 3. We define M by coupling it to N iteratively as follows. For each $k \in \mathbb{N}_0$, we set $M_k^{(1)}$ equal to the kth-generation of N, but replace the spatial coordinates by i.i.d. $\mathrm{Uni}(0,1)$ marks. Next, for each $k \in \mathbb{N}$ and each $(k-1)$th-generation event (s,y) of N, $M_k^{(2)}$ is given by initial streams of events generated by an inhomogeneous Poisson process with intensity

$$[s,\infty) \ni t \mapsto h(t-s)\left(\sup_{y\in[0,1]} D(y) - \int_0^1 W(x,y)\,\mathrm{d}x\right),$$

marked by i.i.d. $\mathrm{Uni}(0,1)$ spatial coordinates; use Assumption 7. Next, for each event (s',y') in $M_k^{(2)}$, we let $M_k^{(3)}$ be a univariate Hawkes cluster of events with intensity

$$[s',\infty) \ni t \mapsto h(t-s') \sup_{y\in[0,1]} D(y),$$

marked by i.i.d. $\mathrm{Uni}(0,1)$ spatial coordinates. We set

$$M^{(1)} := \bigcup_{k\ge 0} M_k^{(1)}; \quad M^{(i)} := \bigcup_{k\ge 1} M_k^{(i)} \text{ for } i = 2,3; \quad M := M^{(1)} \cup M^{(2)} \cup M^{(3)}.$$

Note that it is possible to select $M_k^{(2)}$ and $M_k^{(3)}$ in such a way that $M[0,1]$ is stationary under $\mathbb{P}$, since we couple the history before time 0 of the stationary process N to $M^{(1)}$. Since N is stationary, so is the cluster center process (see [18]) corresponding to N: those centers arrive according to a stationary homogeneous Poisson point process of rate $\|\lambda_\infty\|_{L^1[0,1]}$ with i.i.d. $\mathrm{Uni}(0,1)$ marks; the same holds for M. Compared to N, the clusters of M are enlarged by $M^{(2)}$ and $M^{(3)}$. Here, M is a univariate Hawkes process with baseline intensity $\|\lambda_\infty\|_{L^1[0,1]}$, excitation function $h(\cdot)\sup_{y\in[0,1]} D(y)$ and i.i.d. $\mathrm{Uni}(0,1)$ spatial coordinates; by Assumption 7 it satisfies the stability conditions of Theorem 1 (see Remark 4). Note that any stationary distribution for the univariate Hawkes process consists of a stationary cluster center process, to which mutually independent component processes are attached. This is exactly how it is done for M.

To compare the conditional variances of N and M, note that those conditional variances are w.r.t. different σ-algebras. To overcome this problem, let $\mathcal{F}_0^{(N,M)}$ be the σ-algebra generated by the histories of N, M before time 0. Since $M_k^{(2)}$ and $M_k^{(3)}$ are drawn conditionally independently of N, it holds that

$$\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}[0,1]-\mu_{[0,1]}|\mathcal{F}_0^{(N)}\right]^2\right]=\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}[0,1]-\mu_{[0,1]}|\mathcal{F}_0^{(N,M)}\right]^2\right]. \tag{51}$$

Next, since the effect of the arrival of a kth-generation event on the evolution of a Hawkes process does not depend on $k\in\mathbb{N}_0$,

$$\mathbb{E}\left[\mathbb{E}\left[M_{(n,n+1]}[0,1]-\mu_{[0,1]}^{(M)}|\mathcal{F}_0^{(M)}\right]^2\right]=\mathbb{E}\left[\mathbb{E}\left[M_{(n,n+1]}[0,1]-\mu_{[0,1]}^{(M)}|\mathcal{F}_0^{(N,M)}\right]^2\right]. \tag{52}$$

Note that N is not independent of $(M_k^{(2)},M_k^{(3)})$, as there is a positive covariation between N and $M_k^{(2)},M_k^{(3)}$. Indeed, higher kth-generation offspring for N leads to an excited conditional intensity for $M_k^{(2)}$, which in turn induces an excited conditional intensity for $M_k^{(3)}$. Therefore, the conditional variance of $N_{(n,n+1]}[0,1]$ given $\mathcal{F}_0^{(N,M)}$ can be bounded by that of $M_{(n,n+1]}[0,1]$. Upon combining this with (51)–(52), it follows from the proof of [63], Theorem 1, that

$$\begin{aligned}&\sum_{n\geq1}\left(\mathbb{E}\left[\mathbb{E}\left[N_{(n,n+1]}[0,1]-\mu_{[0,1]}|\mathcal{F}_0^{(N)}\right]^2\right]\right)^{1/2}\\ &\leq\sum_{n\geq1}\left(\mathbb{E}\left[\mathbb{E}\left[M_{(n,n+1]}[0,1]-\mu_{[0,1]}^{(M)}|\mathcal{F}_0^{(M)}\right]^2\right]\right)^{1/2}<\infty.\end{aligned} \tag{53}$$

Combining (50) and (53), [63], Theorem 3, gives that for some $\sigma_A\in\mathbb{R}_+$,

$$\left(\frac{N_{\lfloor Tv\rfloor}(A)-\lfloor Tv\rfloor\mu_A}{\sqrt{T}}\right)_{v\in[0,1]}\longrightarrow\sigma_A B(v), \tag{54}$$

as $T\to\infty$, weakly on $D[0,1]$ equipped with the Skorokhod topology. Again by bounding by M as in (50)–(53), we deduce that $\mathbb{E}[N_{[0,1]}(A)^2]<\infty$, hence we may proceed as in [63], eqn. (2.19), to conclude that (48) holds.

We still need to identify σ_A. To this end, note that by [5], Corollary 1, (48) holds for graphon Hawkes processes with piecewise constant parameters on the sets of a partition $\mathcal{P}$ of $[0,1]$ such that $A\in\mathcal{P}$. □

6. Fixed-point theorems in the transform domain and an application. Given the definition of a finite-dimensional linear Hawkes process as a Poisson cluster process, it is possible to derive fixed-point equations for Hawkes-fed birth-death processes in the transform domain, which can be used to approximate transforms numerically; see [38, 41] for more details. In this section, we utilize Definition 3 to establish fixed-point theorems in the infinite-dimensional setting. As argued at the start of Section 5, we can restrict ourselves to $\mathscr{X}=[0,1]$ without losing generality. As before, this is for legibility only: all arguments only rely on $\mathscr{X}$ being a compact subset of an Euclidean space $\mathbb{R}^m$, and in particular do not require $\mathscr{X}$ to be one-dimensional. We finish this section by applying the fixed-point theorems we establish to prove that, starting with a d-dimensional Hawkes birth-death process $\tilde{Q}^d(t)$, the limits $d\to\infty$ and $t\to\infty$ commute.

6.1. *Transform characterization.* Besides a graphon Hawkes pure-birth process N, i.e., a counting process, we are interested in a model allowing for deaths, i.e., departures from the system. Suppose that each particle in spatial coordinate $x\in\mathscr{X}$ has a stochastic lifetime distributed as J_x, having a distribution only dependent upon the spatial coordinate, independent of the lifetimes of other particles.

DEFINITION 7. *Let* $(\mathscr{X}, \mathcal{B}(\mathscr{X}), \mu)$ *be a topological* σ*-finite measure space. Let* $N = (N_t(x))_{t\in[0,\infty)\times\mathscr{X}}$ *be a spatiotemporal point process, where to each event* (t,x) *a* J_x*-distributed lifetime* $w_{t,x}$ *is attached, independent of everything else. Define the spatiotemporal* birth-death process Q *through the formula*

$$Q_t(A) = \left|\{(t',x) \in N, x \in A : t' \leq t < t' + w_{t',x}\}\right|, \quad A \in \mathcal{B}(\mathscr{X}).$$

ASSUMPTION 8. *Consider a topological* σ*-finite measure space* $(\mathscr{X}, \mathcal{B}(\mathscr{X}), \mu)$*. Let* $N = (N_t(x))_{t\in[0,\infty)\times\mathscr{X}}$ *be a* linear *spatiotemporal point process, where to each event* (t,x) *a* J_x*-distributed lifetime* $w_{t,x}$ *is attached, independent of everything else. Assume that the lifetimes* $(J_x)_{x\in\mathscr{X}}$ *are i.i.d., distributed as* J*, which is a stochastic process* $\mathscr{X} \to \mathbb{R}_+$*, separable w.r.t. the class* $\mathcal{U}$ *of open subsets of* $\mathscr{X}$*.*

The definition of *separability* of a stochastic process can be found in [51], §III.4.

We aim to characterize the distribution of the linear graphon Hawkes process *with departures* through a suitable transform. Since we work with a continuum of spatial coordinates, we consider a Laplace functional instead of a Laplace transform.

DEFINITION 8. *Let* Z *be a spatiotemporal point or birth-death process, where each event consists of a spatial coordinate* $x \in \mathscr{X}$ *and a temporal coordinate* $t \in \mathbb{R}_+$*. Let* $\mathrm{BM}_+(\mathscr{X})$ *be the space of bounded measurable functions* $f : \mathscr{X} \to \mathbb{R}_+$*. For* $t \in \mathbb{R}_+$*, we define its* Laplace functional $\mathcal{L}_N(\cdot,t) : \mathrm{BM}_+(\mathscr{X}) \to \mathbb{R}_+$ *by*

$$\mathcal{L}_Z(f,t) := \mathbb{E}\left[\exp(-f_Z^t)\right] := \mathbb{E}\left[\exp\left(-\sum_{\substack{(s,y)\in Z\\ s<t}} f(y)\right)\right], \tag{55}$$

where the summation is over all events of Z *at times* s *strictly before* t*, at location* y*. Here, the expectation is w.r.t. the natural filtration generated by* Z *at time* 0*. We denote the space of such transforms by* $\mathbb{L}$*.*

Note that a characteristic functional defines a distribution uniquely, by a functional version of the Minlos-Bochner theorem. See [9], Theorem 2.1, and [56], Theorem 2.

REMARK 11. *By taking* $f \equiv f(z) : \mathscr{X} \to \mathbb{R}_+ : x \to z\mathbf{1}_A(x)$ *for* $z \in \mathbb{R}$ *and* $A \subset \mathscr{X}$*, or, more generally,* $f \equiv f(\mathbf{z}) : \mathscr{X} \to \mathbb{R} : x \to \sum_{i=1}^K z_i \mathbf{1}_{A_i}(x)$ *for* $\mathbf{z} \in \mathbb{R}^K$*,* $K \in \mathbb{N}$*, and* $A_1, \ldots, A_K \subset \mathscr{X}$*, one obtains usual Laplace transforms, which can be used to determine moments of* $N(A)$ *or* $N(A_1)\cdots N(A_K)$ *by differentiation. By ranging over different* K *and* $A_1, \ldots, A_K$*, the Laplace functional allows one to construct moment measures.*

Our goal is to use the first step of the cluster representation, (i) in Definition 3, to relate the transform $\mathcal{L}_Q$, where Q is a linear graphon Hawkes birth-death process, to the cluster processes constituted by each of the (immigrant) arrivals. Suppose that the lifetime random variables $(J_x)_{x\in[0,1]}$ satisfy Assumption 8. Denote the cluster process resulting from an arrival in coordinate $x \in [0,1]$ by S_x. This keeps track of all surviving offspring of the initial particle in coordinate x, including that particle itself, if it has not yet expired. We denote the Laplace functional of S_x by $\eta_x = \mathcal{L}_{S_x}$, i.e., $\eta_x(f,u) = \mathbb{E}\left[-\exp(f_{S_x}^u)\right]$. With this characterization of $\mathcal{L}_Q$ in terms of $(\eta_x)_{x\in[0,1]}$, we use the iterative step in the cluster representation, (ii)–(iii) in Definition 3, to derive a fixed-point equation for η_x, where we express η_x as a function of $(\eta_x)_{x\in[0,1]}$, the collection of Laplace functionals of cluster processes $(S_x)_{x\in[0,1]}$. This asks for a transform of $\mathscr{X}$-dimensional, $\mathscr{X}$-valued point processes.

DEFINITION 9. *Let $Z = (Z_x)_{x\in\mathscr{X}}$ be an $\mathscr{X}$-dimensional spatiotemporal point or birth-death process, where for $x \in \mathscr{X}$, Z_x is a spatiotemporal point or birth-death process in the sense of Definition 8. Let $\mathscr{Z}$ be the space of such processes. We define the Laplace functional of $Z \in \mathscr{Z}$ as the mapping*

$$\mathcal{L}_Z(\cdot,t) : \mathrm{BM}_+(\mathscr{X}) \to \mathbb{R}_+^{\mathscr{X}} : f \mapsto (\mathcal{L}_{Z_x}(f,t))_{x\in\mathscr{X}}\,.$$

We denote the space of such transforms by $\mathbb{L}^{\mathscr{X}}$.

Fix a time horizon $t \in \mathbb{R}_+$ and some $p \in [1,\infty]$. On $\mathbb{L}^{\mathscr{X}}$, we consider the metric

$$d_{\mathbb{L}^{\mathscr{X}},p} : \mathbb{L}^{\mathscr{X}} \times \mathbb{L}^{\mathscr{X}} \to \mathbb{R}_+ : (\mathcal{L},\mathcal{L}') \mapsto \sup_{\substack{u\in[0,t]\\ f\in \mathrm{BM}_+(\mathscr{X})}} \|\mathcal{L}_x(f,u) - \mathcal{L}'_x(f,u)\|_{L^p(\mathscr{X})}. \tag{56}$$

Our first result relates the transform of Q to the transform of the cluster processes $(S_x)_{x\in[0,1]}$. It implies that in order to calculate the transform of Q, it suffices to calculate $(\eta_x)_{x\in[0,1]}$. The proof of this result can be found in Appendix D.

THEOREM 7. *Grant Assumptions 1 and 8. Let $\mathscr{X} = [0,1]$. The transform of Q satisfies*

$$\mathcal{L}_Q(f,t) = \exp\left(\int_0^1 \int_0^t (\eta_x(f,u) - 1)\, \lambda_\infty(x)\, \mathrm{d}u\, \mathrm{d}x\right). \tag{57}$$

The next step is to derive a fixed-point equation for the cluster transforms η_x, by operationalizing the iterative step from the cluster representation. This iterative step can be transformed into a distributional equality for S_x. Given an arrival in x, translate the time frame such that this arrival occurs at $t = 0$. Conditionally upon the outcome of B, let $K_x(\cdot)$ be an inhomogeneous Poisson process with intensity $\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}h(\cdot)$, counting the children of the initial event in x. Let the locations x_i of the children born at time T_i be drawn i.i.d. according to the law having Lebesgue density

$$\mathbb{P}(x_i \in \mathrm{d}z) = \frac{B_{zx}(0)W(z,x)\,\mathrm{d}z}{\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}}.$$

Then, with $\delta_x(y) = \mathbf{1}\{y = x\}$ the Kronecker delta, the following distributional equality holds:

$$S_x(u) \overset{\mathcal{D}}{=} \delta_x \cdot \mathbf{1}\{J_x > u\} + \sum_{i=1}^{K_x(u)} S_{x_i}(u - T_i). \tag{58}$$

This distributional equality can be used to prove a fixed-point result. Denote the survival function and c.d.f. corresponding to the lifetime random variable J_x by

$$\mathscr{J}_x(u) = \mathbb{P}(J_x > u) \quad \text{and} \quad \bar{\mathscr{J}}_x(u) = \mathbb{P}(J_x \le u) = 1 - \mathscr{J}_x(u).$$

DEFINITION 10. *Let β_x be the Laplace functional of the stochastic process $B_{\cdot x}(0)$. For $\mathscr{X} = [0,1]$, we define the operator $\Phi : \mathbb{L}^{[0,1]} \to \mathbb{L}^{[0,1]}$ as follows. Let*

$$\gamma_x(f,u) := \bar{\mathscr{J}}_x(u) + \mathscr{J}_x(u)e^{-f(x)}. \tag{59}$$

For $\xi \in \mathbb{L}^{[0,1]}$, $f \in \mathrm{BM}_+[0,1]$ and $u \in \mathbb{R}_+$, let

$$\Phi_x(\xi)(f,u) := \gamma_x(f,u)\beta_x\left(\left\{y \mapsto \int_0^u (1 - \xi_y(f,u-s))\, W(y,x)h(s)\, \mathrm{d}s\right\}\right). \tag{60}$$

Before proving that η is a fixed point of Φ, we prove that Φ is a well-defined mapping. The proofs of the next four results can be found in Appendix D.

LEMMA 4. *Grant Assumptions 1 and 8. The mapping* $\Phi : \mathbb{L}^{[0,1]} \to \mathbb{L}^{[0,1]}$ *is well-defined, i.e.,* $\Phi(\xi) \in \mathbb{L}^{[0,1]}$ *for* $\xi \in \mathbb{L}^{[0,1]}$.

THEOREM 8. *Grant Assumptions 1 and 8. The* $[0,1]$*-dimensional spatiotemporal cluster process* η *is a fixed point of* Φ*: we have* $\eta = \Phi(\eta)$*, i.e., for each* $x \in [0,1]$, $\eta_x = \Phi_x(\eta)$.

LEMMA 5. *Grant Assumptions 1, 2 and 8. Fix a time horizon* $t \in \mathbb{R}_+$ *and* $p \in [1,\infty]$*. The mapping* Φ *is continuous w.r.t. the topology induced by the metric* $d_{\mathbb{L}^{[0,1]},p}$.

Next we show that Φ is a *contraction*. To state the result, for $\xi^{(0)} \in \mathbb{L}^{[0,1]}$, we denote its iterates under Φ by $(\xi^{(n)})_{n\in\mathbb{N}_0}$, i.e., $\xi^{(n)} = \Phi(\xi^{(n-1)})$ for $n \in \mathbb{N}$.

LEMMA 6. *Grant Assumptions 1, 2 and 8. Let* $\xi^{(0)}, \zeta^{(0)} \in \mathbb{L}^{[0,1]}$*. Then there exists* $C \in \mathbb{R}_+$ *such that, for any* $u \in [0,t]$, $f \in \mathrm{BM}_+[0,1]$ *and* $n \in \mathbb{N}$*, it holds that*

$$\left\|\xi_x^{(n)}(f,u) - \zeta_x^{(n)}(f,u)\right\|_{L^p[0,1]} \le \frac{C^n u^n}{n!}. \tag{61}$$

We are now ready to state the main result of this subsection, saying that iterates under Φ of any transform in $\mathbb{L}^{[0,1]}$ converge to η.

THEOREM 9. *Grant Assumptions 1, 2 and 8. Fix* $t \in \mathbb{R}_+$*. For any* $\xi^{(0)} \in \mathbb{L}^{[0,1]}$*, the sequence* $(\xi^{(n)})_{n\in\mathbb{N}_0}$ *converges pointwise to the unique fixed point* η *in* $\mathbb{L}^{[0,1]}$ *of* Φ*. More specifically, for any* $u \le t$ *and* $f \in \mathrm{BM}_+[0,1]$,

$$\xi^{(n)}(f,u) \to \eta(f,u). \tag{62}$$

PROOF. By Lemma 6, iterates $\xi^{(n)}, \zeta^{(n)}$ of two transforms $\xi^{(0)}, \zeta^{(0)} \in \mathbb{L}^{[0,1]}$ converge pointwise and have the same limit $\mathscr{L}$. By Lemma 4, we know that those iterates are in $\mathbb{L}^{[0,1]}$ for any $n \in \mathbb{N}_0$. Next, for each $x \in [0,1]$, we have convergence of $\xi_x^{(n)}$ to $\mathscr{L}_x$, pointwise, as $n \to \infty$. This convergence holds in particular for all test functions f in the Schwartz space on $[0,1]$. By treating a $[0,1]$-marked point process as a functional on $[0,1]$, we may apply a functional form of Lévy's continuity theorem, [9], Theorem 2.3, to conclude that there exists a graphon point process Z_x such that $\mathscr{L}_x = \mathcal{L}_{Z_x}$, i.e., $\mathscr{L}_x \in \mathbb{L}$. In particular, for $Z = (Z_x)_{x\in[0,1]} \in \mathscr{Z}$, we have $\mathscr{L} = \mathcal{L}_Z$, meaning that $\mathscr{L} \in \mathbb{L}^{[0,1]}$. Since Φ is continuous by Lemma 5, we have

$$\mathscr{L} = \lim_{n\to\infty} \xi^{(n+1)} = \lim_{n\to\infty} \Phi\left(\xi^{(n)}\right) = \Phi\left(\lim_{n\to\infty} \xi^{(n)}\right) = \Phi(\mathscr{L}),$$

i.e., $\mathscr{L}$ is a fixed point of Φ. From Theorem 8, η is also a fixed point of Φ. As iterates of η converge to $\mathscr{L}$, it follows that $\eta = \mathscr{L}$. □

6.2. *Commuting large-time and high-dimension limits.* In this subsection, we study large-time behavior of the graphon Hawkes birth-death process. Suppose that we start with multivariate Hawkes population processes $\tilde{Q}^d(\cdot)$ converging to a graphon Hawkes population process $Q(\cdot)$, where we first let $t \to \infty$, and then $d \to \infty$. We would like to know whether the resulting limit graphon process is the same as the stationary version of Q, i.e., whether we can interchange limits as illustrated by Figure 2. To answer this question, we view the prelimit as a piecewise constant graphon Hawkes process, so that we can employ the characterizations of the Laplace functional from the previous subsection to describe its distribution.

$$\begin{array}{ccc} \tilde{Q}^d(t) & \xrightarrow{d\to\infty} & Q(t) \\ {\scriptstyle t\to\infty}\big\downarrow & & \big\downarrow{\scriptstyle t\to\infty} \\ \tilde{Q}^d(\infty) & \xrightarrow[d\to\infty]{} & Q(\infty) \end{array}$$

Fig 2: The limits $d \to \infty$ and $t \to \infty$, applied to $\tilde{Q}^d(t)$, commute.

ASSUMPTION 9. *Suppose that the baseline intensity λ_∞ is bounded below by some c_∞ on the set $\{x : \lambda_\infty(x) > 0\} = \{x : \lambda_\infty(x) \geq c_\infty\}$. Suppose that $\int_0^\infty th(t)\,\mathrm{d}t < \infty$. Finally, assume that there is a single lifetime distribution J, independent of the particle location $x \in [0,1]$, satisfying $\mathbb{E}[J] < \infty$. Write $\mathscr{J}$, $\bar{\mathscr{J}}$ for the corresponding survival function and c.d.f., respectively.*

We impose the (mild) assumption on the lower bound on λ_∞ in Theorem 10 below for technical reasons. Furthermore, we require the assumption of finite expected time between parent and child, $\int_0^\infty th(t)\,\mathrm{d}t < \infty$, because we use boundedness of the cluster durations. Finally, we use the assumption on the lifetime distributions in order to couple lifetimes of prelimit and graphon Hawkes processes.

THEOREM 10. *Grant Assumptions 1, 2, 3, 4, 8 and 9, and suppose that the stability conditions of Theorem 1 hold. Take a linear graphon Hawkes birth-death process $Q(t)$ and consider its multivariate projections $\tilde{Q}^d(t)$, using partitions $\mathcal{P}^d$ of $[0,1]$ with $\mathrm{mesh}(\mathcal{P}^d) \to 0$, as $d \to \infty$. Assume that the lifetime distributions for $\tilde{Q}^d$ are the same as those of Q, with coupled lifetimes for simultaneous events. Then the stationary version $\tilde{Q}^d(\infty)$ of $\tilde{Q}^d(t)$ converges weakly in the space of all point measures on $[0,\infty) \times [0,1]$ to $Q(\infty)$, as $d \to \infty$.*

PROOF. We prove weak convergence of $\tilde{Q}^d(\infty)$ to $Q(\infty)$ by proving pointwise convergence of the corresponding Laplace functionals $\tilde{\mathcal{L}}_Q^d(f,\infty), \mathcal{L}_Q(f,\infty)$, using [18], Proposition 11.1.VIII. We do this by invoking Theorems 7 and 8, after which we bound the difference $|\tilde{\mathcal{L}}_Q^d(f,\infty) - \mathcal{L}_Q(f,\infty)|$ by two integral terms, which we treat separately.

Using Theorems 7 and 8, we characterize the transform of the prelimit for $t \to \infty$ as

$$\tilde{\mathcal{L}}_Q^d(f,\infty) = \exp\left(\sum_{k=1}^d \lambda_{\infty,k}^d \int_{\mathscr{X}_k^d} \int_0^\infty (\tilde{\eta}_x^d(f,u) - 1)\,\mathrm{d}u\,\mathrm{d}x\right), \tag{63}$$

where for $x \in \mathscr{X}_m^d$ we have, by recognizing the Laplace functional β_{km}^d of the marks B_{km}^d,

$$\begin{aligned} \tilde{\eta}_x^d(f,u) &= \gamma_x(f,u)\mathbb{E}\left[\exp\left(\sum_{k=1}^d B_{km}^d(0) W_{km}^d \int_{\mathscr{X}_k^d} \int_0^u (\tilde{\eta}_y^d(f,u-s) - 1)h(s)\,\mathrm{d}s\,\mathrm{d}y\right)\right] \\ &= \gamma_x(f,u) \prod_{k=1}^d \beta_{km}^d \left(W_{km}^d \int_{\mathscr{X}_k^d} \int_0^u (\tilde{\eta}_y^d(f,u-s) - 1)h(s)\,\mathrm{d}s\,\mathrm{d}y\right). \end{aligned} \tag{64}$$

Here, $\gamma_x(f,u) = \bar{\mathscr{J}}(u) + \mathscr{J}(u)e^{-f(x)}$, cf. (59). From the convergence of the processes $\tilde{N}^d$ to N, as $d \to \infty$, for finite time horizons, we conclude the convergence $\tilde{\eta}_x^d(f,u) \to \eta_x(f,u)$ pointwise, as $d \to \infty$, since the transforms up to time u only depend on smaller times $0 \leq s \leq u$; or, alternatively, since the convergence for single clusters is implied by the convergence of the whole process (on bounded time intervals).

We show that $|\tilde{\mathcal{L}}_Q^d(f,\infty) - \mathcal{L}_Q(f,\infty)|$ tends to 0 as $d \to \infty$, for each f, since pointwise convergence of Laplace functionals is equivalent to weak convergence of random measures, see [18], Proposition 11.1.VIII. It suffices to restrict ourselves to positive, continuous functions $f \in C_+[0,1]$ instead of $f \in \mathrm{BM}_+[0,1]$, see, e.g., [49], Theorem 9.1.4. Such a function f is uniformly continuous on $[0,1]$ and uniformly bounded by some $C_f \in \mathbb{R}_+$.

Note that by the mean value theorem and the triangle inequality

$$\begin{aligned}
&|\tilde{\mathcal{L}}_Q^d(f,\infty) - \mathcal{L}_Q(f,\infty)| \\
&= \Bigg| \exp\left(\int_0^1 \sum_{k=1}^d \mathbf{1}\{x \in \mathscr{X}_k^d\} \lambda_{\infty,k}^d \int_0^\infty (\tilde{\eta}_x^d(f,u) - 1) \,\mathrm{d}u \,\mathrm{d}x \right) \\
&\quad - \exp\left(\int_0^1 \int_0^\infty (\eta_x(f,u) - 1)\lambda_\infty(x) \,\mathrm{d}u \,\mathrm{d}x \right) \Bigg| \\
&\le \left| \int_0^1 \int_0^\infty \left[\sum_{k=1}^d \mathbf{1}\{x \in \mathscr{X}_k^d\} \lambda_{\infty,k}^d (\tilde{\eta}_x^d(f,u) - 1) - \lambda_\infty(x)(\eta_x(f,u) - 1) \right] \mathrm{d}u \,\mathrm{d}x \right| \\
&\le \int_0^1 \left| \sum_{k=1}^d \mathbf{1}\{x \in \mathscr{X}_k^d\} \lambda_{\infty,k}^d - \lambda_\infty(x) \right| \int_0^\infty \left| \tilde{\eta}_x^d(f,u) - 1 \right| \,\mathrm{d}u \,\mathrm{d}x \qquad (65) \\
&\quad + \int_0^1 \int_0^\infty \lambda_\infty(x) \left| \eta_x(f,u) - \tilde{\eta}_x^d(f,u) \right| \,\mathrm{d}u \,\mathrm{d}x. \qquad (66)
\end{aligned}$$

We bound the integrals appearing in (65) and (66) separately.

For (65), under the stability conditions, the collection of cluster durations $(D_x)_{x\in[0,1]}$ of the clusters $(S_x)_{x\in[0,1]}$ is a uniformly tight collection of random variables, since the expected cluster sizes are uniformly bounded by $\mathfrak{K}$, by Lemma 2, and since the excitation function is not dependent on the spatial coordinate. Here, we use our assumption $\int_0^\infty th(t)\,\mathrm{d}t < \infty$ in order to have times between the birth of a parent and the birth of a child — conditionally upon the creation of a child — of bounded expectation. Under the conditions of Theorem 1, we have a well-defined, stable system, assuring that $\tilde{\mathcal{L}}_Q^d(f,\infty) > 0$ for d sufficiently large, by Lemma 2. From the formula (57) for $\tilde{\mathcal{L}}_Q^d(f,\infty)$, it follows that $\int_0^\infty \left|\tilde{\eta}_x^d(f,u) - 1\right| \,\mathrm{d}u$ is convergent a.e. on $\{x : \lambda_\infty(x) > 0\}$, since by Assumption 9 the baseline intensity λ_∞ is bounded below by some c_∞ on the set $\{x : \lambda_\infty(x) > 0\}$; here we use that f is a *positive* function. By this reasoning, we can even find a bound on $\int_0^\infty \left|\tilde{\eta}_x^d(f,u) - 1\right| \,\mathrm{d}u$ uniformly over $x \in \{y : \lambda_\infty(y) > 0\}$. It then follows by the L^1 convergence of $\tilde{\lambda}_\infty^d$ to λ_∞ (Lemma 3) that (65) tends to 0, as $d \to \infty$.

Next, to bound (66) by ϵ, it suffices to bound $\int_0^\infty \left|\eta_x(f,u) - \tilde{\eta}_x^d(f,u)\right| \,\mathrm{d}u$ by ϵ/α, uniformly over $x \in [0,1]$. By the mean value theorem,

$$\begin{aligned}
\int_0^\infty \left|\eta_x(f,u) - \tilde{\eta}_x^d(f,u)\right| \,\mathrm{d}u &= \int_0^\infty \left| \mathbb{E}\left[\exp(-f_{S_x}^u)\right] - \mathbb{E}[\exp(-f_{\tilde{S}_x^d}^u)] \right| \,\mathrm{d}u \\
&\le \int_0^\infty \mathbb{E}\left| \exp(-f_{S_x}^u) - \exp(-f_{\tilde{S}_x^d}^u) \right| \,\mathrm{d}u \\
&\le \int_0^\infty \mathbb{E}\left| f_{S_x}^u - f_{\tilde{S}_x^d}^u \right| \,\mathrm{d}u.
\end{aligned}$$

Using Lemma 2, we select $D \in \mathbb{N}$, $\mathfrak{K} \in \mathbb{R}_+$ such that the expected cluster sizes of N, $\tilde{N}^d$ are uniformly bounded by $\mathfrak{K}$, uniformly over $d \ge D$, $x \in [0,1]$. Then the expected discrepancy in $\mathfrak{d}$-distance between a cluster S_x and its coupled cluster $\tilde{S}_x^d$ of $\tilde{N}^d$ can be bounded as in (ii) from the proof of Theorem 2: we can find a $d_1 \geq D$ such that for $d \geq d_1$ the expected discrepancy in $\mathfrak{d}$-distance between the two coupled clusters is bounded by

$$\epsilon' := \frac{\epsilon}{4\alpha C_f \mathfrak{K}\mathbb{E}[J]},$$

hence with probability of at least $1-\epsilon'$, the two clusters S_x and $\tilde{S}^d_x$ only consist of simultaneous events. Now we use the uniform continuity of f to find $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon'' := \frac{\epsilon}{2\alpha\mathfrak{K}\mathbb{E}[J]}$$

for $|x-y| < \delta$. Next, we select $d_2 \geq d_1$ such that $\mathrm{mesh}(\mathcal{P}^{d_2}) < \delta$.

Given that we only have simultaneous events within the coupled clusters S_x and $\tilde{S}^d_x$, which is an event with probability of at least $1-\epsilon'$ for $d \geq d_2$, $\int_0^\infty \mathbb{E}|f(S_x,u) - f(\tilde{S}^d_x,u)|\,\mathrm{d}u$ can be bounded by the product of: the expected number of events within the cluster S_x, the expected duration of an event, and ϵ'' — since the simultaneous events are within the same element $\mathscr{X}^d_n$ of the partition $\mathcal{P}^d$. Here, we use the uniform continuity argument from the previous paragraph and the fact that simultaneous events are in the same $\mathscr{X}^d_n$, which follows by the construction presented in Section 3. Hence, given this event with probability of at least $1-\epsilon'$, $\int_0^\infty \mathbb{E}|f(S_x,u) - f(\tilde{S}^d_x,u)|\,\mathrm{d}u$ can be bounded by $\mathfrak{K}\mathbb{E}[J]\epsilon'' < \epsilon/(2\alpha)$.

Next, with probability of at most ϵ', the clusters S_x and $\tilde{S}^d_x$ are not equal, in which case, similarly as before, for $d \geq D$, $\int_0^\infty \mathbb{E}\left|f^u_{S_x} - f^u_{\tilde{S}^d_x}\right|\,\mathrm{d}u$ can be bounded by $2C_f\mathfrak{K}\mathbb{E}[J]$. To conclude, for $d \geq d_2$, $\int_0^\infty \left|\eta_x(f,u) - \tilde{\eta}^d_x(f,u)\right|\,\mathrm{d}u$ can be bounded by ϵ/α.

This proves that (66) is bounded by ϵ, for $d \geq d_2$. We already showed that (65) tends to 0, as $d \to \infty$, hence pointwise convergence of $\tilde{\mathcal{L}}^d_Q(f,\infty)$ to $\mathcal{L}_Q(f,\infty)$ follows, for $f \in C_+[0,1]$. Invoking again [49], Theorem 9.1.4, it follows that $\tilde{Q}^d(\infty) \to Q(\infty)$ weakly in the space of all point measures on $[0,\infty)\times[0,1]$, as $d \to \infty$. □

7. Concluding remarks. In this paper, we have introduced the *graphon Hawkes process*. Working on a corresponding uncountable space introduces the need for functional-analytic techniques that are unnecessary in the discrete space setting. After establishing its existence, uniqueness and stability, this graphon Hawkes process is shown to occur as the limit of suitably chosen multivariate Hawkes processes. Along the way, we have established technical Lemmas 2 and 3, which may be interesting in their own right. Next, we have provided a dichotomy in large-time behavior depending on whether $\rho(T_{\mathrm{hom}}) < 1$ or $\rho(T_{\mathrm{hom}}) > 1$, and we have established an FLLN and an FCLT in the stable case. By reducing the problem to a finite-dimensional setting, our proofs of the FLLN and FCLT become more tractable, facilitating the derivation of functional limit theorems for infinite-dimensional systems. This method, given its effectiveness, may also benefit other infinite-dimensional models. Finally, we have characterized the distribution of the graphon Hawkes process by fixed-point equations in the transform domain, and we have used this characterization to establish that, for convergent multivariate Hawkes population processes $\tilde{Q}^d(t)$ in the stable regime, the limits $d \to \infty$ and $t \to \infty$ commute, an application that we believe is both interesting and insightful.

Several follow-up questions and extensions are conceivable.

- As we have shown, it is possible to define the graphon Hawkes process for nonlinear conditional intensity density functions. We wonder whether the convergence results of Theorems 2 and 3 carry over to the nonlinear case. This would, however, require an intrinsically different approach than the one followed in this work, since one loses the branching representation and the additive structure of the conditional intensity density. If we consider nonlinear models with $f_x(z) = f(\lambda_\infty(x) + z)$ for a *single* Lipschitz continuous function

f, then by a straightforward modification of the proof of Theorem 2 — first bounding by $c_x \equiv c$, the Lipschitz constant of f —, we are able to prove convergence of the multivariate processes, obtained by averaging parameters, to the corresponding graphon Hawkes processes. We note that nonlinearity of the form $c_x \equiv c$ is often assumed in related literature; see e.g., [2].

- The functional limit theorems in the stable regime exploit the convergence results of Section 4 and related results for multivariate Hawkes processes. We would be interested in direct proofs for our FLLN and FCLT as well, because those proofs might be interesting in their own rights.
- In this paper, we have established both an FLLN and an FCLT. Given those limit theorems, one may be interested in establishing *large deviation principles* for the graphon Hawkes process. Since we are working with an infinite-dimensional model, we expect that this requires more sophisticated, functional-analytic tools than those required for finite-dimensional Hawkes processes ([39]). We intend to pursue this in future work.

Appendix A: Relegated proofs of and supplement to Sections 2 and 3.

PROOF OF LEMMA 1. Denote the random spatial coordinate by X. Under procedure (i),

$$\begin{aligned}\mathbb{P}(X \in \mathrm{d}x) &= \frac{\|\lambda^1(\cdot)\|_{L^1(\mathscr{X})}}{\|\lambda^1(\cdot)+\lambda^2(\cdot)\|_{L^1(\mathscr{X})}} \frac{\lambda^1(x)\mathrm{d}x}{\|\lambda^1(\cdot)\|_{L^1(\mathscr{X})}} + \frac{\|\lambda^2(\cdot)\|_{L^1(\mathscr{X})}}{\|\lambda^1(\cdot)+\lambda^2(\cdot)\|_{L^1(\mathscr{X})}} \frac{\lambda^2(x)\mathrm{d}x}{\|\lambda^2(\cdot)\|_{L^1(\mathscr{X})}} \\ &= \frac{\lambda^1(x)\mathrm{d}x+\lambda^2(x)\mathrm{d}x}{\|\lambda^1(\cdot)+\lambda^2(\cdot)\|_{L^1(\mathscr{X})}} = \frac{\lambda(x)\mathrm{d}x}{\|\lambda(\cdot)\|_{L^1(\mathscr{X})}},\end{aligned}$$

which is the law used under procedure (ii). □

PROOF OF THEOREM 1. First, the proof constructs a Picard scheme to establish existence. Next, for the uniqueness part, we only prove the key modifications required in our general setting compared to [48].

Existence. We use Picard's iteration to construct mappings $\{\lambda^n(t,z)\}$ and spatiotemporal point processes N^n, for $n \in \mathbb{N}_0$. Here, $z = (z_1, z_2, z_3) = (B, i, U)$ denotes the mark random variable: the first component denotes the random mark process B; the second component denotes the site index $z_2 = i$, defined on $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \sigma)$, where σ denotes the counting measure; and the final component denotes the sequences of random variables $(U_i)_{i\in\mathbb{N}}$ determining the spatial coordinate; see the construction in Section 3. Let $L \ni z$ denote the mark space. For the Picard scheme, we set $\lambda^0(t,z,x) \equiv 0$, $\Lambda^0(t,z) \equiv 0$, we let every N^n coincide with N on $\mathbb{R}_- \times L$, and for $n \in \mathbb{N}_0$, $t > 0$ and $z = (B,i,U) \in L$,

$$N^n(\mathrm{d}t \times \mathrm{d}z) = \bar{N}^i\left(\mathrm{d}t \times \mathrm{d}(B \times U) \times [0, \Lambda^n(t,i)]\right); \tag{67}$$

$$\begin{aligned}\Lambda^{n+1}(t,i) &= \int_{\mathscr{X}_i} \lambda^{n+1}(t,x)\,\mathrm{d}x \\ &= \int_{\mathscr{X}_i} f_x \left(\lambda_\infty(x) + \sum_{\substack{(s,y,B_{xy}) \in N^n \\ s<t}} B_{xy}(s) W(x,y) h(t-s) \right) \mathrm{d}\mu(x),\end{aligned} \tag{68}$$

with the understanding that the locations y are determined by the site index and by the random marks U, as outlined in Section 3. Then it holds that

$$\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E} \int_{\mathscr{X}_i} \left|\lambda^{n+1}(t,x) - \lambda^n(t,x)\right| \mathrm{d}\mu(x)$$

$$
\begin{aligned}
&\leq \sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i}\int_{(0,t)}\int_{\mathscr{X}} c_x B_{xy}(s)W(x,y)h(t-s)\,|N_s^n - N_s^{n-1}|(\mathrm{d}s\times\mathrm{d}y)\,\mathrm{d}\mu(x) \\
&\leq \sup_{s\geq 0, i\in\mathbb{N}} \int_{\mathscr{X}_i} \|h\|_{L^1(\mathbb{R}_+)} c_x \int_{\mathscr{X}} \mathbb{E}[B_{xy}]W(x,y)\mathbb{E}\left|\lambda^n(s,y)-\lambda^{n-1}(s,y)\right|\,\mathrm{d}y\,\mathrm{d}\mu(x) \\
&= \sup_{s\geq 0, i\in\mathbb{N}} \int_{\mathscr{X}_i} T_{\mathrm{hom}}\left(\mathbb{E}\left|\lambda^n(s,y)-\lambda^{n-1}(s,y)\right|\right)(x)\,\mathrm{d}\mu(x) \\
&\leq \cdots \leq \sup_{t\geq 0, i\in\mathbb{N}} \int_{\mathscr{X}_i} T_{\mathrm{hom}}^n\left(\mathbb{E}\lambda^1(t,\cdot)\right)(x)\,\mathrm{d}\mu(x) \\
&= \sup_{t\geq 0, i\in\mathbb{N}} \int_{\mathscr{X}_i} T_{\mathrm{hom}}^n(f_\cdot(\lambda_\infty(\cdot)+\eta(t,\cdot)))(x)\,\mathrm{d}\mu(x) \\
&= \sup_{t\geq 0, i\in\mathbb{N}} \|T_{\mathrm{hom}}^n(f_\cdot(\lambda_\infty(\cdot)+\eta(t,\cdot)))\|_{L^1(\mathscr{X}_i,\mu)}
\end{aligned}
$$

$$
\leq \|T_{\mathrm{hom}}^n\| \sup_{t\geq 0, i\in\mathbb{N}} \|f_\cdot(\lambda_\infty(\cdot)+\eta(t,\cdot))\|_{L^1(\mathscr{X}_i,\mu)} \to 0 \tag{69}
$$

exponentially, as $n\to\infty$, using Gelfand's formula and our assumption that $\rho(T_{\mathrm{hom}})<1$. More specifically, by Gelfand's formula we can find some $r\in(\rho(T_{\mathrm{hom}}),1)$ and some $K\in\mathbb{N}$ such that for all $n\geq K$ it holds that $\|T_{\mathrm{hom}}^n\|\leq r^n$. For such large $n\geq K$, it thus holds that

$$
\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i}\left|\lambda^{n+1}(t,x)-\lambda^n(t,x)\right|\,\mathrm{d}\mu(x)\leq Cr^n.
$$

Letting $r'\in(r,1)$, Markov's inequality gives that

$$
\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{P}\left(\int_{\mathscr{X}_i}\left|\lambda^{n+1}(t,x)-\lambda^n(t,x)\right|\,\mathrm{d}\mu(x)\geq (r')^n\right)\leq C\left(\frac{r}{r'}\right)^n.
$$

By an application of the Borel-Cantelli lemma, it holds that, with probability 1, only finitely many of those events

$$
\left\{\int_{\mathscr{X}_i}\left|\lambda^{n+1}(t,x)-\lambda^n(t,x)\right|\,\mathrm{d}\mu(x)\geq (r')^n\right\}
$$

occur, hence for any $t\geq 0, i\in\mathbb{N}$, it holds that, with probability 1, $(\lambda^n(t,\cdot))_{n\in\mathbb{N}_0}$ is a Cauchy sequence in $L^1(\mathscr{X}_i,\mu)$; by completeness, this sequence has some (t-uniform) $L^1(\mathscr{X}_i,\mu)$-limit $\lambda(t,\cdot)$.

Note that our limit process $\lambda(t,x)$ is positive a.s. Furthermore, it satisfies

$$
\begin{aligned}
\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i}\lambda(t,x)\,\mathrm{d}\mu(x) &\leq \sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\left[\sum_{n\geq 0}\int_{\mathscr{X}_i}|\lambda^{n+1}(t,x)-\lambda^n(t,x)|\,\mathrm{d}\mu(x)\right] \\
&\leq (1-r)^{-1}\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\left[\int_{\mathscr{X}_i}|\lambda^K(t,x)-\lambda^0(t,x)|\,\mathrm{d}\mu(x)\right] \\
&= (1-r)^{-1}\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i}\lambda^K(t,x)\,\mathrm{d}\mu(x) \\
&\leq (1-r)^{-1}\|T_{\mathrm{hom}}\|^{K-1}\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i}\lambda^1(t,x)\,\mathrm{d}\mu(x)
\end{aligned}
$$

$$
\leq (1-r)^{-1}\|T_{\mathrm{hom}}\|^{K-1}\cdot C<\infty. \tag{70}
$$

This means that the conditional intensity for the arrival of some point anywhere in $\mathscr{X}_i$ is a.s. bounded, uniformly over $t \in \mathbb{R}_+, i \in \mathbb{N}$.

Next, take some set $D \in \mathcal{L}_1 \otimes \mathcal{P}(\mathbb{N}) \otimes \mathcal{L}_2$ with $(\mathcal{Q}_1 \times \sigma \times Q_2)(D) < \infty$. Then,

$$\sum_{n\geq 0} \mathbb{P}\left(|N^{n+1} - N^n|((0,T]\times D) \neq 0\right) \leq \sum_{n\geq 0} \mathbb{E}|N^{n+1} - N^n|((0,T]\times D)$$

$$\leq \mathbb{Q}_2(D) \sum_{n\geq 0} \int_{(0,T]} \sup_{i\in\mathbb{N}} \mathbb{E} \int_{\mathscr{X}_i} |\lambda^{n+1}(t,x) - \lambda^n(t,x)| \,\mathrm{d}\mu(x)\,\mathrm{d}t$$

$$= \mathbb{Q}_2(D) \int_{(0,T]} \sum_{n\geq 0} \sup_{i\in\mathbb{N}} \mathbb{E} \int_{\mathscr{X}_i} |\lambda^{n+1}(t,x) - \lambda^n(t,x)| \,\mathrm{d}\mu(x)\,\mathrm{d}t$$

$$\leq T\mathbb{Q}_2(D) \sum_{n\geq 0} \sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E} \int_{\mathscr{X}_i} |\lambda^{n+1}(t,x) - \lambda^n(t,x)| \,\mathrm{d}\mu(x) < \infty,$$

by the exponential convergence in (69). The interchange of the summation and integration is justified by monotone convergence. By Borel-Cantelli, it follows that, with probability 1, on $(0,T]\times D$, the processes N^n are eventually constant, and in this sense they converge to some limiting point process N, as $n \to \infty$. From our Picard scheme, it follows that on $(0,T]\times D$, with probability 1, the mappings λ^n are eventually constant as well. This means that the spatial coordinate as determined from λ^n, for sufficiently large n, corresponds to the one determined by λ. Furthermore, by Lemma 1, we can define spatial coordinates in a meaningful way even when we only have λ^n.

The limiting process N satisfies

$$N(\mathrm{d}t \times \mathrm{d}(B\times i\times U)) = \bar{N}^i\left(\mathrm{d}t \times \mathrm{d}(B\times U) \times \left[0, \int_{\mathscr{X}_i} \lambda(t,x)\,\mathrm{d}\mu(x)\right]\right). \tag{71}$$

Indeed, for $T>0$ and $D \in \mathcal{L}_1 \otimes \mathcal{P}(\mathbb{N}) \otimes \mathcal{L}_2$ with $(\mathcal{Q}_1 \times \sigma \times Q_2)(D) < \infty$, by Fatou's lemma,

$$\mathbb{E}\int_{(0,T]\times D} \left| N(\mathrm{d}t \times \mathrm{d}(B\times i\times U)) - \bar{N}^i\left(\mathrm{d}t \times \mathrm{d}(B\times U) \times \left[0, \int_{\mathscr{X}_i} \lambda(t,x)\,\mathrm{d}\mu(x)\right]\right)\right|$$

$$\leq \lim_{n\to\infty} \mathbb{E}\int_{(0,T]\times D} \left| \bar{N}^i\left(\mathrm{d}t \times \mathrm{d}(B\times U) \times \left[0, \int_{\mathscr{X}_i} \lambda^n(t,x)\,\mathrm{d}\mu(x)\right]\right) - \bar{N}^i\left(\mathrm{d}t \times \mathrm{d}(B\times U) \times \left[0, \int_{\mathscr{X}_i} \lambda(t,x)\,\mathrm{d}\mu(x)\right]\right)\right|$$

$$\leq T\mathbb{Q}_2(D) \lim_{n\to\infty} \sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\int_{\mathscr{X}_i} |\lambda^n(t,x) - \lambda(t,x)|\,\mathrm{d}\mu(x) = 0. \tag{72}$$

Finally, we verify that the limiting process λ is indeed the intensity density of N. Let

$$\breve{\lambda}(t,x) := f_x\left(\lambda_\infty(x) + \sum_{\substack{(s,y,B_{xy})\in N \\ s<t}} B_{xy}(s) W(x,y) h(t-s)\right)$$

be the intensity density that we expect to find. For each $i \in \mathbb{N}$, we have

$$\mathbb{E}\int_{\mathscr{X}_i} |\lambda(t,x) - \breve{\lambda}(t,x)|\,\mathrm{d}\mu(x)$$

$$\leq \mathbb{E}\int_{\mathscr{X}_i} |\lambda(t,x)-\lambda^n(t,x)|\,\mathrm{d}\mu(x)$$

$$+\,\mathbb{E}\int_{\mathscr{X}_i}\int_{(0,t)}\int_{\mathscr{X}} c_x B_{xy}(s)W(x,y)h(t-s)\,|N_s-N_s^{n-1}|(\mathrm{d}s\times\mathrm{d}y)\,\mathrm{d}\mu(x)$$

$$\leq \mathbb{E}\int_{\mathscr{X}_i} |\lambda(t,x)-\lambda^n(t,x)|\,\mathrm{d}\mu(x)$$

$$+\,\|h\|_{L^1(\mathbb{R}_+)}\int_{\mathscr{X}_i} c_x\mathbb{E}[B_{xy}]W(x,y)\,\mathrm{d}\mu(x)\sup_{t\geq 0}\mathbb{E}\int_{\mathscr{X}}|\lambda(t,y)-\lambda^{n-1}(t,y)|\,\mathrm{d}\mu(y). \tag{73}$$

The right-hand side goes to 0 uniformly in t and in $i\in\mathbb{N}$, as $n\to\infty$, hence (N,λ) is a solution to (6), as desired. It satisfies $\sup_{t\geq 0, i\in\mathbb{N}}\mathbb{E}\psi(S_tN_-,i)<\infty$ by (70), hence this solution is strongly regular, as desired.

Uniqueness. We need a counterpart to [48], Lemma 4, after which we may simply follow the proof of the uniqueness part of [48], Theorem 2. More specifically, we seek a strictly positive function g on L such that $\int_L g(z)\,\mathcal{Q}(\mathrm{d}z)<\infty$ and such that the inequality

$$\int_{\mathbb{R}_+\times L} c_{x(z')}h(t)B_{x(z')y(z)}(z)W(x(z'),y(z))g(z)\,\mathrm{d}t\,\mathcal{Q}(\mathrm{d}z)\leq rg(z')$$

holds. We focus on proving this inequality. In the previous display, the notations $x(z')$, $y(z)$ and $B(z)$ emphasize the dependency of the location and mark on the random $z'\in L$. Let g_0 be the identity function on L, and for $n\in\mathbb{N}_0$, set

$$g_{n+1}(z')=\int_{\mathbb{R}_+\times L} c_{x(z')}h(t)B_{x(z')y(z)}(z)W(x(z'),y(z))g_n(z)\,\mathrm{d}t\,\mathcal{Q}(\mathrm{d}z).$$

By a calculation similar to (69), combined with Gelfand's formula, given $\tilde r\in(\rho(T_{\mathrm{hom}}),1)$, we can find $N\in\mathbb{N}$, such that for all $n\geq N$ it holds that $\int_L g_{n+1}(z)\,\mathcal{Q}(\mathrm{d}z)\leq\tilde r\int_L g_n(z)\,\mathcal{Q}(\mathrm{d}z)$. Hence, for $n\geq N$,

$$\int_L g_n(z)\,\mathcal{Q}(\mathrm{d}z)\leq\tilde r^{n-N}\|T_{\mathrm{hom}}^N\|\int_L g_0(z)\,\mathcal{Q}(\mathrm{d}z)<\infty.$$

Let $r\in(\tilde r,1)$. Set $g(z):=\sum_{n\geq N} r^{-n}g_n(z)$, which converges in $L^1(\mathcal{Q})$. Then it holds that

$$\int_{\mathbb{R}_+\times L} c_{x(z')}h(t)B_{x(z')y(z)}(z)W(x(z'),y(z))g(z)\,\mathrm{d}t\,\mathcal{Q}(\mathrm{d}z)$$
$$=\sum_{n\geq N} r^{-n}g_{n+1}(z')=r\sum_{n\geq N+1} r^{-n}g_n(z')\leq rg(z').$$

Using this g, the reader can readily verify that the uniqueness part of the proof of [48], Theorem 2, applies. ☐

PROPOSITION 2. *Work in the setting of Theorem 1. Assume that the measure corresponding to the Poisson point processes $\bar N^i$ is compatible w.r.t. the left shifts $\{\theta_t\}$. Then there exists a unique stationary solution N to* (6) *such that $\sup_{i\in\mathbb{N}}\mathbb{E}\psi(S_0N_-,i)<\infty$. Also, if N' is the strongly regular non-stationary solution to* (6) *corresponding to an initial condition satisfying*

$$\sup_{t>0,i\in\mathbb{N}}\int_{\mathscr{X}_i}\eta(t,x)\,\mathrm{d}\mu(x)<\infty,\quad\textit{and}\quad\forall i\in\mathbb{N},\ \lim_{t\to\infty}\int_{\mathscr{X}_i}\eta(t,x)\,\mathrm{d}\mu(x)\to 0,$$

then the law of S_tN' converges weakly to the stationary law.

PROPOSITION 3. *Grant Assumption 1. Consider a (possibly nonlinear) graphon Hawkes process N on a σ-finite measure space $(\mathscr{X}, \mathscr{A}, \mu)$, driven by the conditional intensity density specified by (6). Assume that we have a partition of $\mathscr{X}$ into finite, non-null measure sets $\bigsqcup_{i\in\mathbb{N}} \mathscr{X}_i$ that is such that the following conditions hold:*

$$\rho := \|h\|_{L^1(\mathbb{R}_+)} \sup_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} \sup_{y\in\mathscr{X}_j} \int_{\mathscr{X}_i} c_x \mathbb{E}[B_{xy}] W(x,y)\, \mathrm{d}\mu(x) < 1; \tag{74}$$

$$\alpha := \sup_{i\in\mathbb{N}} \int_{\mathscr{X}_i} f_x(\lambda_\infty(x))\, \mathrm{d}\mu(x) < \infty. \tag{75}$$

Assume furthermore that $\epsilon(t,i)$ satisfies: $\sup_{t>0,i\in\mathbb{N}} \epsilon(t,i) < \infty$ and $\lim_{t\to\infty} \epsilon(t,i) \to 0$ for all $i \in \mathbb{N}$, where $\epsilon(t,i)$ describes the effect of the initial condition on site i at time t:

$$\epsilon(t,i) := \sum_{j\in\mathbb{N}} \mathbb{E}\left[\int_{(-\infty,0)\times L_1} h(t-s) \sup_{y\in\mathscr{X}_j} \int_{\mathscr{X}_i} c_x B_{xy} W(x,y)\, \mathrm{d}\mu(x) N(\mathrm{d}s \times \mathrm{d}B)\right]. \tag{76}$$

Then there exists a unique strongly regular solution N to (6) such that

$$\sup_{t\geq 0, i\in\mathbb{N}} \mathbb{E}\psi(S_t N_-, i) < \infty.$$

PROOF. We are working in the general, nonlinear framework described before Theorem 1, with dynamics given by (6). We fit this into the framework of [48] as follows. Take a process N admitting as its $\mathcal{F}_t^N$-intensity kernel

$$\nu(t, \mathrm{d}B \times \mathrm{d}i \times \mathrm{d}U) = \psi(S_t N_-, i) \mathcal{Q}(\mathrm{d}B \times \mathrm{d}i \times \mathrm{d}U),$$

where, with B^i_{xy} and U^i as defined before,

$$\mathcal{Q}(\mathrm{d}B \times \mathrm{d}i \times \mathrm{d}U) = \mathcal{Q}_1(\mathrm{d}B^i_{xy}) \mathcal{Q}_2(\mathrm{d}U^i).$$

Using this notation, we apply [48], Theorem 2.

By the Lipschitz condition for f_x, and by Fubini's theorem, the Lipschitz condition from [48], eqn. (6), here reads: for each $i \in \mathbb{N}$,

$$\begin{aligned}
&|\psi(S_0 N_-, i) - \psi(S_0 N'_-, i)| \\
&= \Bigg| \int_{\mathscr{X}_i} \Bigg[f_x\Bigg(\sum_{j\in\mathbb{N}} \sum_{\substack{(s, B_{xX^j_s(U_s)}, U_s)\in N^j \\ s<0}} B_{xX^j_s(U_s)} W(x, X^j_s(U_s)) h(-s) \Bigg) \\
&\qquad - f_x\Bigg(\sum_{j\in\mathbb{N}} \sum_{\substack{(s, B_{xX^j_s(U_s)}, U_s)\in (N')^j \\ s<0}} B_{xX^j_s(U_s)} W(x, X^j_s(U_s)) h(-s) \Bigg) \Bigg] \mathrm{d}\mu(x) \Bigg| \\
&\leq \int_{(-\infty,0)\times L} \int_{\mathscr{X}_i} c_x B_{xX^j_s(U_s)} W(x, X^j_s(U_s)) h(-s) |N - N'|(\mathrm{d}s \times \mathrm{d}B \times \mathrm{d}U \times \mathrm{d}j). \qquad (77)
\end{aligned}$$

This equation is valid for all $i \in \mathbb{N}$, hence the Lipschitz dominating function from [48], eqn. (6), reads

$$\bar{h}(t, (B, j, U), i) = h(t) \int_{\mathscr{X}_i} c_x B_{xX^j_{-t}(U_{-t})} W(x, X^j_{-t}(U_{-t}))\, \mathrm{d}\mu(x). \tag{78}$$

Now condition (10) from [48], Theorem 2, is satisfied if $\rho < 1$. Next, condition (11) from [48], Theorem 2, requires that $\alpha < \infty$. Finally, the function involved in [48], eqn. (12), can be bounded by $\epsilon(t,i)$. □

Appendix B: Relegated proofs of Section 4.

PROOF OF LEMMA 2. We work with $\mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}]$, a graphon Hawkes process N with corresponding integral operator

$$T_{\text{hom}} : L^1(\mathscr{X}) \to L^1(\mathscr{X}) : f(\cdot) \mapsto \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} \mathbb{E}[B_{\cdot y}] W(\cdot, y) f(y) \, \mathrm{d}y,$$

and prelimit processes $\tilde{N}^d$ with averaged parameters as described in Section 4.2, with corresponding integral operators

$$\tilde{T}^{(d)}_{\text{hom}} : L^1(\mathscr{X}) \to L^1(\mathscr{X}) : f(\cdot) \mapsto \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} \mathbb{E}[\tilde{B}^d_{\cdot y}] \tilde{W}^d(\cdot, y) f(y) \, \mathrm{d}y.$$

We first prove that $\|T_{\text{hom}} - \tilde{T}^{(d)}_{\text{hom}}\| \to 0$ as $d \to \infty$. By the total variation bound of Lemma 3, and by Assumptions 3, 4, we can select d sufficiently large, i.e., $\operatorname{mesh}(\mathcal{P}^d)$ sufficiently small, such that both

$$\sup_{y \in \mathscr{X}} \mathbb{E}\|B_{\cdot y} - \tilde{B}^d_{\cdot, y}\|_{L^1(\mathscr{X})} \quad \text{and} \quad \sup_{y \in \mathscr{X}} \|W(\cdot, y) - \tilde{W}^d(\cdot, y)\|_{L^1(\mathscr{X})}$$

are bounded by

$$\frac{\epsilon}{\|h\|_{L^1(\mathbb{R}_+)}(C_B + C_W)\mathrm{Leb}^m(\mathscr{X})}.$$

Now let $f \in L^1(\mathscr{X})$ with $\|f\|_{L^1(\mathscr{X})} = 1$. Then, by Assumption 2,

$$\begin{aligned}
&\|(T_{\text{hom}} - \tilde{T}^{(d)}_{\text{hom}}) f\|_{L^1(\mathscr{X})} \\
&= \left\| \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} \Big(\mathbb{E}[B_{\cdot y}] W(\cdot, y) - \mathbb{E}[\tilde{B}^d_{\cdot y}] \tilde{W}^d(\cdot, y) \Big) f(y) \, \mathrm{d}y \right\|_{L^1(\mathscr{X})} \\
&\leq \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} |f(y)| \int_{\mathscr{X}} \Big| \mathbb{E}[B_{xy}] W(x, y) - \mathbb{E}[\tilde{B}^d_{xy}] W(x, y) \Big| \, \mathrm{d}x \, \mathrm{d}y \\
&+ \|h\|_{L^1(\mathbb{R}_+)} \int_{\mathscr{X}} |f(y)| \int_{\mathscr{X}} \Big| \mathbb{E}[\tilde{B}^d_{xy}] W(x, y) - \mathbb{E}[\tilde{B}^d_{xy}] \tilde{W}^d(x, y) \Big| \, \mathrm{d}x \, \mathrm{d}y \\
&< \epsilon, \qquad (79)
\end{aligned}$$

whence $\|T_{\text{hom}} - \tilde{T}^{(d)}_{\text{hom}}\| \leq \epsilon$.

Now note that for $|\lambda| > \rho(T_{\text{hom}}) + \delta$, the resolvent $R(\lambda; T_{\text{hom}})$ exists since $\lambda \notin \sigma(T_{\text{hom}})$. Hence, for $\|T_{\text{hom}} - \tilde{T}^{(d)}_{\text{hom}}\| < 1/\|R(\lambda; T_{\text{hom}})\|$,

$$\tilde{T}^{(d)}_{\text{hom}} - \lambda I = T_{\text{hom}} - \lambda I + \tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}} = \Big(I - (\tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}}) R(\lambda; T_{\text{hom}}) \Big) (T_{\text{hom}} - \lambda I)$$

is a product of invertible operators, hence is invertible itself. We aim to show that for all d sufficiently large such, $\lambda \notin \sigma(\tilde{T}^{(d)}_{\text{hom}})$ for $|\lambda| > \rho(T_{\text{hom}}) + \delta$. To this end, we show that

$$\inf_{|\lambda| > \rho(T_{\text{hom}}) + \delta} \left\{ \frac{1}{\|R(\lambda; T_{\text{hom}})\|} \right\} > 0. \qquad (80)$$

By a Neumann series, we can write

$$R(\mu; T_{\text{hom}}) = \sum_{n \geq 0} R(\lambda; T_{\text{hom}})^{n+1} (\lambda - \mu)^n,$$

hence for $|\lambda - \mu| < 1/(2\|R(\lambda; T_{\text{hom}})\|)$, it follows that $\|R(\mu; T_{\text{hom}})\| \leq 2\|R(\lambda; T_{\text{hom}})\|$. Now it follows from the resolvent identity

$$R(\lambda; T_{\text{hom}}) - R(\mu; T_{\text{hom}}) = (\lambda - \mu) R(\lambda; T_{\text{hom}}) R(\mu; T_{\text{hom}})$$

that the resolvent $R(\lambda; T_{\text{hom}})$ of a bounded linear operator T_{hom} is a continuous function of λ, with operator norm tending to 0, as $\lambda \to \infty$. Indeed, again by a Neumann series, for λ sufficiently large,

$$\|R(\lambda; T_{\text{hom}})\| = \left\| \sum_{n \geq 0} T_{\text{hom}}^n / \lambda^{n+1} \right\| \leq \frac{1}{\lambda - \|T_{\text{hom}}\|}.$$

Now (80) follows. Find D' large enough such that

$$\|T_{\text{hom}} - \tilde{T}_{\text{hom}}^{(d)}\| < \inf_{|\lambda| > \rho(T_{\text{hom}}) + \delta} \left\{ \frac{1}{\|R(\lambda; T_{\text{hom}})\|} \right\}$$

for all $d \geq D'$. It follows that $\tilde{T}_{\text{hom}}^{(d)} - \lambda I$ is invertible for all $|\lambda| > \rho(T_{\text{hom}}) + \delta$, hence $\rho(\tilde{T}_{\text{hom}}^{(d)}) \leq \rho(T_{\text{hom}}) + \delta < 1$, for all $d \geq D'$.

We now prove that we can find some $D \in \mathbb{N}$ sufficiently large such that the expected cluster sizes $\mathbb{E}[Z_x^d]$ of $\tilde{N}^d$ of a cluster generated by a particle in $x \in \mathscr{X}$ are uniformly bounded over $x \in \mathscr{X} = [\boldsymbol{a}, \boldsymbol{b}]$ and $d \geq D$. Note that the maximum expected cluster size over $x \in \mathscr{X}$ for $\tilde{N}^d$ can be bounded by

$$\sup_{x \in \mathscr{X}} \mathbb{E}[Z_x^d] \leq \sum_{n \geq 0} \|(\tilde{T}_{\text{hom}}^{(d)})^n\|, \tag{81}$$

hence it suffices to find $D \in \mathbb{N}$ such that the right-hand side of (81) is uniformly bounded over $d \geq D$. To this end, it suffices to prove that

$$\text{for all } n, \quad \limsup_{d \to \infty} \|(\tilde{T}_{\text{hom}}^{(d)})^n\| \leq \|T_{\text{hom}}^n\|. \tag{82}$$

Indeed, fix $\epsilon > 0$. Since $\rho(T_{\text{hom}}) < 1$, we can find $N \in \mathbb{N}$ such that $\|T_{\text{hom}}^N\| < 1 - \epsilon$. Assuming (82), we can find D such that for all $d \geq D$ and $n \leq N$, $\|(\tilde{T}_{\text{hom}}^{(d)})^n\| \leq \|T_{\text{hom}}^n\| + \epsilon/2$. Hence,

$$\begin{aligned}
\sum_{n \geq 0} \|(\tilde{T}_{\text{hom}}^{(d)})^n\| &= \sum_{k \geq 0} \sum_{n=0}^{N-1} \|(\tilde{T}_{\text{hom}}^{(d)})^{kN+n}\| \\
&\leq \sum_{k \geq 0} \|(\tilde{T}_{\text{hom}}^{(d)})^N\|^k \sum_{n=0}^{N-1} \|(\tilde{T}_{\text{hom}}^{(d)})^n\| \\
&\leq \sum_{k \geq 0} (1 - \epsilon/2)^k \sum_{n=0}^{N-1} (\|T_{\text{hom}}^n\| + \epsilon/2) < \infty,
\end{aligned}$$

providing a uniform bound over $d \geq D$.

To prove (82), it suffices to prove

$$\text{for all } d, n \in \mathbb{N}, \quad \|(\tilde{T}_{\text{hom}}^{(d)})^n\| \leq \|(\tilde{T}_{\text{hom}}^{(d)})^{n-1} T_{\text{hom}}\|; \tag{83}$$

$$\text{for all } n \in \mathbb{N}, \quad \|(\tilde{T}_{\text{hom}}^{(d)})^{n-1} T_{\text{hom}} - T_{\text{hom}}^n\| \to 0, \quad \text{as } d \to \infty. \tag{84}$$

For (83), let $P_d : L^1(\mathscr{X}) \to L^1(\mathscr{X})$ be the conditional expectation operator onto the σ-algebra generated by the partition $\mathcal{P}^d$ used to define $\tilde{N}^d$ and $\tilde{T}^{(d)}_{\text{hom}}$. Then $\|P_d\| \le 1$, see e.g., [61], Theorem 27.11(ii). Note that we can write $\tilde{T}^{(d)}_{\text{hom}} = P_d T_{\text{hom}} P_d$, hence it follows that

$$\|(\tilde{T}^{(d)}_{\text{hom}})^n\| = \|(\tilde{T}^{(d)}_{\text{hom}})^{n-1} P_d T_{\text{hom}} P_d\| \le \|(\tilde{T}^{(d)}_{\text{hom}})^{n-1} P_d T_{\text{hom}}\| = \|(\tilde{T}^{(d)}_{\text{hom}})^{n-1} T_{\text{hom}}\|.$$

To prove (84), write

$$(\tilde{T}^{(d)}_{\text{hom}})^{n-1} T_{\text{hom}} - T^n_{\text{hom}} = \sum_{k=1}^{n-1} (\tilde{T}^{(d)}_{\text{hom}})^{n-k-1} (\tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}}) T^k_{\text{hom}}.$$

By Assumption 2, $\|\tilde{T}^{(d)}_{\text{hom}}\|, \|T_{\text{hom}}\| \le C := \|h\|_{L^1(\mathbb{R}_+)} C_B C_W < \infty$. Hence, for all $k \in [n-1]$,

$$\begin{aligned} &\|(\tilde{T}^{(d)}_{\text{hom}})^{n-k-1} (\tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}}) T^k_{\text{hom}}\| \\ &\le \|\tilde{T}^{(d)}_{\text{hom}}\|^{n-k-1} \|\tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}}\| \|T_{\text{hom}}\|^k \\ &\le C^{n-1} \|\tilde{T}^{(d)}_{\text{hom}} - T_{\text{hom}}\| \to 0, \end{aligned}$$

as $d \to \infty$, by (79). □

PROOF OF LEMMA 3. We prove the result for $f \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}^m)$. If we work on equivalence classes of functions, i.e., $f \in L^1_{\text{loc}}(\mathbb{R}^m)$, take a representant g of f, apply the bound (88) to g, and note that $\|f - f^d\|_{L^1(\mathscr{X})} = \|g - g^d\|_{L^1(\mathscr{X})}$. Taking the infimum over $g \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}^m)$ yields the result for equivalence classes of functions as well.

The assumptions made in the lemma imply that $\mathcal{P}^d$ partitions $\mathscr{X}$ into $K = K(d)$ hyperrectangles $\mathscr{X}^d_1, \dots, \mathscr{X}^d_K$. We first prove the bound for each such hyperrectangle $\mathscr{X}^d_n$, $n \in [K]$. To this end, fix $\Omega := \mathscr{X}^d_n$, and set $u : \Omega \to \mathbb{R} : x \mapsto f(x) - f^d(x)$; notice that $\int_\Omega u = 0$. We use [45], Theorem 13.9, to approximate u by a sequence of smooth functions

$$(\hat{u}_k)_{k \in \mathbb{N}} \subset W^{1,1}(\Omega) \cap C^\infty(\Omega) \quad \text{such that} \quad \hat{u}_k \xrightarrow{L^1(\Omega)} u, \quad \operatorname{Var}(\hat{u}_k, \Omega) \to \operatorname{Var}(u, \Omega),$$

as $k \to \infty$. Here, $W^{1,1}(\Omega) \subset L^1(\Omega)$ denotes the Sobolev space of functions having weak derivatives in $L^1(\Omega)$. Set

$$u_k = \hat{u}_k - \int_\Omega \hat{u}_k \in W^{1,1}(\Omega) \cap C^\infty(\Omega).$$

Since $\hat{u}_k \xrightarrow{L^1(\Omega)} u$ and $\int_\Omega u = 0$, it follows by the reverse triangle inequality on $L^1(\Omega)$ that $\|\hat{u}_k\|_{L^1(\Omega)} \to 0$ as $k \to \infty$. Hence,

$$\|u - u_k\|_{L^1(\Omega)} \le \|u - \hat{u}_k\|_{L^1(\Omega)} + \|\hat{u}_k\|_{L^1(\Omega)} \to 0 \quad \text{as } k \to \infty,$$

i.e., $u_k \xrightarrow{L^1(\Omega)} u$. After replacing $(u_k)_{k \in \mathbb{N}}$ by a subsequence, if necessary, we may assume that $u_k \to u$ a.s. Furthermore, since the total variation does not change by adding a constant to a function, $\operatorname{Var}(u_k, \Omega) = \operatorname{Var}(\hat{u}_k, \Omega) \to \operatorname{Var}(u, \Omega)$, as $k \to \infty$. Therefore, we may assume without loss of generality that our approximating sequence satisfies $\int_\Omega u_k = 0$ for all $k \in \mathbb{N}$.

We prove a bound for the approximating smooth functions $(u_k)_{k \in \mathbb{N}}$, which we then extend to u itself. To this end, note that u_k belongs to the Sobolev space $W^{1,1}$, being smooth. Since we are working on a convex domain Ω and since $\int_\Omega u_k = 0$, [1], Theorem 3.2, gives that

$$\|u_k\|_{L^1(\Omega)} \le \frac{1}{2} \operatorname{diam}(\Omega) \|\nabla u_k\|_{L^1(\Omega)}, \tag{85}$$

where ∇ denotes the gradient operator. It holds that $W^{1,1} \subset \mathrm{BV}(\Omega)$, and for $u_k \in W^{1,1}$ we have $\|\nabla u_k\|_{L^1(\Omega)} = \mathrm{Var}(u_k, \Omega)$, see [45], §13.2. Now Fatou's lemma and (85) imply

$$\|u\|_{L^1(\Omega)} \le \liminf_{k\to\infty} \|u_k\|_{L^1(\Omega)} \le \lim_{k\to\infty} \frac{1}{2}\mathrm{diam}(\Omega)\|\nabla u_k\|_{L^1(\Omega)}$$

$$= \lim_{k\to\infty} \frac{1}{2}\mathrm{diam}(\Omega)\mathrm{Var}(u_k, \Omega) = \frac{1}{2}\mathrm{diam}(\Omega)\mathrm{Var}(u, \Omega). \tag{86}$$

Again, since the total variation is insensitive to adding a constant to a function, this bound implies that

$$\|f - f^d\|_{L^1(\Omega)} \le \frac{1}{2}\mathrm{diam}(\Omega)\mathrm{Var}(f - f^d, \Omega) = \frac{1}{2}\mathrm{diam}(\Omega)\mathrm{Var}(f, \Omega). \tag{87}$$

Next, because the L^1-norm is additive and the total variation is superadditive over disjoint sets,

$$\|f - f^d\|_{L^1(\mathscr{X})} = \sum_{i=1}^{K} \|f - f^d\|_{L^1(\mathscr{X}_i^d)} \le \sum_{i=1}^{K} \frac{1}{2}\mathrm{diam}(\mathscr{X}_i^d)\mathrm{Var}(f, \mathscr{X}_i^d)$$

$$\le \frac{1}{2}\mathrm{mesh}(\mathcal{P}^d)\sum_{i=1}^{K} \mathrm{Var}(f, \mathscr{X}_i^d) \le \frac{1}{2}\mathrm{mesh}(\mathcal{P}^d)\mathrm{Var}(f, \mathscr{X}), \tag{88}$$

since we defined the mesh of a partition as the largest diameter among its elements. □

LEMMA 7. *Let $[a, b] \subset \mathbb{R}$ be some closed interval. Assume that a, b are endpoints of intervals $\mathscr{X}_n^d, \mathscr{X}_m^d$ from the partition $\mathcal{P}^d$ of $\mathbb{R}$ into intervals; if this is not the case, replace $\mathcal{P}^d$ by its coarsest refinement containing a, b as endpoints. Assume that $\mathbb{R}$ is equipped with the Borel σ-algebra and a measure μ dominated by the Lebesgue measure. Consider a piecewise constant approximation f^d of the μ-locally integrable function $f : \mathbb{R} \to \mathbb{R}$, where f^d is constant on each $\mathscr{X}_n^d$, with f^d calculated as a measurable functional of $f|_{\mathscr{X}_n^d}$, and*

$$f^d(\mathscr{X}_n^d) \in \left[\operatorname*{ess\,inf}_{x\in\mathscr{X}_n^d} f(x), \operatorname*{ess\,sup}_{x\in\mathscr{X}_n^d} f(x)\right].$$

Let $\mathrm{mesh}(\mathcal{P}^d)$ be the maximum μ-measure of all elements of $\mathcal{P}^d$. Then it holds that $f^d \in L^1([a, b], \mu)$, $f^d \to f$ a.e. as $\mathrm{mesh}(\mathcal{P}^d) \to 0$, and finally, with $\|\cdot\|_{\mathrm{TV}[a,b]}$ the usual total variational norm over $[a, b]$,

$$\|f^d - f\|_{L^1([a,b],\mu)} \le \|f\|_{\mathrm{TV}[a,b]}\mathrm{mesh}(\mathcal{P}^d). \tag{89}$$

PROOF. The first two claims are easy. For the final claim, the total variation norm of $f \in L^1([a, b], \mu)$ may depend on the representant $g \in [f] := \{\bar{f} \in \mathcal{L}([a, b], \mu) : \bar{f} = f\ \mu\text{-a.e.}\}$. Therefore, we set $\|f\|_{\mathrm{TV[a,b]}} := \inf_{g\in[f]} \mathrm{var}_{[a,b]}(g)$, where

$$\mathrm{var}_{[a,b]}(g) := \sup\left\{\sum_{i=1}^{n} |g(x_{i+1}) - g(x_i)| : n \in \mathbb{N}, a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\right\}.$$

Select a $g \in [f]$. Then, $g^d \equiv f^d$ on $[a, b]$. Note that for $x \in \mathscr{X}_n^d$, it holds that

$$|g^d(x) - g(x)| \le \sup_{u,v\in\mathscr{X}_n^d} |g(u) - g(v)| =: \Delta_n^d.$$

It follows that

$$
\begin{aligned}
\|f^d - f\|_{L^1([a,b],\mu)} &= \int_a^b |f^d(x) - f(x)| \, \mathrm{d}\mu(x) = \int_a^b |g^d(x) - g(x)| \, \mathrm{d}\mu(x) \\
&= \sum_{n\in\mathbb{N}} \int_{\mathscr{X}_n^d \cap [a,b]} |g^d(x) - g(x)| \, \mathrm{d}\mu(x) \le \sum_{n\in\mathbb{N}} \int_{\mathscr{X}_n^d \cap [a,b]} \Delta_n^d \, \mathrm{d}\mu(x) \\
&= \sum_{\substack{n\in\mathbb{N} \\ \mu(\mathscr{X}_n^d \cap [a,b]) \neq 0}} \mu(\mathscr{X}_n^d) \Delta_n^d \le \operatorname{mesh}(\mathcal{P}^d) \sum_{\substack{n\in\mathbb{N} \\ \mu(\mathscr{X}_n^d \cap [a,b]) \neq 0}} \Delta_n^d \\
&\le \operatorname{mesh}(\mathcal{P}^d) \operatorname{var}_{[a,b]}(g).
\end{aligned}
$$

Taking the infimum over $g \in [f]$ over both sides of this inequality, (89) follows. □

PROOF OF COROLLARY 1. We indicate the adjustments needed in the proof of Theorem 2, which consists of bounding the contributions (i)–(iii) introduced there. For (i), convergence of $\breve{\lambda}_\infty^d$ to λ_∞ in $L^1(\mathscr{X})$ is required, and this is ensured by condition (b) of Corollary 1. For (ii), the argument must be refined using conditions (c)–(d) of Corollary 1. Below we describe the necessary modifications for (ii); contribution (iii) can then be handled analogously.

In contribution (ii), the expected offspring count of a particle located at $y \in \mathscr{X}$ was previously bounded from above by taking a supremum over $y \in \mathscr{X}$. A sharper estimate is obtained by observing that the spatial coordinate y is random. A particle at y may arise either as an immigrant or as an offspring. In the immigrant case, its spatial density is $g_I(y) := \lambda_\infty(y)/\alpha$, which is well-defined and satisfies $\|g_I\|_{L^\infty(\mathscr{X})} < \infty$ by Assumption 2. In the offspring case, the particle is generated by a parent event at $z \in \mathscr{X}$ and time s, in which case y has spatial density

$$
g_{O,z}(y) := \frac{B_{yz}(s) W(y,z)}{\int_{\mathscr{X}} B_{xz}(s) W(x,z) \, \mathrm{d}x},
$$

and this satisfies $\sup_{z\in\mathscr{X}} \|g_{O,z}\|_{L^\infty(\mathscr{X})} \le C_B C_W / C < \infty$ by Assumption 2 and condition (d) of Corollary 1. Consequently, the spatial density g_t of an arriving particle at time t obeys

$$
\|g_t\|_{L^\infty(\mathscr{X})} \le \|g_I\|_{L^\infty(\mathscr{X})} \vee \sup_{z\in\mathscr{X}} \|g_{O,z}\|_{L^\infty(\mathscr{X})} =: C^* < \infty,
$$

uniformly for all $t \in [0, \infty)$.

In bounding contribution (ii) in the proof of Theorem 2, consider a simultaneous event at time t with spatial coordinate y distributed according to $h(y)$. Here, h is either g_I or $g_{O,z}$, depending on whether the event is an immigration event or an offspring event with parent in $z \in \mathscr{X}$. For such an event, the spatial densities of the expected number of offsprings in x for the first and second process are $\int_{\mathscr{X}} \mathbb{E}[B_{xy}] W(x,y) h(y) \, \mathrm{d}y$ and $\int_{\mathscr{X}} \mathbb{E}[\breve{B}_{xy}^d] \breve{W}^d(x,y) h(y) \, \mathrm{d}y$, respectively. After stochastically coupling these offspring processes for each parent event that is either an immigrant or results from a grandparent at location z, the difference in parameters leads to an expected discrepancy in offspring counts bounded by

$$
\begin{aligned}
&\int_{\mathscr{X}} \left| \int_{\mathscr{X}} \Big(\mathbb{E}[B_{xy}] W(x,y) - \mathbb{E}[\breve{B}_{xy}^d] \breve{W}^d(x,y) \Big) h(y) \, \mathrm{d}y \right| \mathrm{d}x \\
&\le \|h\|_{L^\infty(\mathscr{X})} \sup_{\|\phi\|_{L^\infty(\mathscr{X})}=1} \int_{\mathscr{X}} \left| \int_{\mathscr{X}} \Big(\mathbb{E}[B_{xy}] W(x,y) - \mathbb{E}[\breve{B}_{xy}^d] \breve{W}^d(x,y) \Big) \phi(y) \, \mathrm{d}y \right| \mathrm{d}x \\
&\le C^* \cdot \|\mathbb{E}[B] W - \mathbb{E}[\breve{B}^d] \breve{W}^d\|_{L^\infty(\mathscr{X}) \to L^1(\mathscr{X})},
\end{aligned}
$$

where h is either g_I or $g_{O,z}$. The quantity on the right-hand side converges to 0 by condition (c) of Corollary 1 and [47], Lemma 8.11.

From this point on, we can proceed as in the proof of Theorem 2. □

PROOF OF PROPOSITION 1. Suppose that the partitions $(\mathcal{P}^d)_{d\in\mathbb{N}}$ are induced by the first $d-1$ hyperplanes of a random sample $(\mathcal{V}_i)_{i\in\mathbb{N}}$ of $(m-1)$-dimensional hyperplanes orthogonal to one of the coordinates. More specifically, the $\mathcal{V}_i$s are sampled as follows. First, the coordinate $n \in [m]$ to which $\mathcal{V}_i$ is orthogonal is picked uniformly at random, i.e., $n \sim \mathrm{Uni}(\{1,\dots,m\})$. Next, we draw $z_n \sim \mathrm{Uni}(a_n, b_n)$, and set $\mathcal{V}_i := \{\mathbf{x} \in \mathbb{R}^m : x_n = z_n\}$. For $d = 1$, we simply split $\mathscr{X}$ into two parts, obtaining $\mathcal{P}^2$. Next, for $d \geq 2$, $\mathcal{V}_{d-1}$ intersects $\mathscr{X}$, splitting $S \in [d-1]$ sets $\mathscr{X}^{d-1}_{n(1)}, \dots, \mathscr{X}^{d-1}_{n(S)}$ of $\mathcal{P}^{d-1}$ into two parts ($S \geq 1$ a.s.). Now obtain $\mathcal{P}^d$ as a refinement of $\mathcal{P}^{d-1}$ by splitting only the set $\mathscr{X}^{d-1}_{n(i)}$ ($i \in [S]$) having the largest diameter into two parts; in case of a tie, choose randomly between the ties.

Next, take $d \in \mathbb{N}$, and for each $i \in [d]$, select a point $x_i \in \mathscr{X}^d_i$. Let $M_d = \{x_i : i \in [d]\} \subset \mathscr{X}$ be the set of those points. Then $(M_d)_{d\in\mathbb{N}}$ constitutes a sequence of well-distributed sets in the sense of [47], p. 185; i.e., M_d converges weakly to a uniform measure on $\mathscr{X}$. By a straightforward modification of the proof of [47], Lemma 11.33 — for which we need Assumption 5, i.e., a.s. continuity of W —, it follows that the random connectivity graph $\mathbb{G}^d$ on $[d]$ with edge set $\{Z^d_{ij} : i,j \in [d]\}$ converges to W in the cut norm, a.s. Since we use possibly non-symmetric graphons and in any case sample looped directed graphs, we use the framework of [42, 46], which allows to go beyond simple, symmetric graphs and graphons.

For the ucp convergence, to invoke the proof of Theorem 3, we only need $\mathrm{mesh}(\mathcal{P}^d) \to 0$ as $d \to \infty$, which is evident. □

Appendix C: Supplement to Section 5.1. Before Assumption 6, we claimed the following.

PROPOSITION 4. *Let $T_{\mathrm{hom}} \in B(L^1[0,1])$ be a bounded operator with $\rho(T_{\mathrm{hom}}) > 1$. Then there exists some $g \in L^1_+[0,1]$ such that $\|T^n_{\mathrm{hom}} g\|_{L^1[0,1]} \to \infty$, as $n \to \infty$.*

PROOF. Observe that $T^n_{\mathrm{hom}} f$ records the spatial density of expected nth-generation offspring in a cluster with an immigrant distributed according to density $f \in L^1[0,1]$. Informally, letting δ_x be the Dirac delta function centered at x, $T^n_{\mathrm{hom}}\delta_x$ records the spatial density of eventual expected nth-generation offspring within a cluster generated by an immigrant in x. If $\rho(T_{\mathrm{hom}}) > 1$, then $\|T^n_{\mathrm{hom}}\| \to \infty$ as $n \to \infty$, hence there exists $(f_n)_{n\in\mathbb{N}} \subset L^1[0,1]$ with $\|f_n\|_{L^1[0,1]} = 1$ for all $n \in \mathbb{N}$, such that $\|T^n_{\mathrm{hom}} f_n\|_{L^1[0,1]} \to \infty$, as $n \to \infty$. Set $\rho' := 1 + \frac{1}{2}(\rho(T_{\mathrm{hom}}) - 1)$ and $\rho'' := 1 + \frac{1}{3}(\rho(T_{\mathrm{hom}}) - 1)$.

We present a construction for a function g, independent of n, such that $\|T^n_{\mathrm{hom}} g\|_{L^1[0,1]} \to \infty$, as $n \to \infty$. For each $n \in \mathbb{N}$, we can find $f_n \in L^1[0,1]$ with $\|f_n\|_{L^1[0,1]} = 1$ such that $\|T^n_{\mathrm{hom}} f_n\|_{L^1[0,1]} \geq (\rho')^n$. Approximate $f^{\pm}_n \in L^1_+[0,1]$ from below by simple functions. By continuity of norms, we can find a simple function $s_n \leq f_n$ pointwise, satisfying $\|T^n_{\mathrm{hom}} s_n\|_{L^1[0,1]} \geq (\rho'')^n$. Consequently, for each n, we can find a set $A_n \in \mathcal{B}[0,1]$ of positive measure, corresponding to a normalized indicator function $\chi_n = (\mathrm{Leb}(A_n))^{-1}\mathbf{1}\{A_n\}$ satisfying $\|T^n_{\mathrm{hom}} \chi_n\|_{L^1[0,1]} \geq (\rho'')^n$. Set $\gamma = (1+\rho'')/(2\rho'') \in (\frac{1}{2}, 1)$; then $\gamma\rho'' > 1$. We define $g \in L^1_+[0,1]$ by $g = (1-\gamma)\sum_{n\geq 1}\gamma^{n-1}\chi_n$; it is easy to see that $\|g\|_{L^1[0,1]} \leq 1$, while for each $n \in \mathbb{N}$,

$$\|T^n_{\mathrm{hom}} g\|_{L^1[0,1]} \geq \frac{1-\gamma}{\gamma}\gamma^n \|T^n_{\mathrm{hom}}\chi_n\|_{L^1[0,1]} \geq \frac{1-\gamma}{\gamma}(\gamma\rho'')^n \to \infty, \quad \text{as } n \to \infty. \qquad \square$$

The existence of g indicates instability of our system. To prove $N_T(A)/T \to \infty$ a.s. using this function g, we need to argue that there is a positive stream of immigrants leading to a point with infinite expected offspring. However, it may be possible that g has an infinite spike (of finite L^1 size) creating the unstable behavior, and there is no way of bounding below the size of a set of $x \in [0,1]$ having infinite offspring given only the existence of such a function g. The next proposition indicates that this approach is not fruitful without using further properties of $(T^n_{\text{hom}})_{n\geq 0}$.

PROPOSITION 5. *There exist a sequence* $(L_n)_{n\in\mathbb{N}} \subset B(L^1[0,1])$ *and a* $g \in L^1[0,1]$ *satisfying* $\|L_n g\|_{L^1[0,1]} \to \infty$, *but such that there does not exist a simple function* $\tilde{g}$ *with* $\|L_n \tilde{g}\|_{L^1[0,1]} \to \infty$.

PROOF. The existence of such a simple function $\tilde{g}$ would imply the existence of some $A \in \mathcal{B}[0,1]$ for which $\|L_n \mathbf{1}_A\|_{L^1[0,1]} \to \infty$. Note that $\|L_n(\cdot)\| \in L^1[0,1]^*$, the dual of $L^1[0,1]$, which is isometrically isomorphic to $L^\infty[0,1]$. In this dual space, to prove our result it suffices to find a sequence $(\phi_n)_{n\in\mathbb{N}} \subset L^\infty[0,1]$ and a $g \in L^1[0,1]$ such that $\int_0^1 \phi_n(x)g(x)\,\mathrm{d}x \to \infty$ as $n \to \infty$, and such that for any indicator function $\mathbf{1}_A$ it holds that $\int_0^1 \phi_n(x)\mathbf{1}_A(x)\,\mathrm{d}x = \int_A \phi_n(x)\,\mathrm{d}x \to 0$. To this end, let $g(x) = 1/(2\sqrt{x})$ and $\phi_n(x) = n^{3/4}\mathbf{1}_{[0,n^{-1}]}$. For any $A \in \mathcal{B}[0,1]$,

$$\int_A \phi_n(x)\,\mathrm{d}x = n^{3/4}\,\mathrm{Leb}([0,n^{-1}]\cap A) \leq n^{-1/4} \to 0, \quad \text{as } n \to \infty. \qquad \square$$

Appendix D: Relegated proofs of Section 6.

PROOF OF THEOREM 7. Conditionally upon having n immigrants in $[0,t]$, those immigrants arrive uniformly over time, since *ignoring the spatial coordinate*, they arrive according to a homogeneous Poisson process of rate $\alpha := \|\lambda_\infty(\cdot)\|_{L^1[0,1]}$. Furthermore, the spatial locations are i.i.d. with Lebesgue densities proportional to $\lambda_\infty(\cdot)$. Let p^n denote the Lebesgue density of the spatial coordinates in $[0,1]^n$ for a sample of n immigrants. We calculate the Laplace functional of Q by conditioning on the number of immigrants and on their spatial coordinates. Let I_t be the number of immigrants that have arrived by time t. Then, with the understanding that $\int_{[0,1]^0} f \equiv 1$,

$$\begin{aligned}
\mathcal{L}_Q(f,t) &= \mathbb{E}\left[\exp\left(-\sum_{z_i\in Q(t)} f(z_i)\right)\right] \\
&= \sum_{n\geq 0} \mathbb{E}\left[\mathbb{E}\left[\exp\left(-\sum_{z_i\in Q(t)} f(z_i)\right)\right]\middle| I_t = n\right]\mathbb{P}(I_t = n) \\
&= \sum_{n\geq 0}\int_{[0,1]^n} \mathbb{E}\left[\mathbb{E}\left[\exp\left(-\sum_{z_i\in Q(t)} f(z_i)\right)\right]\middle| I_t = n,\ \text{locations } x_1,\ldots,x_n\right] \\
&\qquad \times p^n(x_1,\ldots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n\,\mathbb{P}(I_t = n) \\
&= \sum_{n\geq 0}\int_{[0,1]^n}\prod_{j=1}^n\left(\frac{1}{t}\int_0^t \eta_{x_j}(f,u)\,\mathrm{d}u\right)\frac{\lambda_\infty(x_j)}{\|\lambda_\infty(\cdot)\|_{L^1[0,1]}}\,\mathrm{d}x_1\cdots\mathrm{d}x_n\,\frac{e^{-\alpha t}(\alpha t)^n}{n!} \\
&= \sum_{n\geq 0}\frac{e^{-\alpha t}}{n!}\prod_{j=1}^n\left\{\int_0^1\int_0^t \eta_{x_j}(f,u)\,\mathrm{d}u\,\lambda_\infty(x_j)\,\mathrm{d}x_j\right\}
\end{aligned}$$

$$= \sum_{n \geq 0} \frac{e^{-\alpha t}}{n!} \left(\int_0^1 \int_0^t \eta_x(f,u) \lambda_\infty(x) \, \mathrm{d}u \mathrm{d}x \right)^n$$

$$= \exp \left(\int_0^1 \int_0^t \left(\eta_x(f,u) - 1 \right) \lambda_\infty(x) \, \mathrm{d}u \, \mathrm{d}x \right),$$

where $z_i \in Q(t)$ denotes the set of spatial coordinates corresponding to particles $(s_i, z_i) \in Q$ with $s_i < t$. This proves the stated result. $\square$

PROOF OF LEMMA 4. Take $\xi \in \mathbb{L}^{[0,1]}$, which is the transform of some $[0,1]$-dimensional spatiotemporal point process $Z \in \mathscr{Z}$. We aim to define the probabilistic analog $\hat{\Phi}$ in the process domain of the operator Φ in the transform domain, meaning that the diagram depicted in Figure 3 commutes. To this end, for $x \in [0,1]$, let K_x be an inhomogeneous Poisson

$$\begin{array}{ccc} Z & \xrightarrow{\hat{\Phi}} & \hat{\Phi}(Z) \\ \mathcal{L}\downarrow & & \downarrow\mathcal{L} \\ \xi & \xrightarrow{\Phi} & \Phi(\xi) \end{array}$$

Fig 3: Diagram showing the relation between Φ and $\hat{\Phi}$.

process as defined after the proof of Theorem 7. For $Z \in \mathscr{Z}$, $x \in [0,1]$ and $t \in \mathbb{R}_+$, we set

(90) $$(\hat{\Phi}(Z))_x(t) := \delta_x \cdot \mathbf{1}\{J_x > t\} + \sum_{i=1}^{K_x(t)} Z_{x_i}(t - T_i).$$

Then $\hat{\Phi}(Z) \in \mathscr{Z}$. By the same steps given in the proof of Theorem 8 below,

$$\mathcal{L}_{(\hat{\Phi}(Z))_x}(f,t) = \Phi_x(\xi)(f,t),$$

whence $\mathcal{L}_{\hat{\Phi}(Z)} = \Phi(\mathcal{L}_Z)$. As $\Phi(\xi)$ is the transform of $\hat{\Phi}(Z) \in \mathscr{Z}$, $\Phi(\xi) \in \mathbb{L}^{[0,1]}$. $\square$

PROOF OF THEOREM 8. Note that by the multiplicative structure of the excitation term $B_{zx}(0)W(z,x)h(t)$, it follows that conditional on the arrival of a particle in the time interval $[0,u]$, the spatial placement in $[0,1]$ is independent of the temporal placement in $[0,u]$. Let $P_t(s)$ be the probability that an offspring event, conditionally upon being born before time t, is born before time s. Let $H(u) := \int_0^u h(v) \, \mathrm{d}v$ be the integrated excitation function. As in [41], p. 12, it follows that $P_t(s) = H(s)/H(t)$, with density $p_t(s) = P_t'(s) = h(s)/H(t)$.

Let $z_i \in S_x(u)$ denote the set of spatial coordinates corresponding to particles $(s_i, z_i) \in S_x(u)$ with $s_i < u$. The distributional equality (58) for S_x implies that

$$\eta_x(f,u) = \mathbb{E}\left[\exp\left(- \sum_{z_i \in S_x(u)} f(z_i) \right) \right]$$

(91)

$$= \gamma_x(f,u) \mathbb{E}_B \left[\sum_{n \geq 0} \mathbb{E}\left[\exp\left(- \sum_{z_i \in S_x(u), z_i \neq x} f(z_i) \right) \middle| K_x(u) = n, B \right] \mathbb{P}(K_x(u) = n | B) \right],$$

where we exclude x from the first summation since, with probability 1, there is no event with spatial coordinate x, and since we factored out $\gamma_x(f,u) = \big(\bar{\mathscr{I}}_x(u) + \mathscr{I}_x(u)e^{-f(x)}\big)$. By also conditioning on the locations of the n children of the initial event in x, (91) equals

$$\gamma_x(f,u)\mathbb{E}_B\left[\sum_{n\geq 0}\mathbb{P}(K_x(u)=n|B)\int_{[0,1]^n}p^n(y_1,\dots,y_n)\right.$$
$$\left.\times\,\mathbb{E}\left[\exp\left(-\sum_{\substack{z_i\in S_x(u)\\ z_i\neq x}}f(z_i)\right)\middle|K_x(u)=n;\ \text{locations } y_1,\dots y_n;B\right]\,\mathrm{d}y_1\cdots\mathrm{d}y_n\right]$$
$$=\gamma_x(f,u)\mathbb{E}_B\left[\sum_{n\geq 0}e^{-\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}H(u)}\frac{\|B_{\cdot x}(0)W(\cdot,x)\|^n_{L^1[0,1]}H(u)^n}{n!}\right.$$
$$\left.\times\int_{[0,1]^n}\prod_{j=1}^n\eta^{(1)}_{y_j}(f,u)\frac{B_{y_jx}(0)W(y_j,x)}{\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}}\,\mathrm{d}y_1\cdots\mathrm{d}y_n\right],\tag{92}$$

where $\eta^{(1)}_y$ is the transform of a cluster in y that is born between time 0 and u; its input is u, which is the time that has elapsed since the event in x occurred, and is therefore different from η_y, which takes $u-v$ as input when the event in y is born at time v. The distribution of the temporal location of such an event is given by $P_t(s)$ defined in the first part of this proof. Hence, (92) equals

$$\gamma_x(f,u)\mathbb{E}_B\left[\sum_{n\geq 0}e^{-\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}H(u)}\frac{\|B_{\cdot x}(0)W(\cdot,x)\|^n_{L^1[0,1]}H(u)^n}{n!}\right.$$
$$\left.\times\prod_{j=1}^n\int_0^1\int_0^u p_u(s)\eta_{y_j}(f,u-s)\,\mathrm{d}s\frac{B_{y_jx}(0)W(y_j,x)}{\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}}\,\mathrm{d}y_j\right]$$
$$=\gamma_x(f,u)\mathbb{E}_B\left[e^{-\|B_{\cdot x}(0)W(\cdot,x)\|_{L^1[0,1]}H(u)}\right.$$
$$\left.\times\sum_{n\geq 0}\frac{1}{n!}\left(\int_0^1\int_0^u h(s)\eta_y(f,u-s)\,\mathrm{d}sB_{yx}(0)W(y,x)\,\mathrm{d}y\right)^n\right]$$
$$=\gamma_x(f,u)\mathbb{E}\left[\exp\left(\int_0^1\int_0^u(\eta_y(f,u-s)-1)B_{yx}(0)W(y,x)h(s)\,\mathrm{d}s\mathrm{d}y\right)\right]$$
$$=\gamma_x(f,u)\beta_x\left(\left\{y\mapsto\int_0^u(1-\eta_y(f,u-s))W(y,x)h(s)\,\mathrm{d}s\right\}\right),\tag{93}$$

recognizing the Laplace functional β_x of $B_{\cdot x}$. □

PROOF OF LEMMA 5. We present the proof for $p\in[1,\infty)$; the proof for $p=\infty$ follows along the same lines, and is no more difficult.

Take $\xi,\tilde{\xi}\in\mathbb{L}^{[0,1]}$, and fix $\epsilon>0$. We aim to find $\delta>0$ such that

$$d^p_{\mathbb{L}^{[0,1]}}(\xi,\tilde{\xi})<\delta\implies d^p_{\mathbb{L}^{[0,1]}}(\Phi(\xi),\Phi(\tilde{\xi}))<\epsilon,$$

so we even prove uniform continuity. By the definition of $d_{\mathbb{L}^{[0,1]},p}$, it holds that

$$d^p_{\mathbb{L}^{[0,1]},p}(\Phi(\xi),\Phi(\tilde{\xi})) = \sup_{\substack{u\in[0,t]\\ f\in \mathrm{BM}_+[0,1]}} \int_0^1 |\Phi_x(\xi)(f,u) - \Phi_x(\tilde{\xi})(f,u)|^p \,\mathrm{d}x. \tag{94}$$

We first bound the integrand. Since f is positive-valued, $\bar{\mathscr{J}}_x(u) + \mathscr{J}_x(u)e^{-f(x)} \in [0,1]$. Furthermore, ξ_y is a Laplace functional, hence $\xi_y(f,u-s)\in[0,1]$ as well. Therefore, by the mean value theorem applied to $x\mapsto e^x$,

$$\begin{aligned}
&|\Phi_x(\xi)(f,u) - \Phi_x(\tilde{\xi})(f,u)| \\
&\le \left|\mathbb{E}\left[\exp\left(\int_0^1\int_0^u (\xi_y(f,u-s)-1)B_{yx}(0)W(y,x)h(s)\,\mathrm{d}s\mathrm{d}y\right)\right.\right.\\
&\quad \left.\left.-\exp\left(\int_0^1\int_0^u (\tilde{\xi}_y(f,u-s)-1)B_{yx}(0)W(y,x)h(s)\,\mathrm{d}s\mathrm{d}y\right)\right]\right|\\
&\le \left|\mathbb{E}\left[\int_0^1\int_0^u (\xi_y(f,u-s)-\tilde{\xi}_y(f,u-s))B_{yx}(0)W(y,x)h(s)\,\mathrm{d}s\mathrm{d}y\right]\right|\\
&\le u\|h\|_\infty C_B C_W \int_0^1 |\xi_y(f,u)-\tilde{\xi}_y(f,u)|\,\mathrm{d}y,
\end{aligned}$$

using Assumption 2, stating that $\mathbb{E}[B_{yx}]\le C_B$ a.s., and $W\le C_W$. Since $p\ge 1$, the function $x\mapsto x^p$ is convex. Jensen's inequality gives

$$\int_0^1 |\Phi_x(\xi)(f,u)-\Phi_x(\tilde{\xi})(f,u)|^p\,\mathrm{d}x \le u^p\|h\|^p_\infty C_B^p C_W^p \int_0^1\int_0^1 |\xi_y(f,u)-\tilde{\xi}_y(f,u)|^p\,\mathrm{d}y\mathrm{d}x.$$

It now follows that

$$\sup_{\substack{u\in[0,t]\\ f\in \mathrm{BM}_+[0,1]}} \left(\int_0^1 |\Phi_x(\xi)(f,u)-\Phi_x(\tilde{\xi})(f,u)|^p\,\mathrm{d}x\right)^{1/p} \le t\|h\|_\infty C_B C_W \delta.$$

Hence, selecting $\delta < \epsilon/(t\|h\|_\infty C_B C_W)$ guarantees $d_{\mathbb{L}^{[0,1]},p}(\Phi(\xi),\Phi(\tilde{\xi}))<\epsilon$. □

PROOF OF LEMMA 6. We present the proof for $p\in[1,\infty)$, the proof for $p=\infty$ is no more difficult.

By the bounds appearing in the proof of Lemma 5, it holds that

$$\|\Phi_x(\xi)(f,u)-\Phi_x(\zeta)(f,u)\|_{L^p[0,1]} \le u\|h\|_\infty C_B C_W \|\xi_x(f,u)-\zeta_x(f,u)\|_{L^p[0,1]}.$$

Since the Laplace functional takes on values in $[0,1]$,

$$\left\|\xi_x^{(1)}(f,u)-\zeta_x^{(1)}(f,u)\right\|_{L^p[0,1]} \le 2u\|h\|_\infty C_B C_W,$$

verifying (61) for $n=1$ and $C := 2\|h\|_\infty C_B C_W$. Now suppose the result holds for $n\in\mathbb{N}$. Then again by the bounds appearing in the proof of Lemma 5, and using Jensen's inequality,

$$\begin{aligned}
\left|\xi_x^{(n+1)}(f,u)-\zeta_x^{(n+1)}(f,u)\right|^p &\le \|h\|^p_\infty C_B^p C_W^p \int_0^1\left(\int_0^u \left|\xi_y^{(n)}(f,s)-\zeta_y^{(n)}(f,s)\right|\,\mathrm{d}s\right)^p\mathrm{d}y\\
&\le C^p\int_0^1\left(\int_0^u \frac{C^n s^n}{n!}\,\mathrm{d}s\right)^p\mathrm{d}y = \left(\frac{C^{n+1}u^{n+1}}{(n+1)!}\right)^p.
\end{aligned}$$

Integrating over $x\in[0,1]$ proves the bound for $n+1$. The stated result follows by induction.
□

Acknowledgments. We are very grateful to the Editor and a referee for their careful reading and valuable comments, which have substantially improved the clarity and quality of the paper. We are also grateful to workshop participants in the program "Self-interacting processes" at the Isaac Newton Institute of Cambridge University for their comments and suggestions.

Funding. The research of JB and RL was funded in part by the Netherlands Organization for Scientific Research under an NWO VICI grant (2020–2027). The research of MM was funded in part by the NWO Gravitation project NETWORKS, grant number 024.002.003.

REFERENCES

[1] G. ACOSTA and R. G. DURÁN (2003). An optimal Poincaré inequality in L^1 for convex domains. *Proceedings of the American Mathematical Society* **132**, pp. 195–202.
[2] Z. AGATHE-NERINE (2022). Multivariate Hawkes processes on inhomogeneous random graphs. *Stochastic Processes and their Applications* **152**, pp. 86–148.
[3] Y. AÏT-SAHALIA, J. A. CACHO-DIAZ and R. J. A. LAEVEN (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* **117**, pp. 585–606.
[4] A. AURELL, R. CARMONA and M. LAURIÈRE (2022). Stochastic graphon games: II. The linear-quadratic case. *Applied Mathematics & Optimization* **85**, 39.
[5] E. BACRY, S. DELATTRE, M. HOFFMANN and J. F. MUZY (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications* **123**, pp. 2475–2499.
[6] E. BACRY and J. F. MUZY (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance* **14**, pp. 1147–1166.
[7] E. BAYRAKTAR, S. CHAKRABORTY and R. WU (2023). Graphon mean field systems. *The Annals of Applied Probability* **33**, pp. 3587–3619.
[8] G. BET, F. COPPINI and F. R. NARDI (2024). Weakly interacting oscillators on dense random graphs. *Journal of Applied Probability* **61**, pp. 255–278.
[9] H. BIERMÉ, O. DURIEU and Y. WANG (2018). Generalized random fields and Lévy's continuity theorem on the space of tempered distributions. *Communications on Stochastic Analysis* **26**, pp. 427–445.
[10] P. BILLINGSLEY (1999). *Convergence of Probability Measures*. John Wiley & Sons.
[11] H. P. BOSWIJK, R. J. A. LAEVEN and X. YANG (2018). Testing for self-excitation in jumps. *Journal of Econometrics* **203**, pp. 256–266.
[12] P. BRÉMAUD and L. MASSOULIÉ (1996). Stability of nonlinear Hawkes processes. *The Annals of Probability* **24**, pp. 1563–1588.
[13] P. E. CAINES and M. HUANG (2021). Graphon mean field games and their equations. *SIAM Journal on Control and Optimization* **59**, pp. 4373–4399.
[14] R. CARMONA, D. B. COONEY, C. V. GRAVES and M. LAURIÈRE (2022). Stochastic graphon games: I. The static case. *Mathematics of Operations Research* **47**, pp. 750–778.
[15] W. CHIANG, X. LIU and G. MOHLER (2022). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting* **38**, pp. 505–520.
[16] J. B. CONWAY (2007). *A Course in Functional Analysis*, Graduate Texts in Mathematics, Springer-Verlag, New York.
[17] L. R. CUI, A.G. HAWKES and H. YI (2020). An elementary derivation of moments of Hawkes processes. *Advances in Applied Probability* **52**, pp. 102–137.
[18] D. J. DALEY and D. VERE-JONES (2003). *An Introduction to the Theory of Point Processes*, Vol I and II, 2nd ed., Springer-Verlag, New York.
[19] A. DAW (2024). Conditional uniformity and Hawkes processes. *Mathematics of Operations Research* **49**, pp. 40–57.
[20] A. DAW and J. PENDER (2023). Matrix calculations for moments of Markov processes. *Advances in Applied Probability* **55**, pp. 126–150.
[21] S. DELATTRE, N. FOURNIER and M. HOFFMANN (2016). Hawkes processes on large networks. *The Annals of Applied Probability* **26**, pp. 216–261.
[22] N. DU, Y. WANG, L. SONG, H. ZHANG and L. MA (2013). Hawkes processes for clickstream data and the emergence of collective attention. *Proceedings of the 22nd international conference on World Wide Web*, pp. 609–620.

[23] S. EROL, F. PARISE and A. TEYTELBOYM (2023). Contagion in graphons. *Journal of Economic Theory* **211**, 105673.

[24] R. FIERRO, V. LEIVA and J. MOLLER (2015). The Hawkes process with different exciting functions and its asymptotic behavior. *Journal of Applied Probability* **52**, pp. 37–54.

[25] E. W. FOX, F. P. SCHOENBERG and J. S. GORDON (2016). Spatially inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. *The Annals of Applied Statistics* **10**, pp. 1725–1756.

[26] A. GALVES and E. LÖCHERBACH (2013). Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *Journal of Statistical Physics* **151**, pp. 896–921.

[27] F. GAO and L. ZHU (2023). Large deviations for the mean-field limit of Hawkes processes. Preprint. Available at https://arxiv.org/abs/2301.07585.

[28] G. GRIPENBERG, S. O. LONDEN and O. STAFFANS (1990). *Volterra Integral and Functional Equations*. Cambridge University Press, Cambridge.

[29] N. R. HANSEN, P. REYNAUD-BOURET and V. RIVOIRARD (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21**, pp. 83–143.

[30] A. G. HAWKES (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, pp. 83–90.

[31] A. G. HAWKES and D. OAKES (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability* **11**, pp. 493–503.

[32] U. HORST and W. XU (2021). Functional limit theorems for marked Hawkes point measures. *Stochastic Processes and their Applications* **134**, pp. 94–131.

[33] M. IKEFUJI, R. J. A. LAEVEN, J. R. MAGNUS and Y. YUE (2022). Earthquake risk embedded in property prices: Evidence from five Japanese cities. *Journal of the American Statistical Association* **117**, pp. 82–93.

[34] F. ILHAN and S. S. KOZAT (2020). Modeling of spatio-temporal Hawkes processes with randomized kernels. *IEEE Transactions on Signal Processing* **68**, pp. 4946–4958.

[35] T. JAISSON and M. ROSENBAUM (2015). Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability* **25**, pp. 600–631.

[36] T. JAISSON and M. ROSENBAUM (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability* **26**, pp. 2860–2882.

[37] D. KARABASH and L. ZHU (2015). Limit theorems for marked Hawkes processes with application to a risk model. *Stochastic Models* **31**, pp. 433–451.

[38] R. KARIM, R. J. A. LAEVEN and M. MANDJES (2021). Exact and asymptotic analysis of general multivariate Hawkes processes and induced population processes. Available at https://arxiv.org/abs/2106.03560.

[39] R. KARIM, R. J. A. LAEVEN and M. MANDJES (2025). Compound multivariate Hawkes processes: Large deviations and rare event simulation. *Bernoulli* **31**, pp. 3113–3138.

[40] M. KIRCHNER (2017). An estimation procedure for the Hawkes process. *Quantitative Finance* **17**, pp. 571–595.

[41] D. T. KOOPS, M. SAXENA, O. J. BOXMA and M. MANDJES (2018). Infinite-server queues with Hawkes input. *Journal of Applied Probability* **55**, pp. 920–943.

[42] D. KUNSZENTI-KOVÁCS, L. LOVÁSZ and B. SZEGEDY (2022). Multigraph limits, unbounded kernels, and Banach space decorated graphs. *Journal of Functional Analysis* **288**, pp. 1–44.

[43] J. KWON, Y. ZHENG and M. JUN (2023). Flexible spatio-temporal Hawkes process models for earthquake occurrences. *Spatial Statistics* **54**, pp. 1–23.

[44] J. LARGE (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* **10**, pp. 1–25.

[45] G. LEONI (2009). *A First Course in Sobolev Spaces*, 1st ed., Graduate Studies in Mathematics, Vol. 105, American Mathematical Society.

[46] L. LOVÁSZ and B. SZEGEDY (2010). Limits of compact decorated graphs. Available at http://arxiv.org/abs/1010.5155.

[47] L. LOVÁSZ (2012). *Large Networks and Graph Limits*, Colloquium Publications Vol. 60, American Mathematical Society.

[48] L. MASSOULIÉ (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications* **75**, pp. 1–30.

[49] T. MIKOSCH (2009). *Non-Life Insurance Mathematics. An Introduction with the Poisson Process*, 2nd ed., Springer Berlin, Heidelberg.

[50] N. NALDI and G. PATANÈ (2022). A graph-based modelling of epidemics: Properties, simulation, and continuum limit. Available at https://arxiv.org/abs/2208.07559.

[51] J. NEVEU (1965). *Mathematical Foundations of the Calculus of Probability*, 1st ed. Holden-Day.

[52] Y. OGATA (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes dynamics, and applications. *Stochastic Processes and their Applications* **75**, pp. 1–30.

[53] Y. OGATA (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* **83**, pp. 9–27.

[54] J. OLINDE and M. SHORT (2020). A self-limiting Hawkes process: Interpretation, estimation, and use in crime modeling. *2020 IEEE International Conference on Big Data*.

[55] G. PANG and É. PARDOUX (2023). Spatially dense stochastic epidemic models with infection-age dependent infectivity. Available at https://arxiv.org/abs/2304.05211.

[56] Y. V. PROHOROV (1961). The method of characteristic functionals. *Berkeley Symposium on Mathematical Statistics and Probability*, pp. 403–419.

[57] A. REINHART (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science* **33**, pp. 299–318.

[58] P. REYNAUD-BOURET, R. LAMBERT, C. TULEAU-MALOT, T. BESSAIH, V. RIVOIRARD, Y. BOURET and N. LERESCHE (2018). Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *Journal of Neuroscience Methods* **297**, pp. 9–21.

[59] M. RIZOIU, Y. LEE, S. MISHRA and L. XIE (2017). A tutorial on Hawkes processes for events in social media. In book: *Frontiers of Multimedia Research*, pp. 191–218.

[60] H. H. SCHAEFER (1974). *Banach Lattices and Positive Operators*, Grundlehren der mathematischen Wissenschaften, Springer Berlin, Heidelberg.

[61] R. L. SCHILLING (2017). *Measures, Integrals and Martingales*, Cambridge University Press, 2nd edition.

[62] H. XU, D. LUO and H. ZHA (2021). Hawkes processes on graphons. Available at https://arxiv.org/abs/2102.02741.

[63] L. ZHU (2013). Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability* **50**, pp. 760–771.

[64] L. ZHU (2013). Nonlinear Hawkes Processes. Ph.D. thesis, New York University. Available at https://arxiv.org/abs/1304.7531.