Electronic Journal of Probability

Electron. J. Probab. **28** (2023), article no. 145, 1–42.
ISSN: 1083-6489 https://doi.org/10.1214/23-EJP1006

Long-term stability of interacting Hawkes processes on random graphs

Zoé Agathe-Nerine*

Abstract

We consider a population of Hawkes processes modeling the activity of N interacting neurons. The neurons are regularly positioned on the segment $[0,1]$, and the connectivity between neurons is given by a random possibly diluted and inhomogeneous graph where the probability of presence of each edge depends on the spatial position of its vertices through a spatial kernel. The main result of the paper concerns the long-time stability of the synaptic current of the population, as $N \to \infty$, in the subcritical regime in case the synaptic memory kernel is exponential, up to time horizons that are polynomial in N.

Keywords: multivariate nonlinear Hawkes processes; mean-field systems; neural field equation; spatially extended system; W-random graph.
MSC2020 subject classifications: 60F15; 60G55; 44A35; 92B20.
Submitted to EJP on August 3, 2022, final version accepted on August 20, 2023.
Supersedes arXiv:2207.13942.
Supersedes HAL:hal-03739494.

1 Introduction

1.1 Hawkes processes in neuroscience

In the present paper we study the large time behavior of a population of interacting and spiking neurons, as the size of the population N tends to infinity. We model the activity of a neuron by a point process where each point represents the time of a spike: $Z_{N,i}(t)$ counts the number of spikes during the time interval $[0,t]$ of the ith neuron of the population. Its intensity at time t conditioned on the past $[0,t)$ is given by $\lambda_{N,i}(t)$, in the sense that

$$\mathbf{P}\left(Z_{N,i} \text{ jumps between} (t,t+dt) | \mathcal{F}_t\right) = \lambda_{N,i}(t)dt,$$

where $\mathcal{F}_t := \sigma\left(Z_{N,i}(s), s \leq t, 1 \leq i \leq N\right)$.

For the choice of $\lambda_{N,i}$, we want to account for the dependence of the activity of a neuron on the past of the whole population: the spike of one neuron can trigger others'

*Université Paris Cité, Laboratoire MAP5 (UMR CNRS 8145), 75270 Paris, France & FP2M, CNRS FR 2036. E-mail: zoe.agathe-nerine@u-paris.fr

spikes. *Hawkes processes* are then a natural choice to emphasize this interdependency. A generic choice is

$$\lambda_{N,i}(t) = \mu(t, x_i) + f\left(v(t, x_i) + \frac{1}{N}\sum_{j=1}^{N} w_{ij}^{(N)} \int_0^{t-} h(t-s)dZ_{N,j}(s)\right). \tag{1.1}$$

Here, with the ith neuron at position $x_i = \frac{i}{N} \in I := [0,1]$, $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ represents the (possible) non linear synaptic integration, $\mu(t,\cdot) : I \longrightarrow \mathbb{R}_+$ a spontaneous activity of the neuron at time t, $v(t,\cdot) : I \longrightarrow \mathbb{R}$ a past activity and $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ a memory function which models how a past jump of the system affects the present intensity. The term $w_{ij}^{(N)}$ represents the random inhomogeneous interaction between neurons i and j, that will be modeled here in terms of the realization of a random graph.

Since the seminal works of [31, 32], there has been a renewed interest in the use of Hawkes processes, especially in neuroscience. A common simplified framework is to consider an interaction on the complete graph, that is taking $w_{ij}^{(N)} = 1$ in (1.1), as done in [23]. In this case, a very simple instance of (1.1) concerns the so called *linear case*, when $f(x) = x$, $\mu(t,x) = \mu$ and $v = 0$, that is $\lambda_{N,i}(t) = \lambda_N(t) = \mu + \frac{1}{N}\sum_{j=1}^{N}\int_0^{t-} h(t-s)dZ_{N,j}(s)$, with $h \geq 0$ (see [23]). The biological evidence [11, 38] of a spatial organisation of neurons in the brain has led to more elaborate Hawkes models with spatial interaction, possibly including inhibition (see [45, 26, 15]). This would correspond in (1.1) to take $w_{ij}^{(N)} = W(x_i, x_j)$, where W is a macroscopic interaction kernel, usual examples being the exponential distribution on $\mathbb{R}$, $W(x,y) = \frac{1}{2\sigma}\exp\left(-\frac{|x-y|}{\sigma}\right)$ or the "Mexican hat" distribution $W(x,y) = e^{-|x-y|} - Ae^{\frac{-|x-y|}{\sigma}}$, $A \in \mathbb{R}$, $\sigma > 0$. The macroscopic limit of the multivariate Hawkes process (1.1) is then given by a family of spatially extended inhomogeneous Poisson processes whose intensities $(\lambda_t(x))_{x\in I}$ solve the convolution equation

$$\lambda_t(x) = \mu_t(x) + f\left(v_t(x) + \int_I W(x,y)\int_0^t h(t-s)\lambda_s(y)dsdy\right). \tag{1.2}$$

A crucial example is the exponential case, that is when $h(t) = e^{-\alpha t}$ for some $\alpha > 0$. In this case, the Hawkes process with intensity (1.1) is Markovian (see [26]). Denoting in (1.2) $u_t(x) := v_t(x) + \int_I W(x,y)\int_0^t h(t-s)\lambda_s(y)dsdy$ as the potential of a neuron (the synaptic current) localised in x at time t (so that (1.2) becomes $\lambda_t(x) = f(u_t(x))$), an easy computation (see [15]) gives that, when $v_t(x) = e^{-\alpha t}v_0(x)$ for some v_0, u solves the *Neural Field Equation* (NFE)

$$\frac{\partial u_t(x)}{\partial t} = -\alpha u_t(x) + \int_I W(x,y)f(u_t(y))dy + I_t(x), \tag{1.3}$$

with source term $I_t(x) := \int_I W(x,y)\mu_t(y)dy$. Equation (1.3) has been extensively studied in the literature, mostly from a phenomenological perspective [46, 2], and is an important example of macroscopic neural dynamics with non-local interactions (we refer to [13] for an extensive review on the subject).

In a previous work [1], we give a microscopic interpretation of the macroscopic kernel W in terms of an inhomogeneous graph of interaction. We consider $w_{ij}^{(N)} = \xi_{ij}^{(N)}\kappa_i$ in (1.1), where $\left(\xi_{ij}^{(N)}\right)_{1\leq i,j\leq N}$ is a collection of independent Bernoulli variables, with individual parameter $W(x_i, x_j)$: the probability that two neurons are connected depends on their spatial positions. The term κ_i is a suitable local renormalisation parameter, to ensure that the interaction remains of order 1. This modeling constitutes a further difficulty in the analysis as we are no longer in a mean-field framework: contrary to the case $w_{ij}^{(N)} = 1$, the interaction (1.1) is no longer a functional of the empirical measure of the particles $(Z_{N,1},\dots,Z_{N,N})$. A recent interest has been shown to similar issues in the case of diffusions interacting on random graphs (first in the homogeneous Erdős-Rényi case [24, 21, 22, 20], and secondly for inhomogenous random graph [36, 3, 5]). See also [42] where the interaction is random (either excitatory or inhibatory) on the complete graph with a diffusive scaling in $1/\sqrt{N}$ when the excitation and inhibition are balanced.

A common motivation between [1] in the case of Hawkes processes and [36, 3, 5] in the case of diffusions is to understand how the inhomogeneity of the underlying graph may or may not influence the long time dynamics of the system. An issue common to all mean-field models (and their perturbations) is that there is, in general, no possibility to interchange the limits $N\to\infty$ and $t\to\infty$. More precisely, restricting to Hawkes processes, a usual propagation of chaos result (see [23, Theorem 8], [15, Theorem 1], [1, Theorem 3.10]) may be stated as follows: for fixed $T>0$, there exists some $C(T)>0$ such that

$$\sup_{1\leq i\leq N}\mathbf{E}\left(\sup_{s\in[0,T]}\left|Z_{N,i}(s)-\overline{Z}_i(s)\right|\right)\leq\frac{C(T)}{\sqrt{N}}, \tag{1.4}$$

where $\overline{Z}_i$ is a Poisson process with intensity $(\lambda_t(x_i))_{t\geq 0}$ defined in (1.2) suitably coupled to $Z_{N,i}$, see the above references for details. Generically, $C(T)$ is of the form $\exp(CT)$, such that (1.4) remains only relevant up to $T\sim c\log N$ with c sufficiently small. In the pure mean-field linear case ($w_{ij}^{(N)}=1$, $f(x)=x$), there is a well known phase transition [23, Theorems 10, 11] when $\|h\|_1=\int_0^\infty h(t)dt<1$ (*subcritical case*), $\lambda_t\xrightarrow[t\to\infty]{}\frac{\mu}{1-\|h\|_1}$, whereas when $\|h\|_1>1$ (*supercritical case*), $\lambda_t\xrightarrow[t\to\infty]{}\infty$. This phase transition has been extended to the inhomogeneous case in [1]. In the subcritical case, one can actually improve (1.4) in the sense that $C(T)$ is now linear in T so that (1.4) remains relevant up to $T=o(\sqrt{N})$. A natural question is to ask if this approximation remains valid beyond this time scale. The purpose to the present work is to address this question: we show that, in the whole generality of (1.1), in the subcritical regime and exponential case (see details below), the macroscopic intensity (1.2) converges to a finite limit when $t\to\infty$ and that the microscopic system remains close to this limit up to polynomial times in N.

1.2 Notation

We denote by $C_{\text{parameters}}$ a constant $C>0$ which only depends on the parameters inside the lower index. These constants can change from line to line or inside a same equation, we choose just to highlight the dependency they contain. When it is not relevant, we just write C. For any $d\geq 1$, we denote by $|x|$ and $x\cdot y$ the Euclidean norm and scalar product of elements $x,y\in\mathbb{R}^d$. For $(E,\mathcal{A},\mu)$ a measured space, for a function g in $L^p(E,\mu)$ with $p\geq 1$, we write $\|g\|_{E,\mu,p}:=\left(\int_E|g|^p d\mu\right)^{\frac{1}{p}}$. When $p=2$, we denote by $\langle\cdot,\cdot\rangle$ the Hermitian scalar product in $L^2(E)$. Without ambiguity, we may omit the subscript (E,μ) or μ. For a real-valued bounded function g on a space E, we write $\|g\|_\infty:=\|g\|_{E,\infty}=\sup_{x\in E}|g(x)|$.

For (E,d) a metric space, we denote by $\|g\|_L=\sup_{x\neq y}|g(x)-g(y)|/d(x,y)$ the Lipschitz seminorm of a real-valued function g on E. We denote by $\mathcal{C}(E,\mathbb{R})$ the space of continuous functions from E to $\mathbb{R}$, and $\mathcal{C}_b(E,\mathbb{R})$ the space of continuous bounded ones. For any $T>0$, we denote by $\mathbb{D}\left([0,T],E\right)$ the space of càdlàg (right continuous with left limits) functions defined on $[0,T]$ and taking values in E. For any integer $N\geq 1$, we denote by $[\![1,N]\!]$ the set $\{1,\dots,N\}$. For any $p\in[0,1]$, $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter p.

1.3 The model

First, let us focus on the interaction between the particles. The graph of interaction for (1.1) is constructed as follows:

Definition 1.1. *On a common probability space* $\left(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{P}\right)$*, we consider a family of random variables* $\xi^{(N)} = \left(\xi_{ij}^{(N)}\right)_{N\geq 1, i,j\in [\![1,N]\!]}$ *on* $\widetilde{\Omega}$ *such that under* $\mathbb{P}$*, for any* $N \geq 1$ *and* $i, j \in [\![1, N]\!]$*,* $\xi^{(N)}$ *is a collection of mutually independent Bernoulli random variables such that for* $1 \leq i, j \leq N$*,* $\xi_{ij}^{(N)}$ *has parameter* $W_N(\frac{i}{N}, \frac{j}{N})$*, where*

$$W_N(x, y) := \rho_N W(x, y), \tag{1.5}$$

with ρ_N *some dilution parameter and* $W : I^2 \to [0, 1]$ *a macroscopic interaction kernel. We assume that the particles in* (1.1) *are connected according to the oriented graph* $\mathcal{G}_N = \left(\{1, \dots, N\}, \xi^{(N)}\right)$*. For any* i *and* j*,* $\xi_{ij}^{(N)} = 1$ *encodes for the presence of the edge* $j \to i$ *and* $\xi_{ij}^{(N)} = 0$ *for its absence. The interaction in* (1.1) *is fixed as*

$$w_{ij}^{(N)} = \frac{\xi_{ij}^{(N)}}{\rho_N}, \tag{1.6}$$

so that the interaction term remains of order 1 as $N \to \infty$*.*

The class (1.5) of inhomogenous graphs falls into the framework of W-random graphs, see [35, 9, 10]. One distinguishes the **dense case** when $\lim_{N\to\infty} \rho_N = \rho > 0$ and the **diluted case** when $\rho_N \to 0$.

We fix these sequences, and work on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P}\right)$ rich enough for all the following processes can be defined. We denote by $\mathbf{E}$ the expectation under $\mathbf{P}$ and $\mathbb{E}$ the expectation w.r.t. $\mathbb{P}$. In the following definitions, N is fixed and the particles are regularly located on the segment $I = [0, 1]$. We denote by $x_i = \frac{i}{N}$ the position of the i-th neuron in the population of size N. We also divide I in N segments of equal length, denoted by

$$B_{N,i} := \left(\frac{i-1}{N}, \frac{i}{N}\right). \tag{1.7}$$

We can now formally define our process of interest.

Definition 1.2. *Let* $(\pi_i(ds, dz))_{1\leq i\leq N}$ *be a sequence of* $(\mathcal{F}_t)$*-adapted i.i.d. Poisson random measures on* $\mathbb{R}_+ \times \mathbb{R}_+$ *with intensity measure* $dsdz$*. The multivariate counting process* $(Z_{N,1}(t), ..., Z_{N,N}(t))_{t\geq 0}$ *defined by, for all* $t \geq 0$ *and* $i \in [\![1, N]\!]$*:*

$$Z_{N,i}(t) = \int_0^t \int_0^\infty \mathbf{1}_{\{z\leq \lambda_{N,i}(s)\}} \pi_i(ds, dz) \tag{1.8}$$

where

$$\lambda_{N,i}(t) = F(X_{N,i}(t-), \eta_t(x_i)), \tag{1.9}$$

and

$$X_{N,i}(t) = \sum_{j=1}^N \frac{w_{ij}^{(N)}}{N} \int_0^t h(t-s) dZ_{N,j}(s), \tag{1.10}$$

$\eta_t : I \longrightarrow \mathbb{R}^d$ *for any* $t \in [0, +\infty)$ *for some* $d \geq 1$ *and* $F : \mathbb{R} \times \mathbb{R}^d \longrightarrow \mathbb{R}^+$ *is called* a multivariate Hawkes process *with the set of parameters* $\left(N, F, \xi^{(N)}, W_N, \eta, h\right)$*.*

Our main focus is to study the quantity $(X_{N,i})_{1\leq i\leq N}$ defined in (1.10) as $N \to \infty$, and more precisely the random profile defined for all $x \in I$ by:

$$X_N(t)(x) := \sum_{i=1}^N X_{N,i}(t) \mathbf{1}_{x\in B_{N,i}}, \tag{1.11}$$

where $B_{N,i}$ is defined in (1.7).

As $N \to \infty$, an informal Law of Large Numbers (LLN) argument shows that the empirical mean in (1.10) becomes an expectation w.r.t. the candidate limit for $Z_{N,i}$: we can replace the sum in (1.10) by the integral, the microscopic interaction term $w_{ij}^{(N)}$ in (1.10) by the macroscopic term $W(x,y)$ (where y describes the macroscopic distribution of the positions), and the past activity of the neuron $dZ_{N,j}(s)$ by its intensity in large population. In other words, the macroscopic spatial profile will be described by

$$X_t(x) = \int_I W(x,y) \int_0^t h(t-s)\lambda_s(y)ds\ dy, \tag{1.12}$$

where the macroscopic intensity of a neuron at position $x \in I$ denoted by $\lambda_t(x) = F(X_t(x), \eta_t(x))$ solves

$$\lambda_t(x) = F\left(\int_I W(x,y) \int_0^t h(t-s)\lambda_s(y)dsdy, \eta_t(x)\right). \tag{1.13}$$

Such informal law of large number on a bounded time interval has been made rigorous under various settings, we refer for further references to [23, 15] and more especially to [1] which exactly incorporates the present hypotheses.

Remark 1.3. In the expression (1.9) of the intensity $\lambda_{N,i}$, $X_{N,i}$ given in (1.10) accounts for the stochastic influence of the other interacting neurons, whereas η_t represents the deterministic part of the intensity $\lambda_{N,i}$. Having in mind the generic example given in (1.1), a typical choice would correspond to taking $d = 2$ with $\eta := (\mu, v)$ and

$$F(X,\eta) = F(X,\mu,v) = \mu + f(v+X) \tag{1.14}$$

Once again, μ here corresponds to the spontaneous Poisson activity of the neuron and one may see v as a deterministic part in the evolution of the membrane potential of neuron i. Note that we generalize here slightly the framework considered in [15] in the sense that [15] considered (1.14) for $\mu \equiv 0$ and $v_t(x) = e^{-\alpha t}v_0(x)$ for some initial membrane potential $v_0(x)$. In the case of (1.14), one retrieves the expression of the macroscopic intensity $\lambda_t(x)$ given in (1.2). Typical choices of f in (1.14) are $f(x) = x$ (the so-called linear model) or some sigmoïd function. Note that there will be an intrinsic mathematical difficulty in dealing with the linear case in this paper, as f is not bounded in this case. As already mentioned in the introduction, for the choice of $h(t) = e^{-\alpha t}$ and $v_t(x) = e^{-\alpha t}v_0(x)$, a straightforward calculation shows that $u_t(x) := v_t(x) + X_t(x)$ solves the scalar neural field equation (1.3) with source term $I_t(x) = \int_I W(x,y)\mu(t,y)dy$.

We choose here to work with the generic expression (1.9) instead of (1.1) not only for conciseness of notation, but also to emphasize that the result does not intrinsically depend on the specific form of the function F.

Remark 1.4. We have assumed for simplicity in the current definition (1.10) of $X_{N,i}(t)$ that $X_{N,i}(0) = 0$. Define more generally, for any $(\vartheta_i)_{i=1,\dots,N}$, $\vartheta_i \in \mathbb{R}$, $X_{N,i}^{\vartheta_i}(t) = \vartheta_i + \sum_{j=1}^N \frac{w_{i,j}^{(N)}}{N} \int_0^t h(t-s)\mathrm{d}Z_{N,j}(s)$ the same process starting at ϑ_i (here, ϑ_i accounts for the history of the process before $t = 0$). Write then the corresponding intensity (1.9) and process (1.8) as $\lambda_{N,i}^{\vartheta_i}(t)$ and $Z_{N,i}^{\vartheta_i}(t)$ respectively. In particular, when h is exponential (see (2.4)), the process $(X_{N,i}^{\vartheta_i})_{i=1,\dots,N}$ is Markovian (see e.g. [26, Section 5]). The analysis of the profile $X_N^{\vartheta}(t) := \sum_{i=1}^N X_{N,i}^{\vartheta_i}(t)\mathbf{1}_{B_{N,i}}$ remains the same, under the additional hypothesis that $\sum_{i=1}^N \vartheta_i \mathbf{1}_{B_{N,i}} \xrightarrow[N\to\infty]{} \vartheta$ in $L^2(I)$. In the following, the actual dependence in the initial condition ϑ will be dropped, whenever it is clear from the context, for simplicity of notation.

2 Hypotheses and main results

2.1 Hypotheses

Hypothesis 2.1. *We assume that*

- F *is Lipschitz continuous: there exists* $\|F\|_L$ *such that for any* $x, x' \in \mathbb{R}$, $\eta, \eta' \in \mathbb{R}^d$, *we have* $|F(x,\eta) - F(x',\eta')| \leq \|F\|_L (|x - x'| + |\eta - \eta'|)$.
- F *is non decreasing in the first variable, that is for any* $\eta \in \mathbb{R}^d$, *for any* $x, x' \in \mathbb{R}$ *such that* $x \leq x'$, *one has* $F(x,\eta) \leq F(x',\eta)$. *Moreover, we assume that* F *is* $\mathcal{C}^2$ *on* $\mathbb{R}^{d+1}$ *with bounded derivatives. We denote by* $\partial_X F$ *and* $\partial_X^2 F$ *the partial derivatives of* F *w.r.t.* X *and (with some slight abuse of notation)* $\partial_\eta F = (\partial_{\eta_k} F)_{k=1,\ldots d}$ *as the gradient of* F *w.r.t. the variable* $\eta \in \mathbb{R}^d$ *as well as* $\partial^2_{X,\eta} F = \left(\partial^2_{X,\eta_k} F\right)_{k=1,\ldots d}$ *and* $\partial^2_\eta F = \left(\partial^2_{\eta_k,\eta_l} F\right)_{k,l=1,\ldots d}$ *the Hessian of* F *w.r.t. the variable* η.
- $(\eta_t(x))_{t\geq 0, x\in I}$ *is uniformly bounded in* (t,x). *We also assume that there exists* η_∞ *Lipschitz continuous on* I *such that*

$$\delta_t := \sup_{x\in I} |\eta_t(x) - \eta_\infty(x)| \xrightarrow[t\to\infty]{} 0. \tag{2.1}$$

- *The memory kernel* h *is nonnegative and integrable on* $[0, +\infty)$.
- *We assume that* $W : I^2 \to [0,1]$ *is continuous. We refer nonetheless to Section 2.3.4 where we show that the results of the paper remain true under weaker hypotheses on* W.

It has been showed in [1] that the process defined in (1.8) is well-posed, and that the large population limit intensity (1.13) is well defined in the following sense.

Proposition 2.2. *Under Hypothesis 2.1, for a fixed realisation of the family* $(\pi_i)_{1\leq i\leq N}$, *there exists a pathwise unique multivariate Hawkes process (in the sense of Definition 1.2) such that for any* $T < \infty$, $\sup_{t\in[0,T]} \sup_{1\leq i\leq N} \mathbf{E}[Z_{N,i}(t)] < \infty$.

Proposition 2.3. *Let* $T > 0$. *Under Hypothesis 2.1, there exists a unique solution* λ *in* $\mathcal{C}_b([0,T] \times I, \mathbb{R})$ *to* (1.13) *and this solution is nonnegative.*

Both Propositions 2.2 and 2.3 can be found in [1] as Propositions 2.5 and 2.7 respectively, where F is chosen as $\eta = (\mu, v)$ and $F(x,\eta) = f(x + v)$ with f a Lipschitz function. The same proofs work for our general case F. Proposition 2.3 also implies that the limiting spatial profile X_t solving (1.12) is well defined.

Before writing our next hypothesis, we need to introduce the following integral operator.

Proposition 2.4. *Under Hypothesis 2.1, the integral operator*

$$\begin{aligned} T_W : H &\longrightarrow H \\ g &\longmapsto \left(T_W g : x \longmapsto \int_I W(x,y) g(y) dy \right) \end{aligned}$$

is continuous in both cases $H = L^\infty(I)$ *and* $H = L^2(I)$. *When* $H = L^2(I)$, T_W *is compact, its spectrum is the union of* $\{0\}$ *and a discrete sequence of eigenvalues* $(\mu_n)_{n\geq 1}$ *such that* $\mu_n \to 0$ *as* $n \to \infty$. *Denote by* $r_\infty = r_\infty(T_W)$, *respectively* $r_2 = r_2(T_W)$ *the spectral radii of* T_W *in* $L^\infty(I)$ *and* $L^2(I)$ *respectively. Moreover, we have that*

$$r_2(T_W) = r_\infty(T_W). \tag{2.2}$$

The proof can be found in Section 3.1.

Hypothesis 2.5. *In the whole article, we are in the subcritical case defined by*

$$\|\partial_X F\|_\infty \|h\|_1 r_\infty < 1. \tag{2.3}$$

Note that in the complete mean-field case, $W \equiv 1$ and $r_\infty = 1$ so that one retrieves the usual subcritical condition as in [23]. In the linear case $\eta = \mu$ and $F(x, \eta) = \mu + x$, (2.3) is exactly the subcritical condition stated in [1].

The aim of the paper is twofold: firstly, we state a general convergence result as $t \to \infty$ of X_t defined in (1.12) (or equivalently λ_t in (1.13)), see Theorem 2.8. This result is valid for any general kernel h satisfying Hypothesis 2.1. Secondly, we address the long-term stability of the microscopic profile X_N defined in (1.11), see Theorem 2.13. Contrary to the first one, this second result is stated for the particular choice of the exponential kernel h defined as

$$h(t) = e^{-\alpha t}, \text{with } \alpha > 0. \tag{2.4}$$

The parameter $\alpha > 0$ is often called the leakage rate. The main advantage of this choice is that the process X_N then becomes Markovian (see Remark 1.4). This will turn out to be particularly helpful for the proof of Theorem 2.13. As already mentioned in the introduction, (2.4) is the natural framework where to observe the NFE (1.3) as a macroscopic limit, recall Remark 1.3. Note that in the exponential case (2.4), we have that $\|h\|_1 = 1/\alpha$ hence the subcritical case (2.3) reads

$$\|\partial_X F\|_\infty r_\infty < \alpha. \tag{2.5}$$

For our second result (Theorem 2.13), we also need some hypotheses on the dilution of the graph. Recall the definition of ρ_N in Definition 1.1.

Hypothesis 2.6. *The dilution parameter $\rho_N \in [0, 1]$ satisfies the following dilution condition: there exists $\tau \in (0, \frac{1}{2})$ such that*

$$N^{1-2\tau} \rho_N^4 \xrightarrow[N\to\infty]{} \infty. \tag{2.6}$$

If one supposes further that F is bounded, we assume the weaker condition

$$N\rho_N^2 \xrightarrow[N\to\infty]{} \infty. \tag{2.7}$$

Remark 2.7. Hypothesis 2.6 is stronger than $\frac{N\rho_N}{\log N} \xrightarrow[N\to\infty]{} \infty$, which is a dilution condition commonly met in the literature concerning LLN results on bounded time intervals for interacting particles on random graphs: it is the same as in [24, 21] (and slightly stronger than the optimal $N\rho_N \to +\infty$ obtained in [22] in the case of diffusions and as in [1] in the case of Hawkes processes).

2.2 Main results

Our first result, Theorem 2.8, studies the limit as $t \to \infty$ of the macroscopic profile X_t (as an element of $\mathcal{C}(I)$) defined in (1.12). Our second result, Theorem 2.13, focuses on the large time behaviour of $X_N(t)$ defined in (1.11) on any time interval of polynomial length.

2.2.1 Asymptotic behavior of (X_t)

Recall the definition of X_t in (1.12).

Theorem 2.8. *Under Hypotheses 2.1 and 2.5,* *(i) there exists a unique continuous function* $X_\infty : I \mapsto \mathbb{R}^+$ *solution of*

$$X_\infty = \|h\|_1 T_W F\left(X_\infty, \eta_\infty\right). \tag{2.8}$$

(ii) X_t *converges uniformly on* I *when* $t \to \infty$ *towards* X_∞*.*

The proof can be found in Section 3.2.

Remark 2.9. Translating the result of Theorem 2.8 in terms of the macroscopic intensity λ_t defined in (1.13) gives immediately that λ_t converges uniformly to ℓ solution to

$$\ell = F\left(\|h\|_1 T_W \ell, \eta_\infty\right). \tag{2.9}$$

The correspondence between X_∞ and ℓ (recall (1.12)) is simply given by $X_\infty = \|h\|_1 T_W \ell$.

Remark 2.10. In the particular case of an exponential memory kernel (2.4), as a straightforward consequence of the expression of X_t in (1.12) and X_∞ in (2.8), we have the following differential equation

$$\partial_t \left(X_t - X_\infty\right) = -\alpha \left(X_t - X_\infty\right) + T_W \left(F(X_t, \eta_t) - F(X_\infty, \eta_\infty)\right). \tag{2.10}$$

A simple Taylor expansion of X_t around X_∞ shows that the linearised system associated to the nonlinear (2.10) is then

$$\partial_t Y_t = -\alpha Y_t + T_W \left(G Y_t\right), \tag{2.11}$$

where

$$G := \partial_X F(X_\infty, \eta_\infty). \tag{2.12}$$

The subcritical condition (2.5) translates into the existence of a spectral gap for the linear dynamics (2.11), which makes the stationary point X_∞ linearly stable. More precisely,

Proposition 2.11. *Assume that the memory kernel* h *is exponential* (2.5)*. Define the linear operator*

$$\begin{aligned} \mathcal{L} : L^2(I) &\longrightarrow L^2(I) \\ g &\longmapsto \mathcal{L}(g) = -\alpha g + T_W(Gg). \end{aligned} \tag{2.13}$$

Then under Hypotheses 2.1 and 2.5, $\mathcal{L}$ *generates a contraction semi-group on* $L^2(I)$ $\left(e^{t\mathcal{L}}\right)_{t\geq 0}$ *such that for any* $g \in L^2(I)$

$$\|e^{t\mathcal{L}} g\|_2 \leq e^{-t\gamma} \|g\|_2, \tag{2.14}$$

where

$$\gamma := \alpha - r_\infty \left\|\partial_X F\right\|_\infty > 0. \tag{2.15}$$

The proof can be found in Section 3.1.

2.2.2 Long-term stability of the microscopic spatial profile

From now on, we place ourselves in the exponential case (2.4). We first state a convergence result of X_N towards the macroscopic X on a bounded time interval $[0, T]$.

Proposition 2.12. *Let* $T > 0$*. Under Hypotheses 2.1, 2.5 and 2.6,* $\mathbb{P}$*-a.s. for any* $\varepsilon > 0$*,*

$$\mathbf{P}\left(\sup_{t\in[0,T]} \|X_N(t) - X_t\|_2 \geq \varepsilon\right) \xrightarrow[N\to\infty]{} 0. \tag{2.16}$$

The proof can be found in Section 7. Note that Proposition 2.12 slightly generalises [1, Prop. 3.17] (see also [15, Cor. 2] for a similar result) where it is proven that $\mathbf{E}\left[\int_0^T \int_I |X_N(t)(x) - X_t(x)|\, dx\, dt\right] \xrightarrow[N\to\infty]{} 0$ for any $T > 0$. Here, we are more precise as we show uniform convergence of $X_N(t)$ in $L^2(I)$ instead of $L^1(I)$.

We are now in position to state the main result of the paper: the proximity stated in Proposition 2.12 is not only valid on a bounded time interval, but propagates to arbitrary polynomial times in $N\rho_N$.

Theorem 2.13. *Choose some $t_f > 0$ and $m \geq 1$. Then, under Hypotheses 2.1, 2.6 and 2.5, $\mathbb{P}$-a.s. for any $\varepsilon > 0$,*

$$\mathbf{P}\left(\sup_{t\in[t_\varepsilon,(N\rho_N)^m t_f]} \|X_N(t) - X_\infty\|_2 \geq \varepsilon\right) \xrightarrow[N\to\infty]{} 0, \tag{2.17}$$

for some $t_\varepsilon > 0$ independent on N.

The proof can be found in Section 4.

Remark 2.14. The variable t_ε in Theorem (2.13) represents essentially the time for the deterministic dynamics X_t to reach a neighborhood of X_∞ of size ε. The time t_ε is of order $-\log\varepsilon/\gamma$ (where γ is the spectral gap (2.15) given by the mean-field dynamics) and diverges as $\varepsilon \to 0$. Using the fact that on any finite $[0, T]$ (and in particular on $[0, t_\varepsilon]$ for any fixed ε), $X_{N,t}$ converges as $N \to \infty$ to X_t (Proposition 2.12), a simple triangle inequality gives that the following statement is also true: under the Hypotheses of Theorem 2.13, $\mathbb{P}$-a.s., for all $\varepsilon > 0$,

$$\mathbf{P}\left(\sup_{t\in[0,(N\rho_N)^m t_f]} \|X_N(t) - X_t\|_2 \geq \varepsilon\right) \xrightarrow[N\to\infty]{} 0.$$

Since F is Lipschitz and $\lambda_{N,i}(t) = F(X_{N,i}(t-), \eta_t(x_i))$ by (1.9), it is straightforward to derive from Theorem 2.13 a similar result for the profile of intensities

$$\lambda_N(t)(x) := \sum_{i=1}^N \lambda_{N,i}(t)\mathbf{1}_{x\in B_{N,i}},\ x \in I, \tag{2.18}$$

where $B_{N,i}$ is defined in (1.7).

Corollary 2.15. *Recall the definition of ℓ in* (2.9). *Under the same set of hypotheses of Theorem 2.13 and with the same notation,*

$$\mathbf{P}\left(\sup_{t\in[t_\varepsilon,(N\rho_N)^m t_f]} \|\lambda_N(t) - \ell\|_2 \geq \varepsilon\right) \xrightarrow[N\to\infty]{} 0. \tag{2.19}$$

2.3 Examples and extensions

We give here some illustrating examples of our main results.

2.3.1 Mean-field framework

To the best of the knowledge of the author, already in the simple homogeneous case of mean-field interaction, there exists no long-term stability result such as Theorem 2.13. We stress that our result may have an interest of its own in this case. Let us be more specific. When $\rho_N = W_N = 1$ and $\mu_t(x) = \mu \geq 0$, the process introduced in Definition 1.2 reduces to the usual mean-field framework [23]:

$$Z_{N,i}(t) = \int_0^t \int_0^\infty \mathbf{1}_{\{z\leq\lambda_N(s)\}}\pi_i(ds, dz) \tag{2.20}$$

with $\lambda_N(t)$ defined by

$$\lambda_N(t) = F(X_N(t-), \eta), \tag{2.21}$$

where

$$X_N(t) = \sum_{j=1}^{N} \frac{1}{N} \int_0^t h(t-s) dZ_{N,j}(s). \tag{2.22}$$

In this simple case, the spatial framework is no longer useful (in particular the spatial profile defined in (1.11) is constant in x so that the L^2 framework is not relevant, one has only to work in $\mathbb{R}$). The macroscopic intensity and synaptic current (respectively (1.13) and (1.12)) become

$$X_t := \int_0^t h(t-s)\lambda_s ds, \quad \lambda_t := F(X_t, \eta). \tag{2.23}$$

The main results of the paper translate then into

Theorem 2.16. *Under Hypothesis 2.1 and when $\|\partial_X F\|_\infty \|h\|_1 < 1$, there exists a unique $X_\infty \in \mathbb{R}_+$ solution to $X_\infty = \|h\|_1 F(X_\infty, \eta)$, and $(X_t)_{t\geq 0}$ converges when $t \to \infty$ towards X_∞. Respectively, $(\lambda_t)_{t\geq 0}$ converges towards ℓ, the unique solution to $\ell = F(\|h\|_1 \ell, \eta)$. Moreover, under the same hypotheses, in the exponential case* (2.4)*, for any $t_f > 0$ and $m \geq 1$, $\mathbb{P}$-a.s. for any $\varepsilon > 0$, $\mathbf{P}\left(\sup_{t\in[t_\varepsilon, N^m t_f]} |X_N(t) - X_\infty| \geq \varepsilon\right)$ and $\mathbf{P}\left(\sup_{t\in[t_\varepsilon, N^m t_f]} |\lambda_N(t) - \ell| \geq \varepsilon\right)$ tend to 0 as $N \to \infty$ for some $t_\varepsilon > 0$ independent on N.*

Remark 2.17. The previous result applies in particular to the linear case where $\eta = \mu$ and $F(x, \eta) = \mu + x$. We have then that $\ell = \frac{\mu}{1-\|h\|_1}$ in this case, as in [23].

2.3.2 Erdős-Rényi graphs

An immediate extension of the last mean-field case concerns the case of homogeneous Erdős-Rényi graphs: choose $W_N(x, y) = \rho_N$ for all $x, y \in I$. The results of our paper are valid under the dilution Hypothesis 2.6. It is however likely that these dilution conditions are not optimal (compare with the result of [20] with the condition $N\rho_N \to \infty$ in the diffusion case, but a difficulty here is that we deal with a multiplicative noise whereas it is essentially additive in [20]).

2.3.3 Examples in the inhomogeneous case

As already mentionned in Hypothesis 2.1, the results are valid for any W continuous, interesting examples include $W(x, y) = 1 - \max(x, y)$, $W(x, y) = 1 - xy$, see [7, 8]. Note also that we do not suppose any symmetry on W. Another rich class of examples concerns the *Expected Degree Distribution* model [18, 40] where $W(x, y) = f(x)g(y)$ for any continuous functions f and g on I. The specificity of such class is that we have an explicit formulation of r_∞, that is $r_\infty = \int_I f(x)g(x)dx$ when $\int_I g = 1$. In the linear case, we obtain an explicit formula for λ_t in [1, Example 4.3].

2.3.4 Extensions (weaker hypothesis on W)

It is apparent from the proofs below that one can weaken the hypothesis of continuity of W. Under the hypothesis that W is bounded, Proposition 2.3 remains true when $\mathcal{C}_b([0,T] \times I)$ is replaced by $\mathcal{C}([0,T], L^\infty(I))$ (continuity of λ_t and X_t in x may not be satisfied). Supposing further that there exists a partition of I into p intervals $I = \sqcup_{k=1,\ldots,p} C_k$ such that for all $\epsilon > 0$, there exists $\eta > 0$ such that $\int_I |W(x, y) - W(x', y)| dy <$

ϵ when $|x - x'| < \eta$ and $x, x' \in C_k$, then for every k, $\lambda_{|[0,T]\times \mathring{C}_k}$ and $X_{|[0,T]\times \mathring{C}_k}$ are both continuous. When $p = 1$, both λ and X are continuous on $[0, T] \times I$.

Concerning Theorem 2.8, defining for $k \in \{1, 2\}$:

$$R^W_{N,k} := \frac{1}{N} \sum_{i,j=1}^{N} \int_{B_{N,j}} |W(x_i, x_j) - W(x_i, y)|^k \, dy, \tag{2.24}$$

and

$$S^W_N := \sum_{i=1}^{N} \int_{B_{N,i}} \left(\int_I |W(x_i, y) - W(x, y)|^2 \, dy \right) dx, \tag{2.25}$$

Theorem 2.13 remains true when $R^W_{N,1}, R^W_{N,2}, S^W_N \xrightarrow[N\to\infty]{} 0$, see Lemmas 6.4, 6.5, 6.6 and 6.7. These particular conditions are met in the following cases (details of the computation are left to the reader):

- P-nearest neighbor model [39]: $W(x, y) = \mathbf{1}_{d_{\mathcal{S}_1}(x,y)<r}$ for any $(x, y) \in I^2$ for some fixed $r \in (0, \frac{1}{2})$, with $d_{\mathcal{S}_1}(x, y) = \min(|x - y|, 1 - |x - y|)$.
- Stochastic block model [34, 26]: it corresponds to considering p communities $(C_k)_{1\le k\le p}$. An element of the community C_l communicates with an element of the community C_k with probability p_{kl}. This corresponds to the choice of interaction kernel $W(x, y) = \sum_{k,l} p_{kl} \mathbf{1}_{x\in C_k, y\in C_l}$.

2.3.5 Extensions (subcritical case)

The point of this paragraph is to discuss the possibility of relaxing the subcriticality condition given in (2.3). This condition is used at several times in the paper:

(a) as a sufficient condition to ensure the existence of a (unique) fixed-point X_∞ to (2.8) (see Theorem 2.8 (i)),

(b) to prove the convergence of the deterministic X_t to X_∞ (see Theorem 2.8 *(ii)*), for general h, not necessarily exponential,

(c) to prove the long-term stability of X_N around X_∞, in the exponential case (see Theorem 2.13).

The trickiest point is actually the first one (a), i.e. the existence of a fixed-point to (2.8). To fix ideas, let us think of the pure mean-field case seen in Section 2.3.1, for the generic example (1.1) (that is, $F(X, \eta) = \mu + f(X)$). The fixed-point relation (2.8) is then finite-dimensional and it reduces to find $X_\infty \in \mathbb{R}$ solution to

$$X_\infty = \|h\|_1 \left(\mu + f(X_\infty)\right). \tag{2.26}$$

The condition (2.3) is then the same as

$$\|f'\|_\infty \|h\|_1 < 1, \tag{2.27}$$

which is essentially the generic subcriticality condition that one finds in the literature for mean-field nonlinear Hawkes processes [12]. In the linear case (corresponding to $f(x) = x$), condition (2.27) reduces to $\|h\|_1 < 1$ which is optimal. However, it is quite obvious that (2.3)/(2.27) is no longer optimal w.r.t. the existence of a solution to (2.26) for general f: there may very-well be a unique fixed-point to (2.26) whereas (2.27) is violated, for example in the case a sigmoid f (sufficiently close to the Heaviside function $\mathbf{1}_{[\kappa,+\infty)}$ for some $\kappa > 0$): as long as $\kappa \notin [\|h\|_1\mu, \|h\|_1(\mu + 1)]$, there is a unique solution to

(2.26) whereas (2.27) is not true, as $\|f'\|_\infty$ is very large. Not to mention the possibility of having several (three) fixed-points in this sigmoid case, while (2.27) still does not hold. In this homogenous mean-field case, one can compute the solution to (2.26) *by hand*, as it reduces to a simple equation in dimension 1. The situation gets really more complicated in the spatially-extended setting as (2.8) is intrinsically infinite dimensional. It is unclear if there exists a condition (that would be weaker than (2.3)) ensuring the existence of a (possibly non unique) solution to (2.8) for general W.

However, an important point is the following: *provided we have obtained the existence of such X_∞, unique or not, solution to* (2.8) (which is again a straightforward task for (2.26) but may be complicated for (2.8)), points (b) (the convergence $X_t \to X_\infty$ in Theorem 2.8) at least for h exponential and (c) (the long-term stability result in Theorem 2.13) remain valid under the weaker condition

$$\sup_x |\partial_X F(X_\infty(x), \eta_\infty(x))| \, \|h\|_1 \, r_\infty < 1. \tag{2.28}$$

Condition (2.28) is weaker than (2.3) in the sense that it is only local, around X_∞, whereas (2.3) is global (note that in the the homogeneous case (2.28) translates into $|f'(X_\infty)| \, \|h\|_1 < 1$, to compare with (2.27)). The only modification one has to make in the statements of Theorems 2.8 and 2.13 is that they are now essentially local, i.e. valid provided the initial condition X_0 is sufficiently close to X_∞. More precisely, an alternative statement of item *(ii)* of Theorem 2.8 would be:

Proposition 2.18. *Suppose that Hypothesis 2.1 is true and that we are in the exponential case* (2.4)*. Assume the existence of some $X_\infty \in L^2(I)$ solution to* (2.8) *such that* (2.28) *is satisfied. Then, there exists some $\varepsilon_0 > 0$ such that whenever $\|X_0 - X_\infty\|_2 < \varepsilon_0$, one has $X_t \xrightarrow[t\to\infty]{} X_\infty$ in $L^2(I)$.*

However, this extension of point (b) is only valid when h is exponential. A convergence result under (2.28) for general h (not necessarily exponential) remains open: as it is, the proof of Theorem 2.8 uses in an essential way the uniform condition (2.3). In a same way, the corresponding local stability result concerning X_N is then

Theorem 2.19. *Choose some $t_f > 0$ and $m \geq 1$. Assume Hypotheses 2.1 and 2.6. Assume the existence of some $X_\infty \in L^2(I)$ solution to* (2.8) *such that* (2.28) *is satisfied. Let $\varepsilon_0 > 0$ given by Proposition 2.18 and assume that $\|X_0 - X_\infty\|_2 < \varepsilon_0$. Suppose that for all $\varepsilon > 0$, $\mathbf{P}\left(\|X_N(0) - X_0\|_2 > \varepsilon\right) \xrightarrow[N\to\infty]{} 0$. Then $\mathbb{P}$-a.s., for any $\varepsilon > 0$* (2.17) *is true, for some t_ε independent of N.*

Remark in particular that the operator $\mathcal{L}$ in Proposition 2.11 (whose spectral properties are the main key to the long term stability result) is only expressed in terms of $G = \partial_X F(X_\infty, \eta_\infty)$, that is the exact local quantity appearing in (2.28). In particular, under (2.28), the spectral gap γ in (2.15) becomes $\gamma = \alpha - \sup_x |\partial_X F(X_\infty(x), \eta_\infty(x))| > 0$ and the rest of the proof follows from the same arguments. We stress also that this result never requires the fixed-point X_∞ to be unique (it is indeed the case under the present condition (2.3) but it is never used in the proof of the long-term stability result, that is essentially a result of local nature, around X_∞).

2.4 Link with the literature

Several previous works have complemented the propagation of chaos result mentioned in (1.4) in various situations: Central Limit Theorems (CLT) have been obtained in [23, 26] for homogeneous mean-field Hawkes processes (when both time and N go to infinity) or with age-dependence in [14]. One should also mention the functional fluctuation result recently obtained in [33], also in a pure mean-field setting. A result closer to our case with spatial extension is [17], where a functional CLT is obtained for the spatial profile X_N around its limit. Some insights of the necessity of considering stochastic versions of the NFE (1.3) as second order approximations of the spatial profile are in particular given in [17]. Note here that all of these works provide approximation results of quantities such that λ_N or X_N that are either valid on a bounded time interval $[0, T]$ or under strict growth condition on T (see in particular the condition $\frac{T}{N} \to 0$ for the CLT in [26]), whereas we are here concerned with time-scales that grow polynomially with N.

The analysis of mean-field interacting processes on long time scales has a significant history in the case of interacting diffusions. The important issue of uniform propagation of chaos has been especially studied mostly in reversible situations (see e.g. the case of granular media equation [6]) but also more recently in some irreversible situations (see [19]). There has been in particular a growing interest in the long-time analysis of phase oscillators (see [30] and references therein for a comprehensive review on the subject). We do not aim here to be exhaustive, but as the techniques used in this work present some formal similarities, let us nonetheless comment on the analysis of the simplest situation, i.e. the Kuramoto model. One is here interested in the longtime behavior of the empirical measure $\mu_{N,t} := \frac{1}{N}\sum_{i=1}^N \delta_{\theta_{i,t}}$ of the system of interacting diffusions $(\theta_1, \ldots, \theta_N)$ solving the system of coupled SDEs $\mathrm{d}\theta_{i,t} = -\frac{K}{N}\sum_{j=1}^N \sin(\theta_{i,t} - \theta_{j,t})\mathrm{d}t + \mathrm{d}B_{i,t}$. Standard propagation of chaos techniques show that μ_N converges weakly on a bounded time interval $[0, T]$ to the solution μ_t to the nonlinear Fokker-Planck (NFP) equation $\partial_t \mu_t = \frac{1}{2}\partial_\theta^2 \mu_t + K\partial_\theta\Big(\mu_t(\sin * \mu_t)\Big)$. The simplicity of the Kuramoto model lies in the fact that one can easily prove the existence of a phase transition for this model: when $K \leq 1$, $\mu \equiv \frac{1}{2\pi}$ is the only (stable) stationary point of the previous NFP (subcritical case), whereas it coexists with a stable circle of synchronised profiles when $K > 1$ (supercritical case). A series of papers have analysed the longtime behavior of the empirical measure μ_N of the Kuramoto model (and extensions) in both the subcritical and supercritical cases (see in particular [4, 37, 29, 20] and references therein). The main arguments of the mentioned papers lie in a careful analysis of two contradictory phenomena that arise on a long-time scale: the stability of the deterministic dynamics around stationary points (that forces μ_N to remain in a small neighborhood of these points) and the presence of noise in the microscopic system (which makes μ_N diffuse around these points). In particular, the work that is somehow formally closest to the present article is [20], where the long-time stability of μ_N is analysed in both sub and supercritical cases for Kuramoto oscillators interacting on an Erdős-Rényi graph. We are here (at least formally) in a similar situation to the subcritical case of [20]: the deterministic dynamics of the spatial profile X_N (given by (1.11)) has a unique stationary point which possesses sufficient stability properties. The point of the analysis relies then on a time discretization and some careful control on the diffusive influence of noise that competes with the deterministic dynamics. The main difference (and present difficulty in the analysis) with the diffusion case in [20] is that our noise (Poissonnian rather than Brownian) is multiplicative (whereas it is essentially additive in [20]). This explains in particular the stronger dilution conditions that we require in Hypothesis 2.6 (compared to the optimal $N\rho_N \to \infty$ in [20]) and also the fact that we only reach polynomial time scales (compared to the sub-exponential scale in [20]). There is however every reason to believe that the stability result of Theorem 2.13 would remain valid up to this sub-exponential time scale.

Note here that we deal directly with the control of the Poisson noise. Another possibility would have been to use some Brownian approximation of the dynamics of X_N. Some results in this direction have been initiated in [26] for spatially-extended Hawkes processes exhibiting oscillatory behaviors: some diffusive approximation of the dynamics of the (equivalent of) the spatial profile is provided (see [26, Section 5]). Note however that this approximation is based on the comparison of the corresponding semigroups and is not uniform in time. Hence, it is unclear how one could exploit these techniques for our case. Some stronger (pathwise) approximations between Hawkes and Brownian dynamics have been further proposed in [16], based on Komlós, Major and Tusnády (KMT) coupling techniques [28]. Recently, Prodhomme [43] used similar KMT coupling techniques applied to finite dimensional Markov chains and found the Gaussian approximation to remain precise for very large periods of time. However these results are valid for $\mathbb{Z}^d$-valued continous-time Markov chains, and it is unclear how they can be applied in our situation (with infinite dimension and space extension). The proof we propose is direct and does not rely on such Brownian coupling. Another recent work by Erny et al. [27] about Hawkes processes with mean field interactions in a diffusive regime extended also the propagation of chaos to longer time periods, but the scaling used there is different from ours. This diffusion scaling can also be found in [42]. To the author's knowledge, this is the first result on large time stability of Hawkes process (not mentioning the issue of the random graph of interaction, we believe that our results remain also relevant in the pure mean-field case, see Theorem 2.16).

2.5 Strategy of proof and organization of the paper

Section 3 is devoted to prove the convergence result as $t \to \infty$ of Theorem 2.8. This in particular requires some spectral estimates on the operator $\mathcal{L}$ defined in Proposition 2.11 that are gathered in Section 3.1.

The main lines of proof for Theorem 2.13 are given in Section 4. The strategy of proof is sketched here:

1. The starting point of the analysis is a semimartingale decomposition of $Y_N := X_N - X$, detailed in Section 4.1. The point is to decompose the dynamics of Y_N in terms of, at first order, the linear dynamics (2.11) governing the behavior of the deterministic profile X, modulo some drift terms coming from the graph and its mean-field approximation, some noise term and finally some quadratic remaining error coming from the nonlinearity of F.

2. A careful control on each of these terms in the semimartingale expansion on a bounded time interval are given in the remaining of Section 4.1. The proof of these estimates are respectively given in Section 5 (for the noise term) and Section 6 (for the drift term).

3. The rest of Section 4 is devoted to the proof of Theorem 2.13, see Section 4.2. The first point is that for any given $\varepsilon > 0$, one has to wait a deterministic time $t_\varepsilon > 0$, so that the deterministic profile X_t reaches an ε-neighborhood of X_∞. It is easy to see from the spectral gap estimate (2.14) that this t_ε is actually of order $\frac{-\log(\varepsilon)}{\gamma}$. Then, using Proposition 2.12, the microscopic process X_N is itself ε-close to X_∞ with high-probability.

4. The previous argument is the starting point of an iterative procedure that works as follows: the point is to see that provided X_N is initially close to X_∞, it will remain close to X_∞ on some $[0, T]$ for some sufficiently large deterministic $T > 0$. The key argument is that on a bounded time interval, the deterministic linear dynamics dominates upon the contribution of the noise, so that one has only to wait some sufficiently large T so that the deterministic dynamics prevails upon the other contributions.

5. The rest of the proof consists in an iterative procedure from the previous argument, taking advantage of the Markovian structure of the dynamics of X_N. The time horizon at which one can pursue this recursion is controlled by moment estimates on the noise, proven in Section 5.

The rest of the paper is organised as follows: Section 7 collects the proofs for the finite time behavior of Proposition 2.12 whereas some technical estimates are gathered in the appendix.

3 Asymptotic behavior of (X_t)

This section is related to the proof of Theorem 2.8.

3.1 Estimates on the operator $\mathcal{L}$

Proof of Proposition 2.4. The continuity and compactness of T_W come from the boundedness of W. The structure of the spectrum of T_W is a consequence of the spectral theorem for compact operators. The equality between the spectral radii is postponed to Lemma A.8 where a more general result is stated (see also Proposition 4.7 of [1] for a similar result). □

Proof of Proposition 2.11. Let us introduce the operator

$$\begin{aligned} \mathcal{U} :&L^2(I) \longrightarrow L^2(I) \\ &g \longmapsto \mathcal{U}(g) = T_W(Gg), \end{aligned} \tag{3.1}$$

we have then $\mathcal{L} = -\alpha Id + \mathcal{U}$. By Hypothesis 2.1, G is bounded. Then, for any $g \in L^2(I)$ using Cauchy-Schwarz inequality, $\|\mathcal{U}(g)\|_2^2 \leq \|W\|_2^2\|G\|_\infty^2\|g\|_2^2$. The operator $\mathcal{U}$ is then compact and thus has a discrete spectrum. Moreover, $r_2(\mathcal{U}) = r_\infty(\mathcal{U})$, see Lemma A.8, and $r_\infty(\mathcal{U}) \leq r_\infty(T_W)\|G\|_\infty$ as for any $g \in L^\infty$ and $x \in I$, $|\mathcal{U}g(x)| \leq \|T_W\|_\infty\|Gg\|_\infty \leq \|T_W\|_\infty\|G\|_\infty\|g\|_\infty$. Then $\mathcal{L}$ also has a discrete spectrum, which is the same as $\mathcal{U}$ but shifted by α. Since $r_2(\mathcal{U}) = r_\infty(\mathcal{U})$ (see Lemma A.8), for any $\mu \in \sigma(\mathcal{L})\setminus\{0\}$, $|\mu+\alpha| \leq r_\infty(\mathcal{U})$ thus $Re(\mu) \leq -\alpha + r_\infty(\mathcal{U}) \leq -\alpha + r_\infty\|\partial_u F\|_\infty < 0$ by (2.3). The estimate (2.14) follows then from functional analysis (see e.g. Theorem 3.1 of [41]). □

3.2 About the large time behavior of X_t

Proof of Theorem 2.8. We prove that

- there exists a unique function $\ell : I \mapsto \mathbb{R}^+$ solution of (2.9), continuous and bounded on I, and that
- $(\lambda_t)_{t\geq 0}$ converges uniformly when $t \to \infty$ towards ℓ.

It gives then that $X_\infty := \|h\|_1 T_W \ell$ is the unique solution of (2.8). Then, as $X_t(x) = \int_I W(x,y)\int_0^t h(t-s)\lambda_s(y)ds\ dy$, as (λ_t) is uniformly bounded, and as h is integrable and $\lambda_t \to \ell$ uniformly, we conclude by dominated convergence that uniformly on y, $\int_0^t h(t-s)\lambda_s(y)ds \xrightarrow[t\to\infty]{} \|h\|_1\ell(y)$. As T_W is continuous, the result follows: X_t converges uniformly towards X_∞. We show first that (λ_t) is uniformly bounded. Let $\overline{\lambda}_t(x) = \sup_{s\in[0,t]}\lambda_s(x)$, we have then with (1.13), for $s \in [0,t]$

$$\begin{aligned} \lambda_s(x) &\leq F(0,0) + \|F\|_L|\eta_s(x)| + \|\partial_X F\|_\infty \int_I W(x,y)\int_0^s h(s-u)\lambda_u(y)dudy \\ &\leq F(0,0) + \|F\|_L \sup_{s,x}|\eta_s(x)| + \|\partial_X F\|_\infty\|h\|_1 T_W\overline{\lambda}_t(x), \end{aligned}$$

hence $\overline{\lambda}_t(x) \leq C_{F,\eta} + \|\partial_X F\|_\infty\|h\|_1 T_W\overline{\lambda}_t(x)$. An immediate iteration gives then $\overline{\lambda}_t(x) \leq C_{F,\eta,n_0,h} + \|\partial_X F\|_\infty^{n_0}\|h\|_1^{n_0}\left|T_W^{n_0}\overline{\lambda}_t(x)\right|$, so that, by (2.3) and choosing n_0 sufficiently large such that $\|\partial_X F\|_\infty^{n_0}\|h\|_1^{n_0}\|T_W\|^{n_0} < 1$, we obtain that $\|\overline{\lambda}_t\|_\infty < C$, where C is independent of t. Passing to the limit as $t \to \infty$, this implies that $(\lambda_t)_{t>0}$ is then uniformly bounded, i.e. $\sup_{t\geq 0}\sup_{x\in I}|\lambda_t(x)| =: \|\lambda\|_\infty < \infty$.

We show next that (λ_t) converges pointwise. We start by studying the supremum limit of λ_t, denoted by $\overline{\ell}(x) := \limsup_{t\to\infty} \lambda_t(x) = \inf_{r>0}\sup_{t>r}\lambda_t(x) =: \inf_{r>0}\Lambda(r,x)$. Then for any $r>0$ and $t>r$:

$$\begin{aligned}\lambda_t(x) &= F\left(\int_I W(x,y)\int_0^r h(t-s)\lambda_s(y)ds\,dy + \int_I W(x,y)\int_r^t h(t-s)\lambda_s(y)ds\,dy, \eta_t(x)\right)\\ &\leq F\left(\int_I W(x,y)\int_0^r h(t-s)\lambda_s(y)ds\,dy + \int_I W(x,y)\Lambda(r,y)\int_r^t h(t-s)ds\,dy, \eta_t(x)\right)\end{aligned}$$

by monotonicity of F in the first variable and by positivity of W and h. As $\int_r^t h(t-s)ds \leq \|h\|_1$, it gives

$$\lambda_t(x) \leq F\left(\int_I W(x,y)\int_0^r h(t-s)\lambda_s(y)ds\,dy + \|h\|_1\int_I W(x,y)\Lambda(r,y)dy, \eta_t(x)\right),$$

and as $h(t)\to 0$, by dominated convergence $\int_I W(x,y)\int_0^r h(t-s)\lambda_s(y)ds\,dy \xrightarrow[t\to\infty]{} 0$ and by continuity and monotonicity of F, we obtain

$$\overline{\ell}(x) \leq F\left(\|h\|_1\left(T_W\overline{\ell}\right)(x), \eta_\infty(x)\right). \tag{3.2}$$

Note that $\|\overline{\ell}\|_\infty \leq \|\lambda\|_\infty < \infty$, by the first step of this proof. Denote in a same way $\underline{\ell}(x) := \liminf_{t\to\infty}\lambda_t(x) = \sup_{r>0}\inf_{t>r}\lambda_t(x) =: \sup_{r>0} v(r,x)$, for any $t>0$ we have by monotonocity of F in the first variable:

$$\begin{aligned}\lambda_t(x) &= F\left(\int_0^{\frac{t}{2}}\int_I W(x,y)h(t-s)\lambda_s(y)dyds + \int_{\frac{t}{2}}^t\int_I W(x,y)h(t-s)\lambda_s(y)dyds, \eta_t(x)\right)\\ &\geq F\left(\int_{\frac{t}{2}}^t\int_I W(x,y)h(t-s)v\left(\frac{t}{2},y\right)dyds, \eta_t(x)\right)\\ &= F\left(\int_0^{\frac{t}{2}} h(u)du\int_I W(x,y)v\left(\frac{t}{2},y\right)dy, \eta_t(x)\right).\end{aligned}$$

Taking $\liminf_{t\to\infty}$ on both sides, by monotone convergence, we obtain

$$\underline{\ell}(x) \geq F\left(\|h\|_1\left(T_W\underline{\ell}\right)(x), \eta_\infty(x)\right). \tag{3.3}$$

Combining (3.2) and (3.3), setting $H: l\in L^\infty \mapsto F\left(\|h\|_1 T_W l, \eta_\infty\right)\in L^\infty$, we have shown

$$H\underline{\ell} \leq \underline{\ell} \leq \overline{\ell} \leq H\overline{\ell}. \tag{3.4}$$

For any l and l' in $L^\infty(I)$ and any $x\in I$, we have

$$\begin{aligned}|Hl(x) - Hl'(x)| &\leq \left|F\left(\|h\|_1\left(T_W l\right)(x), \eta_\infty(x)\right) - F\left(\|h\|_1\left(T_W l'\right)(x), \eta_\infty(x)\right)\right|\\ &\leq \|\partial_X F\|_\infty \|h\|_1\left|\left(T_W(l-l')\right)(x)\right|.\end{aligned}$$

By iteration we show that $\|H^{n_0}l - H^{n_0}l'\|_\infty \leq \|\partial_u F\|_\infty^{n_0}\|h\|_1^{n_0}\|T_W^{n_0}\|\|l-l'\|_\infty$, so that, choosing again n_0 sufficiently large, H^{n_0} is a contraction mapping by (2.3). Hence, by (3.4), one has necessarily that for all $x\in I$ $\underline{\ell}(x) = \overline{\ell}(x) < +\infty$ thus (λ_t) converges pointwise towards $\ell = \underline{\ell} = \overline{\ell}$ the unique fixed point of H which satisfies (2.9).

We show now that the family $(\lambda_t)_{t\geq 0}$ is equicontinuous so that the pointwise convergence will imply uniform convergence on the compact set I. For any $(x,y)\in I$ and $t\geq 0$, we have

$$\begin{aligned}|\lambda_t(x) - \lambda_t(y)| &= |F(X_t(x),\eta_t(x)) - F(X_t(y),\eta_t(y)|\\ &\leq \|F\|_L\left(|X_t(x)-X_t(y)| + |\eta_t(x)-\eta_t(y)|\right).\end{aligned}$$

With (2.1), we have

$$\begin{aligned}|\eta_t(x)-\eta_t(y)| &\le |\eta_t(x)-\eta_\infty(x)|+|\eta_\infty(x)-\eta_\infty(y)|+|\eta_\infty(y)-\eta_t(y)|\\ &\le 2\delta_t+\|\eta_\infty\|_L|x-y|,\end{aligned}$$

and as λ is bounded, we have

$$\begin{aligned}|X_t(x)-X_t(y)| &= \left|\int_I (W(x,z)-W(y,z))\int_0^t h(t-s)\lambda_s(z)dsdz\right|\\ &\le \|\lambda\|_\infty\|h\|_1\int_I |W(x,z)-W(y,z)|\,dz. \end{aligned}\tag{3.5}$$

Then $|\lambda_t(x)-\lambda_t(y)|\le C_{F,\lambda,h,W}\left(\delta_t+|x-y|+\int_I|W(x,z)-W(y,z)|\,dz\right)$. Fix $\varepsilon>0$, with (2.1), one can find T such that $C_{F,\lambda,h,W}\delta_t\le\frac{\varepsilon}{2}$ for any $t\ge T$, and as W is uniformly continuous on I^2, one can find $\eta>0$ such that $C_{F,\lambda,h,W}\left(|x-y|+\int_I|W(x,z)-W(y,z)|\,dz\right)\le\frac{\varepsilon}{2}$ when $|x-y|\le\eta$. We can divide $[0,1]$ in intervals $[z_k,z_{k+1}]$ such that for any k, $z_{k+1}-z_k\le\eta$. Then, for any $x\in[0,1]$, one can find z_k such that $|z_k-x|\le\eta$, and $|\lambda_t(x)-\ell(x)|\le|\lambda_t(x)-\lambda_t(z_k)|+|\lambda_t(z_k)-\ell(z_k)|+|\ell(z_k)-\ell(x)|$. By pointwise convergence, $|\lambda_t(z_k)-\ell(z_k)|\le\varepsilon$ for t large enough (but independent of the choice of x), and $|\ell(z_k)-\ell(x)|\le\varepsilon$ by taking the limit when $t\to\infty$ in $|\lambda_t(z_k)-\lambda_t(x)|\le\varepsilon$. It gives then $|\lambda_t(x)-\ell(x)|\le 3\varepsilon$ hence $\sup_{x\in I}|\lambda_t(x)-\ell(x)|\xrightarrow[t\to\infty]{}0$, i.e. (λ_t) converges uniformly towards ℓ. Similarly to (3.5), for any $x,x'\in I$,

$$|X_\infty(x)-X_\infty(x')|\le\|h\|_1\|\ell\|_\infty\int_I|W(x,y)-W(x',y)|\,dy$$

which gives, as W is uniformly continous, the continuity of X_∞. □

4 Large time behavior of $X_N(t)$

The aim of the present section is to prove Theorem 2.13. To study the behavior of $\|X_N(t)-X_\infty\|_2$, let

$$Y_N:=X_N-X_\infty. \tag{4.1}$$

The first step is to write the semimartingale decomposition of Y_N, written in a mild form (see Section 4.1). The proper control on the drift and noise terms are given in Propositions 4.2 and 4.3. In Section 4.2, we give the proof of Theorem 2.13, based in particular on the convergence on a bounded time interval in Proposition 2.12.

4.1 Mild formulation

Proposition 4.1. *The process $(Y_N(t))_{t\ge0}$ satisfies the following semimartingale decomposition in $D([0,T],L^2(I))$, written in a mild form: for any $0\le t_0\le t$*

$$Y_N(t)=e^{(t-t_0)\mathcal{L}}Y_N(t_0)+\phi_N(t_0,t)+\zeta_N(t_0,t) \tag{4.2}$$

where (recall (1.7) *the partition of I in N segments of equal length):*

$$\phi_N(t_0,t)=\int_{t_0}^t e^{(t-s)\mathcal{L}}r_N(s)ds \tag{4.3}$$

with

$$r_N(t)(x)=T_W\left(g_N(t)\right)(x)+\\ \sum_{i=1}^N\left(\frac{1}{N\rho_N}\sum_{j=1}^N\xi_{ij}^{(N)}F(X_{N,j}(t-),\eta_t(x_j))-\int_I W(x,y)F(X_N(t,y),\eta_t(y))dy\right)\mathbf{1}_{B_{N,i}}(x), \tag{4.4}$$

$$g_N(t)(y) := \int_0^1 (1-r)\partial_x^2 F\left(X_\infty(y) + rY_N(t)(y), (1-r)\eta_\infty(y) + r\eta_t(y)\right) Y_N(t)(y)^2 dr +$$
$$\int_0^1 (1-r)\left(\eta_t(y) - \eta_\infty(y)\right)\cdot \partial_\eta^2 F\left(X_\infty(y) + rY_N(t)(y), (1-r)\eta_\infty(y) + r\eta_t(y)\right)\left(\eta_t - \eta_\infty\right)(y) dr$$
$$+ \int_0^1 2(1-r)\partial_{x,\eta}^2 F\left(X_\infty(y) + rY_N(t)(y), (1-r)\eta_\infty(y) + r\eta_t(y)\right)\cdot\left(\eta_t(y) - \eta_\infty(y)\right) Y_N(t)(y) dr$$
$$+ \partial_\eta F\left(X_\infty(y), \eta_\infty(y)\right)\cdot\left(\eta_t(y) - \eta_\infty(y)\right), \quad (4.5)$$

and

$$\zeta_N(t_0, t) = \int_{t_0}^t e^{(t-s)\mathcal{L}} dM_N(s) \quad (4.6)$$

with

$$M_N(t) = \sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij}}{N}\left(Z_{N,j}(t) - \int_0^t \lambda_{N,j}(s) ds\right) \mathbf{1}_{B_{N,i}}. \quad (4.7)$$

ϕ_N is the drift term and ζ_N is the noise term coming from the jumps of the process X_N.

Proof of Proposition 4.1. From (1.10) and (1.11), as we are in the exponential case (2.4), we obtain that X_N verifies

$$dX_N(t) = -\alpha X_N(t) dt + \sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij}}{N} dZ_{N,j}(t) \mathbf{1}_{B_{N,i}}. \quad (4.8)$$

The centered noise M_N defined in (4.7) verifies

$$dM_N(t) := \sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij}}{N}\left(dZ_{N,j}(t) - F(X_{N,j}(t-), \eta_t(x_j)) dt\right) \mathbf{1}_{B_{N,i}},$$

and is a martingale in $L^2(I)$. Thus recalling the definition of X_∞ in (2.8) and by inserting the term $\sum_{i=1}^N \sum_{j=1}^N \frac{w_{ij}}{N} F(X_{N,j}(t-), \eta_t(x_j)) dt \mathbf{1}_{B_{N,i}}$ in (4.8), we obtain

$$dY_N(t) = -\alpha Y_N(t) + dM_N(t) + \sum_{i=1}^N \left(\sum_{j=1}^N \frac{w_{ij}}{N} F(X_{N,j}(t-), \eta_t(x_j))\right) \mathbf{1}_{B_{N,i}} dt - T_W F(X_\infty, \eta_\infty) dt. \quad (4.9)$$

A Taylor's expansion gives

$$F(X_N(t,y), \eta_t(y)) - F(X_\infty(y), \eta_\infty(y)) = \partial_X F\left(X_\infty(y), \eta_\infty(y)\right)\left(X_N(t,y) - X_\infty(y)\right) + g_N(t)(y),$$

with g_N given in (4.5). Hence, we have with G defined in (2.12)

$$-T_W F(X_\infty, \eta_\infty)(x) = -\int_I W(x,y) F(X_N(t,y), \eta_t(y)) dy + T_W(GY_N(t)) + T_W g_N(t)(x),$$

hence coming back to (4.9) and recognizing the operator $\mathcal{L}$ (2.13)

$$dY_N(t) = \mathcal{L} Y_N(t) + dM_N(t) + \sum_{i=1}^N \left(\sum_{j=1}^N \frac{w_{ij}}{N} F(X_{N,j}(t-), \eta_t(x_j))\right) \mathbf{1}_{B_{N,i}} dt$$
$$- T_W F(X_N(t,\cdot), \eta_t(\cdot)) dt + T_W g_N(t).$$

We recognize r_N defined in (4.4), and obtain exactly

$$dY_N(t) = \mathcal{L} Y_N(t) dt + r_N(t) dt + dM_N(t). \quad (4.10)$$

Then the mild formulation (4.2) is a direct consequence of Lemma 3.2 of [47]: the unique strong solution to (4.10) is indeed given by (4.2). □

Proposition 4.2 (Noise perturbation)**.** *Let $m \geq 1$ and $T > t_0 \geq 0$. Under Hypotheses 2.1 and 2.6, there exists a constant $C = C(T, m, F, \eta_0) > 0$ such that $\mathbb{P}$-almost surely for N large enough:*

$$\mathbf{E}\left[\sup_{s\leq T} \|\zeta_N(t_0,s)\|_2^{2m}\right] \leq \frac{C}{(N\rho_N)^m}.$$

The proof is postponed to Section 5.

Proposition 4.3 (Drift term)**.** *Under Hypothesis 2.1, for any $t \geq t_0 > 0$, $\mathbb{P}$-almost surely if N is large enough,*

$$\|\phi_N(t_0,t)\|_2 \leq C_{\text{drift}}\left(\int_{t_0}^t e^{-(t-s)\gamma}\|Y_N(s)\|_2^2 ds + G_N + \int_{t_0}^t e^{-\gamma(t-s)}\left(\delta_s^2 + \delta_s\right) ds\right), \quad (4.11)$$

where $C_{\text{drift}} = C_{W,F,\alpha}$, γ is defined in (2.15), δ_s is defined in (2.1) and $G_N = G_N(\xi)$ is an explicit quantity to be found in the proof that tends to 0 as $N \to \infty$.

The proof is postponed to Section 6.

4.2 Proof of the large time behaviour

We prove here Theorem 2.13, based on the results of Section 4.1. The approach followed is somehow formally similar to the strategy of proof developed in [20] for the diffusion case.

Proof of Theorem 2.13. Choose $m \geq 1$ and $t_f > 0$. Let

$$\varepsilon_0 = \frac{\gamma}{6C_{drift}}, \quad (4.12)$$

where γ is defined in (2.15) and the constant C_{drift} comes from Proposition 4.3 above. Note that it suffices to consider ε small enough, such that $\varepsilon < \varepsilon_0$: t_ε defined below increases as $\varepsilon \searrow 0$, so that it suffices to take $t_\varepsilon = t_{\varepsilon_0}$ whenever $\varepsilon \geq \varepsilon_0$. As (X_t) converges uniformly towards X_∞ (Theorem 2.8), there exists $t_\varepsilon^1 < \infty$ such that

$$\|X_t - X_\infty\|_2 \leq \frac{\varepsilon}{4}, \quad t \geq t_\varepsilon^1. \quad (4.13)$$

Moreover, with (2.1), we also have that $\int_0^t e^{-\gamma(t-s)}\left(\delta_s^2 + \delta_s\right) ds \xrightarrow[t\to\infty]{} 0$, hence there exists $t_\varepsilon^2 < \infty$ such that

$$C_{\text{drift}} \int_0^t e^{-\gamma(t-s)}\left(\delta_s^2 + \delta_s\right) ds \leq \frac{\varepsilon}{18}, \quad t \geq t_\varepsilon^2. \quad (4.14)$$

We set now $t_\varepsilon = \max(t_\varepsilon^1, t_\varepsilon^2)$. Let T such that

$$e^{-\gamma T} < \frac{1}{3}, \quad T > t_f. \quad (4.15)$$

The strategy of proof relies on the following time discretisation. The point is to control $\|X_N(t) - X_\infty\|_2$ on $[t_\varepsilon, T_N]$ with

$$T_N := a_N T + t_\varepsilon, \quad \text{with } a_N := \lceil (N\rho_N)^m \rceil, \quad (4.16)$$

which will imply the result (2.17) as $[t_\varepsilon, (N\rho_N)^m t_f] \subset [t_\varepsilon, T_N]$ since $T > t_f$. We decompose below the interval $[t_\varepsilon, T_N]$ into a_N intervals of length T. We define the following events, with $0 \leq t_a \leq t_b$ (recall that $Y_N(t) = X_N(t) - X_\infty$)

$$A_1^N(t,\varepsilon) := \left\{ \|Y_N(t)\|_2 \leq \frac{\varepsilon}{2} \right\} \quad \text{for } t \geq 0, \tag{4.17}$$

$$A_2^N(\varepsilon) := \left\{ \sup_{t\in[t_\varepsilon, t_\varepsilon+T]} \|\zeta_N(t_\varepsilon, t)\|_2 \leq \frac{\varepsilon}{18} \right\}, \tag{4.18}$$

$$E(t_a,t_b) := \left\{ \max\left(2\|Y_N(t_a)\|_2, \sup_{t\in[t_a,t_b]} \|Y_N(t)\|_2, 2\|Y_N(t_b)\|_2 \right) \leq \varepsilon \right\}. \tag{4.19}$$

By (4.13), and as Proposition 2.12 gives that $\mathbf{P}\left(\sup_{t\in[0,t_\varepsilon]} \|Y_N(t)\|_2 > \frac{\varepsilon}{4}\right) \xrightarrow[N\to\infty]{} 0$, we have by triangle inequality

$$\mathbf{P}\left(A_1^N(t_\varepsilon,\varepsilon)\right) \xrightarrow[N\to\infty]{} 1. \tag{4.20}$$

Step 1 We have from the definition (4.19) of $E(t_a,t_b)$ that

$$\mathbf{P}\left(\sup_{t\in[t_\varepsilon,T_N]} \|Y_N(t)\|_2 \leq \varepsilon \right) \geq \mathbf{P}\left(E(t_\varepsilon,T_N)\right) = \mathbf{P}\left(E(t_\varepsilon,T_N)|A_1^N(t_\varepsilon,\varepsilon)\right)\mathbf{P}\left(A_1^N(t_\varepsilon,\varepsilon)\right). \tag{4.21}$$

Moreover,

$$\begin{aligned}\mathbf{P}\left(E(t_\varepsilon,T_N)|A_1^N(t_\varepsilon,\varepsilon)\right) &= \mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+a_NT)|A_1^N(t_\varepsilon,\varepsilon)\right)\\ &\geq \mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+a_NT)\cap E(t_\varepsilon,t_\varepsilon+(a_N-1)T)|A_1^N(t_\varepsilon,\varepsilon)\right)\end{aligned}$$

which is

$$\mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+a_NT)|E(t_\varepsilon,t_\varepsilon+(a_N-1)T)\cap A_1^N(t_\varepsilon,\varepsilon)\right)\mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+(a_N-1)T)|A_1^N(t_\varepsilon,\varepsilon)\right).$$

Recall that we are in the exponential case (2.4), so that $(X_N(t))_t$ is a Markov process. Thus by Markov property

$$\begin{aligned}&\mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+a_NT)|E(t_\varepsilon,t_\varepsilon+(a_N-1)T)\cap A_1^N(t_\varepsilon,\varepsilon)\right)\\ =&\mathbf{P}\left(E(t_\varepsilon+(a_N-1)T,t_\varepsilon+a_NT)|E(t_\varepsilon,t_\varepsilon+(a_N-1)T)\right)\\ =&\mathbf{P}\left(E(t_\varepsilon+(a_N-1)T,t_\varepsilon+a_NT)\,\middle|\left\{\|Y_N(t_\varepsilon+(a_N-1)T)\|_2 \leq \frac{\varepsilon}{2}\right\}\right)\\ =&\mathbf{P}\left(E(t_\varepsilon+(a_N-1)T,t_\varepsilon+a_NT)\,\middle|A_1^N(t_\varepsilon+(a_N-1)T,\varepsilon)\right)\end{aligned}$$

$\mathbf{P}\left(E(t_\varepsilon+(a_N-1)T,t_\varepsilon+a_NT)|A_1^N(t_\varepsilon+(a_N-1)T,\varepsilon)\right)$ means that, under an initial condition at $t_\varepsilon+(a_N-1)T$, we look at the probability that Y_N stays below ε on the interval $[t_\varepsilon+(a_N-1)T, t_\varepsilon+a_NT]$ of size T and comes back below $\frac{\varepsilon}{2}$ at the final time $t_\varepsilon+a_NT$. By Markov's property, it is exactly $\mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+T)|A_1^N(t_\varepsilon,\varepsilon)\right)$. An immediate iteration gives then

$$\mathbf{P}\left(E(t_\varepsilon,T_N)|A_1^N(t_\varepsilon,\varepsilon)\right) \geq \mathbf{P}\left(E(t_\varepsilon,t_\varepsilon+T)|A_1^N(t_\varepsilon,\varepsilon)\right)^{a_N}. \tag{4.22}$$

By (4.20), from now on we consider that we are on this event $A_1^N(t_\varepsilon,\varepsilon)$ and omit this notation for simplicity.

Step 2 We show that

$$A_2^N(\varepsilon) \subset E(t_\varepsilon,t_\varepsilon+T). \tag{4.23}$$

Let us place ourselves in $A_2^N(\varepsilon)$. As we are also under $A_1^N(t_\varepsilon,\varepsilon)$, we have indeed $\|Y_N(t_\varepsilon)\|_2 \leq \frac{\varepsilon}{2}$ for the first condition of $E(t_\varepsilon,t_\varepsilon+T)$. As Y_N verifies (4.1), it can be written for $t \geq t_\varepsilon$

$$Y_N(t) = e^{\mathcal{L}(t-t_\varepsilon)}Y_N(t_\varepsilon) + \phi_N(t_\varepsilon,t) + \zeta_N(t_\varepsilon,t). \tag{4.24}$$

For any $t \in [t_\varepsilon, t_\varepsilon + T]$,

$$\|\phi_N(t_\varepsilon, t)\|_2 \le C_{\text{drift}} \left(\int_{t_\varepsilon}^t e^{-(t-s)\gamma} \|Y_N(s)\|_2^2 ds + G_N + \int_{t_0}^t e^{-\gamma(t-s)} \left(\delta_s^2 + \delta_s\right) ds \right)$$
$$\le C_{\text{drift}} \left(\int_{t_\varepsilon}^t e^{-(t-s)\gamma} \|Y_N(s)\|_2^2 ds \right) + \frac{\varepsilon}{9} \tag{4.25}$$

where the first inequality comes from Proposition 4.3, and the second is true for N large enough using $G_N \to 0$ and (4.14). Coming back to (4.24), using that by Proposition 2.11

$$\left\| e^{\mathcal{L}(t-t_\varepsilon)} Y_N(t_\varepsilon) \right\|_2 \le e^{-\gamma(t-t_\varepsilon)} \|Y_N(t_\varepsilon)\|_2 , \tag{4.26}$$

and using (4.25), we have on $A_1^N(t_\varepsilon, \varepsilon) \cap A_2^N(\varepsilon)$

$$\|Y_N(t)\|_2 \le \frac{\varepsilon}{2} + C_{\text{drift}} \left(\int_{t_\varepsilon}^t e^{-(t-s)\gamma} \|Y_N(s)\|_2^2 ds \right) + \frac{\varepsilon}{9} + \frac{\varepsilon}{18}.$$

Let $\delta > 0$ such that $\delta \le \min\left(\frac{\varepsilon}{6}, \frac{\gamma}{9C_{\text{drift}}}\right)$. Recall that $\|Y_N(\cdot)\|_2$ is not a continuous function, it jumps whenever a spike of the process $(Z_{N,1}, \dots, Z_{N,N})$ occurs, but the size jump never exceeds $\frac{1}{N}$, and for N large enough $\frac{1}{N} \le \frac{\delta}{2}$. Then, one can apply Lemma A.9 and obtain that for all N large enough,

$$\sup_{t \in [t_\varepsilon, t_\varepsilon + T]} \|Y_N(t)\|_2 \le \frac{\varepsilon}{2} + 3\delta \le \varepsilon. \tag{4.27}$$

It remains to prove that $\|Y_N(t_\varepsilon + T)\|_2 \le \frac{\varepsilon}{2}$. We obtain from (4.24), (4.25) and (4.26) for $t = t_\varepsilon + T$ on $A_1^N(t_\varepsilon, \varepsilon) \cap A_2^N(\varepsilon)$

$$\|Y_N(t_\varepsilon + T)\|_2 \le e^{-\gamma T} \frac{\varepsilon}{2} + \frac{\varepsilon}{6} + C_{\text{drift}} \int_{t_\varepsilon}^{t_\varepsilon + T} e^{-(t_\varepsilon + T - s)\gamma} \|Y_N(s)\|_2^2 ds.$$

Using the a priori bound (4.27)

$$\|Y_N(t_\varepsilon + T)\|_2 \le e^{-\gamma T} \frac{\varepsilon}{2} + \frac{\varepsilon}{12} + \varepsilon^2 \frac{C_{\text{drift}}}{\gamma} \le e^{-\gamma T} \frac{\varepsilon}{2} + \frac{\varepsilon}{6} + \frac{\varepsilon}{6} \le \frac{\varepsilon}{2},$$

where we recall the particular choices of T and $\varepsilon < \varepsilon_0$ in (4.15) and (4.12). This concludes the proof of (4.23).

Step 3 We obtain with (4.22) and Markov's inequality,

$$\begin{aligned} \mathbf{P}\left(E(t_\varepsilon, T_N)\right) &\ge \mathbf{P}\left(E(t_\varepsilon, t_\varepsilon + T)\right)^{a_N} \ge \mathbf{P}(A_2^N(\varepsilon))^{a_N} \\ &= \left(1 - \mathbf{P}\left(\sup_{t \in [t_\varepsilon, t_\varepsilon + T]} \|\zeta_N(t_\varepsilon, t)\|_2 > \frac{\varepsilon}{18}\right)\right)^{a_N} \\ &\ge \left(1 - 18^{2m'} \frac{\mathbb{E}\left[\sup_{t \in [t_\varepsilon, t_\varepsilon + T]} \|\zeta_N(t_\varepsilon, t)\|_2^{2m'}\right]}{\varepsilon^{2m'}}\right)^{a_N}, \end{aligned}$$

where we have taken $m' > m$. With Proposition 4.2, it gives

$$\mathbf{P}\left(E(t_\varepsilon, T_N)\right) \ge \left(1 - \frac{C}{\left(\varepsilon^2 N \rho_N\right)^{m'}}\right)^{a_N} = \exp\left(a_N \ln\left(1 - \frac{C}{\left(\varepsilon^2 N \rho_N\right)^{m'}}\right)\right).$$

By definition (4.16), $a_N = o\left(N\rho_N\right)^{m'}$, the right term tends to 1 as N goes to ∞ under Hypothesis 2.6. By (4.21), we conclude that

$$\mathbf{P}\left(\sup_{t\in[t_\varepsilon,T_N]}\|X_N(t)-X_\infty\|_2\leq\varepsilon\right)\xrightarrow[N\to\infty]{}1.$$

This concludes the proof of Theorem 2.13. □

5 Proofs - Noise perturbation

In this section, we prove Proposition 4.2 concerning the control of the noise perturbation $\zeta_N(t_0,t)$ defined in (4.6). For simplicity of notation, we assume that $t_0=0$. Recall the expression of $(Z_{N,j})_{1\leq j\leq N}$ in (1.8). Introduce the compensated measure $\tilde{\pi}_j(ds,dz):=\pi_j(ds,dz)-dsdz$, so that with the linearity of $(e^{t\mathcal{L}})_{t\geq0}$, we obtain that ζ_N can be written as

$$\zeta_N(0,t)=\sum_{j=1}^N\int_0^t\int_0^\infty e^{(t-s)\mathcal{L}}\chi_j(s,z)\tilde{\pi}_j(ds,dz), \tag{5.1}$$

with $\chi_j(s,z):=\left(\sum_{i=1}^N\mathbf{1}_{B_{N,i}}\frac{w_{ij}}{N}\right)\mathbf{1}_{z\leq\lambda_{N,j}(s)}\in L^2(I)$. The proof of Proposition 4.3 relies on a adaptation of an argument given in [47] (Theorem 4.3), where a similar quantity to (5.1) is considered for $N=1$.

5.1 Control of the moments of the process $Z_{N,i}$

Proposition 5.1. *Let $m\geq1$ and $T>0$. Under Hypotheses 2.1 and 2.6, $\mathbb{P}$-almost surely*

$$\sup_{N\geq1}\mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(T)^m\right]<\infty.$$

Proof. Let $N\geq1$. We have for any $i\in[\![1,N]\!]$

$$\begin{aligned}\mathbf{E}\left[Z_{N,i}(T)^m\right]&\leq\mathbf{E}\left[\left(\left(\left(Z_{N,i}(T)\right)-\int_0^T\lambda_{N,i}(t)dt\right)+\int_0^T\lambda_{N,i}(t)dt\right)^m\right]\\&\leq2^{m-1}\mathbf{E}\left[\left(Z_{N,i}(T)-\int_0^T\lambda_{N,i}(t)dt\right)^m\right]+2^{m-1}\mathbf{E}\left[\left(\int_0^T\lambda_{N,i}(t)dt\right)^m\right]\\&\leq2^{m-1}C\mathbf{E}\left[\left(\int_0^T\lambda_{N,i}(t)dt\right)^{\frac{m}{2}}\right]+(2T)^{m-1}\mathbf{E}\left[\int_0^T\lambda_{N,i}(t)^m dt\right],\end{aligned} \tag{5.2}$$

where we used Jensen's inequality and Burkholder-Davis-Gundy Inequality on the martingale $\left(Z_{N,i}(T)-\int_0^T\lambda_{N,i}(t)dt\right)$. Similarly, we obtain

$$\mathbf{E}\left[\left(\int_0^T\lambda_{N,i}(t)dt\right)^{\frac{m}{2}}\right]\leq T^{\frac{m}{2}-1}\mathbf{E}\left[\int_0^T\left(\lambda_{N,i}(t)\right)^{\frac{m}{2}}dt\right].$$

We focus now on the term $\mathbf{E}\left[\int_0^T\lambda_{N,i}(t)^k dt\right]$ for $k\geq1$. From the definition of $\lambda_{N,i}$ (1.9), by Lipschitz continuity of F and with Jensen's inequality

$$\mathbf{E}\left[\int_0^T \lambda_{N,i}(t)^k dt\right] \leq 2^{k-1} T F(0,\eta_t(x_i))^k + 2^{k-1}\|F\|_L^k \mathbf{E}\left[\int_0^T \left(\frac{1}{N}\sum_{j=1}^N \int_0^{t-} w_{ij} e^{-\alpha(t-s)} dZ_{N,j}(s)\right)^k dt\right].$$

Let $S_i := \sum_{j=1}^N \frac{w_{ij}}{N}$. By (A.2), we have that $\mathbb{P}$-almost surely, $\limsup_{N\to\infty} \sup_{1\leq i\leq N} S_i \leq 2$. We obtain with discrete Jensen's inequality that for any $t \geq 0$

$$\left(\frac{1}{N}\sum_{j=1}^N \int_0^{t-} w_{ij} e^{-\alpha(t-s)} dZ_{N,j}(s)\right)^k \leq S_i^k \left(\sum_{j=1}^N \frac{w_{ij}}{N S_i} Z_{N,j}(t)\right)^k \leq S_i^{k-1} \sum_{j=1}^N \frac{w_{ij}}{N} Z_{N,j}(t)^k.$$

We obtain then

$$\mathbf{E}\left[\int_0^T \lambda_{N,i}(t)^k dt\right] \leq C_{T,F,\eta_0,k} + C_{k,F} \sum_{j=1}^N \frac{w_{ij}}{N} \mathbf{E}\left[\int_0^T Z_{N,j}(t)^k dt\right],$$

thus, going back to (5.2), with $C = C_{T,F,\eta_0,m}$

$$\begin{aligned}\mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(T)^m\right] &\leq \frac{C}{N}\sum_{i=1}^N \left(\mathbf{E}\left[\int_0^T \lambda_{N,i}(t)^{\frac{m}{2}} dt\right] + C\mathbf{E}\left[\int_0^T \lambda_{N,i}(t)^m dt\right]\right) \\ &\leq C\left(1 + \sum_{i,j=1}^N \frac{w_{ij}}{N^2} \mathbf{E}\left[\int_0^T Z_{N,j}(t)^{\frac{m}{2}} dt\right] + \sum_{i,j=1}^N \frac{w_{ij}}{N^2} \mathbf{E}\left[\int_0^T Z_{N,j}(t)^m dt\right]\right).\end{aligned}$$

With (A.2), it gives that, $\mathbb{P}$-almost surely for N large enough $\mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(T)^m\right]$ is upper bounded by

$$C\left(1 + \int_0^T \mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(t)^{\frac{m}{2}}\right] dt + \int_0^T \mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(t)^m\right] dt\right).$$

As for any $t \geq 0$

$$\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N Z_{N,i}(t)\right] = \frac{1}{N}\sum_{i=1}^N \mathbf{E}\left[\int_0^t \lambda_{N,i}(s) ds\right] \leq C_{T,\eta_0,F} + C_{T,\eta_0,F} \int_0^t \mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(s)\right] ds,$$

Grönwall's lemma gives that $\sup_{t\leq T} \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N Z_{N,i}(t)\right] < \infty$ (independently of N) and similarly an immediate iteration gives that for any $k \geq 0$, $\sup_{N\geq 1} \mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(T)^{2^k}\right] < \infty$ which concludes the proof. □

5.2 Proof of Proposition 4.2

Proof. We divide the proof in different steps. Fix $m \geq 1$. We prove Proposition 4.2 for the choice $t_0 = 0$, but it remains the same for a general initial time $t_0 \geq 0$.

Step 1 The functional $\phi : L^2(I) \to \mathbb{R}$ given by $\phi(v) = \|v\|_2^{2m}$ is of class $\mathcal{C}^2$ (recall that $\zeta_N \in L^2(I)$) so that by Itô formula on the expression (5.1) we obtain

$$\begin{aligned}\phi(\zeta_N(t)) &= \int_0^t \phi'(\zeta_N(s))\,\mathcal{L}(\zeta_N(s))\,ds + \sum_{j=1}^N \int_0^t \int_0^\infty \phi'(\zeta_N(s-))\,\chi_j(s,z)\tilde{\pi}_j(ds,dz)\\ &+ \sum_{j=1}^N \int_0^t \int_0^\infty \left[\phi(\zeta_N(s-)+\chi_j(s,z)) - \phi(\zeta_N(s-)) - \phi'(\zeta_N(s-))\,\chi_j(s,z)\right]\pi_j(ds,dz)\\ &:= I_0(t) + I_1(t) + I_2(t). \end{aligned} \tag{5.3}$$

We have then for any $v, h, k \in L^2(I)$, $\phi'(v)h = 2m\|v\|_2^{2m-2}\langle v, h\rangle \in \mathbb{R}$ and $\phi''(v)(h,k) = 2m(2m-1)\|v\|_2^{2m-4}\langle v,k\rangle\langle v,h\rangle + 2m\|v\|^{2m-2}\langle h,k\rangle$.

Step 2 We have $I_0(t) = \int_0^t 2m\|\zeta_N(s)\|_2^{2m-2}\langle \zeta_N(s), \mathcal{L}(\zeta_N(s))\rangle ds$. From Proposition 2.11, $\mathcal{L}$ generates a contraction semi-group hence for any $s \geq 0$, $\langle \zeta_N(s), \mathcal{L}(\zeta_N(s))\rangle \leq 0$ by Lumer-Philipps Theorem (see Section 1.4 of [41]). Then for any $t \geq 0$ we have $I_0(t) \leq 0$.

Step 3 About I_1 in (5.3), with $\alpha_j(s,z) := 2m\|\zeta_N(s-)\|_2^{2m-2}\langle \zeta_N(s-), \chi_j(s,z)\rangle \in \mathbb{R}$,

$$I_1(t) = \sum_{j=1}^N \int_0^t \int_0^\infty \alpha_j(s,z)\tilde{\pi}_j(ds,dz).$$

I_1 is then a real martingale. Using that the $(\pi_j)_{1\leq j\leq N}$ are independent so that there are almost surely no simultaneous jumps and hence $[\tilde{\pi}_j, \tilde{\pi}_{j'}] = 0$ if $j \neq j'$,

$$\begin{aligned}[I_1]_t &= \sum_{j=1}^N \int_0^t \int_0^\infty \alpha_j(s,z)^2 \pi_j(ds,dz)\\ &= \sum_{j=1}^N \int_0^t \int_0^\infty \left(2m\|\zeta_N(s-)\|_2^{2m-2}\langle \zeta_N(s-), \chi_j(s,z)\rangle\right)^2 \pi_j(ds,dz)\\ &\leq 4m^2 \sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{4m-2}\right) \sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz). \end{aligned}$$

Then, by Burkholder-Davis-Gundy inequality, for some $C > 0$,

$$\mathbf{E}\left[\sup_{s\leq t}|I_1(s)|\right] \leq C2m\mathbf{E}\left[\sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{2m-1}\right)\left(\sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^{\frac{1}{2}}\right].$$

Applying Hölder inequality with parameter $\frac{2m-1}{2m} + \frac{1}{2m} = 1$ for the random variables $\sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{2m-1}\right)$ and $\left(\sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^{\frac{1}{2}}$, we obtain that $\mathbf{E}\left[\sqrt{[I_1]_t}\right]$ is upper bounded by

$$2m\left(\mathbf{E}\left[\sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right]\right)^{\frac{2m-1}{2m}}\left(\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]\right)^{\frac{1}{2m}}.$$

Let $\varepsilon > 0$ to be chosen later. From Young's inequality, for any $a, b \geq 0$, we can write $ab = \left(\varepsilon^{\frac{2m-1}{2m}}a\right)\left(\varepsilon^{\frac{-(2m-1)}{2m}}b\right) \leq \frac{2m-1}{2m}\left(\varepsilon^{\frac{2m-1}{2m}}a\right)^{\frac{2m}{2m-1}} + \frac{1}{2m}\left(\varepsilon^{\frac{-(2m-1)}{2m}}b\right)^{2m} = \frac{2m-1}{2m}\varepsilon a^{\frac{2m}{2m-1}} +$

$\frac{1}{2m}\varepsilon^{-(2m-1)}b^{2m}$. Then this gives for the choice $a = \left(\mathbf{E}\left[\sup_{0\le s\le t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right]\right)^{\frac{2m-1}{2m}}$ and $b = \left(\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^t\int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]\right)^{\frac{1}{2m}}$:

$$\mathbf{E}\left[\sqrt{[I_1]_t}\right] \le (2m-1)\varepsilon\mathbf{E}\left[\sup_{0\le s\le t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right] + \varepsilon^{-(2m-1)}\mathbf{E}\left[\left(\sum_{j=1}^N\int_0^t\int_0^\infty \|\chi_j(s,z)\|_2^2\pi_j(ds,dz)\right)^m\right].$$

We have then shown that, for the constant C given by Burkholder-Davis-Gundy Inequality,

$$\mathbf{E}\left[\sup_{s\le T}|I_1(s)|\right] \le C(2m-1)\varepsilon\mathbf{E}\left[\sup_{0\le s\le T}\left(\|\zeta_N(s)\|_2^{2m}\right)\right] + C\varepsilon^{-(2m-1)}\mathbf{E}\left[\left(\sum_{j=1}^N\int_0^T\int_0^\infty \|\chi_j(s,z)\|_2^2\pi_j(ds,dz)\right)^m\right]. \quad (5.4)$$

Step 4 Let us focus now on I_2 in (5.3):

$$I_2(t) = \sum_{j=1}^N\int_0^t\int_0^\infty \left[\phi\left(\zeta_N(s-)+\chi_j(s,z)\right) - \phi\left(\zeta_N(s-)\right) - \phi'\left(\zeta_N(s-)\right)\chi_j(s,z)\right]\pi_j(ds,dz).$$

For any jump (s,z) of the Poisson measure π_j, from Taylor's Lagrange formula there exists $\tau_s\in(0,1)$ such that

$$\phi\left(\zeta_N(s-)+\chi_j(s,z)\right) - \phi\left(\zeta_N(s-)\right) - \phi'\left(\zeta_N(s-)\right)\chi_j(s,z) = \frac{1}{2}\phi''\left(\zeta_N(s-)+\tau_s\chi_j(s,z)\right)\left(\chi_j(s,z),\chi_j(s,z)\right).$$

As $\phi''(v)(h,k) = 2m(2m-1)\|v\|_2^{2m-4}\langle v,k\rangle\langle v,h\rangle + 2m\|v\|^{2m-2}\langle h,k\rangle$ for any $v,h,k\in L^2(I)$, one has with Cauchy–Schwarz inequality that

$$\phi''\left(\zeta_N(s-)+\tau_s\chi_j(s,z)\right)\left(\chi_j(s,z)\right)^2 \le 4m^2\|\zeta_N(s-)+\tau_s\chi_j(s,z)\|_2^{2m-2}\|\chi_j(s,z)\|_2^2.$$

But as $\|x+\tau y\|_2^2 \le \max\left(\|x\|_2^2, \|x+y\|_2^2\right)$ for any $x,y\in L^2(I)$ and $\tau\in(0,1)$, we have here

$$\|\zeta_N(s-)+\tau_s\chi_j(s,z)\|_2^{2m-2} \le \max\left(\|\zeta_N(s-)\|_2^{2m-2}, \|\zeta_N(s-)+\chi_j(s,z)\|_2^{2m-2}\right),$$

and as $\|\zeta_N(s-)\|_2^{2m-2} \le \sup_{s\le t}\|\zeta_N(s)\|_2^{2m-2}$ and $\|\zeta_N(s-)+\chi_j(s,z)\|_2^{2m-2} = \|\zeta_N(s)\|_2^{2m-2} \le \sup_{s\le t}\|\zeta_N(s)\|_2^{2m-2}$, thus

$$\mathbf{E}\left[\sup_{s\le t}|I_2(s)|\right] \le 2m^2\mathbf{E}\left[\sup_{s\le t}\|\zeta_N(s)\|_2^{2m-2}\sum_{j=1}^N\int_0^t\int_0^\infty \|\chi_j(s,z)\|_2^2\pi_j(ds,dz)\right].$$

We proceed now similarly as for I_1. From Hölder inequality, as $\frac{2m-2}{2m}+\frac{1}{m}=1$ we know that for any A,B random non-negative variables, $\mathbf{E}[AB] \le \left(\mathbf{E}\left[A^{\frac{2m}{2m-2}}\right]\right)^{\frac{2m-2}{2m}}\left(\mathbf{E}[B^m]\right)^{\frac{1}{m}}$. It leads for the choice $A = \sup_{0\le s\le t}\left(\|\zeta_N(s)\|_2^{2m-2}\right)$ and $B = \sum_{j=1}^N\int_0^t\int_0^\infty\|\chi_j(s,z)\|_2^2\pi_j(ds,dz)$ to $\mathbf{E}\left[\sup_{s\le t}|I_2(s)|\right]$ is upper bounded by

$$2m^2\left(\mathbf{E}\left[\sup_{0\le s\le t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right]\right)^{\frac{2m-2}{2m}}\left(\mathbf{E}\left[\left(\sum_{j=1}^N\int_0^t\int_0^\infty\|\chi_j(s,z)\|_2^2\pi_j(ds,dz)\right)^m\right]\right)^{\frac{1}{m}}.$$

With the same ε introduced for I_1, from Young's inequality, for any $a, b \geq 0$, we can write $ab = \left(\varepsilon^{\frac{2m-2}{2m}} a\right)\left(\varepsilon^{\frac{-(2m-2)}{2m}} b\right) \leq \frac{2m-2}{2m}\left(\varepsilon^{\frac{2m-2}{2m}} a\right)^{\frac{2m}{2m-2}} + \frac{1}{m}\left(\varepsilon^{\frac{-(2m-2)}{2m}} b\right)^m = \frac{2m-2}{2m}\varepsilon a^{\frac{2m}{2m-2}} + \frac{1}{m}\varepsilon^{-(2m-2)} b^m$. For the choice

$$a = \left(\mathbf{E}\left[\sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right]\right)^{\frac{2m-2}{2m}} \quad \text{and}$$
$$b = \left(\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]\right)^{\frac{1}{m}},$$

it gives that $\mathbf{E}\left[\sup_{s\leq t}|I_2(s)|\right]$ is upper bounded by

$$m(2m-2)\varepsilon\mathbf{E}\left[\sup_{0\leq s\leq t}\left(\|\zeta_N(s)\|_2^{2m}\right)\right] + 2m\varepsilon^{-(2m-2)}\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^t \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]. \tag{5.5}$$

Taking the expectation in (5.3) and combining (5.4) and (5.5), we obtain that

$$\begin{aligned}\mathbf{E}\left[\sup_{s\leq T}\|\zeta_N(s)\|_2^{2m}\right] &\leq \varepsilon\left(C(2m-1) + m(2m-2)\right)\mathbf{E}\left[\sup_{0\leq s\leq T}\left(\|\zeta_N(s)\|_2^{2m}\right)\right] \\ &+ \left(C\varepsilon^{-(2m-1)} + 2m\varepsilon^{-(2m-2)}\right)\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^T \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]. \end{aligned}\tag{5.6}$$

Step 5 We can now fix ε such that $\varepsilon\left(C(2m-1) + m(2m-2)\right) \leq \frac{1}{2}$ so that (5.6) leads to

$$\mathbf{E}\left[\sup_{s\leq T}\|\zeta_N(s)\|_2^{2m}\right] \leq 2C\mathbf{E}\left[\left(\sum_{j=1}^N \int_0^T \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right], \tag{5.7}$$

where $C > 0$ depends only on m.

Step 6 Let $A_N := \mathbf{E}\left[\left(\sum_{j=1}^N \int_0^T \int_0^\infty \|\chi_j(s,z)\|_2^2 \pi_j(ds,dz)\right)^m\right]$. We have

$$\|\chi_j(s,z)\|_2^2 = \int_I \left(\sum_{i=1}^N \mathbf{1}_{B_{N,i}}(x)\frac{w_{ij}}{N}\mathbf{1}_{z\leq\lambda_{N,j}(s)}\right)^2 dx = \mathbf{1}_{z\leq\lambda_{N,j}(s)}\sum_{i=1}^N \frac{\xi_{ij}}{N^3\rho_N^2},$$

which leads to, with the definition of $Z_{N,j}$ in (1.8)

$$\begin{aligned}A_N &= \mathbf{E}\left[\left(\sum_{i,j=1}^N \int_0^T \int_0^\infty \mathbf{1}_{z\leq\lambda_{N,j}(s)}\frac{\xi_{ij}}{N^3\rho_N^2}\pi_j(ds,dz)\right)^m\right] \\ &\leq \left(\frac{1}{N\rho_N}\right)^m \mathbf{E}\left[\left(\sum_{i,j=1}^N \frac{\xi_{ij}}{N^2\rho_N} Z_{N,j}(T)\right)^m\right].\end{aligned}$$

With (A.2), Jensen's discrete inequality and (5.7), it leads to

$$\begin{aligned}A_N &\leq \left(\frac{1}{N\rho_N}\right)^m \mathbf{E}\left[\left(\sum_{j=1}^N \frac{1}{N}\left(\sup_j \sum_{i=1}^N \frac{\xi_{ij}}{N\rho_N}\right) Z_{N,j}(T)\right)^m\right] \\ &\leq \frac{C}{(N\rho_N)^m}\mathbf{E}\left[\frac{1}{N}\sum_{j=1}^N Z_{N,j}(T)^m\right],\end{aligned}$$

hence the result with Proposition 5.1. □

6 Proofs – drift term

In this section, we prove Proposition 4.3 concerning the control of the drift term perturbation $\phi_N(t_0,t)$ defined in (4.3).

6.1 Notation

We introduce the following constants

$$\Theta_{t,i,1} := \frac{1}{N\rho_N}\sum_{j=1}^{N}\left(\xi_{ij}^{(N)} - \rho_N W(x_i,x_j)\right) F(X_{N,j}(t-),\eta_t(x_j)), \tag{6.1}$$

$$\Theta_{t,i,2} := \frac{1}{N}\sum_{j=1}^{N} W(x_i,x_j)\left(F(X_{N,j}(t-),\eta_t(x_j)) - F(X_{N,j}(t),\eta_t(x_j))\right), \tag{6.2}$$

$$\Theta_{t,i,3} := \frac{1}{N}\sum_{j=1}^{N} W(x_i,x_j)F(X_{N,j}(t),\eta_t(x_j)) - \int_I W(x_i,y)F(X_N(t,y),\eta_t(y))dy, \tag{6.3}$$

and the auxiliary function in $L^2(I)$

$$\Theta_{t,i,4}(x) := \int_I \left(W(x_i,y) - W(x,y)\right) F(X_N(t,y),\eta_t(y))dy. \tag{6.4}$$

From the expression of r_N in (4.4), we have then

$$r_N(t) = \sum_{i=1}^{N}\left(\sum_{k=1}^{4}\Theta_{t,i,k}\right)\mathbf{1}_{B_{N,i}} + T_W\left(g_N(t)\right), \tag{6.5}$$

and we can divide ϕ_N defined in (4.3) in several terms $\phi_N(t) = \sum_{k=0}^{4}\phi_{N,k}(t)$ with

$$\phi_{N,0}(t) := \int_{t_0}^{t} e^{(t-s)\mathcal{L}} T_W\left(g_N(s)\right) ds, \tag{6.6}$$

$$\phi_{N,k}(t) := \int_{t_0}^{t} e^{(t-s)\mathcal{L}} \sum_{i=1}^{N}\frac{1}{N}\Theta_{s,i,k}\mathbf{1}_{B_{N,i}} ds \quad \text{for } k \in [\![1,4]\!]. \tag{6.7}$$

6.2 Preliminary results

Lemma 6.1. *Denoting by $\tilde{Y}_N(s)(v) := Y_N(s)\left(\frac{\lceil Nv\rceil}{N}\right)$, we have*

$$\sup_{s\geq 0}\|\tilde{Y}_N(s) - Y_N(s)\|_2 \xrightarrow[N\to\infty]{} 0. \tag{6.8}$$

Proof. A direct computation gives, for any $s \geq 0$,

$$\|\tilde{Y}_N(s) - Y_N(s)\|_2^2 = \sum_j \int_{B_{N,j}} \left(X_{N,j}(s) - X_\infty(x_j) - X_N(s)(y) + X_\infty(y)\right)^2 dy.$$

By definition of $X_N(s)$ in (1.11), $X_N = X_{N,j}$ on $B_{N,j}$ hence

$$\|\tilde{Y}_N(s) - Y_N(s)\|_2^2 = \sum_j \int_{B_{N,j}} \left(X_\infty(y) - X_\infty(x_j)\right)^2 dy.$$

Then (6.8) is a straightforward consequence of the uniform continuity of X_∞ on the compact I (see Theorem 2.8). It still holds under the hypotheses of Section 2.3.4 by decomposing the sum on each interval C_k. ☐

We will often use

$$\frac{1}{N}\sum_{j=1}^{N}|Y_N(s)(x_j)|^2 = \|\tilde{Y}_N(s)\|_2^2 \le \frac{1}{2}\left(1+\|\tilde{Y}_N(s)\|_2^4\right) \le \frac{1}{2}\left(2+\|Y_N(s)\|_2^4\right), \tag{6.9}$$

the last inequality being true for N large enough (independently of s) using Lemma 6.1.

Lemma 6.2. *Under Hypothesis 2.1,*

$$R_{N,k}^W \xrightarrow[N\to\infty]{} 0, \quad k\in\{1,2\}, \quad S_N^W \xrightarrow[N\to\infty]{} 0, \tag{6.10}$$

where $R_{N,k}^W$ and S_N^W are respectively defined in (2.24) *and* (2.25).

Proof. Fix $\varepsilon > 0$. As W is uniformly continuous on I, there exists $\eta > 0$ such that $|W(x,y) - W(x,z)| \le \epsilon$ for any $(x,y,z)\in I^3$ with $|y-z|\le\eta$. Then, for N large enough (such that $\frac{1}{N}\le\eta$, we have directly that $R_{N,1}^W \le \epsilon$ and $R_{N,2}^W\le\epsilon$ hence the result. We can do the same for S_N^W. □

Lemma 6.3. *Under Hypothesis 2.1, for any $t > t_0 \ge 0$,*

$$\|\phi_{N,0}(t)\|_2 \le C_{F,W}\int_{t_0}^t e^{-\gamma(t-s)}\left(\|Y_N(s)\|_2^2+\delta_s+\delta_s^2\right)ds. \tag{6.11}$$

Proof of Lemma 6.3. By Proposition 2.11 we have $\|\phi_{N,0}(t)\|_2 \le \int_{t_0}^t e^{-\gamma(t-s)}\|T_W g_N(s)\|_2\,ds$. As for any $x\in I$, $|T_W g_N(s)(x)| \le \int_I W(x,y)\,|g_N(s)(y)|\,dy$, and as

$$\begin{aligned}|g_N(s)(y)| \le{}& \left\|\partial_X^2 F\right\|_\infty Y_N(t)(y)^2 + \left\|\partial_\eta^2 F\right\|_\infty |\eta_t(y)-\eta_\infty(y)|^2\\ &+2\left\|\partial_{X,\eta}^2 F\right\|_\infty |Y_N(t)(y)|\,|\eta_t(y)-\eta_\infty(y)| + \|\partial_\eta F\|_\infty |\eta_t(y)-\eta_\infty(y)|,\end{aligned}$$

with Hypothesis 2.1 it gives

$$\begin{aligned}\|T_W g_N(s)\|_2^2 &= \int_I\left(\int_I W(x,y)g_N(s)(y)dy\right)^2 dx\\ &\le C_F\int_I\left(\int_I W(x,y)\left(Y_N(s)(y)^2+\delta_s^2+Y_N(s)(y)\delta_s+\delta_s\right)dy\right)^2dx\\ &\le C_{F,W}\left(\|Y_N(s)\|_2^4+\|Y_N(s)\|_2^2\delta_s^2+\delta_s^2+\delta_s^4\right)\\ &\le C_{F,W}\left(\frac{3}{2}\|Y_N(s)\|_2^4+\frac{3}{2}\delta_s^2+\delta_s^4\right)\end{aligned}$$

as W is bounded. We obtain then, as $\sqrt{a+b}\le\sqrt{a}+\sqrt{b}$,

$$\|\phi_{N,0}(t)\|_2 \le C_{F,W}\int_{t_0}^t e^{-\gamma(t-s)}\left(\|Y_N(s)\|_2^2+\|Y_N(s)\|_2\delta_s+\delta_s+\delta_s^2\right)ds.$$

Then (6.11) follows as $\|Y_N(s)\|_2 \le \frac{1}{2}\left(1+\|Y_N(s)\|_2^2\right)$ and $\sup_s\delta_s<\infty$. □

Lemma 6.4. *Under Hypotheses 2.1 and 2.6, $\mathbb{P}$-almost surely for N large enough and for any $t > t_0 \ge 0$,*

$$\|\phi_{N,1}(t)\|_2 \le C_F\int_{t_0}^t e^{-(t-s)\gamma}\|Y_N(s)\|_2^2\,ds + G_{N,1}, \tag{6.12}$$

where $G_{N,1} = G_{N,1}(\xi)$ is explicit in N and tends to 0 as $N\to\infty$. Moreover, if we suppose F bounded, we have a better bound

$$\sup_{t>0}\|\phi_{N,1}(t)\|_2 \le \frac{C_F}{\sqrt{N\rho_N^2}}. \tag{6.13}$$

Proof of Lemma 6.4. Proposition 2.11 gives that

$$\|\phi_{N,1}(t)\|_2 \leq K \int_{t_0}^{t} e^{-(t-s)\gamma} \|\gamma_N(s)\|_2 ds \tag{6.14}$$

with

$$\gamma_N(s) := \sum_{i=1}^{N} \Theta_{i,s,1} \mathbf{1}_{B_{N,i}} = \sum_{i,j=1}^{N} \frac{1}{N\rho_N} \overline{\xi_{ij}} F(X_{N,j}(s-), \eta_s(x_j)) \mathbf{1}_{B_{N,i}}, \tag{6.15}$$

where we have used the notation

$$\overline{\xi_{ij}} = \xi_{ij}^{(N)} - W_N(x_i, x_j). \tag{6.16}$$

Forgetting about the term $F(X_{N,j}(s-), \eta_s(x_j))$ in (6.15), γ_N is essentially an empirical mean of the independent centered variables $\overline{\xi_{ij}}$ and thus should be small as $N \to \infty$. One difficulty here is that concentration bounds (e.g. Bernstein inequality) for weighted sums such as $\sum_j \overline{\xi_{ij}} u_{i,j}$ (for some deterministic fixed weight $u_{i,j}$) are not directly applicable, as $u_{i,j} = F(X_{N,j}(s-), \eta_s(x_j)) \mathbf{1}_{B_{N,i}}$ depends in a highly nontrivial way on the variables $\xi_{i,j}^{(N)}$ themselves. A strategy would be to use Grothendieck inequality (see Theorem A.1). We refer here to [20, 22] where the use of such Grothendieck inequality (and extensions) has been implemented in a similar context of interacting diffusions on random graphs. However here, a supplementary difficulty lies in the fact that F need not be bounded (recall that a particular example considered here concerns the linear case where $F(x, \eta) = x + \mu$). Hence the application of Grothendieck inequality is not straightforward when F is unbounded. For this reason, we give below two different controls on γ_N: a general one, without assuming that F is bounded and a second (sharper) one, when F is bounded (using Grothendieck inequality). In the first case, we get around the difficulty of unboundedness of F by introducing $F(X_\infty(x_j), \eta_\infty(x_j))$ which is bounded, since X_∞ is.

First begin with the general control on γ_N: we can write

$$\begin{aligned}\gamma_N(s) &= \sum_{i,j=1}^{N} \frac{1}{N\rho_N} \overline{\xi_{ij}} \left(F(X_{N,j}(s-), \eta_s(x_j)) - F(X_\infty(x_j), \eta_\infty(x_j)) \right) \mathbf{1}_{B_{N,i}} \\ &\quad + \sum_{i,j=1}^{N} \frac{1}{N\rho_N} \overline{\xi_{ij}} F(X_\infty(x_j), \eta_\infty(x_j)) \mathbf{1}_{B_{N,i}} =: \gamma_{N,1}(s) + \gamma_{N,2}.\end{aligned} \tag{6.17}$$

Denoting by $\Delta F_j := F(X_{N,j}(s-), \eta_s(x_j)) - F(X_\infty(x_j), \eta_\infty(x_j))$, we have, as $\langle \mathbf{1}_{B_{N,i}}, \mathbf{1}_{B_{N,i'}} \rangle = \frac{\mathbf{1}_{i=i'}}{N}$ and with $S_{jj'} := \frac{1}{N} \sum_{i=1}^{N} \overline{\xi_{ij}}\, \overline{\xi_{ij'}}$, $\|\gamma_{N,1}(s)\|_2^2 = \frac{1}{N^2 \rho_N^2} \sum_{j,j'=1}^{N} \Delta F_j \Delta F_{j'} \frac{1}{N} S_{jj'}$. Define the following quantity $S_N^{\max} := \sup_{1 \leq j \neq j' \leq N} |S_{jj'}|$. The purpose of Lemma A.5 is exactly to control $S_N^{\max}$, see in particular (A.3). We have

$$\begin{aligned}\|\gamma_{N,1}(s)\|_2^2 &= \left(\frac{1}{N^2 \rho_N^2} \sum_{j \neq j'=1}^{N} \Delta F_j \Delta F_{j'} \frac{S_{jj'}}{S_N^{\max}} \right) S_N^{\max} + \frac{1}{N^3 \rho_N^2} \sum_{i,j=1}^{N} \Delta F_j^2 \overline{\xi_{ij}}^2 \\ &\leq S_N^{\max} \left(\frac{1}{N\rho_N^2} \sum_{j=1}^{N} |\Delta F_j|^2 \right) + \frac{1}{N^2 \rho_N^2} \sum_{j=1}^{N} \Delta F_j^2.\end{aligned}$$

As $|\Delta F_j\| \leq \|F\|_L \left(|Y_N(s-)(x_j)| + \delta_s \right)$, we obtain as $s \mapsto \delta_s$ is bounded

$$\|\gamma_{N,1}(s)\|_2^2 \leq C_F \left(\left\| \tilde{Y}_N(s-) \right\|_2^2 + 1 \right) \left(\frac{S_N^{\max}}{\rho_N^2} + \frac{1}{N\rho_N^2} \right),$$

hence using (6.9) and the fact that $\|Y_N(s-)\|_2 \leq \|Y_N(s)\|_2 + \frac{C}{N\rho_N}$,

$$\|\gamma_{N,1}(s)\|_2^2 \leq C_F \left(\|Y_N(s)\|_2^4 + 1 \right) \left(\frac{S_N^{\max}}{\rho_N^2} + \frac{1}{N\rho_N^2} \right). \tag{6.18}$$

For the second term of (6.17), we have

$$\begin{aligned}\|\gamma_{N,2}\|_2^2 &= \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{N\rho_N} \sum_{j=1}^N \overline{\xi_{ij}} F(X_\infty(x_j), \eta_\infty(x_j)) \right)^2 \\ &= \frac{1}{N^3\rho_N^2} \sum_{i=1}^N \sum_{j,j'=1}^N \overline{\xi_{ij}}\ \overline{\xi_{ij'}} F(X_\infty(x_j), \eta(x_j)) F(X_\infty(x_{j'}), \eta_\infty(x_{j'})).\end{aligned}$$

Let $\alpha_{i,j,j'} := \frac{F(X_\infty(x_j),\eta_\infty(x_j))F(X_\infty(x_{j'}),\eta_\infty(x_{j'}))}{\|F(X_\infty,\eta_\infty)\|_\infty^2} \in [0,1]$, $R_k := \sum_{\substack{i,j,j'=1 \\ j\neq j'}}^k \alpha_{i,j,j'} \overline{\xi_{ij}}\ \overline{\xi_{ij'}}$, and $\mathcal{F}_k = \sigma(\xi_{ij}, 1 \leq i,j \leq k)$. We have then $\|\gamma_{N,2}\|_2^2 = \frac{C_{F,X_\infty}}{N^3\rho_N^2} \sum_{i,j=1}^N \alpha_{i,j,j} \overline{\xi_{ij}}^2 + \frac{C_{F,X_\infty}}{N^3\rho_N^2} R_N \leq \frac{C_{F,X_\infty}}{N\rho_N^2} + \frac{C_{F,X_\infty}}{N^3\rho_N^2} R_N$. We show next that $(R_k)_{1\leq k\leq N}$ is a $(\mathcal{F}_k)$-martingale. Let $\Delta R_k = R_{k+1} - R_k$. For any $k \geq 1$ (note that $R_1 = 0$), we have

$$\Delta R_k = R_{k+1} - R_k = \sum_{\substack{j,j'=1 \\ j\neq j'}}^k \alpha_{k+1,j,j'} \overline{\xi_{k+1,j}}\ \overline{\xi_{k+1,j'}} + \sum_{\substack{1\leq i\leq k+1 \\ 1\leq j\leq k}} (\alpha_{i,j,k+1} + \alpha_{i,k+1,j'}) \overline{\xi_{i,k+1}}\ \overline{\xi_{ij}},$$

and thus $\mathbb{E}[\Delta R_k | \mathcal{F}_k] = 0$ as $\mathbb{E}\left[\overline{\xi_{ij}}\ \overline{\xi_{ij'}} | \mathcal{F}_k\right] = 0$ if $j \neq j'$ and at least one of the indexes i, j, j' is equal to $k+1$ by independence of the family of random variables $(\xi_{ij})_{i,j}$. Moreover, as each $|\overline{\xi_{i,j}}| \leq 1$ and $|\alpha_{i,j,k}| \leq 1$, it gives $|\Delta R_k| \leq 3k^2 + k$. Theorem A.2 gives then that

$$\begin{aligned}\mathbb{P}\left(\left| \frac{C_{F,X_\infty}}{N^3\rho_N^2} R_N \right| \geq x \right) &= \mathbb{P}\left(|R_N| \geq \frac{xN^3\rho_N^2}{C_{F,X_\infty}} \right) \\ &\leq 2\exp\left(- \frac{\left(\frac{xN^3\rho_N^2}{C_{F,X_\infty}} \right)^2}{2\sum_{k=1}^N (3k^2+k)^2} \right) \\ &= 2\exp\left(- \frac{x^2N^6\rho_N^4}{C_{F,X_\infty}^2 P(N)} \right),\end{aligned}$$

with $P(N) = 2N(N+1)\left(\frac{9}{5}\left(N + \frac{1}{2}\right)\left(N^2 + N - \frac{1}{3}\right) + \frac{3N(N+1)}{2} + \frac{2N+1}{6} \right) \sim_{N\to\infty} \frac{18}{5}N^5$. For the choice $x^2 = \frac{C_{F,X_\infty}^2 P(N)}{N^{6-2\tau}\rho_N^4}$ with τ in (2.6), ($x^2 \propto \frac{1}{N^{1-2\tau}\rho_N^4}$) it gives

$$\mathbb{P}\left(\left| \frac{C_{F,X_\infty}}{N^3} R_N \right| \geq \sqrt{\frac{C_{F,X_\infty}^2 P(N)}{N^{6-2\tau}\rho_N^4}} \right) \leq 2\exp\left(-N^{2\tau}\right),$$

which is summable hence by Borel-Cantelli Lemma, there exists $\mathcal{O} \in \mathcal{F}$ such that $\mathbb{P}(\mathcal{O}) = 1$ and on $\mathcal{O}$, there exists $\widetilde{N} < \infty$ such that if $N \geq \widetilde{N}$, $\left| \frac{C_{F,X_\infty}}{N^3} R_N \right| \leq \sqrt{\frac{C_{F,X_\infty}^2 P(N)}{N^{6-2\tau}\rho_N^4}} \propto \frac{1}{N^{1/2-\tau}\rho_N^2}$, hence $\mathbb{P}$-a.s. for N large enough

$$\|\gamma_{N,2}\|_2^2 \leq C\left(\frac{1}{N\rho_N^2} + \frac{1}{N^{1-2\tau}\rho_N^4} \right). \tag{6.19}$$

Coming back to (6.17), combining (6.18) and (6.19) and a control of $S_N^{\max}$ from Lemma A.5, we have $\mathbb{P}$-a.s. for N large enough

$$\|\gamma_N(s)\|_2^2 \leq C_F \left(\|Y_N(s)\|_2^4 + 1\right)\left(\frac{1}{N^{1/2-\tau}\rho_N^2} + \frac{1}{N\rho_N^2}\right) + C_F\left(\frac{1}{N\rho_N^2} + \frac{1}{N^{1-2\tau}\rho_N^4}\right),$$

hence taking the square root and using (6.14),

$$\|\phi_{N,1}(t)\|_2 \leq C_F \int_{t_0}^{t} e^{-(t-s)\gamma}\, \|Y_N(s)\|_2^2\, ds + G_{N,1},$$

where $G_{N,1} = C_F\left(\frac{1}{N\rho_N^2} + \frac{1}{N^{1-2\tau}\rho_N^4} + \frac{1}{N^{1/2-\tau}\rho_N^2}\right) \to 0$ under Hypothesis 2.6.

Let us now turn to the sharper control on γ_N defined in (6.15) when F is bounded. Coming back to (6.17), we have

$$\begin{aligned}
\|\gamma_N(s)\|_2^2 &= \int \left(\sum_{i,j=1}^{N} \frac{1}{N\rho_N}\overline{\xi_{ij}} F\left(X_{N,j}(s-), \eta_s(x_j)\right) \mathbf{1}_{B_{N,i}}(x)\right)^2 dx\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N} \frac{1}{N\rho_N}\overline{\xi_{ij}} F\left(X_{N,j}(s-), \eta_s(x_j)\right)\right)^2\\
&= \frac{1}{N^3\rho_N^2}\sum_{i,j,k=1}^{N} \overline{\xi_{ij}}\;\overline{\xi_{ik}} F\left(X_{N,j}(s-), \eta_s(x_j)\right) F\left(X_{N,k}(s-), \eta_s(x_k)\right)\\
&= \left(\frac{\|F\|_\infty}{N\rho_N}\right)^2 \frac{1}{N}\sum_{j,k=1}^{N} \alpha_{jk} F_j F_k,
\end{aligned}$$

with $\alpha_{jk} := \sum_{i=1}^{N} \overline{\xi_{ij}}\;\overline{\xi_{ik}}$ and $F_j := \frac{F(X_{N,j}(s-),\eta_s(x_j))}{\|F\|_\infty}$. Grothendieck inequality (see Theorem A.1) gives then that there exists $K > 0$ such that

$$\begin{aligned}
\|\gamma_N(s)\|_2^2 &\leq K\frac{1}{N}\left(\frac{\|F\|_\infty}{N\rho_N}\right)^2 \sup_{s_j, t_k = \pm 1} \sum_{j,k} \alpha_{jk} s_j t_k\\
&\leq \frac{C_F}{N^3\rho_N^2} \sup_{s_j, t_k = \pm 1} \sum_{i,j,k=1}^{N} \overline{\xi_{ij}}\;\overline{\xi_{ik}} s_j t_k.
\end{aligned}$$

Fix some vectors of signs $s = (s_i)_{1\leq i\leq N}$ and $t = (t_j)_{1\leq j\leq N}$. Let $A = \left(\overline{\xi_{ij}}\right)_{1\leq i,j\leq N}$, then $\sum_{i,j,k=1}^{N} \overline{\xi_{ij}}\;\overline{\xi_{ik}} s_j t_k = \langle t, A^*As\rangle$ where $\langle , \rangle$ denotes the scalar product in $\mathbb{R}^N$ and A^* the transpose of A. As for any sign vector t, $\|t\|^2 = \sum_{k=1}^{N} t_k^2 = N$, and $\|A^*A\| = \|A\|_{\text{op}}^2$, we obtain as $|\langle t, A^*As\rangle| \leq \|t\|\|A^*As\| \leq N\|A\|_{\text{op}}^2$:

$$\|\gamma_N(s)\|_2^2 \leq \frac{C_F}{N^3\rho_N^2} N\|A\|_{\text{op}}^2 = \frac{C_F}{N^2\rho_N^2}\|A\|_{\text{op}}^2.$$

From Theorem A.3, there exist C_a and C_b positive constants such that for any $x \geq C_a$,

$$\mathbb{P}\left(\|A\|_{\text{op}} > x\sqrt{N}\right) \leq C_a \exp\left(-C_b x N\right).$$

We apply it for $x = C_a$, hence, by Borel-Cantelli Lemma as $\exp(-CN)$ is summable, there exists $\widetilde{\mathcal{O}} \in \mathcal{F}$ such that $\mathbb{P}(\widetilde{\mathcal{O}}) = 1$ and on $\widetilde{\mathcal{O}}$, there exists $\widetilde{N} < \infty$ such that if $N \geq \widetilde{N}$, $\|A\|_{\text{op}} \leq C_a\sqrt{N}$. We obtain then that

$$\|\gamma_N(s)\|_2^2 \leq \frac{C_F}{N\rho_N^2}$$

$\mathbb{P}$-a.s. for N large enough, which concludes the proof in the bounded case with (6.14) as $\int_{t_0}^t e^{-(t-s)\gamma} ds \leq \frac{1}{\gamma}$. ☐

Lemma 6.5. *Under Hypothesis 2.1, there exists $C_F > 0$ such that for any $t > t_0 \geq 0$,*

$$\|\phi_{N,2}(t)\|_2 \leq \frac{C_F}{N\rho_N}.$$

Proof of Lemma 6.5. Recall the definition of $\phi_{N,2}(t)$ in (6.7) and $\Theta_{t,i,2}$ in (6.2), it directly comes from the Lipschitz continuity of F and the fact that $Z_{N,1}, \ldots, Z_{N,N}$ do not jump simultaneously. ☐

Lemma 6.6. *Under Hypothesis 2.1, for any $t > t_0 \geq 0$,*

$$\|\phi_{N,3}(t)\|_2 \leq C_{F,X_\infty,\eta,W} \int_{t_0}^t e^{-(t-s)\gamma} \left(\|Y_N(s)\|_2^2 + \delta_s \right) ds + G_{N,2}, \tag{6.20}$$

where $G_{N,2}$ is explicit in N and tends to 0 as $N \to \infty$. Moreover, if we suppose F bounded, we have

$$\|\phi_{N,3}(t)\|_2 \leq C \left(\int_{t_0}^t e^{-(t-s)\gamma} \delta_s ds + \sqrt{R_{N,2}^W} + \frac{1}{N} \right), \tag{6.21}$$

with $R_{N,2}^W$ defined in (2.24).

Proof of Lemma 6.6. We have, with $\Theta_{s,i,3}$ defined in (6.3), $\Theta_{s,i,3} \leq e_{s,i,1} + e_{s,i,2} + e_{s,i,3}$ with

$$\begin{aligned}
e_{s,i,1} &:= \sum_{j=1}^N \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y)) \left(F\left(X_N(s,x_j), \eta_s(x_j)\right) - F\left(X_\infty(x_j), \eta_s(x_j)\right) \right) dy, \\
e_{s,i,2} &:= \sum_{j=1}^N \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y)) F\left(X_\infty(x_j), \eta_s(x_j)\right) dy, \\
e_{s,i,3} &:= \sum_{j=1}^N \int_{B_{N,j}} W(x_i,y) \left(F\left(X_N(s,x_j), \eta_s(x_j)\right) - F\left(X_N(s,x_j), \eta_s(y)\right) \right) dy.
\end{aligned}$$

We upper-bound each term. We have as F is Lipschitz continuous

$$\begin{aligned}
e_{s,i,1} &\leq \sum_{j=1}^N \left| F\left(X_N(s,x_j), \eta_s(x_j)\right) - F\left(X_\infty(x_j), \eta_s(x_j)\right) \right| \left| \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y))\, dy \right| \\
&\leq \sum_{j=1}^N \|F\|_L \left| Y_N(s)(x_j) \right| \left| \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y))\, dy \right| \\
&\leq C_F \left(\sum_{j=1}^N \left| \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y))\, dy \right| \right)^{\frac{1}{2}} \\
&\qquad \left(\sum_{j=1}^N \left| Y_N(s)(x_j) \right|^2 \left| \int_{B_{N,j}} (W(x_i,x_j) - W(x_i,y))\, dy \right| \right)^{\frac{1}{2}}
\end{aligned}$$

by discrete Jensen's inequality. We have $N\left|\int_{B_{N,j}}\left(W(x_i,x_j)-W(x_i,y)\right)dy\right|\leq C$ as W is bounded, hence

$$\begin{aligned}e_{s,i,1}&\leq C_{F,W}\left(\sum_{j=1}^{N}\left|\int_{B_{N,j}}\left(W(x_i,x_j)-W(x_i,y)\right)dy\right|\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{j=1}^{N}|Y_N(s)(x_j)|^2\right)^{\frac{1}{2}}\\&\leq C_F\left(\sum_{j=1}^{N}\left|\int_{B_{N,j}}\left(W(x_i,x_j)-W(x_i,y)\right)dy\right|\right)^{\frac{1}{2}}\left\|\tilde{Y}_N(s)\right\|_2.\end{aligned}$$

We have then

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}e_{s,i,1}^2&\leq\frac{C_F}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}\left|\int_{B_{N,j}}\left(W(x_i,x_j)-W(x_i,y)\right)dy\right|\right)\left\|\tilde{Y}_N(s)\right\|_2^2\\&\leq C_F R_{N,1}^W\left\|\tilde{Y}_N(s)\right\|_2^2,\end{aligned}$$

where $R_{N,1}^W$ is defined in (2.24).

For the second term, we have as $x\mapsto\sup_s F(X_\infty(x),\eta_s(x))$ is bounded

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}e_{s,i,2}^2&=\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}\int_{B_{N,j}}\left(W(x_i,x_j)-W(x_i,y)\right)F\left(X_\infty(x_j),\eta_s(x_j)\right)dy\right)^2\\&\leq\frac{C_F}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_{B_{N,j}}|W(x_i,x_j)-W(x_i,y)|^2\,dy\leq C_F R_{N,2}^W,\end{aligned}$$

where $R_{N,2}^W$ is defined in (2.24).

For the third term, as F is Lipschitz continuous

$$\begin{aligned}e_{s,i,3}&\leq\sum_{j=1}^{N}\int_{B_{N,j}}W(x_i,y)\|F\|_L|\eta_s(x_j)-\eta_s(y)|dy\\&\leq\sum_{j=1}^{N}\int_{B_{N,j}}W(x_i,y)\|F\|_L\left(|\eta_s(x_j)-\eta_\infty(x_j)|+|\eta_\infty(x_j)-\eta_\infty(y)|\right)dy\\&\leq C_{F,X,W}\left(\delta_s+\frac{1}{N}\right).\end{aligned}$$

We obtain then with (6.9)

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\Theta_{s,i,2}^2&\leq\frac{3}{N}\sum_{i=1}^{N}\left(e_{s,i,1}^2+e_{s,i,2}^2+e_{s,i,3}^2\right)\\&\leq C_{F,X_\infty,X,W}\left(R_{N,1}^W\left(1+\|Y_N(s)\|_2^4\right)+R_{N,2}^W+\delta_s^2+\frac{1}{N^2}\right).\end{aligned}$$

With (6.7) and Proposition 2.11, $\|\phi_{N,3}(t)\|_2\leq\int_{t_0}^t e^{-(t-s)\gamma}\|\sum_{i=1}^N\Theta_{s,i,3}\mathbf{1}_{B_{N,i}}\|_2 ds$, and as $\|\sum_{i=1}^N\Theta_{s,i,3}\mathbf{1}_{B_{N,i}}\|_2^2=\frac{1}{N}\sum_{i=1}^N\Theta_{s,i,3}^2$, the result follows with

$$G_{N,2}=\sqrt{R_{N,1}^W+R_{N,2}^W}+\frac{1}{N},$$

and Lemma 6.2.

When F is bounded, similarly we show that

$$\frac{1}{N}\sum_{i=1}^{N}\Theta_{s,i,3}^{2}\leq C_{F,X_\infty,\eta,W}\left(R_{N,2}^{W}+\delta_s^2+\frac{1}{N^2}\right),$$

hence the result. □

Lemma 6.7. *Under Hypothesis 2.1, for any $t>t_0\geq 0$,*

$$\|\phi_{N,4}(t)\|_2\leq C_{F,X_\infty,W}\int_{t_0}^{t}e^{-(t-s)\gamma}\left(\|Y_N(s)\|_2^2+\delta_s\right)ds+G_{N,3}, \tag{6.22}$$

where $G_{N,3}$ is explicit in N and tends to 0 as $N\to\infty$. Moreover, if we suppose F bounded, we have

$$\sup_{t\geq 0}\|\phi_{N,4}(t)\|_2\leq\sqrt{S_N^W}, \tag{6.23}$$

where S_N^W is defined in (2.25)*.*

Proof. We have

$$\begin{aligned}\left\|\sum_{i=1}^{N}\Theta_{s,i,4}\mathbf{1}_{B_{N,i}}\right\|_2^2&=\int_I\left(\sum_{i=1}^{N}\Theta_{s,i,4}(x)\mathbf{1}_{B_{N,i}}(x)\right)^2dx\\&=\sum_{i=1}^{N}\int_{B_{N,i}}\left(\int_I\left(W(x_i,y)-W(x,y)\right)F\left(X_N(s,y),\eta_s(y)\right)dy\right)^2dx\\&\leq\sum_{i=1}^{N}\int_{B_{N,i}}\left(\int_I\left(W(x_i,y)-W(x,y)\right)^2dy\right)\left(\int_I\left(F\left(X_N(s,y),\eta_s(y)\right)\right)^2dy\right)dx,\end{aligned}$$

with Cauchy Schwarz's inequality. We can recognize S_N^W defined in (2.25), and we have that, as F is Lipschitz continuous and $y\mapsto F\left(X_\infty(y),\eta_\infty(y)\right)$ is bounded,

$$\begin{aligned}&\int_I F\left(X_N(s,y),\eta_s(y)\right)^2dy\\&\leq\int_I\left(F\left(X_N(s,y),\eta_s(y)\right)-F\left(X_\infty(y),\eta_s(y)\right)\right)^2dy+\int_I F\left(X_\infty(y),\eta_s(y)\right)^2dy\\&\leq\|F\|_L^2\int_I Y_N(s)(y)^2dy+\|F(X_\infty,\eta_\infty)\|_\infty^2\leq C_{F,W}\left(\|Y_N(s)\|_2^2+1\right)\leq C_{F,W}\left(\|Y_N(s)\|_2^4+1\right).\end{aligned}$$

As before, (6.7) and Proposition 2.11 give that

$$\|\phi_{N,4}(t)\|_2\leq\int_{t_0}^{t}e^{-(t-s)\gamma}\|\sum_{i=1}^{N}\Theta_{s,i,4}\mathbf{1}_{B_{N,i}}\|_2ds$$

and $\|\sum_{i=1}^N\Theta_{s,i,4}\mathbf{1}_{B_{N,i}}\|_2^2\leq C_{F,W}S_N^W\left(\|Y_N(s)\|_2^4+1\right)$ hence the result with Lemma 6.2 and $G_{N,3}=C_{F,W}\sqrt{R_{N,3}}$. When F is bounded, we directly have $\|\sum_{i=1}^N\Theta_{s,i,4}\mathbf{1}_{B_{N,i}}\|_2^2\leq S_N^W$ hence (6.23) as $\int_{t_0}^t e^{-(t-s)\gamma}ds\leq\frac{1}{\gamma}$. □

6.3 Proof of Proposition 4.3

Proposition 4.3 is then a direct consequence of (6.6) and (6.7), of the controls given by Lemmas 6.3, 6.4, 6.5, 6.6 and 6.7, with $G_N=G_{N,1}+G_{N,2}+G_{N,3}$, and of Lemma 6.2 to have $G_N\to 0$.

7 About the finite time behavior

In this section, we prove Proposition 2.12.

7.1 Main technical results

In the following, we denote by $\widehat{Y}_N(t) := X_N(t) - X_t$.

Proof of Proposition 2.12. Let $t \leq T$. Recall the definition of $X_N(t)$ in (1.11) and X_t in (1.12). Proceeding exactly as in the proof of Proposition 4.1, and recalling the definition of $M_N(t)$ in (4.7), we have

$$d\widehat{Y}_N(t) = -\alpha \widehat{Y}_N(t)dt + dM_N(t) + \sum_{i,j=1}^{N} \mathbf{1}_{B_{N,i}} \frac{w_{ij}}{N} F\left(X_{N,j}(t-), \eta_t(x_j)\right) dt - T_W F\left(X_t, \eta_t\right) dt$$

$$= -\alpha \widehat{Y}_N(t)dt + dM_N(t) + \sum_{k=1}^{4} \sum_{i=1}^{N} \Theta_{t,i,k} \mathbf{1}_{B_{N,i}} dt + T_W \left(F\left(X_{N,j}(t-), \eta_t(x_j)\right) - F\left(X_t, \eta_t\right)\right) dt$$

with the notations introduced in (6.1)–(6.4). It gives then, as $\widehat{Y}_N(0) = 0$,

$$\widehat{Y}_N(t) = \int_0^t e^{-\alpha(t-s)} \widehat{r}_N(s) ds + \int_0^t e^{-\alpha(t-s)} dM_N(s) =: \widehat{\phi}_N(t) + \widehat{\zeta}_N(t)$$

with

$$\widehat{r}_N(t) = \sum_{k=1}^{4} \sum_{i=1}^{N} \Theta_{t,i,k} \mathbf{1}_{B_{N,i}} + T_W \left(F\left(X_N(t-), \eta_t\right) - F\left(X_t, \eta_t\right)\right).$$

Note that we obtain a similar expression as for Y_N in Proposition 4.1, but with $e^{-\alpha t}$ instead of the semi-group $e^{t\mathcal{L}}$. We use then the two following results, similar to Propositions 4.2 and 4.3.

Proposition 7.1. *Let $T > 0$. Under Hypothesis 2.1, there exists a constant $C = C(T, F, \|\eta\|_\infty) > 0$ such that $\mathbb{P}$-almost surely for N large enough:*

$$\mathbf{E}\left[\sup_{s \leq T} \|\widehat{\zeta}_N(s)\|_2\right] \leq \frac{C}{\sqrt{N\rho_N}}.$$

Proposition 7.2. *Under Hypotheses 2.1 and 2.6, for any $t > 0$,*

$$\|\widehat{\phi}_N(t)\|_2 \leq C\left(\int_0^t e^{-\alpha(t-s)} \|\widehat{Y}_N(s)\|_2 ds + \widehat{G}_N\right), \tag{7.1}$$

where $\widehat{G}_N$ is an explicit quantity to be found in the proof that tends to 0 as $N \to \infty$.

Their proofs are postponed to the following subsection. Hence we obtain

$$\left\|\widehat{Y}_N(t)\right\|_2 \leq C\left(\widehat{G}_N + \left\|\widehat{\zeta}_N(t)\right\|_2 + \int_0^t e^{-\alpha(t-s)} \left\|\widehat{Y}_N(s)\right\|_2 ds\right),$$

which gives with Grönwall lemma

$$\sup_{t \leq T} \left\|\widehat{Y}_N(t)\right\|_2 \leq C\left(\widehat{G}_N + \sup_{t \leq T} \left\|\widehat{\zeta}_N(t)\right\|_2\right).$$

With Proposition 7.1, it leads to

$$\mathbf{E}\left[\sup_{t \leq T} \left\|\widehat{Y}_N(t)\right\|_2\right] \leq C\left(\widehat{G}_N + \frac{1}{\sqrt{N\rho_N}}\right),$$

hence the result (2.16) as (2.6) implies $\frac{1}{\sqrt{N\rho_N}} \to 0$ and $\widehat{G}_N \to 0$. □

7.2 Proofs of Propositions 7.1 and 7.2

Proof of Proposition 7.1. We do as for Proposition 4.2, and apply Îto's formula on

$$\widehat{\zeta}_N(t) = \sum_{j=1}^{N} \int_0^t \int_0^\infty e^{-\alpha(t-s)} \chi_j(s,z) \tilde{\pi}_j(ds,dz).$$

The term $I_0(t)$ in (5.3) becomes $-\alpha \int_0^t \left\| \widehat{\zeta}_N(s) \right\|_2 ds$ which is still non-positive. About $I_1(t)$ and $I_2(t)$, the proof remains the same aside from the fact that we now consider $\widehat{\zeta}_N$ instead of ζ_N. □

To prove 7.2, we introduce an auxilliary quantity as in Lemma 6.1.

Lemma 7.3. *Let $\overline{Y}_N(s)(v) := \widehat{Y}_N(s)\left(\frac{\lceil Nv \rceil}{N}\right)$. Then for any $T \geq 0$*

$$\sup_{0 \leq s \leq T} \|\overline{Y}_N(s) - \widehat{Y}_N(s)\|_2 \xrightarrow[N\to\infty]{} 0. \tag{7.2}$$

Proof. It plays the role of $\tilde{Y}_N(s)$ introduced in Lemma 6.1. Similarly to what has been done before, we have

$$\begin{aligned}
\left\| \widehat{Y}_N(s) - \overline{Y}_N(s) \right\|_2^2 &= \sum_{j=1}^{N} \int_{B_{N,j}} \left(\widehat{Y}_N(s)(y) - \overline{Y}_N(s)(y) \right)^2 dy \\
&= \sum_{j=1}^{N} \int_{B_{N,j}} \left(X_s(x_j) - X_s(y) \right)^2 dy
\end{aligned}$$

which tends to 0 by uniform continuity of X on $[0,T] \times I$. It still holds under the hypotheses of Section 2.3.4 by decomposing the sum on each interval C_k. □

Proof of Proposition 7.2. We divide $\widehat{\phi}$ as in (6.7) and study each contribution. About $\widehat{\phi}_{N,0}(t) := \int_0^t e^{-\alpha(t-s)} T_W \left(F(X_N(s), \eta_s) - F(X_s, \eta_s) \right) ds$, we have

$$\begin{aligned}
\|T_W \left(F(X_N(s), \eta_s) - F(X_s, \eta_s) \right)\|_2^2 &\leq C_{W,F} \left(\int_I \|F\|_L \left| X_N(s)(y) - X_s(y) \right| dy \right)^2 \\
&\leq C_{W,F} \left\| \widehat{Y}_N(s) \right\|_2^2,
\end{aligned}$$

which gives

$$\left\| \widehat{\phi}_{N,0}(t) \right\|_2 \leq C_{W,F} \int_0^t e^{-\alpha(t-s)} \left\| \widehat{Y}_N(s) \right\|_2 ds.$$

About $\widehat{\phi}_{N,1}(t) := \int_0^t e^{-\alpha(t-s)} \sum_{i=1}^N \frac{\Theta_{s,i,1}}{N} \mathbf{1}_{B_{N,i}} ds$, we do as in Lemma 6.4. Instead of inserting the terms $F(X_\infty(x_j), \eta_\infty(x_j))$ in (6.17) we insert the terms $F(X_s(x_j), \eta_s(x_j))$, that is

$$\begin{aligned}
\gamma_N(s) \leq &\sum_{i,j=1}^{N} \frac{1}{N} \kappa_{N,i} \overline{\xi_{ij}} \left(F(X_{N,j}(s-), \eta_s(x_j)) - F(X_s(x_j), \eta_s(x_j)) \right) \mathbf{1}_{B_{N,i}} \\
&+ \sum_{i,j=1}^{N} \frac{1}{N} \kappa_{N,i} \overline{\xi_{ij}} F(X_s(x_j), \eta_s(x_j)) \mathbf{1}_{B_{N,i}} =: \widehat{\gamma}_{N,1}(s) + \widehat{\gamma}_{N,2}(s).
\end{aligned}$$

The treatment of $\widehat{\gamma}_{N,1}$ is similar of $\gamma_{N,1}$: we make $\overline{Y}_N(s-)$ appear instead of $\tilde{Y}_N$ and obtain $\|\widehat{\gamma}_{N,1}(s-)\|_2^2 \leq C_F \left(\left\| \widehat{Y}_N(s) \right\|_2^2 + 1 \right) \left(\frac{S_N^{\max}}{\rho_N^2} + \frac{1}{N\rho_N^2} \right)$ with (7.2). About $\widehat{\gamma}_{N,2}$, we do as

$\gamma_{N,2}$ as $\sup_{t\in[0,T],x\in I} F(X_t(x),\eta_t(x)) < \infty$ and obtain that $\mathbb{P}$-almost surely if N is large enough, $\|\widehat{\gamma}_{N,2}\|_2^2 \leq C\left(\frac{1}{N\rho_N^2} + \frac{1}{N^{1-2\tau}\rho_N^4}\right)$. We have then that, $\mathbb{P}$-almost surely if N is large enough,

$$\left\|\widehat{\phi}_{N,1}(t)\right\|_2 \leq C_F \int_{t_0}^t e^{-\alpha(t-s)} \left\|\widehat{Y}_N(s)\right\|_2 ds + G_{N,1},$$

where $G_{N,1} \to 0$.

About $\widehat{\phi}_{N,2}(t)$, we proceed as Lemma 6.5 to show that $\|\widehat{\phi}_{N,2}(t)\|_2 \leq \frac{C_F}{N\rho_N}$. About $\widehat{\phi}_{N,k}(t) := \int_0^t e^{-\alpha(t-s)} \sum_{i=1}^N \frac{\Theta_{s,i,k}}{N} \mathbf{1}_{B_{N,i}} ds$ for $k \in \{3,4\}$, we proceed similarly, doing as in Lemmas 6.6 and 6.7 but instead of inserting the terms $F(X_\infty(x_j),\eta_\infty(x_j))$ we insert the terms $F(X_s(x_j),\eta_s(x_j))$: then there is no δ_s terms. We obtain then

$$\|\widehat{\phi}_{N,3}(t)\|_2 \leq C \int_{t_0}^t e^{-\alpha(t-s)} \left\|\widehat{Y}_N(s)\right\|_2 ds + G_{N,2},$$

and

$$\|\widehat{\phi}_{N,4}(t)\|_2 \leq C \int_{t_0}^t e^{-\alpha(t-s)} \|Y_N(s)\|_2 ds + G_{N,3},$$

where both $G_{N,2}$ and $G_{N,3}$ tends to 0. Note that we can obtain better bounds when F is bounded. By putting all the terms $\widehat{\phi}_{N,k}$ together, we get (7.1). □

A Auxiliary results

A.1 Concentration results

Theorem A.1 (Grothendieck's inequality as in [20]). *Let $\{a_{ij}\}_{i,j=1,\dots,n}$ be a $n\times n$ real matrix such that for all s_i, $t_j \in \{-1,1\}$*

$$\sum_{i,j=1}^n a_{ij} s_i t_j \leq 1.$$

Then, there exists a constant $K_R > 0$, such that for every Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ and for all S_i and T_j in the unit ball of H

$$\sum_{i,j=1}^n a_{ij} \langle S_i, T_j\rangle_H \leq K_R.$$

Theorem A.2 (Azuma–Hoeffding inequality). *Let (M_n) be a martingale with $M_0 = 0$. Assume that for all $1 \leq k \leq n$, $|\Delta M_k| \leq c_k$ a.s. for some constants (c_k). Then for all $x \geq 0$*

$$\mathbb{P}\left(|M_n| \geq x\right) \leq 2\exp\left(-\frac{x^2}{2\sum_{k=1}^n c_k^2}\right). \tag{A.1}$$

Theorem A.3 (Upper tail estimate for iid ensembles, Corollary 2.3.5 of [44]). *Suppose that $M = (m_{ij})_{1\leq i,j\leq n}$, where n is a (large) integer and the m_{ij} are independent centered random variables uniformly bounded in magnitude by 1. Then there exist absolute constants $C, c > 0$ such that*

$$\mathbb{P}\left(\|M\|_{op} > x\sqrt{n}\right) \leq C\exp\left(-cxn\right)$$

for any $x \geq C$.

Lemma A.4. *Under Hypothesis 2.6, we have $\mathbb{P}$-almost surely if N is large enough:*

$$\sup_{1\leq j\leq N}\left(\sum_{i=1}^N \frac{\xi_{ij}^{(N)}}{N\rho_N}\right) \leq 2, \quad \sup_{1\leq i\leq N}\left(\sum_{j=1}^N \frac{\xi_{ij}^{(N)}}{N\rho_N}\right) \leq 2. \tag{A.2}$$

Proof. It is a direct consequence of Corollary 8.2 of a previous work [1], in the case $w_N = \rho_N$, $\kappa_N = \frac{1}{\rho_N}$, $W_N(x_i, x_j) = \rho_N W(x_i, x_j)$ with W bounded. □

Lemma A.5. *Let $N \geq 1$, for $j \neq j'$ in $[\![1, N]\!]$, let $S_{jj'} := \frac{1}{N}\sum_{i=1}^N \overline{\xi_{ij}}\ \overline{\xi_{ij'}}$ with ξ defined in Definition 1.1, and $S_N^{max} := \sup_{1\leq j\neq j'\leq N} |S_{jj'}|$. Then, under Hypothesis 2.6, $\mathbb{P}$-a.s.*

$$\limsup_{N\to\infty} S_N^{max} \leq N^{\tau - \frac{1}{2}} \tag{A.3}$$

where $\tau \in (0, \frac{1}{2})$ comes from Hypothesis 2.6.

Proof. When j and j' are fixed and $j \neq j'$, $\left(X_i := \overline{\xi_{ij}}\ \overline{\xi_{ij'}}\right)_{1\leq i\leq N}$ is a family of independent random variables with $|X_i| \leq 1$, $\mathbf{E}[X_i] = 0$ and $\mathbf{E}[X_i^2] \leq 1$. Bernstein's inequality gives then for any $t > 0$

$$\mathbf{P}\left(\left|\sum_{i=1}^N \overline{\xi_{ij}}\ \overline{\xi_{ij'}}\right| > t\right) \leq 2\exp\left(-\frac{1}{2}\frac{t^2}{N + \frac{t}{3}}\right)$$

hence for the choice $t = N^{\frac{1}{2}+\tau}$ with $\tau \in (0, \frac{1}{2})$,

$$\mathbf{P}\left(\left|\sum_{i=1}^N \overline{\xi_{ij}}\ \overline{\xi_{ij'}}\right| > N^{\frac{1}{2}+\tau}\right) \leq 2\exp\left(-\frac{1}{2}\frac{N^{2\tau}}{1 + \frac{1}{3}N^{-\frac{1}{2}+\tau}}\right) \leq 2\exp\left(-\frac{1}{4}N^{2\tau}\right)$$

as $1 + \frac{1}{3}N^{-\frac{1}{2}+\tau} \leq 2$. With an union bound

$$\mathbf{P}\left(\sup_{j\neq j'} |S_{jj'}| > \frac{1}{N^{\frac{1}{2}-\tau}}\right) \leq 2N^2 \exp\left(-\frac{1}{4}N^{2\tau}\right).$$

We apply then Borel Cantelli's lemma and obtain (A.3). □

Lemma A.6. *Fix $N > 1$ and $(Y_l)_{l=1,\dots,n}$ real valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that there exists $\nu > 0$ such that, almost surely, for all $l = 1, \dots, n-1$, $Y_l \leq 1$, $\mathbb{E}\left[Y_{l+1}\,|Y_l\right] = 0$ and $\mathbb{E}\left[Y_{l+1}^2\,|Y_l\right] \leq \nu$. Then*

$$\mathbb{P}\left(n^{-1}(Y_1 + \cdots + Y_n) \geq x\right) \leq \exp\left(-n\frac{x^2}{2\nu}B\left(\frac{x}{\nu}\right)\right)$$

for all $x \geq 0$, where

$$B(u) := u^{-2}\left((1+u)\log(1+u) - u\right). \tag{A.4}$$

Proof. A direct application of [25, Corollary 2.4.7] gives that

$$\mathbb{P}\left(n^{-1}(Y_1 + \cdots + Y_n) \geq x\right) \leq \exp\left(-nH\left(\frac{x+v}{1+v}|\frac{v}{1+v}\right)\right),$$

where $H(p|q) := p\log(p/q) + (1-p)\log((1-p)/(1-q))$ for $p, q \in [0,1]$. Then, the inequality $H\left(\frac{x+v}{1+v}|\frac{v}{1+v}\right) \geq \frac{x^2}{2v}B\left(\frac{x}{v}\right)$ (see [25, Exercise 2.4.21]) gives the result. □

Corollary A.7. *Let $(Z_{ij})_{i,j}$ be a family of independent Bernoulli variables, with $\mathbb{E}[Z_{ij}] = m_{ij}$. Let $(\beta ij)_{ij}$ be a sequence such that for any i, j, $\beta_{i,j} \in (0,1]$. Then, for all $x \geq 0$*

$$\mathbb{P}\left(\frac{1}{N^2}\sum_{i,j=1}^N \beta_{ij}\left((Z_{ij} - m_{ij})^2 - \mathbb{E}\left(Z_{ij} - m_{ij}\right)^2\right) \geq x\right) \leq \exp\left(-\frac{N^2x^2}{2}B(x)\right).$$

Proof. Fix a bijection $\phi_N : [\![1, N^2]\!] \to [\![1, N]\!] \times [\![1, N]\!]$. For any $k \in [\![1, N^2]\!]$ and $(i, j) = \phi_N(k)$, let $R_k = \beta_{ij} \left((Z_{ij} - m_{ij})^2 - \mathbb{E}\,(Z_{ij} - m_{ij})^2 \right)$. As the $(m_{ij})_{i,j}$ are independent, the family of randon variables $(R_k)_{1\le k\le N^2}$ is also independent. As $R_k \le 1$ a.s., $\mathbb{E}\,[R_{k+1}|R_k] = 0$ and $\mathbb{E}\left[R_{k+1}^2|R_k\right] \le 1$, Lemma A.6 implies that for any $x \ge 0$,

$$\mathbb{P}\left(\frac{1}{N^2} \sum_{k=1}^{N^2} R_k \ge x \right) \le \exp\left(-\frac{N^2 x^2}{2} B(x) \right)$$

where B is defined in (A.4). □

A.2 Other technical results

Lemma A.8. *Let K be a kernel from $I^2 \to \mathbb{R}_+$ such that $\sup_{x\in I} \int_I K(x, y)^2 dy < \infty$. Let $T_K : g \mapsto T_K g := \left(x \to \int_I K(x, y)dy\right)$ be the operator associated to K, that can be defined from $L^2(I) \to L^2(I)$ and from $L^\infty(I) \to L^\infty(I)$. We assume that $T_K^2 : L^2(I) \to L^2(I)$ is compact. Then*

$$r_2(T_K) = r_\infty(T_K).$$

Proof. First note that for both $r = r_2$ and $r = r_\infty$, we have, for all $p \ge 1$, $r(T_K^p)^{\frac{1}{p}} = \left(\lim_{n\to\infty} \|T_K^{pn}\|^{\frac{1}{n}}\right)^{\frac{1}{p}} = \lim_{n\to\infty} \|T_K^{pn}\|^{\frac{1}{pn}} = r(T_K)$, so that $r(T_K^p) = r(T_K)^p$. Hence $r_2(T_K^2) = r_\infty(T_K^2)$ gives $r_2(T_K) = r_\infty(T_K)$.

Denote by $\sigma_\infty(T_K^2)$ and $\sigma_2(T_K^2)$ the corresponding spectrum of T_K^2 (in $L^\infty(I)$ and $L^2(I)$ respectively). Let us prove that $r_2(T_K^2) = r_\infty(T_K^2)$ by proving $\sigma_\infty(T_K^2) = \sigma_2(T_K^2)$. To do so, first note that $T_K^2 : L^\infty(I) \to L^\infty(I)$ is compact: consider $(f_n)_n$ a bounded sequence of $L^\infty(I)$. It is then also bounded in $L^2(I)$, and as $T_K : L^2(I) \to L^2(I)$ is compact, there exists a subsequence $\left(f_{\phi(n)}\right)$ such that $T_K f_{\phi(n)}$ converges in $L^2(I)$ to a certain g. Then for any $x \in I$,

$$|T_K^2 f_{\phi(n)} - T_K g|(x) \le \int_I K(x, y) \left|T_K f_{\phi(n)}(y) - g(y)\right| dy \le C_K \|T_K f_{\phi(n)} - g\|_2 \xrightarrow[n\to\infty]{} 0,$$

thus $T_K^2 : L^\infty(I) \to L^\infty(I)$ is compact. Now we prove that $\sigma_\infty(T_K^2) = \sigma_2(T_K^2)$: let $\mu \in \sigma_2(T_K^2) \setminus \{0\}$, there exists $g \in L^2(I)$ such that $\mu g = T_K^2 g$. As

$$\left|T_K^2 g(x)\right| = \left| \int_I K(x, y) \int_I K(y, z) g(z)\ \nu(dz)\nu(dy) \right| \le C_K \|g\|_2 < \infty,$$

$g = \frac{1}{\mu} T_K^2 g \in L^\infty(I)$ and $\mu \in \sigma_\infty(T_K^2)$. Conversely, let $\mu \in \sigma_\infty(T_K^2) \setminus \{0\}$, there exists $g \in L^\infty(I)$ such that $\mu g = T_K^2 g$. As $L^\infty(I) \subset L^2(I)$, $\mu \in \sigma_2(T_K^2)$. Hence $r_2(T_K^2) = r_\infty(T_K^2)$ and (2.2) follows. □

Lemma A.9 (Quadratic Grönwall's lemma)**.** *Let f be a non-negative function piecewise continuous with finite number of distinct jumps of size inferior to θ on $[t_0, T]$, let g be a non-negative continuous function and $h \in L^1([t_0, T])$. For any $t \in [t_0, T]$, assume f satisfies*

$$f(t) \le f(t_0) + g(t) + \int_{t_0}^t h(t - s) f(s)^2 ds.$$

Then, for $\delta < \frac{1}{9\|h\|_1}$, if $\theta \le \frac{\delta}{2}$ and if $\sup_{t\in[t_0,T]} g(t) \le \delta$, we have

$$\sup_{t\in[t_0,T]} f(t) \le f(t_0) + 3\delta.$$

Proof. Let $A = \{t \in [t_0, T], f(t) > f(t_0) + 3\delta\}$, suppose $A \neq \emptyset$. Let $t^* = \inf\{t \in [t_0, T], f(t) > f(t_0) + 3\delta\}$. If there is no jump at t_0, by the initial conditions $t^* > t_0$, and if there is a jump, $f(t_0^+) \leq f(t_0) + \frac{\delta}{2}$ hence we also have $t^* > t_0$. Moreover, for all $t \in [t_0, t^*)$, $f(t) \leq f(t_0) + \delta + 9\delta^2 \int_{t_0}^{t} h(t-s)ds \leq f(t_0) + 2\delta$. If there is a jump at t^*, it is of amplitude $\theta \leq \frac{\delta}{2}$ hence $f(t^*) \leq f(t_0) + \frac{5\delta}{2} < f(t_0) + 3\delta$ which is a contradiction. If there is no jump at t^*, by local continuity we have $f(t^*) \leq f(t_0) + \delta + 9\delta^2 \int_{t_0}^{t^*} h(t-s)ds \leq f(t_0) + 2\delta$ which is also a contradiction. We conclude then that $\sup_{t \in [t_0, T]} f(t) \leq f(t_0) + 3\delta$. □

References

[1] Z. Agathe-Nerine. Multivariate Hawkes processes on inhomogeneous random graphs. *Stochastic Process. Appl.*, 152:86–148, 2022. MR4450462

[2] S.-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87, 1977. MR0681526

[3] E. Bayraktar, S. Chakraborty, and R. Wu. Graphon mean field systems. *Ann. Appl. Probab.*, 33(5):3587–3619, 2023. MR4514837

[4] L. Bertini, G. Giacomin, and C. Poquet. Synchronization and random long time dynamics for mean-field plane rotators. *Probab. Theory Related Fields*, 160(3-4):593–653, 2014. MR3278917

[5] G. Bet, F. Coppini, and F. R. Nardi. Weakly interacting oscillators on dense random graphs, 2020. arXiv:2006.07670.

[6] F. Bolley, I. Gentil, and A. Guillin. Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.*, 208(2):429–445, 2013. MR3035983

[7] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi. Limits of randomly grown graph sequences. *European Journal of Combinatorics*, 32(7):985–999, Oct. 2011. MR2825531

[8] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.*, 46(1):337–396, 2018. MR3758733

[9] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008. MR2455626

[10] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)*, 176(1):151–219, 2012. MR2925382

[11] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17(6):2112–2127, Mar. 1997.

[12] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996. MR1411506

[13] P. C. Bressloff. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*, 45(3):033001, Dec. 2011. MR2871421

[14] J. Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Process. Appl.*, 127(12):3870–3912, 2017. MR3718099

[15] J. Chevallier, A. Duarte, E. Löcherbach, and G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. *Stochastic Processes and their Applications*, 129(1):1–27, 2019. MR3906989

[16] J. Chevallier, A. Melnykova, and I. Tubikanec. Diffusion approximation of multi-class Hawkes processes: theoretical and numerical analysis. *Adv. in Appl. Probab.*, 53(3):716–756, 2021. MR4322401

[17] J. Chevallier and G. Ost. Fluctuations for spatially extended Hawkes processes. *Stochastic Process. Appl.*, 130(9):5510–5542, 2020. MR4127337

[18] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, Nov. 2002. MR1955514

[19] L. Colombani and P. L. Bris. Chaos propagation in mean field networks of FitzHugh-Nagumo neurons, 2022. arXiv:2206.13291.

[20] F. Coppini. Long time dynamics for interacting oscillators on graphs. *Ann. Appl. Probab.*, 32(1):360–391, 2022. MR4386530

[21] F. Coppini, H. Dietert, and G. Giacomin. A Law of Large numbers and Large Deviations for interacting diffusions on Erdős–Rényi graphs. *Stochastics and Dynamics*, 20(02):2050010, July 2019. MR4080158

[22] F. Coppini, E. Luçon, and C. Poquet. Central limit theorems for global and local empirical measures of diffusions on Erdős–Rényi graphs, 2022. arXiv:2206.06655.

[23] S. Delattre, N. Fournier, and M. Hoffmann. Hawkes processes on large networks. *Ann. Appl. Probab.*, 26(1):216–261, 02 2016. MR3449317

[24] S. Delattre, G. Giacomin, and E. Luçon. A note on dynamical models on random graphs and Fokker–Planck equations. *Journal of Statistical Physics*, 165(4):785–798, Nov. 2016. MR3568168

[25] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998. MR1619036

[26] S. Ditlevsen and E. Löcherbach. Multi-class oscillating systems of interacting neurons. *Stochastic Process. Appl.*, 127(6):1840–1869, 2017. MR3646433

[27] X. Erny, E. Löcherbach, and D. Loukianova. Mean field limits for interacting Hawkes processes in a diffusive regime. *Bernoulli*, 28(1):125–149, 2022. MR4337700

[28] S. N. Ethier and T. G. Kurtz, editors. *Markov Processes*. John Wiley & Sons, Inc., Mar. 1986. MR0838085

[29] G. Giacomin, K. Pakdaman, X. Pellegrin, and C. Poquet. Transitions in active rotator systems: invariant hyperbolic manifold approach. *SIAM Journal on Mathematical Analysis*, 44(6):4165–4194, 2012. MR3023444

[30] G. Giacomin and C. Poquet. Noise, interaction, nonlinear dynamics and the origin of rhythmic behaviors. *Brazilian Journal of Probability and Statistics*, 29(2):460–493, 2015. MR3336876

[31] A. G. Hawkes. Point spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. MR0278410

[32] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, Sept. 1974. MR0378093

[33] S. Heesen and W. Stannat. Fluctuation limits for mean-field interacting nonlinear Hawkes processes. *Stochastic Process. Appl.*, 139:280–297, 2021. MR4273097

[34] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983. MR0718088

[35] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, Nov. 2006. MR2274085

[36] E. Luçon. Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *Stochastic Process. Appl.*, 130(11):6783–6842, 2020. MR4158803

[37] E. Luçon and C. Poquet. Long time dynamics and disorder-induced traveling waves in the stochastic Kuramoto model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(3):1196–1240, 2017. MR3689966

[38] V. Mountcastle. The columnar organization of the neocortex. *Brain*, 120(4):701–722, Apr. 1997.

[39] I. Omelchenko, B. Riemenschneider, P. Hövel, Y. Maistrenko, and E. Schöll. Transition from spatial coherence to incoherence in coupled chaotic systems. *Physical Review E*, 85(2), Feb. 2012.

[40] S. Ouadah, S. Robin, and P. Latouche. Degree-based goodness-of-fit tests for heterogeneous random graph models: Independent and exchangeable cases. *Scandinavian Journal of Statistics*, 47(1):156–181, Oct. 2019. MR4075233

[41] A. Pazy. *Semi-groups of linear operators and applications to partial differential equations*. University of Maryland, Department of Mathematics, College Park, Md., 1974. Department of Mathematics, University of Maryland, Lecture Note, No. 10. MR0512912

[42] P. Pfaffelhuber, S. Rotter, and J. Stiefel. Mean-field limits for non-linear Hawkes processes with excitation and inhibition. *Stochastic Process. Appl.*, 153:57–78, 2022. MR4463083

[43] A. Prodhomme. Strong Gaussian approximation of metastable density-dependent Markov chains on large time scales. *Stochastic Process. Appl.*, 160:218–264, 2023. MR4564537

[44] T. Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. MR2906465

[45] J. Touboul. Propagation of chaos in neural fields. *The Annals of Applied Probability*, 24(3):1298–1328, June 2014. MR3199987

[46] H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24, Jan. 1972.

[47] J. Zhu, Z. a. Brzeźniak, and E. Hausenblas. Maximal inequalities for stochastic convolutions driven by compensated Poisson random measures in Banach spaces. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(2):937–956, 2017. MR3634281

Acknowledgments. This is a part of my PhD thesis. I would like to thank my PhD supervisors Eric Luçon and Ellen Saada for introducing this subject, for their useful advices and for their encouragement. This research has been conducted within the FP2M federation (CNRS FR 2036), and is supported by ANR-19-CE40-0024 (CHAllenges in MAthematical NEuroscience) and ANR-19-CE40-0023 (Project PERISTOCH). I would like to warmly thank two anonymous Referees for their useful and precise comments, which considerably helped to improve the paper.