



HAL
open science

Weakly interacting diffusions on graphs

Fabio Coppini

► **To cite this version:**

Fabio Coppini. Weakly interacting diffusions on graphs. Probability [math.PR]. Université de Paris, 2020. English. NNT: . tel-02970875

HAL Id: tel-02970875

<https://theses.hal.science/tel-02970875v1>

Submitted on 19 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Paris
École Doctorale de Sciences Mathématiques de Paris Centre – ED 386
Laboratoire de Probabilités, Statistique et Modélisation

Thèse de doctorat

Discipline : Mathématiques Appliquées

Weakly interacting diffusions on graphs

Fabio Coppini

Sous la direction de Giambattista Giacomin

Rapporteurs :

François Delarue Université de Nice Sophia-Antipolis, France
Roberto Imbuzeiro Oliveira Instituto de Matemática Pura e Aplicada, Brésil

Présentée et Soutenue à Paris le 14 octobre 2020 devant le jury composé de

François Delarue	Université de Nice Sophia-Antipolis	Rapporteur
Roberto Imbuzeiro Oliveira	Instituto de Matemática Pura e Aplicada	Rapporteur
Sylvie Méléard	École Polytechnique	Examinatrice
Mireille Bossy	Inria Sophia-Antipolis	Examinatrice
Lorenzo Zambotti	Sorbonne Université	Examineur
Giambattista Giacomin	Université de Paris	Directeur

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665850.

*alla mia famiglia
presente, passata e futura*

Remerciements

Ces humbles remerciements ne pourront pas traduire la reconnaissance et le respect que j'ai eue pour mon directeur de thèse : merci Giambattista pour tout ce que tu m'as donné et pour ta patience durant mes périodes d'hésitation.

C'est avec joie et trépidation que j'ai lu les rapports que François et Roberto ont accepté d'écrire. Merci à vous d'avoir accordé à mes débuts et à mes efforts un peu de votre temps et de votre estime. Merci aux membres du jury d'avoir accepté mon invitation.

Grand est le plaisir d'avoir travaillé avec Francesca, Helge, Gianmarco et Florian. Vous avez enrichi mon rapport à la recherche : je vous dois beaucoup.

Les rencontres et le temps passé au LPSM vivront en moi ; beaucoup de personnes ont été importantes et elles m'excuseront s'il n'y a pas assez de place pour toutes les citer : je tiens tout de même à remercier le directeur du labo Lorenzo ainsi que les secrétaires Nathalie et Valérie pour leur travail et leur aide.

Merci à la ville de Paris et à l'Union Européenne pour avoir rendu possible cette thèse et m'avoir fait rencontrer des personnes exceptionnelles.

Abstract

In the last twenty years, the modeling of complex systems has become a relevant domain of study, not only in applied sciences, but also among mathematicians. The recent improvements in the understanding of interacting particle systems, as well as the new insights coming from graph theory, allow to mathematically tackle new exciting problems in the challenging world of complex phenomena.

This thesis addresses a rather general class of interacting particle systems defined on graph sequences. Notably, it focuses on weakly interacting particles described by differential equations, both deterministic and stochastic, where an extra structure encoding the connections among the particles is present. The mean-field hypothesis under which each particle is connected to all the others and in exactly the same way, is relaxed to a much more general assumption: the connections between the particles are supposed to be encoded by a general network, instead of the trivial complete graph of the mean-field case, meaning that a particle is interacting with another in a way that is proportional to the weight of the edge connecting the twos in the underlying graph.

Several aspects for this class of models appear to be new in current research and demand new tools and techniques, but also new insights to unveil how the complexity behind the underlying network affects the particle dynamics.

The present manuscript poses the focus on three main aspects: the relationship with the mean-field behavior, i.e., on which graph sequences the system behavior is suitably described by the classical mean-field limit; the extensions to inhomogeneous graph sequences and consequent inhomogeneous behaviors; finally, a first study on the long time dynamics for a particular model of interacting diffusions on graphs.

Keywords: interacting particle systems, interacting diffusions, McKean-Vlasov equation, interacting oscillators, Kuramoto model, random graphs, graph limits, random graphons, stochastic differential equations, long-time dynamics, mean-field models, pseudo-random graphs, exchangeability

Résumé

Les systèmes complexes sont devenus un sujet très étudié, non seulement dans les sciences appliquées, mais aussi en mathématiques. Récemment, de nouveaux résultats rigoureux ont été établis à propos de l'étude des systèmes de particules et des graphes aléatoires. Ils ont ainsi ouvert de nouvelles perspectives sur la modélisation des phénomènes complexes en mathématiques.

Cette thèse porte sur une classe assez large de systèmes de particules qui interagissent entre elles, définis sur des séquences de graphes. En particulier, on se focalise sur des systèmes d'équations différentielles stochastiques et déterministes, dans lesquels une extra-structure code les connexions entre les différentes particules. La condition classique de champ moyen qui impose que toute unité soit connectée aux autres de la même façon, est relaxée au profit d'une hypothèse bien plus générale : les connexions dans le système sont définies à l'aide d'un réseau complexe, au lieu d'un graphe complet. En d'autres termes, une particule interagit avec une autre d'une façon proportionnelle au poids de la connexion dans le graphe sous-jacent.

Plusieurs aspects dans cette classe de modèles se présentent comme nouveaux dans la recherche actuelle et nécessitent des techniques et outils plus avancés. Des nouvelles idées sont requises pour dévoiler la complexité intrinsèque de ces systèmes.

Ce travail se focalise sur trois aspects fondamentaux : la relation avec le comportement de champ moyen, l'extension aux séquences des graphes inhomogènes et, pour conclure, l'étude de la dynamique à temps longs d'un modèle de synchronisation défini sur des graphes.

Mots clés : système de particules interagissant, diffusions avec interaction, équation McKean-Vlasov, oscillateurs interagissant, modèle de Kuramoto, graphes aléatoires, limites de graphes, graphons aléatoires, équations stochastiques différentielles, dynamique à temps longs, modèles à champ moyen, pseudo-random graphes, échangeabilité

Contents

Remerciements	iv
Abstract	vi
Résumé	vii
Chapter 1. Introduction	1
1. Weakly interacting particle systems	1
2. Graph sequences and graphons	7
3. Interacting particles on graphs	13
4. Motivations	16
5. Overview of the results and general organization	22
Chapter 2. Interacting diffusions on Erdős-Rényi random graphs	23
1. Introduction	23
2. Main results	27
3. Proof: The Law of Large Numbers	30
4. Proof: Large Deviation Principle	34
Chapter 3. Weakly interacting oscillators on dense random graphs	39
1. Introduction, organization and set-up	39
2. The models and main results	42
3. The non-linear process	49
4. Proof of Theorem 2.3	57
3.A. Graph convergence and random graphons	60
Chapter 4. Long time dynamics for interacting oscillators	63
1. Introduction	63
2. Main results	68
3. Longtime dynamics close to M	74
4. Longtime behavior around $1/2\pi$	84
5. Finite time behavior	88
4.A. Graphs	90
4.B. H_{-1} and Semigroups	92
Chapter 5. A Law of Large Numbers via a mild formulation	97
1. Introduction	97

2. Main result	100
3. Proofs	104
5.A. Hilbert spaces and semigroups	118
5.B. Rough integration associated to semigroup functionals	121
Chapter 6. Perspectives	127
1. From dense graph sequences to the sparse regime	127
2. Irregular graphons and random graphons	127
3. Unbounded graphons	128
4. Long-time dynamics and uniform propagation of chaos	129
Bibliography	131

CHAPTER 1

Introduction

This thesis stems from the interplay between two established theories in mathematics: weakly interacting particle systems and graph sequences. In the first part of the introduction, we present an overview on some known result for interacting particle systems, focusing on Law of Large Numbers and Propagation of Chaos properties. We also give the key notions and tools for the study of (random) graph sequences, mainly addressing the dense regime. Once these ingredients have been introduced, we are able to define a weakly interacting particle system on a graph sequence, the main topic of the following chapters.

The second part of the introduction is devoted to the motivations behind the results coming along with this manuscript and a presentation of the underlying works. The three key aspects that will be addressed and discussed throughout the thesis are: the relationship with the classical mean-field limit, extensions to inhomogeneous graph sequences and, finally, the long-time behavior of a well-known model of interacting oscillators on graphs.

A brief overview on the works within the manuscript, as well as the organization of the chapters, is given at the end of the introduction.

1. Weakly interacting particle systems

Consider a system of n interacting particles on \mathbb{R} , evolving on a finite time interval $[0, T]$, for some arbitrary $T > 0$. The dynamics of the particles $\{\bar{x}_t^{i,n}\}_{i=1,\dots,n}$ is described by a system of n coupled stochastic differential equations of the following form:

$$\begin{cases} d\bar{x}_t^{i,n} = F(\bar{x}_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \Gamma(\bar{x}_t^{i,n}, \bar{x}_t^{j,n})dt + \sigma(\bar{x}_t^{i,n})dB_t^i, \\ \bar{x}_0^{i,n} = x_0^i, \end{cases} \quad (1.1)$$

for $i = 1, \dots, n$ and $t \in [0, T]$. The functions F , Γ and σ are taken to be bounded and uniformly Lipschitz; the sequence $\{x_0^i\}_{i \in \mathbb{N}} \subset \mathbb{R}$ denotes the initial conditions which can be random or deterministic.

In most instances, we suppose the diffusion coefficient σ to be constantly equal to one, i.e., the particles $\{\bar{x}_t^{i,n}\}_{i=1,\dots,n}$ are driven by a sequence of independent and identically distributed (IID) Brownian motions, denoted by $\{B^i\}_{i \in \mathbb{N}}$. Since a few results presented in the sequel hold true also in the deterministic case, i.e., when $\sigma \equiv 0$, or for a non-constant strictly positive diffusion, we work with the general case.

The Brownian motions are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbf{P})$ and are adapted to the filtration \mathcal{F} , which satisfies the usual conditions. Whenever the initial conditions are random, they are chosen to be independent of the Brownian motions.

Under the previous hypothesis, existence and uniqueness of a strong continuous solution $\{\bar{x}_t^{i,n}, t \in [0, T]\}_{i=1, \dots, n}$ to equation (1.1) are a classical result [97]. The space of continuous functions on $[0, T]$ with values in \mathbb{R} is denoted by $\mathcal{C}([0, T], \mathbb{R})$, the solution $\{\bar{x}^{i,n}\}_{i=1, \dots, n}$ to (1.1) is an element of $\mathcal{C}([0, T], \mathbb{R}^n)$.

The statistical information present in (1.1), is encoded in the empirical measure associated to $\{\bar{x}^{i,n}\}_{i=1, \dots, n}$. This last one is a random element with values in the space of probability measures on $\mathcal{C}([0, T], \mathbb{R})$, which we denote by $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$. The empirical measure $\bar{\mu}^n$ is defined as

$$\bar{\mu}^n := \frac{1}{n} \sum_{j=1}^n \delta_{\bar{x}^{j,n}}.$$

In the sequel, we will often consider its time-marginal projection $\bar{\mu}^n = \{\bar{\mu}_t^n\}_{t \in [0, T]}$, which is a continuous trajectory with values on the probability measures $\mathcal{P}(\mathbb{R})$, i.e., an element of $\mathcal{C}([0, T], \mathcal{P}(\mathbb{R}))$. This last object is equivalently defined as

$$\bar{\mu}_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{\bar{x}_t^{j,n}} \quad (1.2)$$

for $t \in [0, T]$. The precise relation between $\bar{\mu}^n$ and $\bar{\mu}^n$ is discussed in the next subsections.

1.1. The classical mean-field limit. One of the key aspects that has made the theory of weakly interacting particle systems so studied, is the existence of a rather understood limit description for $\bar{\mu}^n$, as the number of particle grows. Notably, when n tends to infinity it is possible to characterize the limit of $\bar{\mu}^n$, in a suitable topological space, as the unique weak solution to a certain partial differential equation (PDE), usually called McKean-Vlasov or Fokker-Planck equation, see equation (1.4).

Although proving such convergence may be technical under several aspects – we refer to Chapter 5 for an exhaustive point of view on this topic – the formal derivation of the PDE appears to be very simple, provided some confidence in the fundamental tools of stochastic calculus. Its derivation goes as follows.

Let f be a smooth function on \mathbb{R} . The application of Itô's formula to $f(\bar{x}_t^{i,n})$ for some fixed $i = 1, \dots, n$ yields

$$f(\bar{x}_t^{i,n}) = f(x_0^i) + \int_0^t \partial_x f(\bar{x}_s^{i,n}) d\bar{x}_s^{i,n} + \frac{1}{2} \int_0^t \partial_x^2 f(\bar{x}_s^{i,n}) d\langle \bar{x}^{i,n} \rangle_s.$$

Using the explicit expression of $d\bar{x}_s^{i,n}$ in equation (1.1), as well as the fact that the quadratic variation $\langle \bar{x}^{i,n} \rangle_t$ is given by $\int_0^t \sigma^2(\bar{x}_s^{i,n}) ds$, one obtains that

$$\begin{aligned} f(\bar{x}_t^{i,n}) &= f(x_0^i) + \int_0^t \left[F(\bar{x}_s^{i,n}) + \frac{1}{n} \sum_{j=1}^n \Gamma(\bar{x}_s^{i,n}, \bar{x}_s^{j,n}) \right] \partial_x f(\bar{x}_s^{i,n}) ds \\ &\quad + \frac{1}{2} \int_0^t \sigma^2(\bar{x}_s^{i,n}) \partial_x^2 f(\bar{x}_s^{i,n}) ds + M_t^{i,n} \end{aligned}$$

where $M_t^{i,n}$ is the martingale formally given by $\int_0^t \sigma(\bar{x}_s^{i,n}) \partial_x f(\bar{x}_s^{i,n}) dB_s^i$. Summing the last expression over i and dividing by n , on the left-hand side we recognize the action of the empirical measure $\bar{\mu}^n$ onto f at time t , i.e.,

$$\frac{1}{n} \sum_{j=1}^n f(\bar{x}_t^{j,n}) = \langle \bar{\mu}_t^n, f \rangle.$$

Similarly, we rewrite the right-hand side as

$$\langle \bar{\mu}_0^n, f \rangle + \int_0^t \langle \bar{\mu}_s^n, [F + \frac{1}{n} \sum_{j=1}^n \Gamma(\cdot, \bar{x}_s^{j,n})] \partial_x f + \sigma^2 \partial_x^2 f \rangle ds + M_t^n$$

where $M_t^n = \frac{1}{n} \sum_{i=1}^n M_t^{i,n}$ is again a martingale. The Brownian motions $\{B^i\}_{i \in \mathbb{N}}$ are independent, hence the quadratic variation of M^n is given by

$$\langle M^n \rangle_t = \frac{1}{n} \int_0^t \langle \bar{\mu}_s^n, (\sigma \partial_x f)^2 \rangle ds = O\left(\frac{t}{n}\right),$$

the last equality holding \mathbf{P} -almost surely.

Overall, one obtains that for every positive integer n , the empirical measure $\bar{\mu}^n$ satisfies the following equation

$$\langle \bar{\mu}_t^n, f \rangle = \langle \bar{\mu}_0^n, f \rangle + \int_0^t \langle \bar{\mu}_s^n, [F + \Gamma * \bar{\mu}_s^n] \partial_x f + \sigma^2 \partial_x^2 f \rangle ds + M_t^n \quad (1.3)$$

for $t \in [0, T]$ and where we have denoted with $*$ the integration with respect to the second variable, i.e. $(\Gamma * \mu)(x) = \int \Gamma(x, y) \mu(dy)$ for $x \in \mathbb{R}$ and $\mu \in \mathcal{P}(\mathbb{R})$.

Since the quadratic variation of M^n goes to zero uniformly in $t \in [0, T]$ as n tends to infinity, one expects that the limit in probability of $\bar{\mu}^n$, whenever it exists, satisfies equation (1.3) without the noise term M^n . In other words, the limit of $\bar{\mu}^n$ is expected to satisfy the weak formulation of

$$\partial_t \mu_t = \frac{1}{2} \partial_x^2 (\sigma^2 \mu_t) - \partial_x (\mu_t [F + \Gamma * \mu_t]) \quad (1.4)$$

provided that $\bar{\mu}_0^n$ converges, in a sense to be precised, to some $\mu_0 \in \mathcal{P}(\mathbb{R})$ being the initial condition to (1.4).

REMARK 1.1. The crucial property that allows to derive (1.4), is the fact that equation (1.3) is written in closed form with respect to the empirical measure – if one ignores the martingale term which vanishes as n tends to infinity. We will see that this property fails in case system

(1.1) is defined on a graph sequence. This will in turn promote the investigation whether a limit object, written in closed form, does exist or not. We refer to Section 3.

Partial differential equations as (1.4) belong to a particular class of non-linear Fokker-Planck equations which usually goes under the name of McKean-Vlasov equations. Existence and uniqueness for (1.4), when the initial datum is a probability measure, have been severely addressed under very general assumptions on F , Γ and σ and, in particular, they hold in our context, see, e.g., [56, 50]. In the sequel, we will refer to μ , the unique solution to equation (1.4), as to the *mean-field limit*.

Once the possible limits of $\bar{\mu}^n$ have been identified and uniqueness is established for equation (1.4), one way to obtain the weak convergence of $\{\bar{\mu}^n\}_{n \in \mathbb{N}}$ to μ , is to show that this last sequence is tight in the space of probability measures $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$. This is usually possible at the cost of requiring a finite moment condition on the starting measure μ_0 or by requiring IID initial conditions in system (1.1). We refrain from entering into these details, but refer to Chapter 5 where a generous discussion is devoted to this topic.

The precise result on $\bar{\mu}^n$, usually known under the name of Law of Large Numbers for interacting diffusions [80, 88], can be summarized in the following theorem.

THEOREM 1.2. *Suppose that $\bar{\mu}_0^n$ weakly converges to some μ_0 in $\mathcal{P}(\mathbb{R})$ and that μ_0 has finite second moment. Then, $\bar{\mu}^n$ weakly converges in probability to μ , solution to equation (1.4), as n tends to infinity.*

Many results equivalent to Theorem 1.2 are available in literature: the different statements depend not only on the hypothesis required on system (1.1) but also on the topological space where the convergence is established. Moreover, it is possible to work in topologies that can be weaker or stronger than the weak convergence. We refer to [56, 98] for two very different approaches and again to Chapter 5 for a very weak notion of convergence.

Theorem 1.2 represents a Law of Large Numbers, yet many other results for system (1.1) are available in the form of Central Limit Theorems or Large Deviation Principles. The complete literature is difficult to cite, we refer to [20, 45, 95] for Central Limit Theorems and to [99, 44, 23, 29] for Large Deviation Principles and to all the references therein. In this manuscript, we do not address the study of fluctuations, but we will establish a result in the spirit of Large Deviations in Chapter 2.

1.2. The non-linear process description. Another key property related to weakly interacting particle systems that has allowed a deep understanding of systems such as (1.1), is the possibility to construct a stochastic process whose time-marginal laws are precisely given by $\{\mu_t\}_{t \in [0, T]}$, solution to equation (1.4).

The process $x = \{x_t\}_{t \in [0, T]}$, originally introduced by McKean [81], is defined as the solution to the following non-linear stochastic differential equation

$$\begin{cases} dx_t = F(x_t)dt + \int_{\mathbb{R}} \Gamma(x, y) \mu_t(dx)dt + \sigma(x_t)dB_t, \\ \mu_t = \mathcal{L}(x_t), \end{cases} \quad (1.5)$$

for $t \in [0, T]$ and where the initial condition x_0 is some random variable with law $\mathcal{L}(x_0) = \mu_0$ and B is a Brownian motion independent of x_0 .

The precise relation between x solution to (1.5) and μ solution to equation (1.4), is given in the next theorem.

THEOREM 1.3 (Theorem 1.1 of [98]). *There exists a unique solution to equation (1.5). If $\mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$ denotes its law, then its time-marginal projection $\mu_\cdot = \{\mu_t\}_{t \in [0, T]}$, seen as an element of $\mathcal{C}([0, T], \mathcal{P}(\mathbb{R}))$, solves the weak formulation of equation (1.4).*

The duality between μ and μ_\cdot – and the one between x and μ_\cdot – will follow us throughout many parts of the manuscript. It is the same duality that exists between $\bar{\mu}^n \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$ and $\bar{\mu}^n \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{R}))$, recall the definition in (1.2). Depending on the context, it will be more suitable to work in one space, and its consequent topology, or in the other.

1.3. The coupling method and propagation of chaos. Starting from the well-known work of Sznitman on propagation of chaos [98], many subsequent papers on interacting particles (see [85, 83] and references therein) have been using the so-called *coupling method* as technique to show the proximity of system (1.1) to n independent copies of the non-linear process (1.5) and, in particular, a Law of Large Numbers equivalent to Theorem 1.2.

The method consists in coupling the trajectories of system (1.1), to n IID copies of the non-linear process solving (1.5) by choosing the initial conditions to be independent and identically distributed. The coupling is then made by taking the same Brownian motions and the same initial conditions for both systems. Under the only hypothesis of boundness and uniformly Lipschitz continuity on the coefficients in equation (1.1), it is possible to prove that the two systems are close trajectory per trajectory. This in turn implies the convergence of the empirical measure $\bar{\mu}^n$ to the solution to the McKean-Vlasov equation (1.4).

On the one hand, the coupling method is extremely powerful since it asks very little on the dynamics and require no finite moment on μ_0 . On the other hand, it forces to assume independent and identically distributed initial conditions. This requirement is far from being innocent – as stressed by Sznitman himself in [98, Remark 1.3] – and it will be a crucial point of investigation in our first attempts to the study of interacting particles on graphs. As explained in Subsection 4.1, there are situations in which we do not want to assume IID initial conditions since, e.g., they can hide relevant properties of the underlying graph sequence as connectedness. This point of view, originally presented in [36], is further tackled in Chapter 2.

The original idea of considering a chaotic initial datum stems from Boltzmann equation and notably by the works of the physicists Kač and McKean [63, 81]. Suppose that at time 0 the particles are independent and identically distributed, one is interested in understanding if the initial chaoticity can be propagated in time, eventually by taking a system with more and more particles. Indeed, the interaction between two particles in the finite system is weak – of order $O(n^{-1})$ – and it is plausible to prove that the statistical dependence between them goes to zero as the size of the system grows – at least up to some *finite* time. This argument can be obviously extended to any finite number of particles and yields the propagation of chaos related to their joint distribution.

The work of Sznitman [98] puts on a mathematical basis the *propagation of chaos* concepts coming from the insights originally given by Kač and McKean. Even though very little progress has been achieved in the case of Boltzmann equation (see, e.g., [83] and references therein), this strategy has proven to be well suited for other general stochastic mean-field systems as the ones considered here.

In [98], it is proven the following result on the joint law of particles solving (1.1).

THEOREM 1.4. *Suppose that $\{x_0^i\}_{i \in \mathbb{N}}$ is a sequence of independent and identically distributed random variables with law μ_0 . Then $\bar{\mu}^n$ weakly converges in probability to μ , the law of x solving equation (1.5). Furthermore, for every $k \in \mathbb{N}$ it holds that*

$$\lim_{n \rightarrow \infty} \mathcal{L}(x^{1,n}, \dots, x^{k,n}) = \prod_{j=1}^k \mathcal{L}(x) = \prod_{j=1}^k \mu. \quad (1.6)$$

REMARK 1.5. An important consequence of taking independent and identically distributed initial conditions is the exchangeability property of system (1.1) – or symmetry, as expressed in [98]. Under such assumption, the particles in (1.1) are exchangeable random variables and their joint law is thus invariant under permutation of the labels. In particular,

$$\mathcal{L}(x^{1,n}, \dots, x^{k,n}) = \mathcal{L}(x^{\sigma(1),n}, \dots, x^{\sigma(k),n})$$

for every $2 \leq k \leq n$ and every σ permutation on k elements.

Exchangeability turns out to be a crucial property when studying the convergence of empirical measures for systems defined on exchangeable graph sequences. For these sequences, a notion of limit is given by graphon theory – that we recall in the next section – and an explicit description exists.

We refer to Subsections 4.1 and 4.2 for a more detailed viewpoint on both limitations and benefits of taking IID initial conditions. In Chapter 3, the exchangeability property coming from the coupling method and exchangeable random graphs is fully exploited by means of graph limits theory. A propagation of chaos result in the spirit of Theorem 1.4 is also established.

2. Graph sequences and graphons

In the last twenty years, there has been a great interest in graph theory and notably concerning suitable notions of convergence. Several attempts [74, 16] have been made to find a unifying theory of convergence for general (possibly random) graph sequences. Despite the recent progress [16, 17, 18], the two extremal regimes, represented by dense and sparse graphs respectively, require different mathematical tools and seem to provide diverse, although complementary, descriptions.

This manuscript does not investigate the sparse regime – we refer to the only two references available so far [91, 70] and to the perspectives presented in Chapter 6. Nonetheless, it addresses intermediate regimes between the dense case, where the connections are in magnitude equal to the square of the number of particles, and very diluted graph sequences, where almost every site has a diverging number of connections, this divergence being possibly very slow.

The following subsections present the general notation for graph sequences used throughout the manuscript. We also give a fair introduction to the results and insights of graph limits theory which will be used in many parts of the manuscript. For instance, Chapter 3 adopts many of the ingredients recalled here.

We work with both random and deterministic graph sequences. We point out that throughout this thesis, a sequence of graphs will always be considered convergent in the sense of graph limits theory [73, 74].

2.1. A general graph sequence and dilution parameters. A general graph sequence is denoted by $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$. For every $n \in \mathbb{N}$, $\xi^{(n)}$ represents a directed graph $(V^{(n)}, E^{(n)})$ where $V^{(n)}$ denotes the vertex set, always composed of n vertices labeled from 1 to n , and $E^{(n)} \subset V^{(n)} \times V^{(n)}$ represents the edge set. Particle $i \in V^{(n)}$ points to particle $j \in V^{(n)}$ if and only if the couple (i, j) belongs to $E^{(n)}$.

We always consider labeled graph sequences and denote the labels by the letters i, j and k . The notation $[n] := \{1, \dots, n\}$ is often adopted, so that $V^{(n)} = [n]$ for every $n \in \mathbb{N}$.

The adjacency matrix of a graph $\xi^{(n)}$ is denoted by $\xi^{(n)}$ itself and it is given by the $n \times n$ square matrix $\{\xi_{ij}^{(n)}\}_{i, j \in [n]}$, where $\xi_{ij}^{(n)}$ is a binary variable for all $i, j = 1, \dots, n$ and takes value 1 whenever i points to j and 0 otherwise. From the adjacency matrix $\xi^{(n)}$, the edge set can be coherently defined by

$$E^{(n)} = \left\{ (i, j) \in V^{(n)} \times V^{(n)} : \xi_{ij}^{(n)} \neq 0 \right\}.$$

Occasionally, we will consider multigraphs, i.e., allowing for multiedges between vertices, with consequent notation $\xi_{ij}^{(n)} \in \{0, 1, 2, \dots\}$ and ignoring the edge set $E^{(n)}$. In the whole manuscript, selfloops do not play any relevant role and can be ignored: we may set $\xi_{ii}^{(n)} = 0$ for every $n \in \mathbb{N}$, $i \in [n]$ and any graph sequence.

Whenever the graph sequences ξ is random, we denote its probability measure by \mathbb{P} . In most instances, we will suppose the randomness present in the graph sequences to be independent

of the one in the initial conditions – and always independent of the Brownian motions – so that we will actually work with $\mathbf{P} \times \mathbb{P}$. However, there are cases in which the initial conditions may be taken dependent of the graph structure and we will use \mathbb{P} to denote the conditional probability with respect to \mathbf{P} .

The most common example of random graph sequence is given by Erdős-Rényi random graphs. A possible definition is given in the next example.

Symmetric Erdős-Rényi random graphs. For $n \in \mathbb{N}$, let $p_n \in [0, 1]$. Define $\xi^{(n)}$ by

$$\xi_{ij}^{(n)} \sim \text{Ber}(p_n), \quad \text{independently for } 1 \leq i < j \leq n,$$

meaning that $\xi_{ij}^{(n)}$ is a Bernoulli random variables of parameter p_n . If we set $\xi_{ji}^{(n)} = \xi_{ij}^{(n)}$ for every $i, j \in [n]$, this means that particle i is connected to particle j with probability p_n . In particular, $\{\xi_{ij}^{(n)}\}_{1 \leq i < j \leq n}$ are IID random variables. We say that $\xi^{(n)}$ is a *symmetric Erdős-Rényi random graph* with parameter p_n and, more generally, that $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of symmetric Erdős-Rényi random graphs with parameters $\{p_n\}_{n \in \mathbb{N}}$.

We will see that such sequence is highly homogeneous, see the next Remark 2.1. An equivalent construction by means of W -random graphs is given in the next subsection. \square

For a general random graph sequence $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$, we say that ξ is *symmetric*, corresponding to undirected graphs, when for every n and \mathbb{P} -almost surely

$$\xi_{ij}^{(n)} = \xi_{ji}^{(n)}, \quad 1 \leq i < j \leq n.$$

Furthermore, we say that $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is

- *dense*, when the expected number of edges is of order n^2 , i.e.,

$$\mathbb{E} [|E^{(n)}|] = \Theta(n^2);$$

- *sparse*, when the expected number of edges is sublinear in n , i.e.,

$$\mathbb{E} [|E^{(n)}|] = O(n).$$

- in the *intermediate regime* if there exists a sequence $\{p_n\}_{n \in \mathbb{N}} \subset (0, 1]$ such that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} [|E^{(n)}|]}{n^2 p_n} = p, \tag{1.7}$$

for some $p \in (0, 1]$. The coefficient p_n is called the *dilution parameter* associated to $\xi^{(n)}$. In the literature, it is sometime referred to as the sparsity parameter.

Observe that dense and sparse graph sequences are a priori included in the intermediate regime since no conditions are required on the asymptotic of p_n . To a dense sequence it usually corresponds $p_n = O(1)$ and $p_n = O(n^{-1})$ to a sparse sequence. In other words, the dilution parameter p_n represents the average edge density in the graph $\xi^{(n)}$ and, equivalently, np_n represents the degree of a vertex taken uniformly at random in the vertex set $V^{(n)}$. For instance, a sequence which is neither dense nor sparse, is in the intermediate regime if the average mean

degree satisfies $1 \ll np_n \ll n$ in the limit for large n . In the sequel, we will exclude sparse graph sequences when considering the intermediate regimes.

Focusing on equation (1.7), one can always take $p = 1$, e.g., by taking $p_n = \frac{1}{n^2} \mathbb{E} |E^{(n)}|$ for every n . However, an arbitrary $p \in (0, 1]$ turns out to be more suitable when the graph sequence is dense. In this last case, we take $p_n \equiv 1$ and p represents the limit average degree density.

It is not difficult to see that a sequence of Erdős-Rényi random graphs with parameters $\{p_n\}_{n \in \mathbb{N}}$ satisfies equation (1.7) with the same choice of dilution parameters and with $p = 1$.

We usually consider graph sequences where the degrees have all the same magnitude with respect to the number of vertices. However, this condition can be relaxed by means of a suitable renormalization of the interaction between the particles. We have preferred to keep the key ideas as clear as possible, and to present further generalizations as possible developments of the presented works. We refer to the perspectives proposed in Chapter 6.

REMARK 2.1. Throughout the manuscript, we often use the notion of homogeneous and inhomogeneous for a general graph sequence. We keep in mind that the property of being homogeneous is related to having good mixing properties, as the one later expressed in (1.14), and to the notion of mean-field limit, recall equation (1.4). Loosely speaking, the most homogeneous graph sequence consists of complete graphs, these last ones being the graphs on which system (1.1) is implicitly defined for every n .

In the next subsection we will see how it is possible to characterize a notion of inhomogeneity by means of graph limits theory. We refer to Remark 3.1.

2.2. Graphons. A general graph theory with a notion of convergence that is coherent with most of the graph parameters – clique size, perfect matchings, spectrum, etc. – is available for dense graphs starting from the seminal work of Lovász and Szegedi [74]. Whether this theory, also known as graphon theory, is intrinsically related to weakly interacting particle systems, has been a constant debate during the last few years. As it will be clear in Chapter 3, graph limits theory *is* strictly related to the convergence of empirical measures for particle systems on graphs.

We present here a few ingredients that will be useful throughout our analysis and refer to Chapter 3 and the complete monograph [73] for further information. The reader who is familiar with the graphon framework can skip this subsection.

There are many equivalent ways to introduce the notion of graph convergence and the corresponding graph limits, known in the literature as graphons. Since we do not need all the details coming from the original definition of graph homomorphisms, we rather give a few key concepts from the analytic viewpoint, the one we will mostly adopt.

The major difference with respect to the previous subsection is given by the fact that we are now focusing on dense symmetric graph sequences and always assuming $p_n \equiv 1$. We are not interested in pushing the graph analysis close to the sparse regime, but rather in understanding how the degree distribution among the vertices affects the particle dynamics. In this sense,

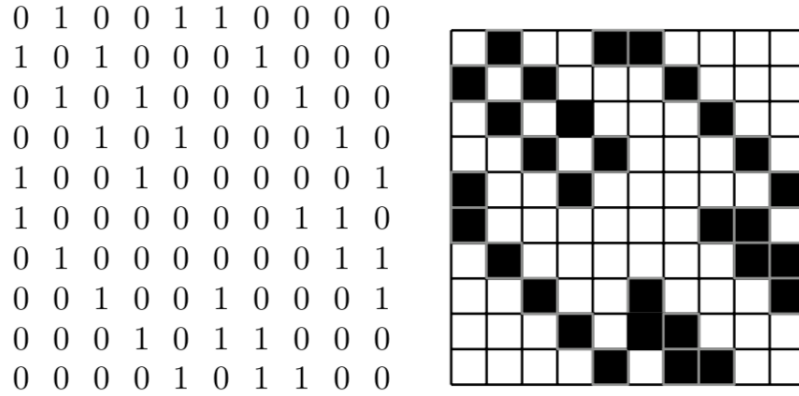


FIGURE 1.1. An adjacency matrix and its pixel picture. Image from [73, Chapter 1].

what interests us – and a substantial part of graph limits theory – is a theory for describing the underlying structure of a graph sequence and defining a suitable notion of inhomogeneity.

Consider a deterministic graph sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$. In order to understand the inhomogeneity present in $\xi^{(n)}$ and to give a meaning to $\lim_n \xi^{(n)}$, we first need to cast the *underlying structure* present in $\xi^{(n)}$ to an object that we can compare for different graphs and somewhat independent of the number of vertices. This is usually done by extracting the *pixel picture*.

The main idea is to represent the adjacency matrix of a finite graph ξ with n vertices, as a black and white grid composed of n^2 pixels of the same size, scaled in the unit square $[0, 1] \times [0, 1] \subset \mathbb{R}^2$. By coloring the pixel corresponding to the i -th row and j -th column of the grid if and only if $\xi_{ij} = 1$, we end up with a $\{0, 1\}$ -valued function with domain $[0, 1]^2$. In other words, to every finite graph ξ with n vertices, we can associate a symmetric measurable function $W_\xi : [0, 1]^2 \rightarrow \{0, 1\}$ defined by

$$W_\xi(x, y) := \sum_{i, j=1}^n \xi_{ij} \mathbf{1}_{\{\frac{i-1}{n} < x \leq \frac{i}{n}\}}(x) \mathbf{1}_{\{\frac{j-1}{n} < y \leq \frac{j}{n}\}}(y), \quad \text{for } x, y \in [0, 1]. \quad (1.8)$$

See Figure 1.1 for a concrete example. Such a function is said to be a *step-graphon* and belongs to the general space

$$\mathcal{W}_0 = \{W : [0, 1]^2 \rightarrow [0, 1] : \text{measurable and symmetric}\}$$

usually called the space of *labeled graphons*. We recall that a function $W : [0, 1]^2 \rightarrow [0, 1]$ is symmetric if it is such that $W(x, y) = W(y, x)$ for almost every (a.e.) x and y in $[0, 1]$. Observe that we tacitly consider two functions in \mathcal{W}_0 to be equal if the set where they differ is of Lebesgue measure zero.

Define the space of *kernels* by $\mathcal{W} = \{W : [0, 1]^2 \rightarrow \mathbb{R} : \text{bounded, measurable and symmetric}\}$. Clearly, a labeled graphon is a kernel and $\mathcal{W}_0 \subset \mathcal{W}$. We now define a distance on \mathcal{W} .

For every two kernels $W, V \in \mathcal{W}$, let

$$d_{\square}(W, V) := \max_{S, T \subset [0,1]} \left| \int_{S \times T} (W(x, y) - V(x, y)) \, dx \, dy \right| \quad (1.9)$$

where the maximum is taken over all measurable subsets S and T of $[0, 1]$. Observe that the right-hand side of (1.9) defines a norm $\|\cdot\|_{\square}$ on \mathcal{W} – usually called the *cut-norm* – so that $d_{\square}(W, V) = \|W - V\|_{\square}$ and $(\mathcal{W}, \|\cdot\|_{\square})$ is a normed space.

We can now measure the distance between two graphs by comparing their distance as step-graphons in $\mathcal{W}_0 \subset \mathcal{W}$, i.e., if ξ and ξ' are two graphs, then we set $d_{\square}(\xi, \xi') := d_{\square}(W_{\xi}, W_{\xi'})$. Loosely speaking, W_{ξ} contains some structure of ξ that we can compare with the one of other graphs. Nonetheless, W_{ξ} depends on the particular labeling of ξ and, if ξ' is a relabeling of ξ , then the associated step-graphons W_{ξ} and $W_{\xi'}$ do not necessarily coincide. Although ξ and ξ' are equal as unlabeled graphs, their d_{\square} -distance is, in general, greater than zero.

REMARK 2.2. The notion of structure we are thinking of is independent of how the vertices are ordered and, in particular, it has to be invariant under relabeling. Even if at a first sight this may seem poor at an abstract level – where the labels might be important – we will see in Chapter 3 that this notion is very natural. For instance, when we consider objects which are invariant under permutations, as joint distributions of exchangeable particles and empirical measures, it is the proper notion of graph structure to be taken.

A way to overcome the label dependency in W_{ξ} , is to consider a suitable equivalence relation on the space \mathcal{W} . Consider the *cut-distance* defined as

$$\delta_{\square}(W, V) := \min_{\varphi \in S_{[0,1]}} d_{\square}(W, V^{\varphi}), \quad (1.10)$$

where the minimum ranges over $S_{[0,1]}$ the space of invertible measure preserving maps from $[0, 1]$ into itself and where $V^{\varphi}(x, y) := V(\varphi(x), \varphi(y))$ for a.e. $x, y \in [0, 1]$. Observe that an element in $S_{[0,1]}$ represents a relabeling of the interval $[0, 1]$.

Clearly, if one defines $\delta_{\square}(\xi, \xi') := \delta_{\square}(W_{\xi}, W_{\xi'})$ for every two graphs ξ and ξ' , then $\delta_{\square}(\xi, \xi') = 0$ whenever ξ' is a relabeling of ξ . It is readily not the only case when this happens since the space $S_{[0,1]}$ is much bigger than finite permutations. More generally, we say that two graphs are equal in structure if their δ_{\square} -distance is zero.

It turns out that this notion is appropriate and allows to describe all the possible graph limits. Observe that two graphs do not need to have the same number of vertices to be equal in structure, e.g., complete graphs have d_{\square} -distance (and thus cut-distance) equal to zero no matter the number of nodes. See also the example in Figure 1.2.

The quotient space $\widetilde{\mathcal{W}}_0$ obtained from \mathcal{W}_0 by identifying two elements with δ_{\square} -distance zero is called the space of *unlabeled graphons*¹ and represents all the possible graph limits. Indeed,

¹Or, simply, the space of *graphons*. In order to avoid any possible confusion, we always explicit whether a graphon is labeled or unlabeled.

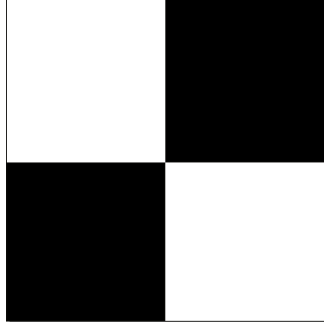


FIGURE 1.2. The graphon in the picture represents a graph with two, disjoint and fully connected components. It is also the step-graphon associated to the graph with two vertices connected by a single link and no self-loop.

a sequence of graphs $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ is said to be convergent if there exists $W \in \widetilde{\mathcal{W}}_0$ such that

$$\lim_{n \rightarrow \infty} \delta_{\square}(W_{\xi^{(n)}}, W) = 0. \quad (1.11)$$

A well-known result of graph limits theory says that $(\widetilde{\mathcal{W}}_0, \delta_{\square})$ is a compact metric space. In particular, we can define the notion of convergence in probability – and, more generally, in distribution – for a random graph sequence, see, e.g., [14]. The convergence expressed in (1.11), as other equivalent definitions, is a crucial notion that will be exploited in Chapter 3.

We end up this brief introduction with an important example that shows how it is possible to construct an exchangeable random graph with a prescribed structure and an arbitrary number of nodes.

W-random graphs. Let W be a graphon in \mathcal{W}_0 and $\{U^i\}_{i \in \mathbb{N}}$ a sequence of IID uniform random variables. For every $n \in \mathbb{N}$, define the random graph $\xi^{(n)}$ by

$$\xi^{(n)} \sim \text{Ber}(W(U^i, U^j)), \quad \text{independently for } 1 \leq i < j \leq n. \quad (1.12)$$

Set $\xi_{ji}^{(n)} = \xi_{ij}^{(n)}$ for every $i, j = 1, \dots, n$. For every $n \in \mathbb{N}$, the graph $\xi^{(n)}$ is called *W-random graph*. If W is not constantly zero, this gives rise to a dense symmetric graph sequence. Observe that the underlying structure of $\xi^{(n)}$ does not depend on the equivalence class of W in $\widetilde{\mathcal{W}}_0$. Namely, if $\bar{\xi}^{(n)}$ is another realization from some \bar{W} in the class of W , then [73, Exercice 10.20]

$$\delta_{\square}(\xi^{(n)}, \bar{\xi}^{(n)}) \leq 22 \log^{-1} n, \quad \text{with probability at least } 1 - 2^{1-n}.$$

In particular, it is well-known that the sequence $\xi^{(n)}$ converges \mathbb{P} -almost surely to W in $\widetilde{\mathcal{W}}_0$. \square

Observe that W -random graphs generalize the definition of symmetric Erdős-Rényi random graphs by allowing for a different probability for every link in the graph. In particular, the Erdős-Rényi case is recovered by taking a constant W .

REMARK 2.3. We are now able to define a notion of homogeneity. A sequence of graphs is said to be *homogeneous* whenever it converges to a deterministic constant graphon, and *inhomogeneous* when it is not the case. The previous example provides a way to construct homogeneous sequences by simply considering W -random graphs with a constant W . In particular, this implies that Erdős-Rényi random graphs are homogeneous, as well as the complete graph². In the literature, a sequence which converges to a constant graphon is known as *pseudo-random* graph sequence or quasirandom graph sequence. It has many equivalent characterizations, see the original work [28] and the recent result [5]. With a slight abuse of notation, we use the adjective pseudo-random for random graph sequences as well.

3. Interacting particles on graphs

Let $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ be a graph sequence which admits the dilution parameters $\{p_n\}_{n \in \mathbb{N}} \subset [0, 1]$, recall definition (1.7). For every $n \in \mathbb{N}$, a weakly interacting particle system on the graph $\xi^{(n)}$ is given by the solution to the following system on \mathbb{R}^n :

$$dx_t^{i,n} = F(x_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \frac{\xi_{ij}^{(n)}}{p_n} \Gamma(x_t^{i,n}, x_t^{j,n})dt + \sigma(x_t^{i,n})dB_t^i, \quad (1.13)$$

for $i = 1, \dots, n$ and $t \in [0, T]$. The functions F , Γ and σ are the same as the ones in equation (1.1), as well as the initial conditions and Brownian motions.

The crucial difference with (1.1) is represented by the weights $\{\xi_{ij}^{(n)}/p_n\}_{j=1, \dots, n}$ appearing in the interaction term. For some $i = 1, \dots, n$, particle i is not interacting with all the others as before, but only with the fraction of them given by $\{j = 1, \dots, n : \xi_{ij}^{(n)} \neq 0\}$. In other words, each particle is represented in $\xi^{(n)}$ by a vertex and its dynamics in (1.13) is influenced only by the particles at which it is pointing to in the underlying graph. In case $\xi^{(n)}$ is symmetric, the interaction in the dynamics is mutual.

Observe that whenever $\xi_{ij}^{(n)} = 0 = \xi_{ji}^{(n)}$, particles i and j are not directly interacting with each other but, if they belong to the same connected component in $\xi^{(n)}$, they are still dependent (in probabilistic sense) of one another.

3.1. Relation with the mean-field model. In equation (1.13), one can recover the mean-field case (1.1) by taking $\xi^{(n)}$ to be the complete graph for each n and the dilution parameter p_n constantly equal to 1. As for the mean-field case, one is interested in studying the empirical measure associated to system (1.13). We now denote it by

$$\mu^n := \frac{1}{n} \sum_{j=1}^n \delta_{x^{j,n}}.$$

We want to investigate the possible limits of μ^n , if any exists, as n tends to infinity. Recall that the empirical measure of the mean-field case (1.1), converges to the mean-field limit μ , solution to equation (1.4).

²which is again a W -random graph with graphon W constantly equal to 1.

Contrary to the derivation of equation (1.4), one is faced with the lack of a closed form for μ^n . Indeed, proceeding as in Subsection 1.1 yields the following expression for μ^n

$$\langle \mu_t^n, f \rangle = \langle \mu_0^n, f \rangle + \int_0^t \langle \mu_s^n, F \partial_x f + \sigma^2 \partial_x^2 f \rangle ds + \int_0^t \frac{1}{n^2} \sum_{i,j=1}^n \frac{\xi^{(n)}}{p_n} \Gamma(x_s^{i,n}, x_s^{j,n}) ds + M_s^n.$$

One can observe that the previous equation can be rewritten as

$$\langle \mu_t^n, f \rangle = \text{mean-field equation} + \int_0^t \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi^{(n)}}{p_n} - 1 \right) \Gamma(x_s^{i,n}, x_s^{j,n}) ds$$

where *mean-field equation* stands for the equation satisfied by $\langle \bar{\mu}_t^n, f \rangle$. A first step in understanding for which sequences $\xi^{(n)}$, the empirical measure converges to the mean-field limit μ , can be made by studying the term

$$\int_0^t \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi^{(n)}}{p_n} - 1 \right) \Gamma(x_s^{i,n}, x_s^{j,n}) ds.$$

If it goes to zero uniformly in $t \in [0, T]$ as n tends to infinity, then the limit of μ^n satisfies the McKean-Vlasov equation (1.4), i.e., we expect μ^n to converge to the mean-field limit μ . However, all the particles are dependent on the graph $\xi^{(n)}$ and it is not clear how to control such perturbation.

A part of the literature has focused on methods to decouple the dynamics of the particles from the graph structure. In particular, the first results look for homogeneity conditions on the sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$, to show that quantities such as

$$\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi^{(n)}}{p_n} - 1 \right) c_{ij} \tag{1.14}$$

converge to zero for a suitable class of coefficients c_{ij} , possibly independent of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$.

Depending on the setting one is working with, the methods to tackle expressions as (1.14) vary as well: we refer to Chapters 2 and 4 for two explicit, and rather different, techniques. Subsection 4.1 underlines how the initial conditions can strongly influence the choice of a suitable decoupling method.

REMARK 3.1. On the one hand, the previous strategy yields the mean-field limit (1.4) whenever the graph sequence is in some sense homogeneous, i.e., such that quantities as (1.14) converge to zero for a large class of coefficients c_{ij} . On the other hand, it does not help in understanding whether a different limit exists for a general – possibly inhomogeneous – graph sequence. This last issue is tackled in the next subsection and further discussed in Subsection 4.2. In Chapters 3 and 4, we will see that the notion of homogeneity given in Remark 2.3, is equivalent to requiring (1.14) to go to zero for a precise choice of the coefficients c_{ij} .

3.2. Inhomogeneous behaviors. Some progress in the direction of inhomogeneous graph sequences has been made in [6, 27, 77, 90]. These works consider graph sequences which

converge – in a topology that depends case by case – to labeled deterministic graphons and show that the empirical measure μ^n is suitably described, as n tends to infinity, by an object depending on the corresponding graph limit.

We briefly present the inhomogeneous non-linear process proposed in the cited works and the main ideas behind its derivation. This description will be addressed in Subsection 4.2 and represents the main object of study in Chapter 3.

Fix a labeled graphon W . For $n \in \mathbb{N}$ and $i = 1, \dots, n$, we represent particle i in system (1.13) as a small interval in $[0, 1]$ of length $1/n$, i.e., by $((i-1)/n, i/n] \subset [0, 1]$. The n particles cover the unit interval and maintain the same ordering. In the limit for n which tends to infinity, the length of each interval shortens and the number of particles grows: we end up by associating to every particle a *label*, i.e., some $x \in [0, 1]$, instead of an index $i \in \mathbb{N}$. We will often call the unit interval $[0, 1]$, the space of labels.

In the resulting infinite system, the particles associated to x and y in $[0, 1]$ are connected with probability $W(x, y)$. In this sense, the function $W : [0, 1]^2 \rightarrow [0, 1]$ represents the connection structure of an infinite particle system indexed by $[0, 1]$. Observe that the unit interval $[0, 1]$ and a step-graphon can represent a finite number of particles on a labeled graph: in this case, we use the n sub-intervals in $[0, 1]$ previously defined, and the step graphon associated to the labeled graph – recall definition (1.8) – to represent the interactions among them.

The previous procedure describes the interaction of a (possibly infinite) system of particles, we now see how to define a dynamics on it. As n tends to infinity, we are able to guess the limit of equation (1.13), provided that the asymptotic of $\xi^{(n)}$ is suitably captured by the graphon W .

Suppose that $\xi^{(n)}$ is a W -random graph, recall the definition (1.12). The n particles $\{x^{i,n}\}_{i=1,\dots,n}$ solves system (1.13), which is now given by

$$dx_t^{i,n} = F(x_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n W(U_i, U_j) \Gamma(x_t^{i,n}, x_t^{j,n})dt + \sigma(x_t^{i,n})dB_t^i, \quad i = 1, \dots, n$$

where $\{U^i\}_{i \in \mathbb{N}}$ is an IID sequence of uniform random variables in $[0, 1]$. As n tends to infinity, this leads to the collection of non-linear processes $\{\theta^x\}_{x \in [0,1]}$ which solve

$$d\theta_t^x = F(\theta_t^x)dt + \int_{[0,1]} W(x, y) \int_{\mathbb{R}} \Gamma(\theta_t^x, \theta') \mu_t^y(d\theta') dy dt + \sigma(\theta_t)dB_t^x, \quad x \in [0, 1] \quad (1.15)$$

where μ_t^x is the law of θ^x at time $t \in [0, T]$. The initial conditions in equation (1.15) are given by an infinite vector $\{\theta_0^x\}_{x \in [0,1]}$ where θ_0^x follows some law $\mu_0^x \in \mathcal{P}(\mathbb{R})$, for every $x \in [0, 1]$. Observe that $\int_{[0,1]} \mu_0^x dx = \mu_0$ is equal to the weak limit of μ_0^n . The sequence $\{B^x\}_{x \in [0,1]}$ consists of IID Brownian motions independent of the initial conditions. We point out that the label x in equation (1.15) has no physical meaning and there is no dynamics along it.

Existence and uniqueness for equation (1.15) can be obtained by extending known techniques to take the label variable x into account and with no assumption on W but measurability. We refer to [6, 79] and to Chapter 3 where a non-linear Fokker-Planck equation associated to the law of the process (1.15) is explicitly derived.

Under suitable hypothesis on the initial conditions and the regularity of W , the works [6, 27, 77, 90] prove that, if the sequence of graphs $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ converges – in some sense depending case by case – to the labeled graphon W , then the empirical measure μ^n converges to the probability measure

$$\bar{\mu} = \int_{[0,1]} \mu^x dx, \quad (1.16)$$

where μ^x is the law of θ^x solving (1.15) for $x \in [0, 1]$.

The collection $\{\theta^x\}_{x \in [0,1]}$ describes a possible inhomogeneous behavior for the asymptotic of system (1.13), recall Remark 1.1 and Remark 3.1. Nonetheless, we emphasize that the cited results are far from giving a satisfactory description of (1.13) for a general sequence $\xi^{(n)}$, and do not completely exploit the properties of (1.15). For instance, even though the limit $\bar{\mu}$ in (1.16) is independent of the labeling of $\{\theta^x\}$ – and thus of the equivalence class of W in $\widetilde{\mathcal{W}}_0$ – existing results do not address this important aspect. Furthermore, some regularity is demanded on W with the consequent difficulty to work in $\widetilde{\mathcal{W}}_0$. These issues are discussed in Subsection 4.2.

4. Motivations

This section presents the motivations behind the results coming along with this manuscript.

4.1. The role of initial conditions. We want to emphasize how little innocent the assumption of IID initial conditions can be, in the context of particle systems on graph sequences. Consider the following theorem taken from [36], where the authors pointed out the issue for the first time.

THEOREM 4.1 (Theorem 1.1 and Corollary 1.2 of [36]). *Consider system (1.13) with independent and identically distributed initial conditions sampled from μ_0 . Suppose that the graph sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ and the dilution parameter p_n satisfy*

$$\lim_{n \rightarrow \infty} \sup_{i=1, \dots, n} \left| \frac{1}{n} \sum_{j=1}^n \frac{\xi_{ij}^{(n)}}{p_n} - 1 \right| = 0 \quad (1.17)$$

and

$$\lim_{n \rightarrow \infty} np_n = \infty. \quad (1.18)$$

Then the empirical measure μ^n of (1.13) weakly converges in probability to μ solution to equation (1.4), as n tends to infinity.

We first observe that conditions (1.17) and (1.18) mean that $\xi^{(n)}$ is in an intermediate regime and has a diverging number of connections per site. More precisely, a graph sequence satisfies equation (1.17) if and only if the normalized degree density of each vertex, given by

$\frac{1}{n} \sum_j \xi_{ij}^{(n)} / p_n$ for vertex i , converges to 1 as n tends to infinity. In other words, condition (1.17) is a homogeneity requirement on the sequence $\xi^{(n)}$ that addresses the degree density in the limit for n which tends to infinity. Roughly speaking, Theorem 4.1 says that the empirical measure μ^n converges to the mean field limit μ whenever the underlying graph sequence is in the intermediate regime and each vertex has precisely np_n connections as n tends to infinity.

Consider the sequence of graphs where two vertices are connected if and only if their labels have the same parity. This sequence satisfies equation (1.17) with $p_n \equiv 1/2$ and, for every $n \in \mathbb{N}$, it is made of two complete graphs with no edge in common. In particular, this shows that condition (1.17) does not require the graph $\xi^{(n)}$ to be connected. However, Theorem 4.1 implies that system (1.13) defined on such sequence behaves as the mean-field system (1.1). How can a system on a disconnected graph behave like the one on a fully connected component?

The apparent contradiction is explained if one observes that, by taking independent and identically distributed initial conditions, the joint law of the system (1.13) is the same in each connected component. Hence, the law of the empirical measure μ^n cannot distinguish the different components in the graph and the limit of μ^n does not depend on the connectedness of the underlying sequence.

As explicitly written in [36], the result of Theorem 4.1 is unsatisfactory with regards to applications and hides many of the graph properties related to the sequence $\xi^{(n)}$. In particular, it leaves the reader with the following unsolved³ issues:

- (1) Is it possible to prove a Law of Large Numbers similar to Theorem 4.1 without asking a chaotic initial datum? For which graph sequences?
- (2) Which homogeneity condition on the graph sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ can replace equation (1.17) for general initial conditions?

The joint work with Helge Dietert and Giambattista Giacomin [31], presented in Chapter 2, partially answers the first question. By focusing on general Erdős-Rényi random graph sequences in the intermediate regimes, we are able to prove that the empirical measure of system (1.13) converges to the mean-field limit for almost every realization. We only require the weak convergence of μ_0^n to μ_0 and a mild condition on the divergence of the average mean degree in $\xi^{(n)}$. Overall, we show that Erdős-Rényi random graphs are highly homogeneous and yield the mean-field limit for (1.13).

The second question is completely solved in the case of dense graph sequences and interacting oscillators, in the joint work with Gianmarco Bet and Francesca Nardi [11], see Chapter 3. The work [30], presented in Chapter 4, partially answers (2) for graph sequences in the intermediate regimes, but it concerns the Kuramoto model only.

³It appears that all the results in the literature on stochastic interacting particles on graphs, with the exception of [30, 31], require independent initial conditions.

The cited works [11, 30] exploit distances in the space of graphons and show that the homogeneous condition sought in (2) coincide with the notion of homogeneous graph sequence given in Remark 2.3.

4.2. Inhomogeneous behavior versus mean-field behavior. In Subsection 3.2, we have seen that a description for μ^n in the case of an inhomogeneous graph sequence is available. Namely, it is given by extending the classical non-linear process (1.5) to a collection of processes indexed by $[0, 1]$ and coupled by means of a labeled graphon W , recall equation (1.15). We have also pointed out that the limit measure $\bar{\mu}$ defined by (1.16) exists for every $W \in \mathcal{W}_0$ and is independent of the particular labeling of $\{\theta^x\}_{x \in [0,1]}$. In other words, although existing proofs seem to require regular labeled graphons, the limit description depends on the equivalent class of W in $\widetilde{\mathcal{W}}_0$ only, i.e., on an unlabeled graphon. Furthermore, the graph convergence considered in [6, 27, 77, 82, 90] is always required to be in stronger topologies than the one in $\widetilde{\mathcal{W}}_0$.

Consider a sequence $\xi^{(n)}$ that converges in $\widetilde{\mathcal{W}}_0$ to some graphon W . We want to understand the following issues:

- (1) Is it possible to prove that the empirical measure μ^n converges to $\bar{\mu}$?
- (2) For which W , is the limit of μ^n mean-field for n which tends to infinity? More precisely, under which hypothesis on the sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$, μ^n is approximately described by the mean-field limit μ solving equation (1.4) as n diverges?
- (3) What can be said if $\xi^{(n)}$ is a random sequence which converges, e.g., in probability, to a random $W \in \widetilde{\mathcal{W}}_0$?

Suppose that we can positively answer to (1). As we will see in Chapter 3, $\bar{\mu}$ solves the weak formulation of the following Fokker-Planck equation weighted by the graphon W

$$\begin{aligned} \partial_t \bar{\mu}_t(\theta) = & \frac{1}{2} \partial_\theta^2 (\sigma^2(\theta) \bar{\mu}_t(\theta)) - \partial_\theta (F(\theta) \bar{\mu}_t(\theta)) \\ & - \partial_\theta \left[\int_{[0,1]^2} W(x, y) \mu_t^x(\theta) \int_{\mathbb{R}} \Gamma(\theta, \theta') \mu_t^y(\theta') d\theta' dy dx \right]. \end{aligned} \quad (1.19)$$

In particular, equation (1.19) is – formally! – the mean-field limit (1.4), if and only if W is constant⁴. This seems to answer (2): only homogeneous graph sequences, as in Remark 3.1, yield the mean-field limit. However, even though the equations satisfied by μ and $\bar{\mu}$ are formally different, this does not mean that μ and $\bar{\mu}$ are different as probability measures.

Within this perspective, question (2) should be reformulated as:

- (2*) For which graphons are $\bar{\mu}$ and μ close as probability measures?

A satisfactory understanding of the solutions to equation (1.19) is missing. In particular, we cannot a priori exclude the possibility that two different unlabeled graphons yield two solutions to (1.19) which are close as probability measures – and one can probably build up a model with this property. Nonetheless, we expect that if two graphons are close in the cut-distance,

⁴the classical mean-field limit is obtained for $W \equiv 1$, but whenever $W \equiv p \in [0, 1]$ it suffices to multiply by p the interaction in system (1.1) to get (1.19).

then the two solutions to equation (1.19) are close in a suitable topology. Pushing the analysis further, we could try to understand whether there exists a metric D_T on $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$, such that the map

$$\begin{aligned} \Psi : (\widetilde{\mathcal{W}}_0, \delta_{\square}) &\rightarrow (\mathcal{P}(\mathcal{C}([0, T], \mathbb{R})), D_T) \\ W &\mapsto \bar{\mu} \text{ solution to (1.19)} \end{aligned} \tag{1.20}$$

is continuous and, hence, partially answering to (2*).

In the case of interacting oscillators, i.e., particles defined on the one-dimensional torus rather than \mathbb{R} , we can give a positive answer to (1) and we are able to exhibit a nice control on (1.19). Notably, we prove that the map Ψ in (1.20) is Hölder-continuous with respect to a natural distance in $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$, i.e., to the Wasserstein metric. As a byproduct, we can estimate the distance between $\bar{\mu}$ and μ , by estimating the distance between W and a constant graphon in $\widetilde{\mathcal{W}}_0$. More generally, we show that if two graphons are close in the cut-distance, the resulting particle system behaviors will be similar as the size of the systems tends to infinity.

Observe that if the map (1.20) is continuous, it is possible to work with random elements in $\widetilde{\mathcal{W}}_0$ and their push-forward in $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R})), D_T$. Notably, we can tackle the issue raised in (3). By extending the classical graph convergence to the convergence in probability in $\widetilde{\mathcal{W}}_0$, the limit description (1.19) is allowed to depend on a random graphon W . This setting appears to be new in the literature and demands for further investigations. For instance, the previous question (2*) needs to be formulated once again and it is not apriori clear which notion of mean-field approximation should be taken into account. This issue, together with related consequences, represents the main object of study in Chapter 3.

4.3. Long-time dynamics. The long-time dynamics associated to an interacting particle system is one of the most interesting domains of current research. It can be hardly predicted by numerical simulations, yet it seems crucial to explain macroscopic phenomena appearing due to the finiteness of real-world systems.

To the author's knowledge, very little is known concerning the long-time behavior of interacting particles described by stochastic differential equations and only two results [10, 78], along with [30] presented in Chapter 4 are available. The cited works address the long-time dynamics of a well-known model for describing synchronization phenomena, the Kuramoto model [68, 69, 96, 71].

The Kuramoto model is a system of n interacting particles on the n -dimensional torus $\mathbb{T}^n := (\mathbb{R}/2\pi\mathbb{Z})^n$ where each oscillator is rotating at some speed and is attracted by the others. Depending on a parameter which regulates the coupling strength, the model can show both synchronized states and incoherent behaviors.

The original Kuramoto model [68, 69] is deterministic and each particle comes into the dynamics with a natural frequency at which it would oscillate if no interaction were present. We consider here the stochastic version without the natural frequencies, see [9, 10, 51].

Let $\{\theta^{i,n}\}_{i=1,\dots,n}$ be the family of oscillators which satisfy

$$\begin{cases} d\theta_t^{i,n} = \frac{K}{n} \sum_{j=1}^n \sin(\theta_t^{j,n} - \theta_t^{i,n}) dt + dB_t^i, & t > 0, \\ \theta_0^{i,n} = \theta_0^i, & i = 1, \dots, n. \end{cases} \quad (1.21)$$

The positive constant $K \geq 0$ represents the coupling strength among the particles. The sequence $\{\theta_0^i\}_{i \in \mathbb{N}} \subset \mathbb{T}$ denotes again the initial conditions.

A rather complete understanding of the McKean-Vlasov equation corresponding to the mean-field limit of system (1.21) is given in [51] for all values of K . It is well known that a phase transition appears at the critical value $K_c = 1$. Notably, whenever $K > K_c$ a continuous compact manifold of stationary synchronized solutions appears. When $K < K_c$ there is a unique stable stationary solution that is given by $1/2\pi$, i.e., particles tend to be uniformly spread around the torus.

From the general theory of weakly interacting particles, the PDE approximation is known to be valid on a finite time interval, in our case represented by $[0, T]$. If the initial empirical measure converges fast enough, the closeness between the particle system and the PDE can be pushed up to times T which can slowly diverge with the size of the system n , e.g., $T = T_n = O(\log n)$. Very little is known on longer time scales, see [10] and references therein.

A substantial progress in this direction is made in [10] for the Kuramoto model (1.21) in the supercritical case $K > K_c$. The authors show that the mean-field approximation given by the PDE is lost on time scales diverging faster than the logarithm, i.e., $T_n \gg \log(n)$. Notably, on the linear time scale $T_n = n$, the empirical measure is shown to perform a Brownian motion around the manifold of synchronized solutions – and a similar characterization is shown to hold for any polynomial time in n . This behavior is in sharp contrast with the solution prescribed by the PDE, which is stuck at a stationary synchronized solution.

When one considers system (1.21) on a graph sequence, it is natural to believe that the resulting long-time dynamics will not be affected if *a few* connections among the particles are removed, e.g., on an Erdős-Rényi graph sequence with p close to 1. However, the long-time dynamics on graphs is challenging under several aspects. Namely:

- (1) Already dropping a few links among the particles breaks the mean-field structure and makes the classical arguments useless as already discussed in Section 3;
- (2) One cannot use a coupling argument with the corresponding non-linear process, since on diverging time scales the proximity to the PDE is lost;
- (3) The correlations arising from the graph structure need to be controlled for long time scales and no a priori estimate is available.

In Chapter 4, we show that the empirical measure of system (1.21) stays close to the – eventually trivial – manifold of stationary solutions for almost exponential time scales, both in the subcritical and in the supercritical case. The graph sequence needs to satisfy a suitable deterministic equation, in the spirit of Remark 2.3, which is written by means of the cut-norm

(1.9). Notably, the homogeneity condition on the graph sequence is equivalent to requiring pseudo-random graphs with a diverging average degree. For instance, it is shown to be satisfied by almost every sequence of Erdős-Rényi random graphs in the intermediate regimes, i.e., with parameters $\{p_n\}_{n \in \mathbb{N}}$ such that $np_n \rightarrow \infty$.

This result, presented in [30], is possible by directly studying the equation satisfied by the empirical measure and without using any coupling on the single trajectories, as pointed out in (2). The derivation of such equation reminds of the one presented in Subsection 3.1, yet it makes sense in a much more general space, i.e., a Hilbert space. The issue raised in (3) is tackled by a detailed study of the dynamics around the stationary solutions: only by exploiting the contractive properties of the linearized operators it is possible to control the graph perturbation up to very long-times.

4.4. A n -dependent equation for the empirical measure. Consider equation (1.1) with $\sigma \equiv 0$, i.e., a deterministic system of weakly interacting particles. It is well known [87, 20, 41] that, with the only hypothesis of the weak convergence of μ_0^n , the empirical measure $\bar{\mu}^n$ weakly converges to the solution of the Vlasov equation, i.e., equation (1.4) without the diffusive term. To the author's knowledge, this result is missing in the stochastic framework where a finite moment condition on μ_0 , or chaotic initial data in (1.1), is always required to show this same convergence.

In collaboration with Florian Bechtold [8], we have investigated whether a similar result holds in the case $\sigma \equiv 1$. Under the only hypothesis of the convergence of μ_0^n , we show that the weak convergence of $\bar{\mu}^n$ to the solution of equation (1.4), can be established in a suitable class of Hilbert spaces. This is possible by giving a meaning to the stochastic term in the n -dependent equation satisfied by the empirical measure, recall equation (1.3), and by using the properties of the heat semigroup. Indeed, equation (1.3) can be rewritten by means of Duamel's formula as

$$\langle \bar{\mu}_t^n, h \rangle = \langle S_t \bar{\mu}_0^n, h \rangle + \left\langle \int_0^t S_{t-s} \partial_x [(F + \Gamma * \bar{\mu}_s^n) \bar{\mu}_s^n] ds, h \right\rangle + \langle w_t^n, h \rangle$$

where h is an element of a Hilbert space H and $\langle \cdot, \cdot \rangle$ represents the action of the dual space of H on H itself. The operator $(S_t)_{t \in [0, T]}$ represents the analytic semigroup associated to the Laplacian operator and w^n is the noise term coming from the martingale M^n in equation (1.3).

It turns out that the noise perturbation w^n can be defined as a stochastic convolution [32, 54, 100]. A uniform bound in n is hard to establish due to the loss of the martingale property and represents the hardest challenge for what we aim at. With the help of rough path theory and maximal inequalities for self-normalized processes, we are able to obtain a satisfactory control on w^n which is the novel ingredient for obtaining the convergence of μ^n to the McKean-Vlasov solution μ . The details are covered in Chapter 5.

5. Overview of the results and general organization

This manuscript addresses the limit behavior of weakly interacting particle systems on homogeneous and inhomogeneous graph sequences (Chapters 2 and 3 respectively), as well as the long-time behavior of the Kuramoto model on pseudo-random graph sequences (Chapter 4). In addition, it investigates the classical Law of Large Numbers for interacting diffusions under very weak assumptions on the initial conditions (Chapter 5).

In Chapter 2 we consider a general system of interacting particles on \mathbb{R} on a sequence of Erdős-Rényi random graphs in the intermediate regimes. We show that, under the only assumption of the weak convergence of the empirical measure at time 0, the system satisfies the same Law of Large Numbers and Large Deviation Principle of its mean-field analogous. Every graph sequence with average degree diverging faster than the logarithm scale satisfies our assumptions.

Chapter 3 focuses on interacting oscillators and dense graph sequences. By means of the coupling method and by exploiting the exchangeability coming from graph limits theory, it is possible to show a Law of Large Numbers for very general graph sequences. The limit describing the system behavior is described by a non-linear process depending only on the graph limit, i.e., by an unlabeled random graphon. A comparison with the mean-field behavior is also discussed as well as a Propagation of Chaos result. It based on the joint work with Gianmarco Bet and Francesca Nardi [11].

Chapter 4 addresses the long-time behavior of the Kuramoto model on graph sequences presented in [30]. Under a deterministic condition on the graph sequence, it is possible to control the macroscopic behavior of the system up to almost exponential time scales, both in the subcritical and supercritical regime. It turns out that such condition is equivalent to requiring the graph sequences to be made of pseudo-random graphs with a possibly very slow diverging average degree, i.e., it suffices that np_n diverges as n tends to infinity.

Chapter 5 establishes a Law of Large Numbers for a class of interacting diffusions on \mathbb{R}^d . The accent is posed on the initial conditions which need to satisfy a milder assumption compared to ones required in the existing literature. Namely, the weak convergence of the empirical measure at time 0 suffices. It exploits rough path theory, semigroup operators and maximal inequalities for self-normalized stochastic processes. It is based on a joint work with Florian Bechtold [8].

Finally, Chapter 6 presents a very brief overview on the open questions and important issues related to the aforementioned results.

Interacting diffusions on Erdős-Rényi random graphs

This chapter presents the result of the joint work with Helge Dietert and Giambattista Giacomin [31].

1. Introduction

1.1. Basic notations, the models and a first look at the main question. Large systems of interacting diffusions with mean field type interactions have been an important research topic in the mathematical community at least since the 60's. The program of identifying the emerging behavior for $n \rightarrow \infty$, where n is the number of interacting *units*, has been fully developed under suitable regularity and boundedness assumptions on the coefficients defining the system. In particular, Law of Large Numbers, Central Limit Theorems and Large Deviation Principles have been established (see for example [98, 85, 99, 34, 23]). A number of important issues remain unsolved, like the generalization to singular interactions (e.g. [61]) or understanding the delicate issue of considering at the same time large n and large time (e.g. [78]). But another direction in which mathematical results are still very limited is about relaxing the *complete graph assumption* for the interaction network – *complete graph* is just a different wording for *mean field* – and going towards more heterogeneous interaction networks. This is an issue that emerges in plenty of applied disciplines and giving a proper account of the available literature would be a daunting task: so we limit ourselves to signaling the recent survey [94] which contains an extended literature.

We are therefore going to study the emerging behavior of interacting diffusion models when, like in complete graphs, every unit interacts with a diverging number of other units. The interaction network is described as a random graph, notably of Erdős-Rényi (ER) type; so we start with the basic notions on graphs.

Let $\xi^{(n)} = \{\xi_{i,j}^{(n)}\}_{i,j \in \{1, \dots, n\}}$ denote the adjacency matrix of a graph $(V^{(n)}, E^{(n)})$ with n vertices ($\xi^{(n)}$ will also denote the graph itself):

$$V^{(n)} := \{1, \dots, n\} \quad \text{and} \quad E^{(n)} := \left\{ (i, j) \in V^{(n)} \times V^{(n)} : \xi_{i,j}^{(n)} = 1 \right\}. \quad (2.1)$$

We consider sequences of asymmetric ER random graphs with self loops with probabilities $p_n \in (0, 1)$ for $n = 2, 3, \dots$. More precisely, we just assume that $\{\xi_{i,j}^{(n)}\}_{i,j \in \{1, \dots, n\}}$ are Independent Identically Distributed (IID) Bernoulli random variables of parameter p_n (with notation $B(p_n)$). The arguments are easily adapted to the case in which $\xi_{j,j}^{(n)} = 0$ for every j and the results are unchanged.

Even if these graphs are not coupled for different values of n , it is practical to work with only one probability space and to couple these adjacency matrices (or random graphs). For example one can start from a sequence $\{U_k\}_{k \in \mathbb{N}}$ of IID $U(0, 1)$ variables and define $\xi_{i,j}^{(n)} = \mathbf{1}_{U_{k(i,j)} < p_n}$, with k an arbitrary bijection from \mathbb{N}^2 to \mathbb{N} . The law of the graph is denoted by \mathbb{P} , with \mathbb{E} the corresponding expectation, and we will just write $\mathbb{P}(d\xi)$ -a.s. meaning "almost surely in the realization of $\{\xi^{(n)}\}_{n=2,3,\dots}$ ".

Given a realization of $\xi^{(n)}$, consider the n -dimensional diffusion $\theta_t^n := \{\theta_t^{i,n}\}_{i=1,\dots,n}$ which solves for every i

$$d\theta_t^{i,n} = F(\theta_t^{i,n}) dt + \frac{1}{n} \sum_{j=1}^n \frac{\xi_{i,j}^{(n)}}{p_n} \Gamma(\theta_t^{i,n}, \theta_t^{j,n}) dt + \sigma(\theta_t^{i,n}) dB_t^i, \quad (2.2)$$

where $\{B_t^i\}_{i \in \mathbb{N}}$ are independent standard Brownian motions (whose law is denoted by \mathbf{P}) and independent also of $\xi^{(n)}$ (so, we are effectively working with $\mathbf{P} \otimes \mathbb{P}$). For simplicity, we consider only deterministic initial conditions; but the results apply to random initial conditions once they are taken independent of Brownian motions and of ξ . Moreover, assume that:

- (1) F , Γ and σ are real valued (uniformly) Lipschitz functions: the corresponding Lipschitz constants are denoted by L_F , L_Γ and L_σ ;
- (2) Γ is bounded, in particular $\|\Gamma\|_\infty := \sup_{x,y \in \mathbb{R}} |\Gamma(x,y)| < \infty$;
- (3) $\sigma_- \leq \sigma(\cdot) \leq \sigma_+$ with σ_\pm two positive constants (non degenerate diffusion). If $\sigma(\cdot)$ is a constant, we include the case $\sigma(\cdot) \equiv 0$.

Fix $T > 0$, the law of the n trajectories $\{\theta_t^n\}_{t \in [0,T]}$ for the *quenched* system is denoted by \mathbf{P}_n^ξ , i.e. $\mathbf{P}_n^\xi \in \mathcal{P}(\mathcal{C}^0([0,T]; \mathbb{R}^n))$, and the associated empirical measure at time t by $\{\mu_t^n\}_{t \in [0,T]}$, i.e.

$$\mu_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{\theta_t^{j,n}} \in \mathcal{P}(\mathbb{R}). \quad (2.3)$$

$\mathcal{P}(\mathbb{R})$ denotes the set of probability measures over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ equipped with the (metrizable) topology of weak convergence: i.e., if $\mu_n \in \mathcal{P}(\mathbb{R})$ for every n , then $\lim_n \mu_n = \mu \in \mathcal{P}(\mathbb{R})$ if $\int h(x) \mu_n(dx) \rightarrow \int h(x) \mu(dx)$ as $n \uparrow \infty$ for every $h(\cdot)$ continuous and bounded function. Note that since ξ is random, μ_t^n is a random variable taking values in $\mathcal{P}(\mathbb{R})$, equipped with the σ -algebra of its Borel subsets.

The solution $\{\theta_t^{i,n}\}_{i=1,\dots,n}$ is going to be tightly linked with $\{\bar{\theta}_t^{i,n}\}_{i=1,\dots,n}$ which solves

$$d\bar{\theta}_t^{i,n} = F(\bar{\theta}_t^{i,n}) dt + \frac{1}{n} \sum_{j=1}^n \Gamma(\bar{\theta}_t^{i,n}, \bar{\theta}_t^{j,n}) dt + \sigma(\bar{\theta}_t^{i,n}) dB_t^i. \quad (2.4)$$

The law of $\{\bar{\theta}_t^n\}_{t \in [0,T]}$ is denoted by \mathbf{P}_n . Moreover $\bar{\mu}_t^n := (1/n) \sum_{j=1}^n \delta_{\bar{\theta}_t^{j,n}}$. Often (2.4) is called *annealed* system: of course (2.4) is obtained from (2.2) by taking the expectation of the drift with respect to \mathbb{P} .

If the empirical measure of the initial conditions converges to a probability μ_0 , i.e.

$$\lim_{n \rightarrow \infty} \bar{\mu}_0^n = \mu_0 \in \mathcal{P}(\mathbb{R}), \quad (2.5)$$

and if $\int_{\mathbb{R}} x^2 \mu_0(dx) < \infty$, then it is well known that, for every $t > 0$, $\bar{\mu}_t^n$ weakly converges in $\mathcal{P}(\mathbb{R})$ to μ_t , the unique weak solution of the following McKean-Vlasov (or Fokker-Planck) equation

$$\partial_t \mu_t(\theta) = \frac{1}{2} \partial_\theta^2 (\sigma^2(\theta) \mu_t(\theta)) - \partial_\theta (\mu_t(\theta) F(\theta)) - \partial_\theta \left(\mu_t(\theta) \int_{\mathbb{R}} \Gamma(\theta, \theta') \mu_t(d\theta') \right). \quad (2.6)$$

The slightly stronger result that is proven is in fact: for every $T > 0$, if one considers μ^n as an element of $C^0([0, T]; \mathcal{P}(\mathbb{R}))$ (a complete separable metric space), then $\lim_n \mu^n = \mu$. (**P**-a.s. when σ is non degenerate). The notion of weak solution $\mu \in C^0([0, T]; \mathcal{P}(\mathbb{R}))$ to (2.6), which can be found for example in [56], is strictly related to the nonlinear diffusion formulation: the stochastic process $\{\varphi_t\}_{t \in [0, T]}$ that solves

$$\begin{cases} d\varphi_t = F(\varphi_t) dt + \int \Gamma(\varphi_t, \varphi) \nu_t(d\varphi) dt + \sigma(\varphi_t) dB_t, \\ \nu_t = \text{Law}(\varphi_t), \quad \text{for all } t \in [0, T], \end{cases} \quad (2.7)$$

with initial condition which is a square integrable random variable independent of the standard Brownian motion B . Existence and uniqueness for this *atypical* stochastic differential equation is not obvious at all, but it is by now well known that if $\nu_0 = \mu_0$, then the unique $\nu \in C^0([0, T]; \mathcal{P}(\mathbb{R}))$ such that ν_t is the law of φ_t for all $t \in [0, T]$, is the unique weak solution of (2.6), i.e. $\nu_t = \mu_t$ for all $t \in [0, T]$. The literature on the results that we have just mentioned is vast, see e.g. [88, 98, 85, 56] for the non degenerate diffusion case and [41, 87] for the $\sigma(\cdot) \equiv 0$ case; in this last case there is no need to assume that $\int_{\mathbb{R}} x^2 \mu_0(dx) < \infty$.

In the sequel, we will also work with probabilities in $\mathcal{P}(C^0([0, T]; \mathbb{R}))$, that is considering the law of $\{\varphi_t\}_{t \in [0, T]}$ seen as a random trajectory on the path space $C^0([0, T]; \mathbb{R})$, rather than its time marginals $\mu_t \in \mathcal{P}(\mathbb{R})$.

REMARK 1.1. Observe that knowing the law of (2.7) gives more information than the solution μ of the McKean-Vlasov equation (2.6). Indeed, call P_φ the law of $\{\varphi_t\}_{t \in [0, T]}$, then P_φ is an element of $\mathcal{P}(C([0, T]; \mathbb{R}))$, whereas $\mu \in C^0([0, T]; \mathcal{P}(\mathbb{R}))$. It is straightforward to obtain μ_t from P_φ by just observing

$$\mu_t(\cdot) = P_\varphi \circ \pi_t^{-1}(\cdot), \quad (2.8)$$

where $\pi_t : C([0, T]; \mathbb{R}) \rightarrow \mathbb{R}$ is the canonical projection at time t . Observe that a reverse statement is not always possible: μ alone does not allow to compute multidimensional time marginals like $\mathbf{P}(\varphi_s \in A, \varphi_t \in B)$, for $s, t \in [0, T]$ and $A, B \subset \mathbb{R}$. Existence, uniqueness and well-posedness of the problem for P_φ can be found in [85] and references therein.

1.2. Aim of the paper. Informally stated, our aim is to study the proximity of μ^n and $\bar{\mu}^n$, for n large. Since $\bar{\mu}^n$ approaches the solution of the McKean-Vlasov equation (2.6), this

turns out to be studying the proximity of μ^n and the solution of the McKean-Vlasov equation. This of course requires (at least) the assumption that

$$\lim_{n \rightarrow \infty} \mu_0^n = \mu_0. \quad (2.9)$$

A result of this type has been already achieved: in the case $\sigma(\cdot) \equiv \sigma \geq 0$, [36] proved a LLN for the trajectories of (2.2) where $\xi^{(n)}$ is a (deterministic) sequence of graphs such that

$$\lim_{n \rightarrow \infty} \sup_{i \in \{1, \dots, n\}} \left| \frac{1}{n} \sum_{j=1}^n \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| = 0, \quad (2.10)$$

and with IID initial conditions (*chaotic* initial datum), that is $\theta_0^{j,n} = \theta_0^j$ for every n and every $j = 1, \dots, n$ where

$$\{\theta_0^j\}_{j \in \mathbb{N}} \text{ is a typical realization of an IID sequence of variables with law } \mu_0. \quad (2.11)$$

Under conditions (2.10) and (2.11), it is proved that $\lim_n \mu^n = \mu$ in \mathbf{P} -probability. We recall that, as stated right after (2.2), $\{\theta_0^j\}_{j \in \mathbb{N}}$ is independent of the driving Brownians and of the graph ξ .

This seems at first rather satisfactory. However in [36] it is discussed at length how this result in reality is, on one hand, surprising and, on the other, that it does not really solve the problem. This can be understood by considering that the *homogeneous degree condition* (2.10) is $\mathbb{P}(d\xi)$ -a.s. verified for ER type graphs when $\liminf_n np_n / \log n$ is larger than a well-chosen constant (see Proposition 1.3 in [36]). But the class of graphs satisfying (2.10) goes well beyond ER graphs: in particular, it is straightforward to construct graphs with an arbitrary number of connected components that satisfy (2.10), see the following remark.

REMARK 1.2. (2.9) and (2.10) are not sufficient to obtain a result in the direction we are aiming at. In fact, if $\xi^{(n)}$ is the graph in which two vertices are connected if and only if they have the same parity (which corresponds to $\lim_n p_n = 1/2$), then, as long as μ_0 is not the uniform measure, one can easily arrange the initial condition in order to have different limit distributions on even and odd sites, or no limit at all. Thus, as $n \rightarrow \infty$, the evolution will not be described by (2.6).

In a nutshell, the results in [36] are obtained under a weak assumption on the graph, but under strong assumptions on the initial condition. And this to the point of obtaining a result that is troublesome: a system with plenty of disconnected components behaves essentially like a totally connected one! Of course the *solution* of this apparent paradox is in the chaotic character of the initial condition that leads to a homogeneous and identical behavior of the initial datum on all components, and the fact that chaos propagates at least on a finite time horizon (see [36] for more on this issue). But there is no reason to expect mean field type behavior, assuming only (2.10) on the graph, without a strong *statistical homogeneity* assumption on the initial datum, as argued in Remark 1.2.

The aim of this paper is to attack the problem assuming only the convergence of the empirical measure of the initial datum, that is (2.9), but assuming that the graph is of ER type. Otherwise said, we want to make a minimal assumption on the initial condition and we try to exploit the *chaoticity* of the graph to achieve the result. We will attack the problem from more than one perspective, not only the direct LLN angle of attack, but also from the Large Deviations (LD) perspective. The vast literature related to our results is presented and discussed after the statements.

2. Main results

Let us denote $d_{\text{bL}}(\cdot, \cdot)$ the bounded Lipschitz distance which endows the weak convergence topology on $\mathcal{P}(\mathbb{R})$ (this choice is somewhat arbitrary: other distances can be used, for example the Wasserstein one, see [41]). By this we mean that $d_{\text{bL}}(\mu, \nu) = \sup_h |\int h d\mu - \int h d\nu|$, where the supremum is taken over $h : \mathbb{R} \rightarrow [0, 1]$ such that $|h(x) - h(y)| \leq |x - y|$.

We are now ready to state the LLN. Recall that μ^n is a random element of $C^0([0, T]; \mathcal{P}(\mathbb{R}))$ and that μ , a non random element of $C^0([0, T]; \mathcal{P}(\mathbb{R}))$, is the unique weak solution of the McKean-Vlasov equation (2.6).

THEOREM 2.1. *Assume that the initial datum is deterministic, that it satisfies (2.9) and, if $\sigma(\cdot) \not\equiv 0$, that it satisfies also that $\int_{\mathbb{R}} x^2 \mu_0(dx) < \infty$. Make the hypothesis that p_n satisfies*

$$\liminf_{n \rightarrow \infty} \frac{p_n^n}{\log n} > 0, \quad (2.12)$$

and either that $0 < \sigma_- \leq \sigma(\cdot) \leq \sigma_+ < \infty$ or $\sigma(\cdot) \equiv 0$. Then $\mathbf{P} \otimes \mathbb{P}$ -a.s. we have that

$$\lim_{n \rightarrow \infty} \mu^n = \mu. \quad \text{in } C^0([0, T]; \mathcal{P}(\mathbb{R})). \quad (2.13)$$

The requirement of deterministic initial data is easily lifted to IID initial conditions under the assumption that they are independent of the graph (and, of course, of the driving Brownians).

From the viewpoint of the proof, Theorem 2.1 may be viewed as two different statements.

- in the case of $\sigma(\cdot) \equiv \sigma \in [0, \infty)$, the proof follows by coupling the system on the ER graph and the system on the complete graph;
- in the case of $0 < \sigma_- \leq \sigma(\cdot) \leq \sigma_+ < \infty$, the result is a corollary of a Large Deviation Principle (LDP) stating that, at the Large Deviations (LD) level, the system on ER graph and the complete graph system are indistinguishable, see Theorem 2.2.

In the next subsection we present the result related to Large Deviations.

2.1. The Large Deviation Principle. Stating the LDP needs some preparation on the general LD approach (classical references are for example [38, 39, 42]).

Given a complete, separable metric space χ , a rate function I is a lower semicontinuous mapping $I : \chi \rightarrow [0, \infty]$ such that each level set $K_l = \{x \in \chi : I(x) \leq l\}$ is compact for all $l \geq 0$

(sometimes I is called a *good* rate function). Given $\{P_n\}_{n \in \mathbb{N}}$ a sequence of probability measures on χ associated with its Borel σ -field, we say that P_n satisfies a LDP (on χ) with rate function I if for every measurable set $A \subset \chi$

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n(A^\circ) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n(\bar{A}) \leq -\inf_{x \in \bar{A}} I(x), \quad (2.14)$$

where A° is the interior of A and \bar{A} is its closure.

Let us now recall that (2.4), or equivalently (2.2) on a complete graph, satisfies a LDP, we refer to Theorem 3.1 in [23]. We choose to state the LDP for the empirical law of the process, that is for

$$\bar{L}_n := \frac{1}{n} \sum_{j=1}^n \delta_{\bar{\theta}^{j,n}} \in \mathcal{P}(C^0([0, T]; \mathbb{R})), \quad (2.15)$$

but other LDP are possible. Namely, Theorem 5.1 in [34] gives a LDP for the empirical measure $\bar{\mu}^n$ seen as an element of $C^0([0, T]; \mathcal{P}(\mathbb{R}))$ (recall (2.8)), yet our result includes this case. In Remark 1.1 we have pointed out the continuity of the projection π_t and how to pass from P_n to $\mu^n \in C^0([0, T]; \mathcal{P}(\mathbb{R}))$, therefore a corollary of a LDP for \bar{L}_n is a LDP on $C^0([0, T]; \mathcal{P}(\mathbb{R}))$ for the law of $\bar{\mu}^n$ with LD functional given by the contraction principle: see for example [33, 76] for the mathematical procedure and [34] for an explicit form of the LD functional in the full generality.

We set $\chi = \mathcal{P}(C^0([0, T]; \mathbb{R}))$; since $C^0([0, T]; \mathbb{R})$ is a metric space, χ is a complete, separable metric space once equipped (among various possibilities) with the bounded Lipschitz distance. Define the probability measure \bar{P}_n on χ by setting $\bar{P}_n(\cdot) := \mathbf{P}(\bar{L}_n \in \cdot)$, of course χ equipped with the σ -algebra of its Borel subsets, then, by Theorem 3.1 in [23], \bar{P}_n satisfies a LDP whose rate function concentrates on $\nu \in \chi$ such that $\nu = \nu \circ \pi_t^{-1} \in C^0([0, T]; \mathcal{P}(\mathbb{R}))$ is solution of the McKean-Vlasov equation (2.6).

We are now ready to state the main result of this subsection. For every realization of the graph ξ define the probability P_n^ξ on χ , by setting $P_n^\xi(\cdot) := \mathbf{P}(L^n \in \cdot)$, where $L^n(\cdot)$ is defined as in (2.15), but replacing $\bar{\theta}^{j,n}$ with $\theta^{j,n}$. In particular, P_n^ξ is the empirical measure of the trajectories $\theta^{j,n}$ solving (2.2).

THEOREM 2.2. *Assume that $\sigma_- > 0$. If ξ is an ER graph that satisfies (2.12) and if the initial datum satisfies (2.9) and $\int x^2 \mu_0(dx) < \infty$, then P_n^ξ satisfies the same LDP of P_n $\mathbb{P}(d\xi)$ -a.s..*

2.2. A look at the literature. We recall that for interacting particle systems on the complete graph, i.e. (2.4), many results on the LLN are available and many of them, as [36], include propagation of chaos properties. However, as already mentioned, propagation of chaos results are very demanding on the initial condition.

The literature is vast and difficult to be properly cited: we mention the seminal contribution [81] and we mention again [88, 98, 85, 56], that are also useful source of more references

and that are not limited to propagation of chaos results, in the sense that also the case of deterministic initial data is treated. For the $\sigma(\cdot) \equiv 0$ case, we mention the important original contributions [41, 87] that gave origin to a vast literature that goes beyond our purposes.

Large deviation properties for mean field diffusions have been studied in the seminal work by Dawson and Gärtner [34], but also in [39, 44, 76] in the so called gradient case. In [23] the problem is attacked in great generality using an approach based on weak convergence and control theory.

The LLN case has already been addressed in the literature, even if few results seem to have been proven so far. As mentioned, [36] proves a LLN for μ^n requiring the initial datum to be a product measure: the case $\sigma(\cdot) \equiv \sigma \geq 0$ is considered. In the same spirit, from the initial datum viewpoint, but for a time-varying graph and for multi-type processes, there is the work of [13]. It is important to mention at this stage that in [13] the interaction is renormalized by the number of neighbors of each site i : we normalize instead by the expected number of neighbors.

Turning to LD results, the recent work of [90] extends the LDP for Hamiltonian systems in random media, presented in [33], to (sparse) random interactions which include symmetric ER random graphs. The convergence of the empirical measure is shown under the assumption $\lim_n np_n = \infty$, without requiring any log divergence. However, they still focus on IID initial conditions and constant diffusion term $\sigma(\cdot) \equiv 1$.

Focusing on the case $\sigma(\cdot) \equiv 0$, we mention the contributions:

- in [21] one finds the stability analysis for the stationary state of an ordinary differential equations system with ER interacting network, requiring a logarithmic divergence of $p_n n$;
- in [27] the Kuramoto model, i.e. $\Gamma(x, y) = \sin(x - y)$ and $F(\cdot)$ is a random constant (*natural frequencies*), is studied with an interaction network that is given by a *graphon*: this leads to a more general limit equation, but their approach includes the case of ER graphs (in this case the graphon is trivial) with p_n that tends to a positive constant. In [82] the case of sparse graphs is considered: for ER graphs the condition is $\lim_n p_n \sqrt{n} = \infty$.

In many of the papers we cite, notably [36, 33, 76, 27, 82], another source of randomness is allowed: for example, in the Kuramoto model this corresponds to the important feature that each oscillator has a priori its own oscillation frequency and, more generally, with this extra source of randomness we can model systems in which the interacting diffusions (or units, agents, ...) are not identical. This source of randomness is chosen independently of the graph and of the dynamical noise. All the results we have presented generalize easily to this case, but at the expense of heavier notations and heavier expressions. We have chosen not to treat this case for sake of conciseness and readability.

The rest of the paper is devoted to the proofs. Section 3 contains the proof of Theorem 2.1 in the case of constant (possibly degenerate) diffusion coefficient. Section 4 contains the proof of Theorem 2.2.

3. Proof: The Law of Large Numbers

This section is devoted to the proof of Theorem 2.1 in the case $\sigma(\cdot) \equiv \sigma \in [0, \infty)$: we recall that the case of non trivial and non degenerate diffusion is a corollary of the LDP (Theorem 2.2). We start with two preliminary lemmas that will be used for Proposition 3.3, from which Theorem 2.1 follows.

LEMMA 3.1. *Let $K > 2$. For all $n \in \mathbb{N}$, it holds*

$$\mathbb{P} \left(\sum_{j=1}^n \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| \geq Kn \right) \leq \exp \left(- \frac{3(K-2)^2}{6 + 2(K-2)} p_n n \right). \quad (2.16)$$

In particular, under hypothesis (2.12) and setting $C := \liminf_{n \rightarrow \infty} \frac{p_n n}{\log n} \in (0, \infty]$, we have that, if $K > K_C := 2 + \frac{2}{3C} + \sqrt{\frac{4}{9C^2} + \frac{4}{C}}$, then $\mathbb{P}(d\xi)$ -a.s. there exists $n_0 = n_0(\xi) < \infty$ such that for $n \geq n_0$

$$\max \left(\sup_{i=1, \dots, n} \sum_{j=1}^n \left| \frac{\xi_{j,i}^{(n)}}{p_n} - 1 \right|, \sup_{i=1, \dots, n} \sum_{j=1}^n \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| \right) \leq Kn. \quad (2.17)$$

PROOF. We use Bernstein's inequality (see for example Corollary 2.11 in [19]) which says that if X_1, \dots, X_n are independent zero-mean random variables such that $|X_j| \leq M$ a.s. for all j , then for all $t \geq 0$

$$\mathbb{P} \left(\sum_{j=1}^n X_j > t \right) \leq \exp \left\{ - \frac{\frac{1}{2}t^2}{\sum_{j=1}^n \mathbb{E}[X_j^2] + \frac{1}{3}Mt} \right\}.$$

Set $X_j = \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| - 2(1 - p_n)$. X_j is a zero-mean random variable and we can bound

$$\mathbb{P} \left(\sum_{j=1}^n \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| \geq Kn \right) \leq \mathbb{P} \left(\sum_{j=1}^n X_j \geq (K-2)n \right). \quad (2.18)$$

We have $|X_j| \leq \max(1/p_n - 3 + 2p_n, 2p_n - 1) \leq 1/p_n =: M$ and $\mathbb{E}[X_j^2] - 1/p_n = -5 + 8p_n - 4p_n^2 \leq -1$, so $\mathbb{E}[X_j^2] \leq 1/p_n$ and Bernstein's inequality together with an union bound show that

$$\mathbb{P} \left(\sup_{i=1, \dots, n} \sum_{j=1}^n \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| \geq Kn \right) \leq n \exp \left(- \frac{3(K-2)^2}{6 + 2(K-2)} p_n n \right). \quad (2.19)$$

The proof is now completed with some elementary computations and by applying the Borel-Cantelli Lemma. \square

LEMMA 3.2. *Assume (2.12) and let*

$$\Delta_i(s) := \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right) \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \right|^2, \quad \text{for every } s \in [0, T]. \quad (2.20)$$

Then, for every realization of the Brownian motions, it holds that

$$\lim_{n \rightarrow \infty} \int_0^T \frac{1}{n} \sum_{i=1}^n \Delta_i(s) ds = 0, \quad \mathbb{P}(d\xi)\text{-a.s.} \quad (2.21)$$

PROOF. First, we rewrite $\int_0^T \Delta_i(s) ds$ as

$$\int_0^T \Delta_i(s) ds = \frac{1}{(np_n)^2} \sum_{j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk}, \quad (2.22)$$

where we have dropped the superscript (n) , the dependency on T and we have introduced the notations

$$\hat{\xi}_{i,j} := \xi_{i,j}^{(n)} - p_n \quad \text{and} \quad d_{ijk} := \int_0^T [\Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{k,n})] ds. \quad (2.23)$$

Observe that $\hat{\xi}_{i,j}$ are centered random variables and $|d_{ijk}| \leq T \|\Gamma\|_\infty^2 =: d_\star$.

Let δ_n be a sequence of positive numbers such that (recall (2.12))

$$\delta_n \gg \frac{1}{p_n n} \quad \text{and} \quad \lim_{n \rightarrow \infty} \delta_n = 0. \quad (2.24)$$

Let Ω_n be the set

$$\Omega_n := \left\{ \xi : \frac{1}{(np_n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} > \delta_n n \right\}, \quad (2.25)$$

We want to show that

$$\sum_{n \in \mathbb{N}} \mathbb{P}(\Omega_n) < \infty. \quad (2.26)$$

Let $K > 2$ and consider the events

$$A_n = \bigcup_{i=1}^n A_{n,i} \quad \text{with} \quad A_{n,i} = \left\{ \xi^{(n)} : \sum_j \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| > Kn \right\}. \quad (2.27)$$

We use

$$\mathbb{P}(\Omega_n) \leq \mathbb{P}(\Omega_n \cap A_n^c) + \mathbb{P}(A_n). \quad (2.28)$$

and Lemma 3.1 ensures that choosing $K > K_C (> 2)$ we have

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty, \quad (2.29)$$

so that one is left with proving that $\sum_{n \in \mathbb{N}} \mathbb{P}(\Omega_n \cap A_n^c) < \infty$. By Markov's inequality applied to $\mathbb{P}(\cdot | A_n^c)$ we see that

$$\mathbb{P}(\Omega_n \cap A_n^c) \leq \exp \left(-n\delta_n + \log \mathbb{E} \left[\mathbf{1}_{A_n^c} \exp \left(\frac{1}{(np_n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] \right). \quad (2.30)$$

Given (2.24), it suffices to show that

$$\log \mathbb{E} \left[\mathbf{1}_{A_n^c} \exp \left(\frac{1}{(np_n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] = n O \left(\frac{1}{p_n n} \right) = O \left(\frac{1}{p_n} \right). \quad (2.31)$$

We exploit the independence w.r.t. i :

$$\mathbb{E} \left[\mathbf{1}_{A_n^c} \exp \left(\frac{1}{(np_n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] = \prod_i \mathbb{E} \left[\mathbf{1}_{A_{n,i}^c} \exp \left(\frac{1}{(np_n)^2} \sum_{j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right], \quad (2.32)$$

and use the inequality $\exp(x) \leq 1 + |x| \exp |x|$ which holds for all $x \in \mathbb{R}$, together with Cauchy-Schwarz and obtain

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{A_{n,i}^c} \exp \left(\frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] \leq \\ & 1 + \mathbb{E} \left[\mathbf{1}_{A_{n,i}^c} \left| \frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right| \exp \left(\left| \frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right| \right) \right] \leq \\ & 1 + \mathbb{E} \left[\left(\frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right)^2 \right]^{1/2} \mathbb{E} \left[\mathbf{1}_{A_{n,i}^c} \exp \left(\frac{2}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right]^{1/2}. \end{aligned} \quad (2.33)$$

Under the condition that we are in $A_{n,i}^c$, it holds that

$$\left| \frac{2}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right| \leq 2K^2 d_* \quad (2.34)$$

so that the exponential expectation can be bounded as $\exp \{2K^2 d_*\}$. Estimating the moment expectation leads to

$$\mathbb{E} \left[\left(\frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right)^2 \right] = \frac{1}{(np_n)^4} \sum_{j,k,p,q} \mathbb{E} \left[\hat{\xi}_{i,j} \hat{\xi}_{i,k} \hat{\xi}_{i,p} \hat{\xi}_{i,q} \right] d_{ijk} d_{ipq} \leq \quad (2.35)$$

$$\leq \frac{d_*^2}{(np_n)^4} [np_n + 3(np_n)^2] \leq \frac{4d_*^2}{(np_n)^2}. \quad (2.36)$$

From (2.33), we get

$$\mathbb{E} \left[\mathbf{1}_{A_{n,i}^c} \exp \left(\frac{1}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] \leq 1 + \frac{2d_*}{np_n} \exp \{2K^2 d_*\}. \quad (2.37)$$

Putting everything back in (2.32), one obtains

$$\mathbb{E} \left[\mathbf{1}_{A_n^c} \exp \left(\frac{1}{(np_n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} d_{ijk} \right) \right] \leq \exp \left\{ \frac{2d_\star}{p_n} \exp \{2K^2 d_\star\} \right\}, \quad (2.38)$$

which gives (2.31). \square

We are now ready for

PROPOSITION 3.3. *If (2.12) holds, then for all $T > 0$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{t \in [0, T]} |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2 = 0, \quad \mathbf{P} \otimes \mathbb{P}\text{-a.s.} \quad (2.39)$$

PROOF. For $i \in \{1, \dots, n\}$, consider

$$\begin{aligned} |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2 &= \\ &2 \int_0^t (\theta_s^{i,n} - \bar{\theta}_s^{i,n}) \left(F(\theta_s^{i,n}) - F(\bar{\theta}_s^{i,n}) + \frac{1}{n} \sum_{j=1}^n \left[\frac{\xi_{i,j}^{(n)}}{p_n} \Gamma(\theta_s^{i,n}, \theta_s^{j,n}) - \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \right] \right) ds \leq \\ &2L_F \int_0^t |\theta_s^{i,n} - \bar{\theta}_s^{i,n}|^2 ds + 2L_\Gamma \frac{1}{n} \sum_{j=1}^n \frac{\xi_{i,j}^{(n)}}{p_n} \int_0^t (|\theta_s^{i,n} - \bar{\theta}_s^{i,n}| + |\theta_s^{j,n} - \bar{\theta}_s^{j,n}|) |\theta_s^{i,n} - \bar{\theta}_s^{i,n}| ds \\ &\quad + 2 \int_0^t \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right) \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \right| |\theta_s^{i,n} - \bar{\theta}_s^{i,n}| ds, \quad (2.40) \end{aligned}$$

which gives

$$\begin{aligned} |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2 &\leq \\ &(2L_F + 1) \int_0^t |\theta_s^{i,n} - \bar{\theta}_s^{i,n}|^2 ds + L_\Gamma \frac{1}{n} \sum_{j=1}^n \frac{\xi_{i,j}^{(n)}}{p_n} \int_0^t [3|\theta_s^{i,n} - \bar{\theta}_s^{i,n}|^2 + |\theta_s^{j,n} - \bar{\theta}_s^{j,n}|^2] ds \\ &\quad + \int_0^t \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right) \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \right|^2 ds. \quad (2.41) \end{aligned}$$

Summing over i and dividing by n , one obtains

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2 &\leq \\ &\leq \left(2L_F + 1 + L_\Gamma \sup_{i=1, \dots, n} \sum_{j=1}^n \frac{3\xi_{i,j}^{(n)} + \xi_{j,i}^{(n)}}{np_n} \right) \int_0^t \frac{1}{n} \sum_{i=1}^n |\theta_s^{i,n} - \bar{\theta}_s^{i,n}|^2 ds \\ &\quad + \frac{1}{n} \sum_{i=1}^n \int_0^t \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right) \Gamma(\bar{\theta}_s^{i,n}, \bar{\theta}_s^{j,n}) \right|^2 ds. \quad (2.42) \end{aligned}$$

In order to bound $\sum_{j=1}^n \frac{3\xi_{i,j}^{(n)} + \xi_{j,i}^{(n)}}{np_n}$ for all $i = 1, \dots, n$, we choose $K > K_C$ and use Lemma 3.1 to obtain that

$$\sup_{i=1, \dots, n} \sum_{j=1}^n \frac{3\xi_{i,j}^{(n)} + \xi_{j,i}^{(n)}}{np_n} \leq 4 + 4K, \quad (2.43)$$

$\mathbb{P}(d\xi)$ -a.s.. The application of Gronwall lemma to

$$S_n(t) = \frac{1}{n} \sum_{i=1}^n |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2, \quad (2.44)$$

leads to

$$S_n(t) \leq \int_0^t \exp\{G(t-s)\} \left(\frac{1}{n} \sum_{i=1}^n \Delta_i(s) \right) ds, \quad (2.45)$$

with $G = 2L_F + 1 + (4 + 4K)L_\Gamma > 0$ and $\Delta_i(s)$ defined in (2.20). Therefore

$$\sup_{t \in [0, T]} S_n(t) \leq \exp\{GT\} \int_0^T \frac{1}{n} \sum_{i=1}^n \Delta_i(s) ds. \quad (2.46)$$

The last estimate is true for all realizations of the Brownian motions. Taking the limit for n which tends to ∞ and integrating the RHS of (2.46), first with respect to \mathbb{P} (recall Lemma 3.2), completes the proof of Proposition 3.3. \square

Proof of Theorem 2.1. Since we already know that $\bar{\mu}^n$ converges \mathbf{P} -a.s. to μ in $C^0([0, T]; \mathcal{P}(\mathbb{R}))$ (see Theorem 1.6 in [56]), it suffices to show that

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq T} d_{\text{bL}}(\mu_t^n, \bar{\mu}_t^n) = 0, \quad \mathbf{P} \otimes \mathbb{P}\text{-a.s.} \quad (2.47)$$

For every $f : \mathbb{R} \rightarrow [0, 1]$ 1-Lipschitz function, we have

$$\left| \int_{\mathbb{R}} f(\theta) (\mu_t^n - \bar{\mu}_t^n) (d\theta) \right| = \left| \frac{1}{n} \sum_{i=1}^n f(\theta_t^{i,n}) - f(\bar{\theta}_t^{i,n}) \right| \leq \frac{1}{n} \sum_{i=1}^n |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|. \quad (2.48)$$

In particular,

$$\sup_{0 \leq t \leq T} d_{\text{bL}}(\mu_t^n, \bar{\mu}_t^n) \leq \sqrt{\frac{1}{n} \sum_{i=1}^n \sup_{0 \leq t \leq T} |\theta_t^{i,n} - \bar{\theta}_t^{i,n}|^2}. \quad (2.49)$$

The proof follows from Proposition 3.3. \square

4. Proof: Large Deviation Principle

The proof of Theorem 2.2 relies on two results that contain most of the work. We first prove Theorem 2.2 assuming these two results, and prove them right after.

Proof of Theorem 2.2. Observe that we can write (2.2) as

$$d\theta_t^{i,n} = F(\theta_t^{i,n}) dt + \frac{1}{n} \sum_{j=1}^n \Gamma(\theta_t^{i,n}, \theta_t^{j,n}) dt + \sigma(\theta_t^{i,n}) c_i(\theta_t^n) dt + \sigma(\theta_t^{i,n}) dB_t^i, \quad (2.50)$$

with

$$c_i(\theta_t^n) := \frac{1}{n\sigma(\theta_t^{i,n})} \sum_{j=1}^n \left(\frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right) \Gamma(\theta_t^{i,n}, \theta_t^{j,n}). \quad (2.51)$$

Recall that \mathbf{P}_n^ξ , respectively \mathbf{P}_n , is the law of the trajectories $\{\theta_t^{i,n}\}_{i=1,\dots,n;t \in [0,T]}$, respectively the law of $\{\bar{\theta}_t^{i,n}\}_{i=1,\dots,n;t \in [0,T]}$. The Radon-Nikodym derivative $d\mathbf{P}_n^\xi/d\mathbf{P}_n$ is $\exp(M_T^n - \langle M^n \rangle_T/2)$ with

$$M_T^n = \sum_{i=1}^n \int_0^T c_i(\theta_t^n) d\theta_t^{i,n} \quad \text{and} \quad \langle M^n \rangle_T = \sum_{i=1}^n \int_0^T c_i^2(\theta_t^n) dt. \quad (2.52)$$

The following lemma is given for every realization of $\xi^{(n)}$ and it has a deterministic nature. Recall that $\chi = \mathcal{P}(C^0([0,T]; \mathbb{R}))$ and \bar{L}_n defined in (2.15). Then $\bar{P}_n(\cdot) := \mathbf{P}(\bar{L}_n \in \cdot)$ is the law of the empirical process associated to (2.4) and $P_n^\xi(\cdot) := \mathbf{P}(L_n \in \cdot)$ is the one associated to (2.2), \bar{P}_n and P_n^ξ are probabilities on χ .

LEMMA 4.1. *Suppose $\bar{P}_n(\cdot)$ satisfies a LDP on χ with rate function I . If, for every $C \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}_n [\exp \{C \langle M^n \rangle_T\}] = 0, \quad (2.53)$$

then $P_n^\xi(\cdot)$ satisfies a LDP on χ with the same rate function as \bar{P}_n .

Since we want the LDP to hold $\mathbb{P}(d\xi)$ -a.s., we need to show that condition (2.53) holds in this sense. To this aim, we redefine the sets Ω_n given in (2.25), as

$$\Omega_n := \left\{ \xi : \frac{1}{n} \log \mathbf{E}_n [\exp \{C_n \langle M^n \rangle_T\}] > \delta_n \right\}, \quad (2.54)$$

where δ_n and $1/C_n$ tend to zero: they have to do so in a slow way and arbitrarily slow will do for us (explicit choices will be given at the end of the proof).

We need also

$$\Omega^* = \left\{ \xi : \text{there exists } n_0 \text{ s.t. } \frac{1}{n} \log \mathbf{E}_n [\exp \{C_n \langle M^n \rangle_T\}] \leq \delta_n \text{ for every } n \geq n_0 \right\}. \quad (2.55)$$

LEMMA 4.2. *Assuming (2.12) we have that $\mathbb{P}(\Omega^*) = 1$.*

One readily sees that Lemma 4.2 provides the missing ingredient and the proof of Theorem 2.2 is complete. \square

Proof of Lemma 4.1. Recall (2.50)-(2.52). We have to show that (2.14) holds. Consider A a measurable set and recall that A° is the interior of A and \bar{A} is its closure.

Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$P_n^\xi(A^\circ) = \mathbf{P}_n^\xi \left(\{\mu_t^n\}_{t \in [0, T]} \in A^\circ \right) = \mathbf{E}_n \left[1_{\{\mu_t^n\}_{t \in [0, T]} \in A^\circ} \exp \left\{ M_T^n - \frac{1}{2} \langle M^n \rangle_T \right\} \right] \quad (2.56)$$

and Hölder inequality gives

$$P_n^\xi(A^\circ) \geq (\bar{P}_n(A^\circ))^p \left(\mathbf{E}_n \left[\exp \left\{ -\frac{q}{p} M_T^n - \frac{1}{2} \frac{q}{p} \langle M^n \rangle_T \right\} \right] \right)^{-\frac{p}{q}}. \quad (2.57)$$

Now observe that Cauchy-Schwarz inequality together with the fact that an exponential martingale has expectation less or equal to 1 (see Theorem 5.2 in [60]) imply

$$\mathbf{E}_n \left[\exp \left\{ -\frac{q}{p} M_T^n - \frac{1}{2} \frac{q}{p} \langle M^n \rangle_T \right\} \right] \leq \mathbf{E}_n \left[\exp \left\{ \left(\frac{2q^2}{p^2} + \frac{q}{p} \right) \langle M^n \rangle_T \right\} \right]^{\frac{1}{2}}. \quad (2.58)$$

Hence

$$P_n^\xi(A^\circ) \geq (\bar{P}_n(A^\circ))^p \left(\mathbf{E}_n \left[\exp \left\{ \left(\frac{2q^2}{p^2} - \frac{q}{p} \right) \langle M^n \rangle_T \right\} \right] \right)^{-\frac{p}{2q}}. \quad (2.59)$$

In particular, one obtains

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n^\xi(A^\circ) \geq -p \inf_{x \in A^\circ} I(x) - \frac{p}{2q} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}_n [\exp \{C \langle M^n \rangle_T\}], \quad (2.60)$$

with $C = \left(\frac{2q^2}{p^2} - \frac{q}{p} \right)$. By hypothesis the second term on the right is zero and since

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_n^\xi(A^\circ) \geq -p \inf_{x \in A^\circ} I(x), \quad (2.61)$$

is true for all $p > 1$, the lower bound in (2.14) is established.

The upper bound is almost the same: let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Similarly

$$P_n^\xi(\bar{A}) \leq (\bar{P}_n(\bar{A}))^{\frac{1}{p}} \left(\mathbf{E}_n \left[\exp \left\{ q M_T^n - \frac{1}{2} q \langle M^n \rangle_T \right\} \right] \right)^{\frac{1}{q}}, \quad (2.62)$$

and, using the properties of exponential martingales as in (2.58), one gets

$$P_n^\xi(\bar{A}) \leq (\bar{P}_n(\bar{A}))^{\frac{1}{p}} \left(\mathbf{E}_n [\exp \{(2q^2 - q) \langle M^n \rangle_T\}] \right)^{\frac{1}{2q}}. \quad (2.63)$$

Finally, the desired inequality reads

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_n^\xi(\bar{A}) \leq -\frac{1}{p} \inf_{x \in \bar{A}} I(x) + \frac{1}{2p} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}_n [\exp \{C \langle M^n \rangle_T\}], \quad (2.64)$$

with $C = 2q^2 - q$. And we conclude as before. \square

Proof of Lemma 4.2. We want to show that

$$\sum_{n \in \mathbb{N}} \mathbb{P}(\Omega_n) < \infty. \quad (2.65)$$

As in the proof of Lemma 3.2, let $K > 2$ and consider the events A_n defined in (2.27),

$$A_n = \bigcup_{i=1}^n A_{n,i} \quad \text{with} \quad A_{n,i} = \left\{ \xi^{(n)} : \sum_j \left| \frac{\xi_{i,j}^{(n)}}{p_n} - 1 \right| > Kn. \right\}.$$

Following the proof of Lemma 3.1, we use $\mathbb{P}(\Omega_n) \leq \mathbb{P}(\Omega_n \cap A_n^c) + \mathbb{P}(A_n)$ and (2.29), i.e. $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$, so that one is left with proving that $\sum_{n \in \mathbb{N}} \mathbb{P}(\Omega_n \cap A_n^c) < \infty$.

By Markov's inequality applied to $\mathbb{P}(\cdot | A_n^c)$ we see that

$$\mathbb{P}(\Omega_n \cap A_n^c) \leq \exp\left(-n\delta_n + \log \mathbf{E}\mathbf{E}[\mathbf{1}_{A_n^c} \exp(C_n \langle M^n \rangle_T)]\right). \quad (2.66)$$

so it suffices to show that

$$\log \mathbf{E}\mathbf{E}[\mathbf{1}_{A_n^c} \exp(C_n \langle M^n \rangle_T)] = o(n\delta_n). \quad (2.67)$$

To lighten the notation we go back to using the centered random variables $\hat{\xi}_{i,j} := \xi_{i,j}^{(n)} - p_n$ (cf. (2.23)). With these notations, $\langle M^n \rangle_T$ can be rewritten as

$$\langle M^n \rangle_T = \frac{1}{(p_n n)^2} \sum_{i,j,k=1}^n \hat{\xi}_{i,j} \hat{\xi}_{i,k} c_{ijk}, \quad (2.68)$$

where

$$c_{ijk} = \int_0^T \frac{1}{\sigma^2 (\theta_t^{i,n})} \Gamma(\theta_t^{i,n}, \theta_t^{j,n}) \Gamma(\theta_t^{i,n}, \theta_t^{k,n}) dt. \quad (2.69)$$

Observe that $|c_{ijk}| \leq c_*$ given the boundness of Γ and the conditions on σ .

The estimation of (2.67) is exactly the same as in (2.31), where d_{ijk} are replaced by $C_n c_{ijk}$ (and d_* by $C_n c_*$). Following the same strategy, we get

$$\prod_i \mathbf{E} \left[\mathbf{1}_{A_{n,i}^c} \exp\left(\frac{C_n}{(np_n)^2} \sum_{j,k} \hat{\xi}_{i,j} \hat{\xi}_{i,k} c_{ijk}\right) \right] \leq \left(1 + \frac{2C_n}{np_n} \exp\{2C_n K^2 c_*\}\right)^n. \quad (2.70)$$

Therefore

$$\log \mathbf{E}\mathbf{E} \left[\mathbf{1}_{A_n^c} \exp(C_n \langle M^n \rangle_T) \right] \leq \frac{2C_n}{p_n} \exp\{2C_n K^2 c_*\}. \quad (2.71)$$

Which gives (2.67) when $C_n = o(\log(np_n))$ and $\frac{1}{\delta_n} = o\left(\frac{np_n}{\exp\{cC_n\}}\right)$ with $c > 2K^2 c_*$: choose, for example, $C_n = \sqrt{\log(np_n)}$ and $\delta_n = \frac{1}{\sqrt{np_n}}$. \square

Weakly interacting oscillators on dense random graphs

This chapter presents the joint work with Gianmarco Bet and Francesca Nardi [11].

1. Introduction, organization and set-up

In the last twenty years there has been a growing interest in complex networks and inhomogeneous particle systems. The classical mean-field framework (e.g., [88, 98]) in which the particles are all connected with each other, has been extended to include interactions described by general networks. In these more general models, the interaction between two particles depends on the weight of the edge connecting the two in the underlying network, see, e.g., [4, 94].

The first mathematically rigorous results appeared only recently [13, 36]. They consider weakly interacting particle systems defined on certain graph sequences. They show that, under suitable conditions on the degrees, the system converges to the classical mean-field behavior in the limit as the number of particles tends to infinity. However, these works leave several relevant questions unanswered: is it possible to characterize the graph sequences for which the system converges to the mean-field limit? How sensitive are the dynamics to the degree inhomogeneity in the underlying graph? How does the graph structure affect the long-time behavior? See also [30, 31, 77].

We address these questions by considering a system of weakly interacting oscillators, i.e., functions taking values in the one-dimensional torus. The interactions between the particles are encoded in a general random graph sequence, meaning that two particles are interacting if and only if they are connected in the underlying graph. Our main object of study is the empirical measure associated to these systems. We rely on the recent graphon theory for the notion of graph convergence and graph limit. Graphons, a generalization of dense graph sequences, have proven to be useful in a variety of contexts, from extremal graph theory to statistical mechanics (see the monograph [73]). Recently, they have also been employed in mean-field games theory, see [24, 25] and references therein.

The main result of this work is a Law of Large Numbers for the empirical measure. More precisely, if the underlying graph sequence converges to some (possibly random) graphon, then we characterize the limit of the empirical measure as the solution to a non-linear Fokker-Planck equation suitably weighted by the corresponding graph limit. We do not impose any regularity condition on the graph sequence, which can be deterministic or random, nor on the limiting graphon, which can also be random. Notably, our analysis includes pseudo-random graphs (see, e.g., [5, 28]) and exchangeable random graphs (see, e.g., [40]).

As a byproduct, we present a characterization of deterministic and random graph sequences for which the behavior of the empirical measure is approximately mean-field. Furthermore, we show that the map associating to each graphon the solution to the corresponding Fokker-Planck equation is Hölder-continuous. The continuity is obtained with respect to the cut-distance on the space of graphons and a classical Wasserstein distance on the space of trajectories.

Weakly interacting particle systems on graph sequences converging to graphons have already been considered in a series of works, both in the stochastic setting [6, 77, 90] and in the deterministic one [27, 82]. However, all the models proposed so far are based on *labeled* graphons and do not address the graph convergence in the natural topology of graph limits theory. Existing proofs always work under somewhat stringent regularity assumptions on the limiting graphon, which in any case has to be deterministic, and they are not able to deal with general graph sequences as we do.

Our work stems from the fact that the empirical measure of a particle system is invariant under relabeling of the particles and, thus, that its law should depend on an *unlabeled* graphon. In fact, unlabeled graphons represent a building block of graph limits theory and are formally obtained as certain equivalence classes of labeled graphons. They are, in general, very irregular objects so that one of the main difficulties towards our result is to deal with functions that are only measurable. By taking independent and identically distributed initial conditions, we are able to exploit the symmetry property of the system together with the key ingredients of graphon theory, i.e., exchangeability and random sampling, and to obtain a convergence result in the natural space of graph limits. To our knowledge, the results presented here appear to be the first in the literature to tackle interacting particle systems on unlabeled graphons. Moreover, we are also able to include the case of random graphons, an aspect that has never been addressed so far.

Finally, whenever exchangeable graphs are considered, we establish a propagation of chaos result. The non-linear process that describes the behavior of a tagged particle is written down explicitly and compared with the existing characterization present in the literature.

1.1. A look at the literature. Weakly interacting particle systems on graphs have first been studied in [13, 36], where the convergence to the classical mean-field system is shown under some homogeneity property of the degrees and under independence of the initial conditions. The work [31] addresses sequences of Erdős-Rényi random graphs and establishes a Law of Large Numbers and a Large Deviation Principle by only assuming that the initial empirical measure converges weakly.

The works [6, 77, 90] deal with more general sequences of graphs and take into account a few notions coming from graph limits theory. Namely, [90] establishes a Large Deviation Principle for the empirical measure of weakly interacting particles on W -random graphs, see (3.13) for the precise definition. The works [6, 77] present Law of Large Numbers results and

consider converging graph sequences in the space of labeled graphons, although with respect to different metrics and including unbounded graphons.

For deterministic particle systems, Medvedev and coauthors consider the Kuramoto model on a variety of graph sequences arising from labeled graphons, we refer to [27, 82] and references therein.

To the authors' knowledge, the only work addressing the long-time behavior of interacting particle systems on graphs is given by [30], where the Kuramoto model defined on pseudo-random graphs is shown to be close to the mean-field behavior on long time scales. See Subsection 2.3 for more on pseudo-random graph sequences.

Recently mean-field game theoretical models defined on graphons have been proposed, we refer to [24, 25] and references therein.

Most of the cited works consider both the dense regime (the number of edges is roughly proportional to the square of the number n of vertices) as well as intermediate regimes between sparse and dense (the number of edges grows strictly faster than n but not necessarily as fast as n^2). Although the results in [6, 27, 77, 82, 90] allow for random graph sequences, it is always assumed that the limiting graphon is deterministic.

1.2. Organization. We now present the set-up and notation used, and recall the distances between probability measures that will be used in the sequel.

In Section 2 we define the interacting particle system and the associated non-linear process. Existence, uniqueness and stability results for the non-linear process are presented right-after, see Propositions 2.1 and 2.2. Our main result, Theorem 2.3, is given in Subsection 2.2. Exchangeable random graphs are then discussed together with a propagation of chaos result; see Corollary 2.6 and Proposition 2.7 respectively. Subsection 2.3 is devoted to the comparison with the classical mean-field behavior and to illustrate a few important consequences of Theorem 2.3; the discussion is supported by two explanatory examples.

In Section 3 we focus on the non-linear process. In particular, we discuss its relationship with other characterizations already known in the literature. The proofs of Propositions 2.1 and 2.2 are given in Subsection 3.4.

Section 4 contains the proof of Theorem 2.3. Finally, in Appendix 3.A we collect the most important known results on graphons and we derive a characterization of convergence in probability for random graph sequences.

1.3. Setting and notations. We consider particle dynamics occurring on a finite time interval, say $[0, T]$, which we fix once and for all. We work on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, P)$, where $\{\mathcal{F}_t\}$ is a filtration satisfying the usual conditions. All Brownian motions that we consider later on are adapted to $\{\mathcal{F}_t\}_{t \geq 0}$ and are independent of the other random variables.

We use two different notations for expressing conditional probabilities: the one referring to Brownian motions and initial conditions is denoted by \mathbf{P} , its expectation by \mathbf{E} ; the one

referring to the randomness in the graph sequences, and/or in its limit object, is denoted by \mathbb{P} , its expectation by \mathbb{E} . When not explicitly written, if a result holds in \mathbf{P} -probability, it means that it holds \mathbb{P} -a.s., and viceversa.

The interval $I := [0, 1]$ represents the space of (continuous) labels. The oscillators are functions with values in the one-dimensional torus $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$, so that their trajectories are random variables defined on the space of continuous functions with values in \mathbb{T} , i.e., on $\mathcal{C}([0, T], \mathbb{T})$, endowed with the supremum norm.

For two probability measures $\bar{\mu}, \bar{\nu} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$, we define their distance by

$$D_T(\bar{\mu}, \bar{\nu}) := \inf_{m \in \gamma(\bar{\mu}, \bar{\nu})} \left\{ \int \sup_{t \in [0, T]} |x_t - y_t|^2 m(dx, dy) \right\}^{1/2}, \quad (3.1)$$

where $\gamma(\bar{\mu}, \bar{\nu})$ is the space of probability measures on $\mathcal{C}([0, T], \mathbb{T}) \times \mathcal{C}([0, T], \mathbb{T})$ with first marginal equal to $\bar{\mu}$ and second marginal equal to $\bar{\nu}$. This definition coincides with the 2-Wasserstein distance between probability measures. The right-hand side of (3.1) can be rewritten as

$$D_T(\bar{\mu}, \bar{\nu}) = \inf_{X, Y} \left\{ \mathbf{E} \left[\sup_{t \in [0, T]} |X_t - Y_t|^2 \right] : \mathcal{L}(X) = \bar{\mu}, \mathcal{L}(Y) = \bar{\nu} \right\}^{1/2} \quad (3.2)$$

where the infimum is taken on all random variables X and Y with values in $\mathcal{C}([0, T], \mathbb{T})$ and law \mathcal{L} equal to $\bar{\mu}$ and $\bar{\nu}$ respectively. From (3.1) we obtain that for every $s \in [0, T]$

$$\sup_f \left| \int_{\mathbb{T}} f(\theta) \bar{\mu}_s(d\theta) - \int_{\mathbb{T}} f(\theta) \bar{\nu}_s(d\theta) \right| \leq D_s(\bar{\mu}, \bar{\nu}), \quad (3.3)$$

where the supremum is taken over all Lipschitz functions from \mathbb{T} to \mathbb{R} . Observe that these definitions make sense also with $T = 0$ and $\mathcal{C}([0, T], \mathbb{T})$ replaced by \mathbb{T} .

For a brief overview of the theory of graphons and graph limits, we refer to Appendix 3.A. We follow the notation of [73], the notions of labeled and unlabeled graphs are taken from [40], as well as the notion of convergence in probability for a sequence of random graphs. Note that a sequence of graphs will always be considered convergent in the sense of graph limits. We emphasize that what is usually referred to in the literature as *graphon* is referred to here as a *labeled graphon*, and its equivalence class, i.e., an unlabeled graphon in the notation of [73], is simply referred to as graphon.

The various constants throughout the paper are always denoted by C or C' and may vary from line to line. An explicit dependence on a parameter α will be denoted by C_α .

2. The models and main results

2.1. The models. We introduce the two main models: a weakly interacting particle system (3.4) and a non-linear process (3.6).

2.1.1. *Weakly interacting oscillators on graphs.* Let $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ be a sequence of undirected, labeled graphs. For $n \in \mathbb{N}$, the adjacency matrix of $\xi^{(n)}$ is given by the $n \times n$ symmetric

matrix $\{\xi_{ij}^{(n)}\}_{i,j=1,\dots,n}$ where $\xi_{ij}^{(n)}$ takes value 1 whenever the vertices i and j are connected and 0 otherwise.

Let $\{\theta^{i,n}\}_{i=1,\dots,n}$ be the family of oscillators on \mathbb{T}^n that satisfy

$$\begin{cases} d\theta_t^{i,n} = F(\theta_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \xi_{ij}^{(n)} \Gamma(\theta_t^{i,n}, \theta_t^{j,n})dt + dB_t^i, & 0 < t < T, \\ \theta_0^{i,n} = \theta_0^i, & i \in \{1, \dots, n\}, \end{cases} \quad (3.4)$$

where F and Γ are bounded uniformly Lipschitz functions and $\{B^i\}_{i \in \mathbb{N}}$ a sequence of independent and identically distributed (IID) Brownian motions on \mathbb{T} . The initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ are IID random variables sampled from some probability distribution $\bar{\mu}_0 \in \mathcal{P}(\mathbb{T})$ which is fixed once for all.

We are interested in studying the empirical measure associated to (3.4). This is defined as the (random) probability measure on \mathbb{T} such that

$$\mu_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{\theta_t^{j,n}}, \quad (3.5)$$

for every $t \in [0, T]$.

Many interesting examples of interacting oscillators such as the Kuramoto model, the plane rotator model and other generalizations fit this framework, see e.g., [36, §1.2], [5] and Subsection 2.3.

2.1.2. The non-linear process. The results of this subsection are proven in Section 3, together with the comparison to other existing formulations in the literature. The graphon framework is briefly recalled in Appendix 3.A.

Fix a graphon $W \in \widetilde{\mathcal{W}}_0$ and a uniform random variable U on I . Consider the solution $\theta = \{\theta_t\}_{t \in [0, T]}$ to the following system

$$\begin{cases} \theta_t = \theta_0 + \int_0^t F(\theta_s)ds + \int_0^t \int_I W(U, y) \int_{\mathbb{T}} \Gamma(\theta_s, \theta) \mu_s^y(d\theta) dy ds + B_t, \\ \mu_t^y = \mathcal{L}(\theta_t | U = y), \quad \text{for } y \in I, t \in [0, T], \end{cases} \quad (3.6)$$

where $\mathcal{L}(\theta_0) = \bar{\mu}_0$ and B is a Brownian motion independent of the previous sequence $\{B^i\}_{i \in \mathbb{N}}$. We take U to be independent of all the randomness in the system and, in particular, of the initial condition θ_0 .

The next proposition establishes the existence of the solution to equation (3.6) and its uniqueness. In Section 3, we prove the well-posedness of equation (3.6) with respect to W , i.e., the law of θ does not depend on the representative of W in the space of labeled graphons \mathcal{W}_0 , see Remark 3.4.

PROPOSITION 2.1. *For every uniform random variable U on I independent from all other randomness, there exists a unique solution to (3.6). If $\bar{\mu} \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$ denotes its law and μ^x the law of θ conditioned on $U = x$, then $\bar{\mu}$ solves the following non-linear Fokker-Planck*

equation in the weak sense

$$\partial_t \bar{\mu}_t(\theta) = \frac{1}{2} \partial_\theta^2 \bar{\mu}_t(\theta) - \partial_\theta [\bar{\mu}_t(\theta) F(\theta)] - \partial_\theta \left[\int_{I \times I} W(x, y) \mu_t^x(\theta) \int_{\mathbb{T}} \Gamma(\theta, \tilde{\theta}) \mu_t^y(d\tilde{\theta}) dy dx \right] \quad (3.7)$$

with initial condition $\bar{\mu}_0 \in \mathcal{P}(\mathbb{T})$.

Recall that δ_\square defines a metric in the space of graphons $\widetilde{\mathcal{W}}_0$, see (3.84). We have the following Hölder continuity result for $\bar{\mu}$ with respect to W .

PROPOSITION 2.2. *Assume that $\Gamma \in \mathcal{C}^{1+\varepsilon}(\mathbb{T}^2)$ for some $\varepsilon > 0$. There exists a positive constant C such that, if $\bar{\mu}^W$ and $\bar{\mu}^V$ denote the laws of the solutions to equation (3.6) associated with graphons W and V respectively, then*

$$D_T(\bar{\mu}^W, \bar{\mu}^V) \leq C \delta_\square(W, V)^{1/2}. \quad (3.8)$$

The proof is postponed to Section 3.4. Note that taking the p -Wasserstein distance in (3.1) for $p \geq 1$ leads to a Hölder exponent as large as $1/p$.

Propositions 2.1 and 2.2 imply that the following mapping is continuous:

$$\begin{aligned} \Psi : (\widetilde{\mathcal{W}}_0, \delta_\square) &\rightarrow (\mathcal{C}([0, T], \mathcal{P}(\mathbb{T})), D_T) \\ W &\mapsto \bar{\mu}^W, \end{aligned} \quad (3.9)$$

where $\bar{\mu}^W$ is the law of θ solving equation (3.6) with graphon W .

In particular, for every random variable W in $\widetilde{\mathcal{W}}_0$ – which is well defined since $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a compact metric space – it corresponds a random variable $\bar{\mu}^W$ with values in $\mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$, i.e., for almost every $\omega \in \Omega$, $\bar{\mu}^W(\omega) = \bar{\mu}^{W(\omega)}$.

2.2. Convergence of empirical measures. We are now able to present our main result. Afterwards, we present an application to exchangeable random graphs and a propagation of chaos result.

THEOREM 2.3. *Let $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ be a sequence of random graphs. Assume that there exists a random variable W in $\widetilde{\mathcal{W}}_0$ to which $\xi^{(n)}$ converges in \mathbb{P} -probability, or equivalently such that*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\delta_\square(\xi^{(n)}, W)] = 0. \quad (3.10)$$

If the initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ are independent of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$, then

$$\mu^n \longrightarrow \bar{\mu}, \quad \text{in } \mathbf{P} \times \mathbb{P}\text{-probability, as } n \rightarrow \infty, \quad (3.11)$$

where the convergence is in $\mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$ and $\bar{\mu}$ is a random variable depending only on the randomness of W , i.e., for almost every $\omega \in \Omega$, $\bar{\mu}(\omega)$ solves equation (3.7) starting from $\bar{\mu}_0$, with graphon $W(\omega)$.

Condition (3.10) extends the convergence of graph sequences to the convergence in probability in $\widetilde{\mathcal{W}}_0$. In particular, Theorem 2.3 also holds in case the graphs are deterministic or take

values in $[0, 1]$ rather than $\{0, 1\}$. The equivalence between condition (3.10) and the convergence in probability for random graph sequences is proven in Lemma 3.A.2.

One may wonder if the convergence of μ^n holds under weaker conditions on the initial data: to the authors' knowledge, exchangeability is a necessary requirement so as to deal with unlabeled graphons; we refer to Section 3 for more on this aspect.

Looking at the proof of Theorem 2.3, we remark that, if the graphon W is deterministic, the initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ can depend on the graph sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$. In other words, Theorem 2.5 remains true if one requires $\{\theta_0^i\}_{i \in \mathbb{N}}$ to be independent of the randomness in W but not necessary on the whole sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$. The relationship between the randomness left in W and the one present in $\xi^{(n)}$ is further discussed in Subsection 2.3.

2.2.1. *Applications to exchangeable graphs.* Recall that an exchangeable random graph $\xi = \{\xi_{ij}\}_{i,j \in \mathbb{N}}$ (see [73]) is a infinite array of binary random variables, such that

$$\mathbb{P}(\xi_{ij} = e_{ij}, 1 \leq i, j \leq n) = \mathbb{P}(\xi_{ij} = e_{\sigma(i)\sigma(j)}, 1 \leq i, j \leq n) \quad (3.12)$$

for all $n \in \mathbb{N}$, all permutations σ on n elements and all $e_{ij} \in \{0, 1\}$. This definition coincides with the definition of jointly exchangeable binary random variables, see [40].

REMARK 2.4. Any finite deterministic graph ξ leads to an exchangeable random graph by performing a uniform random sampling on its associated graphon W_ξ , see (3.82) and [73, §10].

More generally, for $W \in \widetilde{\mathcal{W}}_0$ one may construct an exchangeable random graph ξ^W , usually called W -random graph, defined for i and j in \mathbb{N} by

$$\xi_{ij}^W = W(U_i, U_j), \quad (3.13)$$

where $\{U_i\}_{i \in \mathbb{N}}$ is a sequence of IID uniform random variables on I . The next theorem shows that the converse statement is also true: every exchangeable random graph can be obtained in this way, provided that W is random.

The characterization of exchangeable random graphs is a consequence of the works of Hoover, Aldous and Kallenberg; see [40] and references therein. We recall their main result here.

THEOREM 2.5 ([40, Theorem 5.3] and [73, Theorem 11.52]). *Let $\xi = \{\xi_{ij}\}_{i,j \in \mathbb{N}}$ be an exchangeable random graph. Then, ξ is a W -random graph for some random $W \in \widetilde{\mathcal{W}}_0$.*

Moreover, let $\xi^{(n)} := \{\xi_{ij}\}_{i,j=1,\dots,n}$ for every $n \in \mathbb{N}$. It holds that

$$\xi^{(n)} \longrightarrow W \quad \mathbb{P}\text{-a.s. in } \widetilde{\mathcal{W}}_0, \quad (3.14)$$

as $n \rightarrow \infty$.

We are now ready to state the main corollary of Theorem 2.3, which deals with exchangeable random graphs.

COROLLARY 2.6. *Let $\xi = \{\xi_{ij}\}_{i,j \in \mathbb{N}}$ be an exchangeable random graph and let W be the limit of $\xi^{(n)} := \{\xi_{ij}\}_{i,j=1,\dots,n}$ in the sense of Theorem 2.5. Assume that the initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ are independent of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$, then*

$$\mu^n \longrightarrow \bar{\mu}, \quad \text{in } \mathbf{P} \times \mathbb{P}\text{-probability, as } n \rightarrow \infty, \quad (3.15)$$

where $\bar{\mu}$ is the solution to (3.7) starting from $\bar{\mu}_0$ with graphon W .

2.2.2. *Propagation of Chaos.* Whenever $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of exchangeable graphs¹, the particles $\{\theta^{i,n}\}_{i=1,\dots,n}$ are exchangeable as well and, in particular, their joint distribution is symmetric, i.e., invariant under permutation of the labels. A classical result by Sznitman [98, Proposition 2.2] is that the Law of Large Numbers for the empirical measure of a symmetric joint distribution of particles is equivalent to the propagation of chaos property.

From equation (3.11), we can thus deduce a propagation of chaos statement for the particle system (3.4). This is illustrated in the next proposition.

PROPOSITION 2.7. *If $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of exchangeable graphs, then for every $k \in \mathbb{N}$,*

$$\lim_{n \rightarrow \infty} \mathcal{L}(\theta^{1,n}, \dots, \theta^{k,n}) = \prod_{i=1}^k \mathcal{L}(\theta) = \prod_{i=1}^k \bar{\mu}. \quad (3.16)$$

We omit the proof of Proposition 2.7.

2.3. Mean-field behavior and two explanatory examples. Theorem 2.3 allows for a better understanding of the relationship between random graph sequences and the behavior of the empirical measure. More precisely:

- (1) It highlights the difference between the randomness present in the graph $\xi^{(n)}$ for every $n \in \mathbb{N}$ and the one left in the limit W ;
- (2) It presents a new class of random Fokker-Planck equations as possible limit descriptions for the empirical measure μ^n .

As a byproduct, it allows to derive a precise characterization of the graph sequences for which the empirical measure limit is mean-field. Let us recall what we mean by mean-field limit and first discuss this last issue; we then address (1) and (2) with the help of two examples.

Consider system (3.4) on a sequence of complete graphs, i.e., $\xi_{ij}^{(n)} \equiv 1$ for every i, j and n . It is well known (e.g., [88, 98]) that the empirical measure μ^n converges to the *mean-field limit* $\rho \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$, defined as the unique solution to the following McKean-Vlasov equation:

$$\partial_t \rho_t(\theta) = \frac{1}{2} \partial_\theta^2 \rho_t(\theta) - \partial_\theta [\rho_t(\theta) F(\theta)] - p \partial_\theta \left[\rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \tilde{\theta}) \rho_t(d\tilde{\theta}) \right], \quad (3.17)$$

¹i.e. for each $n \in \mathbb{N}$ the random variables $\{\xi_{ij}^{(n)}\}_{i,j=1,\dots,n}$ are exchangeable. Observe that ξ is not necessarily an exchangeable random graph as in (3.12).

with initial condition $\bar{\mu}_0$ and $p = 1$. Existence and uniqueness for the solution to (3.17) hold under our assumptions on F , Γ and $\bar{\mu}_0$, see e.g., [88, 98].

Suppose that the graph sequence is converging to a deterministic limit; the case of a random limit is discussed after the next example. Theorem 2.3 implies that for every sequence $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ which converges to some flat graphon $W \equiv p \in [0, 1]$, the empirical measure μ^n satisfies equation (3.17) with corresponding p . Since the convergence of $\xi^{(n)}$ to a non-constant graphon gives rise to equation (3.7), which is – at least formally – different from (3.17), we conclude that the limit of μ^n is mean-field if and only if the sequence $\xi^{(n)}$ converges to a constant graphon. The graphs with such asymptotic behavior are known in the literature as pseudo-random graphs, see [5, 28] and [73, §11.8.1].

We now address the issues (1) and (2) with two explanatory examples. The mean-field comparison when the graph limit is random is discussed after the first example.

2.3.1. *Example I: W -random graphs.* Fix $p \in (0, 1)$ and let g be a random variable on $(0, 1)$ with mean \sqrt{p} and distribution function given by F_g . Let $\{g_i\}_{i \in \mathbb{N}}$ be a sequence of IID copies of g . Conditionally on $\{g_i\}_{i \in \mathbb{N}}$, $\xi_{ij}^{(n)}$ is defined as

$$\xi_{ij}^{(n)} \sim \text{Ber}(g_i g_j), \quad \text{independently for each } 1 \leq i < j \leq n. \quad (3.18)$$

The graph $\xi^{(n)}$ is the dense analogue of the inhomogeneous random graph, also known as rank-1 model, see e.g., [12, 15]. In this model, g_i corresponds to the weight associated with particle i and, loosely speaking, the closer g_i is to 1, the more connections particle i forms. We expect that assigning different distributions to g leads to different behaviors for the empirical measure (3.5).

The construction made in (3.18) returns a binary array $\{\xi_{ij}^{(n)}\}_{i,j=1,\dots,n}$ of exchangeable random variables. In particular, they all have the same expected value $\mathbb{E}[\xi_{ij}^{(n)}] = p$, for every distinct pair of i and j . We are interested in comparing the empirical measure of system (3.4) defined on the graph (3.18), to the one of the corresponding *annealed* system. This last one is obtained from (3.4) by replacing $\xi_{ij}^{(n)}$ with their expected values, i.e., it is given by the solution to

$$d\theta_t^{i,n} = F(\theta_t^{i,n})dt + \frac{p}{n} \sum_{j=1}^n \Gamma(\theta_t^{i,n}, \theta_t^{j,n})dt + dB_t^i, \quad (3.19)$$

for which the asymptotic behavior is known to be the mean-field limit (3.17).

Perhaps surprisingly, the behavior of system (1.13) on the graph sequence (3.18) is formally described in the limit by (3.19) only when g is deterministic and $g = \sqrt{p}$. Recall the definition of W -random graph given in (3.13): we see that $\xi^{(n)}$ is a W_g -random graph with

$$W_g(x, y) = F_g^{-1}(x) F_g^{-1}(y), \quad \text{for } x, y \in I, \quad (3.20)$$

where F_g^{-1} is the pseudo inverse of F_g . In particular, the \mathbb{P} -a.s. limit of $\xi^{(n)}$ is given by W_g and thus the limit of μ^n by the solution to equation (3.7) with $W = W_g$. Theorem 2.3 and

Proposition 2.2 imply that the empirical measure of the system associated to $\xi^{(n)}$ is arbitrarily close to the mean-field limit of the annealed system (3.19) if W_g is arbitrarily close to the constant graphon p in the cut-distance, i.e., if $\text{Var}[g] \ll 1$. In this case, $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ is close to an Erdős-Rényi graph sequence, for which the mean-field behavior is already known, see [31]. We point out that we are not able to say whether two different graphons can lead to two solutions which are close as probability measures. This aspect may be model-dependent and needs further developments.

Finally, observe that by choosing a suitable deterministic sequence of the weights $\{g_i\}_{i \in [n]}$, e.g., $g_i = F_g^{-1}(i/n)$ for $i \in [n]$, would lead to a random graph $\xi^{(n)}$ which is not exchangeable. In particular, $\mathbb{E}[\xi_{ij}^{(n)}]$ is not constant and changes for every i and j . Nonetheless, the sequence $\xi^{(n)}$ still converges² to the same limit W_g .

This example illustrates how the randomness related to the exchangeability in the sequence $\xi^{(n)}$ is lost in the limit of μ^n , as it is lost in the graph limit W_g . In this sense, adding exchangeability to system (3.4) does not yield any *averaging property* on the empirical measure μ^n . Moreover, adding the extra randomness through Bernoulli random variables in (3.18) does not alter this fact. In other words, taking $\xi_{ij}^{(n)} = g_i g_j \in [0, 1]$ yields yet again the same limit for μ^n . \square

Until now, we have focused on deterministic limits for the sequence $\xi^{(n)}$. Observe that a characterization of the exchangeable random graphs with such behavior is given in [40]; see also [73, §11.5]. We now address the case when the limit W is random and the relationship with the mean-field limit ρ given in (3.17). One might be led to conjecture that it is possible to recover the mean-field behavior by, e.g., averaging the limit dynamics with respect to the randomness in W . In the next example, we formulate this remark in a rigorous way. We show that this is in general not possible, although it may lead to a new class of asymptotic behaviors which are interesting on their own, as pointed out in (2).

2.3.2. Example II: random mean-field behavior. Consider the growing preferential attachment graph ξ_{pa} constructed iteratively as follows; see also [73, Example 11.44]. Begin with a single node and, assuming that at the n -th step there are already n nodes, create a new node with label $n+1$ and connect it to each node $i \in \{1, \dots, n\}$ with probability $(d_n(i) + 1)/(n+1)$ where $d_n(i)$ is the degree of node i at step n and each connection is made independently of the others. Denote the corresponding random graph by $\xi_{\text{pa}}^{(n+1)}$.

Roughly speaking, the behavior of ξ_{pa} depends crucially on the first steps of the construction and it stabilizes to a homogeneous structure as n grows. This is illustrated in the next proposition.

PROPOSITION 2.8 ([73, Proposition 11.45]). *With probability 1, the sequence $\{\xi_{\text{pa}}^{(n)}\}_{n \in \mathbb{N}}$ converges to a random constant graphon.*

² \mathbb{P} -a.s. in the realization of the Bernoulli random variables and possibly at the cost of requiring some regularity on W_g , see [?, §11.4].

Consider a particle system defined on the graph sequence $\{\xi_{\text{pa}}^{(n)}\}_{n \in \mathbb{N}}$, then the empirical measure converges to the solution of equation (3.17) with a random p . In other words, μ^n converges to a random mean-field limit.

Integrating (3.17) with respect to this randomness and denoting $\mathbb{E}[\rho_t]$ by $\bar{\rho}_t$ for every $t \in [0, T]$, we obtain that $\bar{\rho} \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$ satisfies

$$\partial_t \bar{\rho}_t(\theta) = \frac{1}{2} \partial_\theta^2 \bar{\rho}_t(\theta) - \partial_\theta [\bar{\rho}_t(\theta) F(\theta)] - \partial_\theta \left[\mathbb{E} \left[p \rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \tilde{\theta}) \rho_t(d\tilde{\theta}) \right] \right], \quad (3.21)$$

for $t \in [0, T]$. Note that (3.21) is not written in closed form because of the third term on the right-hand side which is not linear in ρ and p . In this sense, $\bar{\rho}$ does not formally satisfies the mean-field limit, i.e., it is not a solution to (3.17) with some deterministic $p \in [0, 1]$.

To have an intuitive understanding of what $\bar{\rho}$ may look like, consider the stochastic Kuramoto model without natural frequencies [9, 30] defined on the sequence $\xi_{\text{pa}}^{(n)}$. The model is defined as the solution to

$$d\theta_t^{i,n} = \frac{K}{n} \sum_{j=1}^n \xi_{ij}^{(n)} \sin(\theta_t^{j,n} - \theta_t^{i,n}) dt + dB_t^i, \quad (3.22)$$

for $i = 1, \dots, n$ and $t \in [0, T]$. It corresponds to (3.4) with the choices $F \equiv 0$ and $\Gamma(\theta, \psi) = -K \sin(\theta - \psi)$. An application of Theorem 2.3 and Proposition 2.8 implies that the empirical measure of (3.22) converges to the solution of

$$\partial_t \rho_t(\theta) = \frac{1}{2} \partial_\theta^2 \rho_t(\theta) + pK \partial_\theta [\rho_t(\theta) (\sin * \rho_t)(\theta)], \quad (3.23)$$

where $*$ stands for the convolution operator.

It is well-known that equation (3.23) undergoes a phase transition as the coupling strength pK crosses the critical threshold $pK = 1$. Hence, the phase transition for this model occurs at a *random* critical threshold. Depending on the sampled value of p , one obtains stable synchronous solutions in the supercritical regime ($pK > 1$), or uniformly distributed oscillators on \mathbb{T} ($0 \leq pK < 1$). The solution to equation (3.23) can be written down explicitly (see again [9, 30]) and, integrating over the randomness of p , gives a superposition of synchronous and asynchronous states which, in general, is *not* a mean-field solution, i.e., it does not solve (3.23) for some fixed $p \in [0, 1]$. \square

3. The non-linear process

We introduce a non-linear process (3.33) which has already been considered in the literature [6, 27, 77, 79, 90] as the natural candidate in case the particles in (3.4) are not exchangeable and their labels are fixed from the initial condition. This process is interesting for studying the evolution of a tagged particle with a specific profile of connections, as stressed in [77].

Contrary to our setting, some regularity in the – now labeled – graphon is usually assumed to show the convergence of the empirical measure (3.5). We will exploit (3.33) to better understand (3.6) and to establish existence and uniqueness.

Before introducing (3.33), we define some other tools for dealing with empirical measures and graphons. Notably, we introduce an equivalence relation between probability measures on $I \times \mathbb{T}$ inspired by graph limits theory, see (3.29). This will allow us to prove Proposition 2.2, where we establish that the empirical measure is Hölder continuous with respect to the underlying graphon.

3.1. Distances between probability measures. Let \mathcal{M}_T be the space of probability measures on $I \times \mathcal{C}([0, T], \mathbb{T})$ with first marginal equal to the Lebesgue measure λ on I , i.e.,

$$\mathcal{M}_T := \{\mu \in \mathcal{P}(I \times \mathcal{C}([0, T], \mathbb{T})) : p_1 \circ \mu = \lambda\}, \quad (3.24)$$

where p_1 is the projection map associated to the first coordinate. For $\mu \in \mathcal{M}_T$ the following decomposition holds

$$\mu(dx, d\theta) = \mu^x(d\theta)\lambda(dx), \quad x \in I, \quad (3.25)$$

where $\mu^x \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$ for almost every $x \in I$. From now on, we denote the Lebesgue measure $\lambda(dx)$ on I simply by dx .

For $\mu, \nu \in \mathcal{M}_T$, define their distance by

$$d_T(\mu, \nu) := \left(\int_I D_T^2(\mu^x, \nu^x) dx \right)^{1/2}. \quad (3.26)$$

REMARK 3.1. Observe that the previous definitions make sense also with $T = 0$ and $\mathcal{C}([0, T], \mathbb{T})$ replaced by \mathbb{T} . In particular, \mathcal{M}_0 is the space of probability measures on $I \times \mathbb{T}$ with first marginal equal to the Lebesgue measure λ on I , i.e.

$$\mathcal{M}_0 = \{\mu_0 \in \mathcal{P}(I \times \mathbb{T}) : p_1 \circ \mu_0 = \lambda\}, \quad (3.27)$$

and

$$d_0(\mu_0, \nu_0) = \left(\int_I D_0^2(\mu_0^x, \nu_0^x) dx \right)^{1/2}, \quad \text{for } \mu_0, \nu_0 \in \mathcal{M}_0. \quad (3.28)$$

Inspired by the graphon framework, one can define the following relation of equivalence on \mathcal{M}_T (the case $T = 0$ is analogous): for $\mu, \nu \in \mathcal{M}_T$

$$\mu \sim \nu \quad \text{iff there exists } \varphi \in S_I \text{ such that } \mu^x = \nu^{\varphi(x)}, \quad x\text{-a.s.} \quad (3.29)$$

Endow the quotient space \mathcal{M}_T / \sim with the induced distance given by

$$\tilde{d}_T(\mu, \nu) := \inf_{\varphi \in S_I} d_T(\mu, \nu^\varphi), \quad (3.30)$$

where we have used the notation $\nu^\varphi = \{\nu^{\varphi(x)}\}_{x \in I}$.

Observe that if $\mu \sim \nu$, then $\bar{\mu} = \int_I \mu^x dx = \int_I \nu^{\varphi(x)} dx = \int_I \nu^x dx = \bar{\nu}$. In particular, for every $\varphi \in S_I$

$$D_T^2(\bar{\mu}, \bar{\nu}) = D_T^2(\bar{\mu}, \bar{\nu}^\varphi) \leq \int_I D_T^2(\mu^x, \nu^{\varphi(x)}) dx = d_T^2(\mu, \nu^\varphi). \quad (3.31)$$

By taking the infimum with respect to $\varphi \in S_I$, we obtain

$$D_T(\bar{\mu}, \bar{\nu}) \leq \tilde{d}_T(\mu, \nu). \quad (3.32)$$

3.2. The non-linear process with fixed labels. Fix a labeled graphon $W \in \mathcal{W}_0$ together with an initial condition $\mu_0 \in \mathcal{M}_0$. Consider the process $\theta = \{\theta^x\}_{x \in I}$ that solves the system

$$\begin{cases} \theta_t^x = \theta_0^x + \int_0^t F(\theta_s^x) ds + \int_0^t \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta_s^x, \theta) \mu_s^y(d\theta) dy ds + B_t^x, \\ \mu_t^x = \mathcal{L}(\theta_t^x), \quad \text{for } x \in I, t \in [0, T], \end{cases} \quad (3.33)$$

where $\{\theta_0^x\}_{x \in I}$ is a random vector such that $\mathcal{L}(\theta_0^x) = \mu_0^x$ for $x \in I$ and $\{B_t^x\}_{x \in I}$ a sequence of IID Brownian motions independent of $\{\theta_0^x\}_{x \in I}$.

The following proposition shows existence and uniqueness for the solution of (3.33). The proof follows a classical argument by Sznitman [98] and is postponed to Section 3.3.

PROPOSITION 3.2. *There exists a unique solution $\theta = \{\theta^x\}_{x \in I}$ to (3.33). The law $\nu^x \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$ of θ^x for $x \in I$ satisfies the following non-linear Fokker-Planck equation in the weak sense*

$$\partial_t \mu_t^x(\theta) = \frac{1}{2} \partial_\theta^2 \mu_t^x(\theta) - \partial_\theta [\mu_t^x(\theta) F(\theta)] - \partial_\theta \left[\mu_t^x(\theta) \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta, \theta') \mu_t^y(d\theta') dy \right] \quad (3.34)$$

with initial condition $\mu_0^x \in \mathcal{P}(\mathbb{T})$.

The process $\{\theta^x\}_{x \in I}$ is indexed by the space of labels I . For two different labels x and y in I , the behavior of particles θ^x and θ^y may vary depending on their connection profile encoded in W and the two marginals μ^x and μ^y may vary as well. Similar results in different settings have already been shown in [6, 24, 77, 79, 90].

It is interesting to know that the law $\mu = \{\mu^x\}_{x \in I} \in \mathcal{M}_T$ is continuous with respect to the cut-norm (or equivalently in d_\square -distance) in \mathcal{W}_0 , as already remarked in [6, Theorem 2.1] for much more general systems than the ones we consider here. Exploiting the compactness of \mathbb{T} and some extra regularity of Γ , we are able to prove that the map $W \mapsto \mu^W$ is Hölder-continuous, as shown in the next proposition.

PROPOSITION 3.3. *Suppose that $\Gamma \in \mathcal{C}^{1+\varepsilon}(\mathbb{T}^2)$ for some $\varepsilon > 0$. There exists a positive constant C such that, if μ^W and μ^V denote the laws of the solutions to (3.33) with $W \in \mathcal{W}_0$ and $V \in \mathcal{W}_0$ respectively, then*

$$d_T(\mu^W, \mu^V) \leq C \|W - V\|_\square^{1/2}. \quad (3.35)$$

The proof is again postponed to Subsection 3.3. As for Proposition 2.2, different p -Wasserstein metrics for $p \geq 1$ yield a Hölder exponent as large as $1/p$.

3.2.1. *Relationship with the non-linear process* (3.6). Consider a probability distribution $\mu_0 \in \mathcal{M}_0$ such that $\int_I \mu_0^x dx = \bar{\mu}_0$. The solution to (3.7) is given by $\bar{\mu} = \int_I \mu^x dx$, where μ^x is the law of θ^x solving (3.33) with initial condition μ_0^x and labeled graphon W . In other words, θ has the same law of θ^U solution to (3.33), where U is a uniform random variable in I independent of the other randomness in the system. As the following remark shows, the law $\bar{\mu}$ of θ does not depend neither on the representative W , nor on μ_0 .

REMARK 3.4. Let $\varphi \in S_I$, i.e., φ is an invertible measure preserving map from I to itself, and $\nu = \{\nu^x\}_{x \in I}$ the law of $\{\theta^{\varphi(x)}\}_{x \in I}$ solving (3.33). By a change of variable, $\theta^{\varphi(x)}$ solves

$$\theta_t^{\varphi(x)} = \theta_0^{\varphi(x)} + \int_0^t F(\theta_s^{\varphi(x)}) ds + \int_0^t \int_I W(\varphi(x), \varphi(y)) \int_{\mathbb{T}} \Gamma(\theta_s^{\varphi(x)}, \theta) \mu_s^{\varphi(y)}(d\theta) dy ds + B_t^{\varphi(x)} \quad (3.36)$$

and can be rewritten with $V = W^\varphi$ and $\psi^x = \theta^{\varphi(x)}$ as

$$\psi_t^x = \theta_0^{\varphi(x)} + \int_0^t F(\psi_s^x) ds + \int_0^t \int_I V(x, y) \int_{\mathbb{T}} \Gamma(\psi_s^x, \theta) \nu_s^y(d\theta) dy ds + B_t^{\varphi(x)}, \quad (3.37)$$

which has the same law as (3.33) with labeled graphon V and initial conditions $\{\theta^{\varphi(x)}\}_{x \in I}$.

Observe that the laws ν and μ associated to (3.37) and (3.33) respectively, differ only in the labeling of the vertices but their distance in \mathcal{M}_T is not negligible due to the initial conditions and the fact that $\|W - V\|_{\square} = \|W - W^\varphi\|_{\square}$ is, in general, different from zero. However, if one looks at $\bar{\mu} = \int_I \mu^x dx$ and $\bar{\nu} = \int_I \nu^x dx$, they coincide as probability measures in the sense that $D_T(\bar{\mu}, \bar{\nu}) = 0$. In particular, the law of the solution to equation (3.6) is also equivalent to ψ^U , where ψ^x solves (3.37), and U is uniformly distributed on I .

3.3. Proofs for the non-linear process (3.33) with fixed labels.

PROOF OF PROPOSITION 3.2. The proof follows a classical argument given in [98, Lemma 1.3]. Consider $\nu \in \mathcal{M}_T$ and $\{\theta^{x,\nu}\}_{x \in I}$ solving

$$\theta_t^{x,\nu} = \theta_0^x + \int_0^t F(\theta_s^{x,\nu}) ds + \int_0^t \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta_s^{x,\nu}, \theta) \nu_s^y(d\theta) dy ds + B_t^x, \quad (3.38)$$

where the initial conditions and the Brownian motions are the same of (3.33). Since F and Γ are bounded Lipschitz functions, there exists a unique solution to (3.38), which we denote by $\Phi(\nu) \in \mathcal{M}_T$. Thus, the map

$$\begin{aligned} \Phi : (\mathcal{M}_T, d_T) &\rightarrow (\mathcal{M}_T, d_T) \\ \nu &\rightarrow \Phi(\nu) \end{aligned} \quad (3.39)$$

is well defined. A solution to (3.33) is a fixed point of Φ and any fixed point of Φ is a solution to (3.33).

For $\mu, \nu \in \mathcal{M}_T$, consider the processes $\theta^{x,\mu}$ and $\theta^{x,\nu}$, with $x \in I$. We estimate their distance as

$$\begin{aligned} |\theta_t^{x,\mu} - \theta_t^{x,\nu}|^2 &\leq C \int_0^t |F(\theta_s^{x,\mu}) - F(\theta_s^{x,\nu})|^2 ds \\ &+ C \int_0^t \left| \int_I W(x, y) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,\mu}, \theta) \mu_s^y(d\theta) - \int_{\mathbb{T}} \Gamma(\theta_s^{x,\nu}, \theta) \nu_s^y(d\theta) \right) dy \right|^2 ds \end{aligned}$$

Adding and subtracting in the second integral the quantity $\int_{\mathbb{T}} \Gamma(\theta_s^{x,\mu}, \theta) \nu_s^y(d\theta)$ and using that F and Γ are Lipschitz-continuous functions and that F , Γ and W are bounded, we get

$$\leq C \int_0^t |\theta_s^{x,\mu} - \theta_s^{x,\nu}|^2 ds + C \int_0^t \int_I \left| \int_{\mathbb{T}} \Gamma(\theta_s^{x,\mu}, \theta) [\mu_s^y - \nu_s^y](d\theta) \right|^2 dy ds, \quad (3.40)$$

From (3.3) we obtain

$$\left| \int_{\mathbb{T}} \Gamma(\theta_s^{x,\mu}, \theta) (\mu_s^y - \nu_s^y)(d\theta) \right| \leq D_s(\mu^y, \nu^y) \quad (3.41)$$

from which, using (3.26), we deduce

$$|\theta_t^{x,\mu} - \theta_t^{x,\nu}|^2 \leq C \int_0^t |\theta_s^{x,\mu} - \theta_s^{x,\nu}|^2 ds + C \int_0^t d_s^2(\mu, \nu) ds. \quad (3.42)$$

The definition of D_T (3.2) and an application of Gronwall's lemma lead to

$$d_T^2(\Phi(\mu), \Phi(\nu)) \leq \int_I \mathbf{E} \left[\sup_{t \in [0, T]} |\theta_t^{x,\mu} - \theta_t^{x,\nu}|^2 \right] dx \leq C \int_0^T d_s^2(\mu, \nu) ds. \quad (3.43)$$

From the last relation we obtain the uniqueness of solutions to (3.33).

We prove that a solution exists by iterating (3.43). Indeed, for $k \geq 1$ and $\mu \in \mathcal{M}_T$, one gets

$$d_T^2(\Phi^{k+1}(\mu), \Phi^k(\mu)) \leq C^k \frac{T^k}{k!} \int_0^T d_t^2(\Phi(\mu), \mu) dt. \quad (3.44)$$

In particular, $\{\Phi^k(\mu)\}_{k \in \mathbb{N}}$ is a Cauchy sequence for k large enough, and its limit is the fixed point of Φ . Note that $d_t(\Phi(\mu), \mu) < \infty$ since we are working on the compact space \mathbb{T} .

For the second part of Proposition 3.2, apply Itô's formula to $f(\theta_t^x)$ with $f \in \mathcal{C}_0^\infty$ to get

$$\begin{aligned} f(\theta_t^x) &= f(\theta_0^x) + \frac{1}{2} \int_0^t \partial_\theta^2 f(\theta_s^x) ds + \int_0^t \partial_\theta f(\theta_s^x) F(\theta_s^x) ds \\ &+ \int_0^t \partial_\theta f(\theta_s^x) \left[\int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta_s^x, \theta) \mu_s^y(d\theta) dy \right] ds + \int_0^t \partial_\theta f(\theta_s^x) dB_s^x. \end{aligned} \quad (3.45)$$

Integrating with respect to \mathbf{P} yields the weak formulation of (3.34). \square

Next we move to the proof of Proposition 3.3.

PROOF OF PROPOSITION 3.3. Let $\{\theta^{x,W}\}_{x \in I}$ and $\{\theta^{x,V}\}_{x \in I}$ be the two non-linear processes associated to W and V respectively. We compare the two solutions: as done in the proof of Proposition 3.2, by adding and subtracting in the integrals the term $W(x, y)\Gamma(\theta_r^{x,V}, \theta)(\mu_r^{y,W} -$

$\mu_r^{y,V}$) we get

$$\begin{aligned}
|\theta_s^{x,W} - \theta_s^{x,V}|^2 &\leq C \int_0^s |F(\theta_r^{x,W}) - F(\theta_r^{x,V})|^2 dr \\
&+ C \int_0^s \left| \int_I W(x,y) \int_{\mathbb{T}} (\Gamma(\theta_r^{x,W}, \theta) - \Gamma(\theta_r^{x,V}, \theta)) \mu_r^{y,W}(d\theta) dy \right|^2 dr \\
&+ C \int_0^s \left| \int_I W(x,y) \int_{\mathbb{T}} \Gamma(\theta_r^{x,V}, \theta) (\mu_r^{y,W} - \mu_r^{y,V})(d\theta) dy \right|^2 dr \\
&+ C \int_0^s \left| \int_I (W(x,y) - V(x,y)) \int_{\mathbb{T}} \Gamma(\theta_r^{x,V}, \theta) \mu_r^{y,V}(d\theta) dy \right|^2 dr.
\end{aligned} \tag{3.46}$$

Using that F and Γ are Lipschitz-continuous functions and that F , Γ and W are bounded, we get

$$\begin{aligned}
|\theta_s^{x,W} - \theta_s^{x,V}|^2 &\leq C \int_0^s |\theta_r^{x,W} - \theta_r^{x,V}|^2 dr + C \int_0^s d_r^2(\mu^W, \mu^V) dr \\
&+ \int_0^s \left| \int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta) \mu_r^{y,V}(d\theta) \right) dy \right|^2 dr.
\end{aligned} \tag{3.47}$$

After taking the supremum over $s \in [0, t]$, the expectation \mathbf{E} and integrating with respect to $x \in I$, we are able to apply Gronwall's lemma as in (3.43) to get

$$d_t^2(\mu^W, \mu^V) \leq \int_I \mathbf{E} \left[\sup_{s \in [0, t]} |\theta_s^{x,W} - \theta_s^{x,V}|^2 \right] dx \leq C \left(\int_0^t d_s^2(\mu^W, \mu^V) ds + G \right), \tag{3.48}$$

where G is given by

$$G = \int_0^t \mathbf{E} \left[\int_I \left| \int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta) \mu_s^{y,V}(d\theta) \right) dy \right|^2 dx \right] ds. \tag{3.49}$$

Applying Gronwall's inequality to (3.48) yields

$$d_t^2(\mu^W, \mu^V) \leq CG. \tag{3.50}$$

The proof is concluded provided that $G \leq C' \|W - V\|_{\square}$, for some constant $C' > 0$.

Observe that Γ can be written in Fourier series, i.e.

$$\Gamma(\theta, \psi) = \sum_{k, l \in \mathbb{Z}} \Gamma_{kl} e^{ik\theta} e^{il\psi}, \quad \theta, \psi \in \mathbb{T}, \tag{3.51}$$

where $\Gamma_{kl} = \int_{\mathbb{T}^2} \Gamma(\theta, \psi) e^{i(k\theta + l\psi)} d\theta d\psi$. Since $\Gamma \in \mathcal{C}^{1+\varepsilon}$, classical results on the asymptotic of Fourier series [64, pp. 24-26] imply that

$$C_{\Gamma} := \sum_{k, l \in \mathbb{Z}} (kl)^{1+\varepsilon} |\Gamma_{kl}|^2 < \infty. \tag{3.52}$$

Plugging this expression into (3.49), we obtain that

$$\begin{aligned} & \int_I \left| \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta) \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x \\ &= \int_I \left| \sum_{kl} \Gamma_{kl} e^{ik\theta_s^{x,V}} \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} e^{i\ell\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x. \end{aligned} \quad (3.53)$$

Multiplying and dividing by $(kl)^{(1+\varepsilon)/2}$ one is left with

$$\begin{aligned} & \leq \int_I \left| \sum_{kl} \left((kl)^{(1+\varepsilon)/2} \Gamma_{kl} e^{ik\theta_s^{x,V}} \right) \right. \\ & \quad \left. \left((kl)^{-(1+\varepsilon)/2} \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} e^{i\ell\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right) \right|^2 \mathrm{d}x. \quad (3.54) \\ & \leq C_{\Gamma} \sum_{kl} (kl)^{-1-\varepsilon} \int_I \left| \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} e^{i\ell\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x \end{aligned}$$

where in the second step we have applied Cauchy-Schwartz inequality and (3.52). Using that W and V are bounded, as well as the fact that

$$\left| \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} e^{i\ell\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right| \leq 1, \quad (3.55)$$

we conclude

$$\begin{aligned} G & \leq C \sup_{\|a\|_{\infty}, \|b\|_{\infty} \leq 1} \int_I \left| \int_I (W(x, y) - V(x, y)) (a(y) + ib(y)) \mathrm{d}y \right| \mathrm{d}x \\ & \leq C \|W - V\|_{\infty \rightarrow 1}. \end{aligned} \quad (3.56)$$

Since the norm $\|\cdot\|_{\infty \rightarrow 1}$ is equivalent to the cut-norm (3.80), the proof is concluded. \square

3.4. Proofs for the non-linear process (3.6).

PROOF OF PROPOSITION 2.1. The first part follows directly from Proposition 3.2 and Remark 3.4. The proof of (3.7) is similar to the proof of (3.34), but note that we are now integrating with respect to the randomness in U as well. \square

PROOF OF PROPOSITION 2.2. Let $\theta^{U,W}$ and $\theta^{U,V}$ be the two solutions to (3.6) associated to W and V respectively, coupled by taking the same uniform random variable U . Let $\mu^{x,W}$ and $\mu^{x,V}$ represent the laws of $\theta^{U,W}$ and $\theta^{U,V}$ conditioned on $U = x$, for $x \in I$.

Consider $\varphi \in S_I$ an invertible measure preserving map. Recall that $\theta^{\varphi(U),V}$ also satisfies equation (3.6) with V^{φ} , see Remark 3.4. We compare the trajectories $\theta^{U,W}$ and $\theta^{\varphi(U),V}$.

Consider the difference between the equations satisfied by $\theta^{U,W}$ and $\theta^{\varphi(U),V}$, add and subtract the term $W(U, y)\Gamma(\theta_r^{\varphi(U),V}, \theta)(\mu_r^{y,W} - \mu_r^{\varphi(y),V})$ to obtain that

$$\begin{aligned}
|\theta_s^{U,W} - \theta_s^{\varphi(U),V}|^2 &\leq C \int_0^s |F(\theta_r^{U,W}) - F(\theta_r^{\varphi(U),V})|^2 dr \\
&+ C \int_0^s \left| \int_I W(U, y) \int_{\mathbb{T}} \left(\Gamma(\theta_r^{\varphi(U),W}, \theta) - \Gamma(\theta_r^{\varphi(U),V}, \theta) \right) \mu_r^{\varphi(y),W}(d\theta) dy \right|^2 dr \\
&+ C \int_0^s \left| \int_I W(U, y) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta) (\mu_r^{y,W} - \mu_r^{\varphi(y),V})(d\theta) dy \right|^2 dr \\
&+ C \int_0^s \left| \int_I (W(U, y) - V^\varphi(U, y)) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta) \mu_r^{\varphi(y),V}(d\theta) dy \right|^2 dr.
\end{aligned} \tag{3.57}$$

The first two integrals on the r.h.s. are bounded by $C \int_0^s |\theta_r^{U,W} - \theta_r^{\varphi(U),V}|^2 dr$, using that F and Γ are Lipschitz-continuous. While the third integral in the r.h.s. can be estimated using (3.3) and the fact that $0 \leq W \leq 1$. Thus we get

$$\begin{aligned}
&\left| \int_I W(U, y) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta) (\mu_r^{y,W} - \mu_r^{\varphi(y),V})(d\theta) dy \right|^2 \\
&\leq \int_I D_r^2(\mu^{y,W}, \mu^{\varphi(y),V}) dy = d_r^2(\mu^W, (\mu^V)^\varphi),
\end{aligned} \tag{3.58}$$

where we have used the notation $(\mu^V)^\varphi$ for $\{\mu^{\varphi(y),V}\}_{y \in I}$.

Taking the supremum over $s \in [0, t]$ and the expectation with respect to the Brownian motions, the initial conditions and the random variable U , we obtain

$$\begin{aligned}
\int_I \mathbf{E} \left[\sup_{s \in [0, t]} |\theta_s^{x,W} - \theta_s^{\varphi(x),V}|^2 \right] dx &\leq C \int_0^t \int_I \mathbf{E} \left[\sup_{r \in [0, s]} |\theta_r^{x,W} - \theta_r^{\varphi(x),V}|^2 \right] dx ds \\
&+ C \int_0^t d_s^2(\mu^W, (\mu^V)^\varphi) ds + CG,
\end{aligned} \tag{3.59}$$

where G is given by

$$G = \int_0^t \mathbf{E} \left[\int_I \left| \int_I (W(x, y) - V^\varphi(x, y)) \int_{\mathbb{T}} \Gamma(\theta_s^{\varphi(x),V}, \theta) \mu_s^{\varphi(y),V}(d\theta) dy \right|^2 dx \right] ds. \tag{3.60}$$

In the proof of Proposition 3.3 we proved the following estimates:

$$\begin{aligned}
d_t^2(\mu^W, (\mu^V)^\varphi) &\leq \int_I \mathbf{E} \left[\sup_{s \in [0, t]} |\theta_s^{x,W} - \theta_s^{\varphi(x),V}|^2 \right] dx, \\
G &\leq C' \|W - V^\varphi\|_{\square}, \quad \text{for some } C' > 0.
\end{aligned} \tag{3.61}$$

Applying these bounds to (3.59) and using Gronwall's inequality twice as in the previous proof, yields

$$d_t^2(\mu^W, (\mu^V)^\varphi) \leq C \|W - V^\varphi\|_{\square}. \tag{3.62}$$

By taking the infimum with respect to $\varphi \in S_I$ and recalling the definition of the cut-distance (3.84) together with (3.32), we obtain

$$D_t(\bar{\mu}^W, \bar{\mu}^V) \leq \tilde{d}_t(\mu^W, \mu^V) \leq C \delta_{\square}(W, V)^{1/2}. \quad (3.63)$$

The proof is concluded. \square

4. Proof of Theorem 2.3

In order to prove Theorem 2.3, we couple the system (3.4) to a sequence of identically distributed copies of the non-linear process θ , which is obtained by sampling $\{U_i\}_{i \in \mathbb{N}}$ IID uniform random variables and choosing the same initial conditions and Brownian motions of (3.4).

For every $i \in \mathbb{N}$, denote these copies by $\theta^i = \theta(U_i)$. In particular, θ^i is defined as the solution for $t \in [0, T]$ to

$$\theta_t^i = \theta_0^i + \int_0^t F(\theta_s^i) ds + \int_0^t \int_I W(U_i, y) \int_{\mathbb{T}} \Gamma(\theta_s^i, \theta) \mu_s^y(d\theta) dy ds + B_t^i. \quad (3.64)$$

Observe that $\{\theta^i\}_{i \in \mathbb{N}}$ is an exchangeable sequence and, in particular, that the variables θ^i are independent random variables when conditioned on the randomness of W .

Before the proof of Theorem 2.3, we give a trajectorial estimate.

LEMMA 4.1. *Under the hypothesis of Theorem 2.3, it holds that*

$$\lim_{n \rightarrow \infty} \mathbb{E} \times \mathbf{E} \left[\frac{1}{n} \sum_{i=1}^n \sup_{t \in [0, T]} |\theta_t^{i,n} - \theta_t^i|^2 \right] = 0. \quad (3.65)$$

PROOF. As done before, we compare the trajectories $\theta^{i,n}$ and θ^i , by studying the equation satisfied by $|\theta_s^{i,n} - \theta_s^i|^2$, recall (3.4) and (3.64). Add and subtract in the integrals the term $\left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) \Gamma(\theta_r^i, \theta_r^j)$ so as to get

$$\begin{aligned} |\theta_s^{i,n} - \theta_s^i|^2 &\leq C \int_0^s |F(\theta_r^{i,n}) - F(\theta_r^i)|^2 dr \\ &+ C \int_0^s \left| \frac{1}{n} \sum_{j=1}^n \xi_{ij}^{(n)} (\Gamma(\theta_r^{i,n}, \theta_r^{j,n}) - \Gamma(\theta_r^i, \theta_r^j)) \right|^2 dr \\ &+ C \int_0^s \left| \frac{1}{n} \sum_{j=1}^n (\xi_{ij}^{(n)} - W(U_i, U_j)) \Gamma(\theta_r^i, \theta_r^j) \right|^2 dr \\ &+ C \int_0^s \left| \frac{1}{n} \sum_{j=1}^n W(U_i, U_j) \Gamma(\theta_r^i, \theta_r^j) - \int_I W(I_i, y) \int_{\mathbb{T}} \Gamma(\theta_r^i, \theta) \mu_r^y(d\theta) dy \right|^2 dr. \end{aligned} \quad (3.66)$$

We now use the Lipschitz property of Γ and F , sum over i and take the supremum over $s \in [0, t]$, together with the expectation $\mathbb{E} \times \mathbf{E}$, which we just write E for simplicity,

$$\begin{aligned} E \left[\frac{1}{n} \sum_{i=1}^n \sup_{s \in [0, t]} |\theta_s^{i, n} - \theta_s^i|^2 \right] &\leq C \int_0^t E \left[\frac{1}{n} \sum_{i=1}^n \sup_{q \in [0, r]} |\theta_q^{i, n} - \theta_q^i|^2 \right] dr \\ &+ C \int_0^t E \left[\frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n (\xi_{ij}^{(n)} - W(U_i, U_j)) \Gamma(\theta_r^i, \theta_r^j) \right|^2 \right] dr \\ &+ C \int_0^t \frac{1}{n} \sum_{i=1}^n E \left[\left| \frac{1}{n} \sum_{j=1}^n W(U_i, U_j) \Gamma(\theta_r^i, \theta_r^j) - \int_I W(U_i, y) \int_{\mathbb{T}} \Gamma(\theta_r^i, \theta) \mu_r^y(d\theta) dy \right|^2 \right] dr. \end{aligned} \quad (3.67)$$

Observe that the last term is bounded by a constant divided by n since by taking the conditional expectation with respect to θ^j and U^j , one obtains

$$\mathbf{E} [W(U_i, U_j) \Gamma(\theta_s^i, \theta_s^j)] = \int_I W(U_i, y) \int_{\mathbb{T}} \Gamma(\theta_s^i, \theta) \mu_s^y(d\theta) dy \quad (3.68)$$

and, conditionally on W , the random variables $\{\theta^i\}_{i \in \mathbb{N}}$ are IID.

Turning to the second term, we will prove that

$$E \left[\frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n (\xi_{ij}^{(n)} - W(U_i, U_j)) \Gamma(\theta_s^i, \theta_s^j) \right|^2 \right] \leq C \mathbb{E} [\delta_{\square}(\xi^{(n)}, W^{(n)})] + o(1), \quad (3.69)$$

where $W^{(n)} := \{W(U_i, U_j)\}_{i, j=1, \dots, n}$ is a W -random graph with n vertices, see (3.13). This, together with a Gronwall argument implies that

$$E \left[\frac{1}{n} \sum_{i=1}^n \sup_{s \in [0, t]} |\theta_t^{i, n} - \theta_t^i|^2 \right] \leq C \mathbb{E} [\delta_{\square}(\xi^{(n)}, W^{(n)})] + o(1) \quad (3.70)$$

and the claim follows by taking the limit for n which tends to infinity and the fact that $W^{(n)}$ converges \mathbb{P} -a.s. to W , recall Theorem 2.5.

Turning to (3.69), we use an argument similar to (3.49)–(3.52). Recall that since $\Gamma \in \mathcal{C}^{1+\varepsilon}$, it admits a Fourier series (3.51) with coefficients Γ_{kl} such that

$$\sum_{k, l \in \mathbb{Z}} (kl)^{1+\varepsilon} |\Gamma_{kl}|^2 < \infty.$$

Plugging its Fourier expression in the left-hand side of (3.69), multiplying and dividing by $(kl)^{(1+\varepsilon)/2}$, we get

$$\begin{aligned}
& E \left[\frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) \Gamma(\theta_s^i, \theta_s^j) \right|^2 \right] \\
&= E \left[\frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) \sum_{k,l} \Gamma_{kl} e^{i\theta_s^i k} e^{i\theta_s^j l} \right|^2 \right] \\
&\leq CE \left[\sum_{k,l} (kl)^{-1-\varepsilon} \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) e^{i\theta_s^i k} e^{i\theta_s^j l} \right|^2 \right],
\end{aligned} \tag{3.71}$$

where we have used Cauchy-Schwartz inequality as in the proof of Proposition 2.2. Observe that $\sum_{k,l} (kl)^{-1-\varepsilon}$ is convergent and that $|e^{i\theta_s^i k}| \leq 1$ for all k and s : we can thus bound \mathbf{P} -a.s. the previous term by

$$\mathbb{E} \left[\sup_{s_i, t_j \in \{\pm 1\}} \left| \frac{1}{n^2} \sum_{i,j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) s_i t_j \right| \right]. \tag{3.72}$$

Recall that $W^{(n)} = \{W(U_i, U_j)\}_{i,j=1,\dots,n}$ is a W -random graph with n vertices. Since the particles $\{\theta^i\}_{i \in \mathbb{N}}$ are exchangeable, every computation done so far holds no matter the order of $\{\theta^i\}_{i=1,\dots,n}$ and, in particular, of $\{U^i\}_{i=1,\dots,n}$. In particular, the last inequality holds for every relabeling of $W^{(n)}$.

From the definition of $\hat{\delta}_\square$ (3.83), one can thus take the labeling of $\{U_i\}_{i=1,\dots,n}$ for every $n \in \mathbb{N}$, such that

$$\mathbb{E} \left[\sup_{s_i, t_j \in \{\pm 1\}} \left| \frac{1}{n^2} \sum_{i,j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j) \right) s_i t_j \right| \right] = \mathbb{E} \left[\hat{\delta}_\square(\xi^{(n)}, W^{(n)}) \right]. \tag{3.73}$$

Using the asymptotic equivalence of $\hat{\delta}_\square$ with δ_\square , see Remark 3.A.1, the claim is proved and the proof is concluded. \square

PROOF OF THEOREM 2.3. The equivalence between the convergence in \mathbb{P} -probability of $\xi^{(n)}$ and equation (3.14) is proven in Lemma 3.A.2. We turn to the proof of the convergence of μ^n .

It is well known that the bounded Lipschitz distance, recall (3.3), metricizes the weak convergence and defines a distance between probability measures. In particular, in order to show that μ^n converges in $\mathbf{P} \times \mathbb{P}$ -probability to $\bar{\mu}$ in $\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}))$, it is enough to prove that

$$\lim_{n \rightarrow \infty} \mathbb{E} \times \mathbf{E} \left[\int f(\theta) \mu^n(d\theta) - \int f(\theta) \bar{\mu}(d\theta) \right] = 0, \tag{3.74}$$

for every f bounded and Lipschitz function with values in $\mathcal{C}([0, T], \mathbb{R})$.

Using the fact that $\bar{\mu}$ is the law of $\{\theta^i\}_{i \in \mathbb{N}}$ (recall (3.64)), it is enough to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbb{E} \times \mathbf{E} [|f(\theta^{j,n}) - f(\theta^j)|] = 0. \quad (3.75)$$

This is implied by the fact that f is Lipschitz and by Jensen's inequality. Indeed,

$$\frac{1}{n} \sum_{j=1}^n \mathbb{E} \times \mathbf{E} [|f(\theta^{j,n}) - f(\theta^j)|] \leq \mathbb{E} \times \mathbf{E} \left[\frac{1}{n} \sum_{j=1}^n \sup_{t \in [0, T]} |\theta_t^{j,n} - \theta_t^j|^2 \right]^{1/2}, \quad (3.76)$$

which goes to zero as $n \rightarrow \infty$ by Lemma 4.1. \square

3.A. Graph convergence and random graphons

3.A.1. Distance between finite graphs. We denote $[n] := \{1, \dots, n\}$ for $n \in \mathbb{N}$. Let ξ be a labeled graph on n vertices. With an abuse of notation, we let ξ denote its adjacency matrix as well, i.e., $\xi = \{\xi_{ij}\}_{i, j \in [n]}$. We consider simple undirected graphs so that $\xi_{ij} = \xi_{ji}$ and $\xi_{ii} = 0$ for all $1 \leq i \leq j \leq n$.

Let $A = \{A_{ij}\}_{i, j \in [n]}$ be a $n \times n$ real matrix. The *cut-norm* of A is defined as

$$\|A\|_{\square} := \frac{1}{n^2} \max_{S, T \subset [n]} \left| \sum_{i \in S, j \in T} A_{ij} \right|. \quad (3.77)$$

It is well-known that this norm is equivalent to the $l_{\infty} \rightarrow l_1$ norm [3]

$$\|A\|_{\infty \rightarrow 1} := \sup_{s_i, t_j \in \{\pm 1\}} \sum_{i, j=1}^n A_{ij} s_i t_j. \quad (3.78)$$

For two labeled graphs ξ and ξ' on the same set of vertices, we define the distance d_{\square} as

$$d_{\square}(\xi, \xi') := \|\xi - \xi'\|_{\square}. \quad (3.79)$$

3.A.2. Labeled and unlabeled graphons. Recall that $I = [0, 1]$ and let $\mathcal{W} := \{W : I^2 \rightarrow \mathbb{R} \text{ bounded symmetric and measurable}\}$ be the space of kernels, we tacitly consider two kernels to be equal if and only if the subset of I^2 where they differ has Lebesgue measure 0. A *labeled graphon* is a kernel W such that $0 \leq W \leq 1$. Let \mathcal{W}_0 denote the space of labeled graphons. The cut-norm of $W \in \mathcal{W}$ is defined as

$$\|W\|_{\square} := \max_{S, T \subset I} \left| \int_{S \times T} W(x, y) dx dy \right| \quad (3.80)$$

where the maximum is taken over all measurable subsets S and T of I . It is well known that $\|W\|_{\square}$ is equivalent to the norm of W seen as an operator from $L^{\infty}(I) \rightarrow L^1(I)$ [73, Theorem 8.11]. This is defined as

$$\|W\|_{\infty \rightarrow 1} := \sup_{\|g\|_{\infty} \leq 1} \|Wg\|_1, \quad (3.81)$$

where $(Wg)(x) := \int_I W(x, y)g(y)dy$ for $x \in I$ and $g \in L^{\infty}(I)$.

The metric induced by $\|\cdot\|_{\square}$, or equivalently by $\|\cdot\|_{\infty \rightarrow 1}$, in the space of labeled graphons \mathcal{W}_0 is again denoted by $d_{\square}(\cdot, \cdot)$. Definitions (3.77) and (3.80) are consistent in the sense that to each labeled graph ξ is associated a labeled graphon $W_{\xi} \in \mathcal{W}_0$ such that $\|\xi\|_{\square} = \|W_{\xi}\|_{\square}$. The labeled graphon W_{ξ} is usually defined a.e. as

$$W_{\xi}(x, y) = \sum_{i, j=1}^n \xi_{ij} \mathbf{1}_{\left[\frac{i-1}{n}, \frac{i}{n}\right) \times \left[\frac{j-1}{n}, \frac{j}{n}\right)}(x, y), \quad \text{for } x, y \in I. \quad (3.82)$$

Note that W_{ξ} depends on the labeling of ξ . Indeed, different labelings of ξ yield graphs which have large d_{\square} -distance in general. This motivates the definition of the so-called *cut-distance*. For two labeled graphs ξ, ξ' with the same number of nodes, the cut-distance is defined as

$$\hat{\delta}_{\square}(\xi, \xi') := \min_{\xi'} d_{\square}(\xi, \hat{\xi}'), \quad (3.83)$$

where the minimum ranges over all labelings of ξ' . The cut-distance is also defined for graphons as follows. For two labeled graphons $W, V \in \mathcal{W}_0$, their cut-distance is

$$\delta_{\square}(W, V) := \min_{\varphi \in S_I} d_{\square}(W, V^{\varphi}), \quad (3.84)$$

where the minimum ranges over S_I the space of invertible measure preserving maps from I into itself and where $V^{\varphi}(x, y) := V(\varphi(x), \varphi(y))$ for $x, y \in I$.

REMARK 3.A.1. There are at least two ways to compare the graphs ξ, ξ' as unlabeled objects: either by directly computing their distance $\hat{\delta}_{\square}$ or by computing the distance δ_{\square} between W_{ξ} and $W_{\xi'}$. These turn out to be equivalent as the number of vertices tends to infinity [73, Theorem 9.29]. Formally, for every two graphs ξ, ξ' on n vertices, it holds that

$$\delta_{\square}(W_{\xi}, W_{\xi'}) \leq \hat{\delta}_{\square}(\xi, \xi') \leq \delta_{\square}(W_{\xi}, W_{\xi'}) + \frac{17}{\sqrt{\log n}}. \quad (3.85)$$

We always write $\delta_{\square}(\xi, \xi') := \delta_{\square}(W_{\xi}, W_{\xi'})$.

Contrary to d_{\square} , the cut-distance δ_{\square} is a pseudometric on \mathcal{W}_0 since the distance between two different labeled graphons can be zero. This leads to the definition of the *unlabeled* graphon \widetilde{W} associated to W . For a labeled graphon W , \widetilde{W} is defined as the equivalence class of W including all $V \in \mathcal{W}_0$ such that $\delta_{\square}(W, V) = 0$. For notation's sake, we drop both the superscript and the adjective unlabeled when the context is clear. The quotient space obtained in such a way is denoted by $\widetilde{\mathcal{W}}_0$ and we refer to it as the space of graphons. A celebrated result of graph limits theory is that $(\widetilde{\mathcal{W}}_0, \delta_{\square})$ is a compact metric space [73, Theorem 9.23].

We are not going into the details of graph convergence for which we refer to the exhaustive reference [73]. We only recall that a sequence of graphs $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ converges to the graphon $W \in \widetilde{\mathcal{W}}_0$ if and only if $\delta_{\square}(W_{\xi^{(n)}}, W) \rightarrow 0$ as $n \rightarrow \infty$ [73, Theorem 11.22]). We refer to the following subsection for a characterization of the convergence in probability.

3.A.3. Convergence in probability. The characterization of the convergence in distribution for a sequence of graphs has been originally given in [40]. We give here a useful notion

of convergence in $\widetilde{\mathcal{W}}_0$ by means of the cut-distance δ_\square , which is equivalent to the convergence in probability for graph sequences.

LEMMA 3.A.2. *Assume that $\{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of random graphs and W a random graphon in $\widetilde{\mathcal{W}}_0$. Then, $\xi^{(n)}$ converges in \mathbb{P} -probability to W if and only if (3.10) holds, i.e., if and only if*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\delta_\square(\xi^{(n)}, W)] = 0.$$

PROOF. Recall that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a compact metric space, so that the convergence of $\xi^{(n)}$ in probability is equivalent to

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(\delta_\square(\xi^{(n)}, W) > \varepsilon) = 0. \quad (3.86)$$

Observe that the sequence of positive real random variables $\{\delta_\square(\xi^{(n)}, W)\}_{n \in \mathbb{N}}$ is uniformly bounded by 1. Equation (3.86) is then equivalent to the convergence in L^1 , i.e., equivalent to (3.10). \square

Long time dynamics for interacting oscillators

This chapter is based on [30].

1. Introduction

1.1. Synchronization of mean field systems on graphs. In recent years, synchronization of complex networks has become a very important topic for explaining real world phenomena. While in the physics literature the analysis has been pushed quite far and several extended reviews are available (e.g. [43, 94]), from a mathematical point of view these studies and the associated numerical simulations, can be regarded more as heuristic arguments than conclusive proofs.

The mathematical community has started working on particle systems on (random) graphs from the statistical mechanics point of view in the equilibrium regime and, with respect to the graph setting, assuming a locally tree-like structure (e.g [37]). Only in the last few years the attention has been focused on the dynamics of weakly interacting particles, tackling mean field systems on graphs, and their relationship with the corresponding thermodynamical limit (e.g. [13, 36]). These results, and the one presented here, are obtained for graphs in an intermediate regime between the sparse and the dense case, i.e. if G_n has n vertices and np_n represents the average number of edges, then $1 \ll np_n \leq n$. In the case of sparse graphs, i.e. $np_n = O(1)$, the limiting system seems to show a different phenomenology ([70, 91]).

Today, many results on the behavior of the empirical measure of such systems are available ([13, 31, 36, 77, 90]), but there is no agreement on the *weakest* hypothesis the class of graphs should satisfy in order to obtain the classical mean field limit. It turns out that, depending on the setting one is considering, i.e. the normalization chosen in the interaction and/or the hypothesis on the initial data, different requirements on the graph may be asked.

To the author's knowledge, there exists no result on the longtime dynamics of a system defined on a sequence of graphs and the question whether the network is influencing the dynamics on long time scales, is still open and very much awaited with regards to applications.

In this work, we attack these issues by considering a well known model of synchronization defined on a sequence of graphs: we consider the Kuramoto model (e.g. [1]) for which an extensive literature is available and many tools have now been developed ([9, 10, 51]). For the sake of clarity, we study the model without the natural frequencies but our techniques apply as well in the quenched setting. We look for a result of mean field type with the minimal hypothesis on the initial conditions, i.e. the weak convergence of the empirical measure only,

and by proposing a (deterministic) condition on the sequence of graphs which is shown to be satisfied by a large class of *homogeneous* graphs, including Erdős-Rényi random graphs with diverging average degree.

Finally, we show that the condition on the graph is not only sufficient for the system to converge to the mean field limit on bounded time intervals, but also that it is enough to study it on longer time scales. Namely, we push our analysis to the Large Deviation barrier of exponential time scales showing that, if the system synchronizes, then it keeps synchronized for long times.

1.2. The model. For each $n \in \mathbb{N}$, let $\xi^{(n)}$ be the adjacency matrix of a graph $(V^{(n)}, E^{(n)})$ with n vertices:

$$V^{(n)} = \{1, \dots, n\}, \quad E^{(n)} = \left\{ (i, j) \in V^{(n)} \times V^{(n)} : \xi_{ij}^{(n)} \geq 1 \right\}. \quad (4.1)$$

We consider both directed and undirected graphs as well as multigraphs so that $\xi_{ij}^{(n)}$ can take values in $\{1, \dots, n\}$ and not need to be equal to $\xi_{ji}^{(n)}$. We denote the corresponding (multi)graph by $\xi^{(n)}$ itself. Together with $\xi^{(n)}$, we consider a *dilution* parameter $p_n \in (0, 1]$ representing the average density of neighbors per site. The two quantities will be coupled so that it is useful to think of them as one single object, we refer to Subsection 1.4 for the precise condition we require on it.

Given $(\xi^{(n)}, p_n)$, let $\{\theta_t^{i,n}\}_{i=1, \dots, n}$ be the family of oscillators on $\mathbb{T}^n := (\mathbb{R}/2\pi\mathbb{Z})^n$, which satisfy:

$$\begin{cases} d\theta_t^{i,n} = \frac{1}{np_n} \sum_{j=1}^n \xi_{ij}^{(n)} J(\theta_t^{i,n} - \theta_t^{j,n}) dt + dB_t^i, & \text{for } t > 0, \\ \theta_0^{i,n} = \theta_0^i, & \text{for } i \in \{1, \dots, n\}, \end{cases} \quad (4.2)$$

where $J(\cdot) = -K \sin(\cdot)$ with $K \geq 0$. Denote by \mathbf{P} the law induced by $\{B_t^i\}_{i \in \mathbb{N}}$ which are independent and identically distributed (IID) Brownian motions on \mathbb{T} and by $\{\theta_0^i\}_{i \in \mathbb{N}}$ the initial conditions. We consider both deterministic and random initial data and, whenever they are random, they have to be independent of the Brownian motions.

If $\{\xi_{ij}^{(n)}\}_{ij}$ are symmetric, i.e. $\xi_{ij}^{(n)} = \xi_{ji}^{(n)}$ for $1 \leq i < j \leq n$, then the model is reversible (e.g. [9]) with respect to the probability measure on \mathbb{T}^n given by

$$\pi^{(n)}(d\theta) = \frac{1}{Z^{(n)}} \exp \left(-\frac{K}{n} \sum_{i,j=1}^n \xi_{ij}^{(n)} \cos(\theta^i - \theta^j) \right) \lambda_n(d\theta), \quad (4.3)$$

where $Z^{(n)}$ is the normalizing constant and λ_n the uniform probability measure on \mathbb{T}^n .

The main quantity of interest in system (4.2) is the empirical measure μ_t^n associated to $\{\theta_t^{i,n}\}_{i=1, \dots, n}$ and it is defined for all $t \geq 0$ by

$$\mu_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{\theta_t^{j,n}} \in \mathcal{P}(\mathbb{T}), \quad (4.4)$$

the space of probability measure on the torus being denoted by $\mathcal{P}(\mathbb{T})$.

1.3. The reversible Kuramoto model and its mean field limit. When $\xi_{ij}^{(n)} = 1$ for $1 \leq i, j \leq n$ and $p_n \equiv 1$ for all $n \in \mathbb{N}$, i.e. $\xi^{(n)}$ is the complete graph, system (4.2) becomes:

$$\begin{cases} d\bar{\theta}_t^{i,n} = (J * \bar{\mu}_t^n)(\bar{\theta}_t^{i,n})dt + dB_t^i, & \text{for } t > 0, \\ \bar{\theta}_0^{i,n} = \theta_0^i, & \text{for } i \in \{1, \dots, n\}, \end{cases} \quad (4.5)$$

where $\bar{\mu}_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{\bar{\theta}_t^{j,n}}$ is the associated empirical measure and $*$ stands for the convolution. We refer to (4.5) as the reversible Kuramoto model (e.g. [9]).

It is well known (e.g. [9, Proposition 3.1]) that for all fixed time T , $\bar{\mu}_{t \in [0, T]}^n$ seen as a continuous function over $\mathcal{P}(\mathbb{T})$, weakly converges in $\mathcal{C}^0([0, T], \mathcal{P}(\mathbb{T}))$ to a deterministic limit $\mu \in \mathcal{C}^0([0, T], \mathcal{P}(\mathbb{T}))$ that is solution to the following partial differential equation (PDE):

$$\begin{cases} \partial_t \mu_t(\theta) = \frac{1}{2} \partial_\theta^2 \mu_t(\theta) - \partial_\theta [\mu_t(\theta)(J * \mu_t)(\theta)], & \text{for } \theta \in \mathbb{T}, 0 < t \leq T, \\ \mu_t|_{t=0} = \mu_0, \end{cases} \quad (4.6)$$

provided that μ_0^n weakly converges to μ_0 in $\mathcal{P}(\mathbb{T})$. If μ_0 does not have a density, than (4.6) has to be intended in the weak sense; however the regularity properties of the Laplacian operator make μ_t smooth for all $t > 0$ (see again [9, Proposition 3.1]). Equation (4.6) is often called McKean-Vlasov or Fokker-Planck equation and we refer to its solution μ as to the mean field limit of the diffusions solving (4.5).

We recall here the most important results on (4.6), without giving any proof but referring to [51] (and references therein) where a complete analysis of the global dynamics is presented.

As for the mean field limit of the classical Kuramoto model, (4.6) is known to admit a phase transition depending on the coupling strength K : in the subcritical regime, for $0 \leq K < K_c := 1$, the particles behave as they were independently distributed on the circle; in the supercritical regime, for $K > 1$, they tend to synchronize around the same phase. We do not consider the critical case $K = 1$, since it does not add anything to the purpose of this work.

More precisely, in the subcritical regime there is a unique stationary solution which corresponds to the incoherent state $\frac{1}{2\pi}$, the uniform measure on the torus (see [51, Proposition 4.1]). It is globally attractive and the linear operator around it has negative spectrum bounded away from zero: we will make use of this property showing that the fluctuations given by the graph structure are controlled for uniformly in time, whereas the random fluctuations given by the Brownian motions are not and will make the system escape from $1/2\pi$ after some (very long) time, i.e. a Large Deviation phenomenon.

In the supercritical regime, when $K > 1$, there is a manifold of stable stationary solutions corresponding to the synchronous states of the oscillators $\{\theta^{i,n}\}_{i=1, \dots, n}$ (see [51, Subsection 4.3] and [9]). Up to a rotation, all stable stationary solutions of (4.6) are given by

$$q(\theta) = \frac{\exp\{2Kr \cos(\theta)\}}{Z}, \quad (4.7)$$

where Z is the normalizing constant and $r = r(K)$ is the unique solution in $(0, 1)$ of a fixed point equation $r = \Psi(2Kr)$, see [9] for an explicit formula of Ψ . The parameter r is often referred to as the degree of synchronization of the system: r close to 0 indicates that the particles are scattered around the circle, r close to 1 that they are almost fully synchronized. We just recall that whenever $K < 1$, the fixed point equation has a unique solution $r = 0$, which in (4.7) boils down to the uniform measure $1/2\pi$, and whenever $K > 1$ the value $r = 0$ is still a solution but the corresponding measures solving (4.6) are unstable so that we will not consider them.

Let $K > 1$ and $0 < r < 1$. Observe that system (4.5) (and also (4.2)) is invariant under rotations, this property is maintained in the limit (4.6) and the manifold of stationary solutions M can be described as

$$M = \{q_\psi : q_\psi(\cdot) = q(\cdot - \psi), \psi \in \mathbb{T}\}. \quad (4.8)$$

It is possible to show that, unless one starts from the unstable manifold

$$U = \left\{ \mu \in \mathcal{P}(\mathbb{T}) : \int_{\mathbb{T}} \exp(i\theta) \mu(d\theta) = 0 \right\}, \quad (4.9)$$

the measure μ_t solution to (4.6) converges to some $q_\psi \in M$ as t tends to infinity, the phase $\psi \in \mathbb{T}$ depending only on μ_0 . Since each $q \in M$ is a stationary solution, the dynamics of μ_t is fully characterized for all times t .

1.4. The graph's perspective. The aim of this work is to investigate the weakest assumptions on the sequences $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ and $\{p_n\}_{n \in \mathbb{N}}$, such that the long time behavior of (4.2) is well understood: in other words, whenever system (4.2) is comparable to (4.5) or to the mean field limit (4.6), under a proper scale between size of the system n and some horizon time T_n .

The normalization sequence p_n has to be chosen such that the interaction term in (4.2) makes sense. At least, this requires the assumption that the quantity

$$\frac{1}{np_n} \sum_{j=1}^n \xi_{ij}^{(n)} \quad (4.10)$$

is of order one, for almost each vertex i in the graph.

REMARK 1.1. Observe that whenever (4.10) converges to zero or diverges, one should look for a different normalization in order to obtain a proper limit. A control on (4.10) is thus required to exclude degenerate cases, yet it cannot be sufficient for our purpose: whenever one considers a graph composed of two (or more) highly connected components, the degree of each vertex can be correctly defined, but one cannot expect the convergence of the empirical measure since the behavior on each component may differ, depending on the initial conditions! We refer to [31, Remark 1.2] and [36, Remark 1.4] for concrete examples and a precise analysis from this perspective, see also Remark 2.2 in the next section.

For $n \in \mathbb{N}$, define the normalized adjacency matrix $P^{(n)} = \{P_{ij}^{(n)}\}_{i,j=1,\dots,n}$ by

$$P_{ij}^{(n)} := \frac{\xi_{ij}^{(n)}}{p_n}, \quad \text{for } i, j = 1, \dots, n. \quad (4.11)$$

Recall that we do not assume any symmetry on $\xi^{(n)}$ and that it can also represent a multigraph. Define $\mathbf{1}^{(n)}$ as the adjacency matrix associated to the classical mean field model, i.e. $\mathbf{1}_{ij}^{(n)} = 1$ for $i, j = 1, \dots, n$. One would like to compare $P^{(n)}$ to $\mathbf{1}^{(n)}$.

It turns out that a sufficient condition for what we aim at, is given by a control on the difference between $P^{(n)}$ and $\mathbf{1}^{(n)}$ through the $l_\infty \rightarrow l_1$ norm. This norm is defined for a matrix $G = \{G_{ij}\}_{i,j=1,\dots,n}$ as

$$\|G\|_{\infty \rightarrow 1} := \sup_{\|s\|_\infty \leq 1} \|Gs\|_1 = \sup_{s, t \in \{-1, 1\}^n} Gs t^\top = \sup_{s_i, t_j \in \{-1, 1\}} \sum_{i,j=1}^n G_{ij} s_i t_j. \quad (4.12)$$

It has received a lot of attention in the last years: it appears in many applications in computer science (e.g. [57]) and it has been shown to be very useful in graphs concentration (e.g. [55, 72, 89]). Part of this success is because of the equivalence to the cut-norm (e.g. [3]) and, as already remarked in [55, 90], of Grothendieck's Inequality, which is recalled hereafter.

THEOREM 1.2 (Grothendieck's inequality, [93, Theorem 2.4]). *Let $\{a_{ij}\}_{i,j=1,\dots,n}$ be a $n \times n$ real matrix such that for all $s_i, t_j \in \{-1, 1\}$*

$$\sum_{i,j=1}^n a_{ij} s_i t_j \leq 1. \quad (4.13)$$

Then, there exists an universal 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

$$\sum_{i,j=1}^n a_{ij} \langle S_i, T_j \rangle_H \leq K_R. \quad (4.14)$$

It is indeed thanks to this inequality that $l_\infty \rightarrow l_1$ norm turns out to be the natural choice for our setting: an important part of the proof (Lemmas 3.2 and 4.3) consists in showing that the fluctuations due to the graph structure can be described by expressions like (4.14), and thus controlled by $\|\cdot\|_{\infty \rightarrow 1}$.

From now on, the only condition we require on $(\xi^{(n)}, p_n)_{n \in \mathbb{N}}$ is to satisfy:

$$\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1} = o(n^2), \quad (4.15)$$

or, in other words,

$$\lim_{n \rightarrow \infty} \sup_{s_i, t_j \in \{-1, 1\}} \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}^{(n)}}{p_n} - 1 \right) s_i t_j = 0. \quad (4.16)$$

In Proposition 4.A.3 it is shown that Erdős-Rényi random graphs with parameter p_n satisfy condition (4.16) almost surely, provided that $np_n \uparrow \infty$. We also provide a class of deterministic

graphs, Ramanujan graphs, that satisfies (4.16) (see Proposition 4.A.6) and give some link with the theory of graphons.

Appendix 4.A presents such results and includes remarks on the relationship between condition (4.16), the degree condition (4.10) and the connectivity of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$.

1.5. Set-up and notations. The closeness between μ_t^n and μ_t is studied through a norm which controls the bounded Lipschitz (or 1-Wasserstein) distance between probability measures, in an appropriate class of weighted Hilbert spaces $H_{-1,w}$. This class is defined as follows.

Denote by $\mathcal{C}_0^1(\mathbb{T})$ the space of \mathcal{C}^1 functions on the torus with zero mean and consider

$$\mathcal{L}_0^2 = \left\{ f \in \mathcal{L}^2(\mathbb{T}) : \int_{\mathbb{T}} f = 0 \right\}, \quad (4.17)$$

with canonical scalar product $(u, v) := \int_{\mathbb{T}} uv$, for $u, v \in \mathcal{L}_0^2$. Let $w \in \mathcal{C}^1(\mathbb{T}, (0, \infty))$ and V be the closure of $\mathcal{C}_0^1(\mathbb{T})$ with respect to the norm $\|\varphi\|_{H_{1,1/w}} = \sqrt{\int_{\mathbb{T}} \frac{(\varphi')^2}{w}}$ for $\varphi \in \mathcal{C}_0^1(\mathbb{T})$. It is easy to see that V is continuously and densely injected in \mathcal{L}_0^2 (thanks to the compactness of \mathbb{T} and Poincaré inequality). Moreover, one can define an inner product on V which makes it an Hilbert space $H_{1,1/w} := (V, \langle \cdot, \cdot \rangle_{H_{1,1/w}})$ where $\langle \varphi, \psi \rangle_{H_{1,1/w}} = \int_{\mathbb{T}} \frac{\varphi' \psi'}{w}$ for all $\varphi, \psi \in \mathcal{C}_0^1(\mathbb{T})$. The dual space of $H_{1,1/w}$ is denoted by $H_{-1,w}$. Observe that if $u, v \in \mathcal{L}_0^2$ and $v \in H_{1,1/w}$, then $u \in H_{-1,w}$ and

$$u(v) := \langle u, v \rangle_{-1,1} = (u, v), \quad (4.18)$$

where $\langle \cdot, \cdot \rangle_{-1,1}$ denotes the action of $H_{-1,w}$ on $H_{1,1/w}$, we omit the weight w .

The action of a probability measure μ on a test function h is denoted by $\langle \mu, h \rangle = \int h d\mu$: of course whenever u and v are regular enough, one has $u(v) = \langle u, v \rangle_{-1,1} = \langle u, v \rangle = (u, v)$, where we have abused of notation, denoting the density of a probability measure by the probability measure itself.

Finally, observe that different weights w give equivalent norms so that whenever the geometry of the space is not important, we consider the case $w \equiv 1$ and simply note $\|\cdot\|_{-1}$. More information about the construction of $H_{-1,w}$ are given in Appendix 4.B.

Hereafter we drop the dependency on \mathbb{T} , i.e. we write \mathcal{C}_0^1 instead of $\mathcal{C}_0^1(\mathbb{T})$ and so on for the other spaces and integrals.

2. Main results

We present the results in three consecutive subsections: we start by the finite time behavior, then pass to the supercritical regime and, finally, the subcritical case.

In all results, the convergence of empirical measures is stated in the norm $\|\cdot\|_{-1}$. It is not difficult to see that the difference of two probability measures belongs to H_{-1} and that the distance induced on $\mathcal{P}(\mathbb{T})$ controls the bounded Lipschitz distance (or, equivalently, the 1-Wasserstein distance). These details are covered in Appendix 4.B.

Recall that throughout the paper, we only require $(\xi^{(n)}, p_n)_{n \in \mathbb{N}}$ to satisfy condition (4.16) and $\mu_0 \in \mathcal{P}(\mathbb{T})$, no independence between μ_0^n and $\xi^{(n)}$ is demanded.

2.1. The finite time behavior. We give the result and then comment it.

THEOREM 2.1. *Let $K \geq 0$. Suppose that for all $\varepsilon_0 > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\|\mu_0^n - \mu_0\|_{-1} \leq \varepsilon_0 \right) = 1. \quad (4.19)$$

Then, for every fixed time $T > 0$ and for every $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T]} \|\mu_t^n - \mu_t\|_{-1} \leq \varepsilon \right) = 1. \quad (4.20)$$

The finite time behavior of weakly interacting particle systems on graphs is already known under suitable hypothesis on the initial conditions and on the graph sequence, we refer to Subsection 2.5 for a comprehensive literature on the subject. We decide to present Theorem 2.1 because, contrary to all the previous results, it does not require any independence between initial conditions and (the realization of) the sequence of graphs. In particular, even if one accurately assigns the initial conditions for each vertex, the mixing properties of the graph will shuffle all the information and make the empirical measure converge, losing any memory of the initial coupling. This property is crucial for studying the longtime behavior as pointed out in the next subsections.

Observe that Theorem 2.1 implies the existence of a unique giant component in $\{\xi^{(n)}\}_{n \in \mathbb{N}}$, as pointed out in the next remark.

REMARK 2.2. The result is independent of K . First observe that this implies the uniqueness of a giant component: if there are two, then one can accurately prepare the initial conditions so to obtain different behaviors on the twos and loose the proximity to (4.6). Secondly, with the same argument one deduces that the size of the giant component is asymptotically n , i.e. all but $o(n)$ vertices are connected. Finally, the existence comes from the fact that the system cannot synchronize on components of size $o(n)$, no matter the value of K . Lemma 4.A.2 shows that condition (4.10) indeed implies the existence of a giant component of size asymptotically n .

2.2. Long time behavior in the supercritical regime. In the supercritical regime, we suppose to be already close to the manifold M at time 0. However, since we do not assume any independence between graph and initial data, this hypothesis can be weakened by requiring the initial condition μ_0 to be in the domain of attraction of M , i.e. $\mu_0 \in \mathcal{P}(\mathbb{T}) \setminus U$, and using Theorem 2.1. One can then start after some time T with initial condition given now by μ_T^n (and dependent on the graph!): if T is big enough, than μ_T^n will be close to M . Observe that the choice of T depends only on how close to M μ_t has to be, it thus depends only on μ_0 .

Before stating the theorem, we define the distance of a probability measure from M . For $\mu \in \mathcal{P}(\mathbb{T})$, let

$$\text{dist}(\mu, M) := \inf_{\nu \in M} \|\mu - \nu\|_{-1}. \quad (4.21)$$

We are ready for the main result of this section.

THEOREM 2.3. *Let $K > 1$. Suppose there exists $\psi \in \mathbb{T}$ such that for every $\varepsilon_0 > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\|\mu_0^n - q_\psi\|_{-1} \leq \varepsilon_0 \right) = 1. \quad (4.22)$$

Then, for every positive sequence $\{T_n\}_{n \in \mathbb{N}}$ such that $T_n = \exp(o(n))$, and for all $\varepsilon > 0$ small enough

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T_n]} \text{dist}(\mu_t^n, M) \leq \varepsilon \right) = 1. \quad (4.23)$$

Theorem 2.3 implies the proximity of the empirical measure to the manifold of solutions of the McKean-Vlasov equation (4.6) for almost exponential times. On this time scale, Large Deviation phenomena take control of the finite system (e.g. [34, 48, 92]) making it escape from the stationary solutions.

Observe that we do not prove the closeness to the mean field limit μ . Indeed, it is by now well known that, on longtime scales, the mean field limit is not a faithful description of the finite system of n diffusions. In other words, the behavior of $\mu_{T_n}^n$ highly depends on the scale of time T_n under consideration, whereas the dynamics of μ_t is deterministic and completely known for large t , i.e. it sticks to q_ψ .

In [10], a deep analysis of the longtime dynamics for the classical mean field system (4.5) is presented. Namely, it is shown that μ_t solution to the PDE (4.6) is a reasonable approximation of $\bar{\mu}_t^n$ for time scales of order $O(\log n)$. On times proportional to n , the dynamics of the empirical measure can be coupled to a Brownian motion on M with a non trivial diffusion coefficient that can be explicitly computed (see [10, Theorem 1.1]). Whereas the PDE prescribes the system to stay synchronized on a fixed phase, the noise induced by the Brownian motions makes this phase oscillate and it turns out that the oscillations become significant on times proportional to the size of the system n .

We do not show this property, yet extend the closeness to M for exponential times, whereas in [10] this is shown up to polynomial times.

Theorem 2.3, as Theorem 2.1, does not depend on the speed of convergence of the condition on the graph (4.16). The escaping time is indeed only due to the stochastic nature of the system, given by the Brownian motions, and it cannot be improved as explained above. The reason why one can control the perturbation induced by the graph structure for long times is somehow hidden in the martingale properties of μ^n and in the fact that we do not really analyze the dynamics near M (which can, a priori, depend on the graph). We refer to the proof of the subcritical regime for a clear control on the perturbations given by the graph, through the exponential stability of the stationary solution.

2.3. Longtime behavior in the subcritical regime. The subcritical regime is somehow easier than the supercritical regime since there is an unique stable stationary solution. We decide to include this case firstly because, to the author's knowledge, it is missing in the literature

and, secondly, because the proof enlightens some aspects hidden in the supercritical regime. As a byproduct, we obtain the equivalent of maximal inequalities for Ornstein-Uhlenbeck processes in infinite dimensional Hilbert spaces, see Corollary 2.5.

THEOREM 2.4. *Let $0 \leq K < 1$. Suppose that condition (4.19) holds, i.e. for all $\varepsilon_0 > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\|\mu_0^n - \mu_0\|_{-1} \leq \varepsilon_0 \right) = 1. \quad (4.24)$$

Then, for every positive sequence $\{T_n\}_{n \in \mathbb{N}}$ such that $T_n = \exp(o(n))$, and for all $\varepsilon > 0$ small enough

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T_n]} \|\mu_t^n - \mu_t\|_{-1} \leq \varepsilon \right) = 1. \quad (4.25)$$

Observe that, since μ converges as t tends to infinity to $1/2\pi$ for all initial conditions μ_0 , then Theorem 2.4 implies the proximity of μ_t^n to the stable solution up to exponential times.

Of independent interest, we present a corollary of Theorem 2.4 in the limit case $K = 0$. This result seems to be well known, yet the author was unable to find it elsewhere.

COROLLARY 2.5. *Let μ^n be the empirical measure of n independent Brownian motions $\{B^{j,n}\}_{j=1,\dots,n}$ on \mathbb{T} with initial conditions $\{\theta_0^i\}_{1 \leq i \leq n}$ satisfying (4.19). Then, there exist $C > 0$ and $T_0 > 0$ such that for all $T > T_0$, the following maximal inequality holds:*

$$\mathbf{E} \left[\sup_{t \in [T_0, T]} \left\| \mu_t^n - \frac{1}{2\pi} \right\|_{-1}^2 \right] \leq C \frac{\log(1 + T - T_0)}{n}. \quad (4.26)$$

Corollary 2.5 shows a maximal inequality for the empirical measure of n independent Brownian motions on the torus, establishing the SPDE version of the result for Ornstein-Uhlenbeck processes presented in [52] for stochastic ordinary differential equations.

Observe that if the initial conditions and the graph are exchangeable (not necessarily independent), then Corollary 2.5 and a classical result by Sznitman ([98, Proposition 2.2]) imply the creation of chaos for all times $T_n = o(\exp(n))$.

2.4. Organization of the paper. This section ends presenting the existing literature and giving an outline of the proof for the three theorems.

Sections 3, 4 and 5 concern the proofs of the three results. In particular, Section 3 is devoted to the long time dynamics close to M , it starts from the derivation of a mild formulation for the empirical measure, then proceeds with the control on the graph and the noise, and it ends with the proof of Theorem 2.3. Section 4 concerns the subcritical regime where a different control on the perturbations is given. Finally, Section 5 proves Theorem 2.1 by using slight variations of the previous techniques.

Appendix 4.A gives a few examples of graph sequences that satisfy condition (4.16), together with remarks on the degrees and connectivity of such sequences. Appendix 4.B contains information about the Hilbert spaces $H_{-1,\omega}$ and the linear operator L_ψ .

2.5. A glance at the existing literature. The results presented here are at a crossroads of two different research areas: the long time dynamics of stochastic differential equations and the role of a network in a mean field model.

Concerning the long time behavior of weakly interacting particle systems, Theorems 2.3 and 2.4 can be seen as a complement to the previous results presented in [10], filling the gap of the exponential time scale which has not been addressed so far. To the author's knowledge, they also represent the first equivalent, in infinite dimensional Hilbert spaces, to the famous result for stochastic ordinary differential equations in \mathbb{R}^d by Friedlin and Wentzell ([48]).

Looking at variations on the same model, the behavior of the classical Kuramoto model with intrinsic frequencies has been studied in [78], showing that the longtime dynamics is indeed dependent on the quenched setting given by the frequencies. A macroscopic constant speed in the phase appears on time scales of order $O(\sqrt{n})$, making the effects of the noise vanish. The results and the techniques presented here should be easily adaptable to this case showing the proximity to the manifold of solutions for long times, yet losing the precise characterization of the motion on M for which a deeper analysis is needed.

The Kuramoto model is an example of system which admits more than one stable stationary solution, a continuous manifold as already precised, and that's one of the reasons why it shows a rich phenomenology depending on the time scale under consideration. For similar results on different models, one has to dip in the context of stochastic partial differential equations (SPDEs) with vanishing noise. Since the aim of this work is more oriented on the effects of the network rather than the longtime dynamics of SPDEs, the author refers to the bibliography in [10, 78] for a more comprehensive discussion.

Turning to interacting particle systems on graphs, the subject has become an interesting topic in the mathematical community given the several applications to complex systems, in particular regarding the Kuramoto model and synchronization phenomena (e.g. [1, 94]), yet it has always been addressed on a finite time scale or up to times slowly diverging on n , i.e. $T_n = O(\log n)$.

The first articles [13, 36] attack the problem under a propagation of chaos viewpoint, requiring independent and identically distributed initial conditions and also independent of the realization of the graph. In this setting, the condition on the graph boils down to a condition on the degrees only so that very general graphs are allowed (see again [36, Remark 1.4] and [31, Remark 1.2]). Regarding more inhomogeneous settings, [77] extends [36] to graphons and [90] presents a Large Deviation result again in the graphon setting. Observe that [90] already makes use of Grothendieck's inequality and the norm $\|\cdot\|_{\infty \rightarrow 1}$ to control the graphs fluctuations. Up to now, the only result not assuming independence in the initial data is given by [31], where general systems of interacting particles are defined on Erdős-Rényi random graphs and the empirical measure is shown to satisfy a Law of Large Number and a Large Deviation principle, implying the convergence to the respective mean field limit.

In the deterministic setting, the Kuramoto model has been studied on different networks with various hypothesis on the initial conditions, we refer to [27, 82] and references therein.

In all the cited works, the condition on the normalization p_n is slightly stronger, or equivalent, to the one required in (4.16), see in particular Propositions 4.A.4 and 4.A.6 in Appendix 4.A. If np_n is not diverging as n tends to infinity, i.e. the case of sparse graphs, the limiting behavior of the empirical measure seems to be rather different from the mean field limit, see [70, 91].

2.6. Outline of the proofs. The three theorems are proven in a similar way and the main ingredients are given by

- (1) An equation written in mild form satisfied by μ^n for each $n \in \mathbb{N}$;
- (2) The control on the perturbations given by the graph structure through Grothendieck's Inequality;
- (3) The control on the random perturbations given by the Brownian motions through maximal inequalities for self-normalized processes.

The mild formulation will be different in all the three cases and will depend on the linear dynamics around M or $1/2\pi$ or on the properties of μ_t . In Section 3, we give a full derivation of the stochastic partial differential equation satisfied by $\mu_t^n - q_\psi$ in a neighborhood of M .

The control on the graph will also depend whether there is a strong contraction given by the dynamics, or not. Whenever the evolution is contracting in all direction, as in the subcritical case around $1/2\pi$, these perturbations can be controlled uniformly in time, we refer to Lemma 4.3 for a precise statement.

A fine control on the random perturbations turns out to be rather delicate and one has to exploit all the properties associated to the Hilbert structure as well as the ones associated to the linear dynamics around the stationary solutions to get the job done. We give two independent explicit proofs:

- Around M , we study the noise using of a strong result on self-normalized martingales ([35]);
- Around $1/2\pi$, we extend a result on maximal inequalities for Ornstein-Uhlenbeck processes ([52]), to an infinite dimensional setting.

Once the perturbations are controlled, a Gronwall-like lemma is used to bound the difference between μ^n and the relative target. In the supercritical case, we need to set up an (easy) iterative scheme in order to estimate the distance between μ^n and M on bounded time intervals, and then make use of the martingale property of system (4.2) to extend the result up to almost exponential times; in the subcritical case, the result is directly obtained by the bound on the noise, the graph perturbations are indeed controlled for all times.

3. Longtime dynamics close to M

In a neighborhood of M , one can exploit the properties of the linear dynamics around q_ψ . For $u \in \mathcal{L}_0^2$, let

$$L_\psi u := \frac{1}{2} \partial_\theta^2 u - \partial_\theta [u(J * q_\psi) + q_\psi(J * u)], \quad (4.27)$$

be the linear operator at q_ψ , its domain is given by $D(L_\psi) = \{u \in \mathcal{C}^2(\mathbb{T}), \int_{\mathbb{T}} u(\theta) d\theta = 0\}$. The operator L_ψ is self-adjoint in $H_{-1,1/q_\psi}$ and its adjoint L_ψ^* in \mathcal{L}_0^2 has the following expression

$$L_\psi^* u = \frac{1}{2} \partial_\theta^2 u + (J * q_\psi) \partial_\theta u - (J * q_\psi \partial_\theta u) - \int_{\mathbb{T}} (J * q_\psi \partial_\theta u), \quad (4.28)$$

and domain $D(L_\psi^*) = D(L_\psi)$.

We recall here the most important properties of L_ψ , referring to Appendix 4.B for more informations. The linear operator $-L_\psi$ has compact resolvent and its spectrum lies in $[0, \infty)$: the smallest eigenvalue $\lambda_0 := 0$, associated to the eigenfunction $\partial_\theta q_\psi$, is isolated from the rest of the spectrum. In particular, this implies that H_{-1} can be decomposed into a direct sum $T_\psi \oplus N_\psi$, where $T_\psi = \text{Span}(\partial_\theta q_\psi)$; we denote by P_ψ^0 the projection on T_ψ along N_ψ and $P_\psi^s = 1 - P_\psi^0$. Observe that both P_ψ^0 and P_ψ^s commute with L_ψ (e.g. [58]).

Let $\{\lambda_0 < \lambda_1 \leq \dots\} \subset [0, \infty)$ denote the set of eigenvalues and let $\{e_l^\psi\}_{l=0,1,\dots}$ be the correspondent set of eigenfunctions, normalized in $H_{-1,1/q_\psi}$, i.e.

$$-L_\psi e_l^\psi = \lambda_l e_l^\psi, \quad \text{for } l = 0, 1, \dots \quad (4.29)$$

Observe that $e_l^\psi \in \mathcal{C}^\infty(\mathbb{T})$. Moreover, the eigenvalues do not depend on the phase ψ , whereas the eigenfunctions do in a rather simple way given by the rotation symmetry of the system, i.e.

$$e_l^\psi(\cdot) = e_l^\theta(\cdot + \psi - \theta), \quad \text{for } l = 0, 1, \dots \quad (4.30)$$

As a matter of fact, we will study the system only around some q_ψ . The dual eigenfunctions f_l^ψ associated to L_ψ^* will play an important role for studying the noise perturbation, their properties are studied in Proposition 4.B.3.

From the previous properties one deduces that L_ψ (respectively L_ψ^*) generates a strong continuous semigroup on H_{-1} (resp. H_1) that we denote by e^{tL_ψ} (resp. $e^{tL_\psi^*}$). These semigroups have many properties that we will recall and use throughout this section, see Proposition 4.B.2 for a general statement.

A final remark: we use the letter C for all the constants even if they are possibly different, the value of C can change from one line to another if the constant is replaced by another constant and the context is clear.

3.1. The mild formulation around $q_\psi \in M$. As shown in [10], $\bar{\mu}^n$ satisfies a SPDE written in mild form once it is close to M ; in this subsection we extend this formulation to μ_t^n . Let $\nu^n := \mu_t^n - q_\psi$, then

PROPOSITION 3.1. *The process $\nu_t^n \in H_{-1}$ satisfies the following stochastic partial differential equation in $\mathcal{C}^0([0, T], H_{-1})$:*

$$\nu_t^n = e^{tL_\psi} \nu_0^n - \int_0^t e^{(t-s)L_\psi} \partial_\theta [\nu_s^n (J * \nu_s^n)] ds - g_t^n + z_t^n, \quad (4.31)$$

where

$$g_t^n = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right] ds, \quad (4.32)$$

and $z_t^n \in H_{-1}$ is defined for $h \in H_1$ by

$$\langle z_t^n, h \rangle_{-1,1} = \frac{1}{n} \sum_{j=1}^n \int_0^t \left[\partial_\theta e^{(t-s)L_\psi^*} h \right] (\theta_s^{j,n}) dB_s^j. \quad (4.33)$$

PROOF. Let $F = F_t(\theta) \in \mathcal{C}^{1,2}([0, \infty) \times \mathbb{T})$, with $\int F_t = 0$ for all $t \geq 0$. For some $t \geq 0$, an application of Ito formula, together with the definition of L_ψ^* gives

$$\begin{aligned} \langle \mu_t^n - q_\psi, F_t \rangle &= \langle \mu_0^n - q_\psi, F_0 \rangle + \int_0^t \langle \mu_s^n - q_\psi, \partial_s F_s + L_\psi^* F_s \rangle ds + \\ &+ \int_0^t \langle (\mu_s^n - q_\psi)(J * (\mu_s^n - q_\psi)), \partial_\theta F_s \rangle ds + G_t^n(F) + Z_t^n(F), \end{aligned} \quad (4.34)$$

where

$$\begin{aligned} G_t^n(F) &= \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) J(\theta_s^{i,n} - \theta_s^{j,n}) \partial_\theta F_s(\theta_s^{i,n}) ds, \\ Z_t^n(F) &= \frac{1}{n} \sum_{j=1}^n \int_0^t \partial_\theta F_s(\theta_s^{j,n}) dB_s^j. \end{aligned} \quad (4.35)$$

The properties of $e^{tL_\psi^*}$, see Proposition 4.B.2, assure that the function

$$F = F_s(\theta) = e^{(t-s)L_\psi^*} h(\theta), \quad \text{for some } h \in \mathcal{C}^2(\mathbb{T}), \quad \int h = 0, \quad (4.36)$$

is $\mathcal{C}^{1,2}([0, t] \times \mathbb{T})$. But then $\partial_s F_s = -L_\psi^* F_s$ and one obtains

$$\langle \nu_t^n, F_t \rangle = \langle \nu_0^n, e^{tL_\psi^*} h \rangle + \int_0^t \langle \nu_s^n (J * \nu_s^n), \partial_\theta e^{(t-s)L_\psi^*} h \rangle ds + g_t^n(h) + z_t^n(h), \quad (4.37)$$

where we have used the definition of ν_t^n and the notations

$$g_t^n(h) = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) J(\theta_s^{i,n} - \theta_s^{j,n}) \left[\partial_\theta e^{(t-s)L_\psi^*} h \right] (\theta_s^{i,n}) ds, \quad (4.38)$$

$$z_t^n(h) = \frac{1}{n} \sum_{j=1}^n \int_0^t \left[\partial_\theta e^{(t-s)L_\psi^*} h \right] (\theta_s^{j,n}) dB_s^j. \quad (4.39)$$

We aim at proving that (4.37) is the weak formulation of the mild equation (4.31) in H_{-1} .

Let $\{\nu_l\}_{l \geq 1} \subset \mathcal{L}_0^2$ such that $\nu_l \xrightarrow{l \uparrow \infty} \nu_0^n$ in H_{-1} . Then, for $h \in \mathcal{C}^2$

$$\langle \nu_l, e^{tL_\psi^*} h \rangle_{-1,1} = (\nu_l, e^{tL_\psi^*} h) = (e^{tL_\psi} \nu_l, h) = \langle e^{tL_\psi} \nu_l, h \rangle_{-1,1}. \quad (4.40)$$

By continuity of the operators, $e^{tL_\psi} \nu_l$ converges in H_{-1} to $e^{tL_\psi} \nu_0^n$ as $l \uparrow \infty$. Taking the limit for $l \uparrow \infty$ in both sides of (4.40), we deduce

$$\langle \nu_0^n, e^{tL_\psi^*} h \rangle_{-1,1} = \langle e^{tL_\psi} \nu_0^n, h \rangle_{-1,1}. \quad (4.41)$$

We now focus on

$$\omega_s^n := \nu_s^n(J * \nu_s^n). \quad (4.42)$$

Consider $\{\nu_{s,l}\}_{l \geq 1} \subset \mathcal{L}_0^2$ which converges to ν_s^n in H_{-1} as $l \uparrow \infty$, and define

$$\omega_{s,l} := \nu_{s,l}(J * \nu_{s,l}^n). \quad (4.43)$$

For any $l \geq 1$, it holds

$$\begin{aligned} \langle \omega_{s,l}, \partial_\theta e^{(t-s)L_\psi^*} h \rangle_{-1,1} &= \left(\omega_{s,l}, \partial_\theta e^{(t-s)L_\psi^*} h \right) = \\ &= - \left(e^{(t-s)L_\psi} \partial_\theta \omega_{s,l}, h \right) = - \langle e^{(t-s)L_\psi} \partial_\theta \omega_{s,l}, h \rangle_{-1,1}. \end{aligned} \quad (4.44)$$

Using the properties of the semigroup one obtains

$$\begin{aligned} \left| \langle e^{(t-s)L_\psi} \partial_\theta (\omega_{s,l} - \omega_s^n), h \rangle_{-1,1} \right| &\leq \|h\|_1 \left\| e^{(t-s)L_\psi} \partial_\theta (\omega_{s,l} - \omega_s^n) \right\|_{-1} \leq \\ &\leq \|h\|_1 \frac{C}{\sqrt{t-s}} \left\| \partial_\theta (\omega_{s,l} - \omega_s^n) \right\|_{-2} = \|h\|_1 \frac{C}{\sqrt{t-s}} \left\| \omega_{s,l} - \omega_s^n \right\|_{-1}, \end{aligned} \quad (4.45)$$

which implies

$$\left\| e^{(t-s)L_\psi} \partial_\theta (\omega_{s,l} - \omega_s^n) \right\|_{-1} \leq \frac{C}{\sqrt{t-s}} \left\| \omega_{s,l} - \omega_s^n \right\|_{-1}. \quad (4.46)$$

Since h is regular and $\omega_{s,l} \xrightarrow{l \uparrow \infty} \omega_s^n$ in H_{-1} , this implies

$$\langle \omega_s^n, \partial_\theta e^{(t-s)L_\psi^*} h \rangle_{-1,1} = - \langle e^{(t-s)L_\psi} \partial_\theta \omega_s^n, h \rangle_{-1,1}. \quad (4.47)$$

We now observe from (4.46) that

$$\left\| e^{(t-s)L_\psi} \partial_\theta \omega_s^n \right\|_{-1} \leq \frac{C}{\sqrt{t-s}} \quad (4.48)$$

thus the integral in (4.31)

$$\int_0^t e^{(t-s)L_\psi} \partial_\theta [\nu_s^n(J * \nu_s^n)] ds \quad (4.49)$$

is almost surely finite. Using [101, Theorem 1, p.133], we deduce that (4.49) makes sense as a Bochner integral in H_{-1} . The continuity is a direct consequence of the continuity of e^{tL_ψ} .

Assume that $g_t^n(h) = \langle g_t^n, h \rangle_{-1,1}$ and $z_t^n(h) = \langle z_t^n, h \rangle_{-1,1}$ are well defined and continuous with respect to t for all $h \in H_1$; we have shown that

$$\begin{aligned} \langle \nu_t^n, h \rangle_{-1,1} &= \langle e^{tL_\psi} \nu_0^n, h \rangle_{-1,1} + \\ &- \left\langle \int_0^t e^{(t-s)L_\psi} \partial_\theta [\nu_s^n (J * \nu_s^n)] ds, h \right\rangle_{-1,1} - \langle g_t^n, h \rangle_{-1,1} + \langle z_t^n, h \rangle_{-1,1}. \end{aligned} \quad (4.50)$$

Since (4.50) holds for all $h \in H_1$, the identity (4.31) follows. All elements in (4.31) take values in $\mathcal{C}^0([0, T], H_{-1})$ and the proof is then concluded modulo regularity and wellposedness of g^n and z^n . We refer to Lemma 3.2 and Lemma 3.3 which are presented in the next subsection. \square

3.2. Control on the perturbations. Two kinds of perturbations are present in the SPDE (4.31): z^n given by the stochastic nature of the system and g^n given by the presence of a network structure. In this subsection, we exhibit the control over the two perturbations. Observe that all the estimates are independent of ψ .

We start with the control on the graph structure, which uses Grothendieck's Inequality presented in Theorem 1.2.

LEMMA 3.2 (Wellposedness and bounds on g_t^n). *For $n \in \mathbb{N}$ and $t \geq 0$, let g_t^n be given by*

$$g_t^n = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right] ds. \quad (4.51)$$

Then

(1) $g_t^n \in \mathcal{C}^0([0, \infty), H_{-1})$. In particular, for every $h \in H_1$ and $t \geq 0$

$$g_t^n(h) = -\frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) J(\theta_s^{i,n} - \theta_s^{j,n}) \left[\partial_\theta e^{(t-s)L_\psi^*} h \right] (\theta_s^{i,n}) ds. \quad (4.52)$$

(2) There exists $D > 0$ such that

$$\|g_t^n\|_{-1} \leq D\sqrt{t} \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}, \quad \text{for all } t \geq 0. \quad (4.53)$$

PROOF. Fix n large. Consider $\{\phi_l\}_{l \geq 1} \subset \mathcal{C}^\infty$ such that $\phi_l \geq 0$, $\phi_l(\theta) = 0$ for $\theta \in [1/l, 2\pi - 1/l]$, $\int \phi_l = 1$ for every $l \geq 1$ and $\lim_{l \rightarrow \infty} \int F \phi_l = F(0)$ for every $F \in \mathcal{C}^0$. For $i = 1, \dots, n$, define

$$\phi_{s,l}^i := \phi_l * \delta_{\theta_s^{i,n}}. \quad (4.54)$$

We start by establishing (4.52). For each $h \in \mathcal{C}^2$

$$\begin{aligned}
& \left\langle \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \phi_{s,l}^i(J * \delta_{\theta_s^{j,n}}), \partial_\theta e^{(t-s)L_\psi^*} h \right\rangle_{-1,1} = \\
& = \left(\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \phi_{s,l}^i(J * \delta_{\theta_s^{j,n}}), \partial_\theta e^{(t-s)L_\psi^*} h \right) = \\
& = - \left(\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\phi_{s,l}^i(J * \delta_{\theta_s^{j,n}}) \right], h \right) = \\
& = - \left\langle \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\phi_{s,l}^i(J * \delta_{\theta_s^{j,n}}) \right], h \right\rangle_{-1,1}
\end{aligned} \tag{4.55}$$

But $\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \phi_{s,l}^i(J * \delta_{\theta_s^{j,n}})$ converges to $\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \delta_{\theta_s^{i,n}}(J * \delta_{\theta_s^{j,n}})$ since

$$\left\| \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \left(\phi_{s,l}^i - \delta_{\theta_s^{i,n}} \right) (J * \delta_{\theta_s^{j,n}}) \right\|_{-1} \leq \frac{1}{p_n} \sup_{i=1,\dots,n} \left\| \phi_{s,l}^i - \delta_{\theta_s^{i,n}} \right\|_{-1}, \tag{4.56}$$

which tends to zero as l tends to infinity.

Thanks to the properties of the semigroup, the same holds true for

$$\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\left(\phi_{s,l}^i - \delta_{\theta_s^{i,n}} \right) (J * \delta_{\theta_s^{j,n}}) \right]; \tag{4.57}$$

indeed, by Proposition 4.B.2,

$$\begin{aligned}
& \left\| \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\left(\phi_{s,l}^i - \delta_{\theta_s^{i,n}} \right) (J * \delta_{\theta_s^{j,n}}) \right] \right\|_{-1} \leq \\
& \leq \frac{C}{p_n \sqrt{t-s}} \sup_{i=1,\dots,n} \left\| \phi_{s,l}^i - \delta_{\theta_s^{i,n}} \right\|_{-1}.
\end{aligned} \tag{4.58}$$

A similar argument shows that

$$\left\| \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\delta_{\theta_s^{i,n}}(J * \delta_{\theta_s^{j,n}}) \right] \right\|_{-1} \leq \frac{C}{p_n \sqrt{t-s}}, \tag{4.59}$$

which, in turn, implies that

$$\frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\delta_{\theta_s^{i,n}}(J * \delta_{\theta_s^{j,n}}) \right] ds \tag{4.60}$$

is almost surely finite and continuous with respect to t . We deduce (4.52).

For the second part (4.53), observe that

$$\begin{aligned} & \left\langle \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_\psi} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right], h \right\rangle_{-1,1} = \\ & = -\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \langle \delta_{\theta_s^{i,n}}, (J * \delta_{\theta_s^{j,n}}) \partial_\theta e^{(t-s)L_\psi^*} h \rangle_{-1,1}. \end{aligned} \quad (4.61)$$

We claim that this last term can be controlled by $\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}$ through Grothendieck's inequality. By choosing $H = H_{-1}$ and

$$\begin{aligned} a_{ij} &= \left(\frac{\xi_{ij}}{p_n} - 1 \right), \\ S_i &= \delta_{\theta_s^{i,n}}, \\ T_j &= \frac{\sqrt{t-s}}{C} \left(J * \delta_{\theta_s^{j,n}} \right) \partial_\theta e^{(t-s)L_\psi^*} \frac{h}{\|h\|_1}, \end{aligned} \quad (4.62)$$

Theorem 1.2 allows us to bound the expression in (4.61) by

$$K_R \frac{C}{\sqrt{t-s}} \|h\|_1 \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}. \quad (4.63)$$

This shows that

$$\|g_t^n\|_{-1} \leq K_R C \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2} \int_0^t \frac{1}{\sqrt{t-s}} ds = D \sqrt{t} \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}, \quad (4.64)$$

where $D := K_R C / 2 > 0$. The proof is concluded. \square

We now turn to the stochastic term z^n in (4.31). Recall that L_ψ^* is diagonal in the basis $\{f_l^\psi\}_{l \geq 0}$ of H_{1,q_ψ} , with eigenvalues denoted by $\{\lambda_l\}_{\lambda \geq 0}$, see Proposition 4.B.3. We precisely analyze z^n through its coefficients in the orthonormal basis given by L_ψ^* .

LEMMA 3.3 (Wellposedness and bounds on z_t^n). *For $n \in \mathbb{N}$ and $t > 0$, let z_t^n be defined by*

$$z_t^n = \sum_{l \geq 1} \langle z_t^n, e_l^\psi \rangle_{H_{-1,1/q_\psi}} e_l^\psi. \quad (4.65)$$

Then

(1) $z^n \in \mathcal{C}^0([0, \infty), H_{-1})$ almost surely. In particular, for every $h \in H_1$

$$z_t^n(h) = \frac{1}{n} \sum_{j=1}^n \int_0^t \left[\partial_\theta e^{(t-s)L_\psi^*} h \right] (\theta_s^{j,n}) dB_s^j. \quad (4.66)$$

(2) For every $T > 0$, there exists a constant $Z = Z(T) > 0$, such that for n large enough it holds that

$$\forall \eta > 0, \quad \mathbf{P} \left(\sup_{t \in [0, T]} \|z_t^n\|_{-1} > \eta \right) \leq \exp \{-Zn\eta^2\}. \quad (4.67)$$

PROOF. We start by observing that the definition of z_t^n in the basis of $H_{-1,1/q_\psi}$, coincides with the one give in the mild formulation. For $h = \sum_{l \geq 0} \langle h, f_l^\psi \rangle_{H_{1,q_\psi}} f_l^\psi$, one obtains

$$\begin{aligned} \langle z_t^n, h \rangle_{-1,1} &= \sum_{l \geq 0} \langle z_t^n, e_l^\psi \rangle_{H_{-1,1/q_\psi}} \langle h, f_l^\psi \rangle_{H_{1,q_\psi}} = \\ &= \sum_{l \geq 0} z_t^n(f_l^\psi) \langle h, f_l^\psi \rangle_{H_{1,q_\psi}} = z_t^n(h), \end{aligned} \quad (4.68)$$

where we have used the properties of e_l^ψ and f_l^ψ , see Proposition 4.B.3.

Before proving (i), we prove (ii) and this will imply the existence of a continuous version of z^n almost surely.

Concerning (ii), we start by observing that

$$\|z_t^n\|_{-1,1/q_\psi} = \sum_{l \geq 0} \left| z_t^n(f_l^\psi) \right|^2. \quad (4.69)$$

Let $l \geq 1$ and consider $z_t^n(f_l^\psi)$, by the definition of f_l^ψ and the properties of the semigroup, one gets

$$z_t^n(f_l^\psi) = \frac{1}{n} \sum_{j=1}^n \int_0^t \left[\partial_\theta e^{(t-s)L_\psi^*} f_l^\psi \right] (\theta_s^{j,n}) dB_s^j = \frac{1}{n} \sum_{j=1}^n \int_0^t e^{-(t-s)\lambda_l} \left[\partial_\theta f_l^\psi \right] (\theta_s^{j,n}) dB_s^j. \quad (4.70)$$

Set $c = \sup_{l \geq 0} \left| \partial_\theta f_l^\psi \right| < \infty$, see Proposition 4.B.3. We rewrite the last expression as

$$z_t^n(f_l^\psi) = \frac{ce^{-t\lambda_l}}{\sqrt{2\lambda_l n}} A_t, \quad (4.71)$$

where A_t is a continuous martingale given by

$$A_t = \sqrt{\frac{2\lambda_l}{c^2 n}} \sum_{j=1}^n \int_0^t e^{s\lambda_l} \left[\partial_\theta f_l^\psi \right] (\theta_s^{j,n}) dB_s^j \quad (4.72)$$

and quadratic variation bounded by

$$\langle A \rangle_t = \frac{2\lambda_l}{c^2 n} \sum_{j=1}^n \int_0^t e^{2s\lambda_l} \left[\partial_\theta f_l^\psi \right]^2 (\theta_s^{j,n}) ds \leq 2\lambda_l \int_0^t e^{2\lambda_l s} ds \leq e^{2t\lambda_l} - 1. \quad (4.73)$$

From (4.71) and (4.73), one deduces that $z_t^n(f_l^\psi)$ is a self normalized process. For estimating $\mathbf{P} \left(\sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2 > \eta \right)$, we can this use the following result

THEOREM 3.4 ([35, Theorem 4.1 and the following remark]). *Let $T > 0$, $\alpha \in (0, \frac{1}{2})$ and $(A_t)_{t \in [0, T]}$ be a martingale with $A_0 = 0$. There exists $C > 0$, depending only on α , such that*

$$\mathbf{E} \left[\sup_{t \in [0, T]} \exp \left\{ \frac{\alpha A_t^2}{\langle A \rangle_t \log \log (\langle A \rangle_t \vee e^2)} \right\} \right] \leq C. \quad (4.74)$$

By standard computations, we obtain

$$\begin{aligned}
\mathbf{P} \left(\sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2 > \eta \right) &= \mathbf{P} \left(\sup_{t \in [0, T]} e^{-2t\lambda_l} A_t^2 > \frac{2\lambda_l n}{c^2} \eta \right) = \\
&= \mathbf{P} \left(\sup_{t \in [0, T]} \frac{A_t^2}{4e^{2t\lambda_l} \log(2 + 2T\lambda_l)} > \frac{\lambda_l n}{2c^2 \log(2 + 2T\lambda_l)} \eta \right) \leq \\
&\leq \mathbf{E} \left[\sup_{t \in [0, T]} \exp \left\{ \frac{A_t^2}{4e^{2t\lambda_l} \log(2 + 2T\lambda_l)} \right\} \right] \exp \left\{ -\frac{\lambda_l n}{2c^2 \log(2 + 2T\lambda_l)} \eta \right\}.
\end{aligned} \tag{4.75}$$

We now use the fact that

$$\alpha := \sup_{t \in [0, T]} \frac{\langle A \rangle_t \log \log(\langle A \rangle_t \vee e^2)}{4e^{2t\lambda_l} \log(2 + 2T\lambda_l)} \leq \frac{1}{4}, \quad \text{for all } l \geq 1, \tag{4.76}$$

and, by Theorem 3.4, we obtain that there exists $C > 0$, independent of T, n and l , such that

$$\mathbf{P} \left(\sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2 > \eta \right) \leq C \exp \left\{ -\frac{\lambda_l n}{c^2 \log(2 + 2T\lambda_l)} \eta \right\}. \tag{4.77}$$

The case $l = 0$ is somehow easier since

$$z_t^n(f_0^\psi) = \frac{1}{n} \sum_{j=1}^n \int_0^t \left[\partial_\theta f_0^\psi \right] (\theta_s^{j,n}) dB_s^j \tag{4.78}$$

is a standard martingale with bounded quadratic variation: one can use Theorem 3.4 or, more simply, exponential estimates and Doob's inequality to obtain that there exists $c_0 > 0$ (depending on T) such that for all $\eta > 0$

$$\mathbf{P} \left(\sup_{t \in [0, T]} \left| z_t^n(f_0^\psi) \right|^2 > \eta \right) \leq c_0 \exp\{-c_0 n \eta\}. \tag{4.79}$$

The last part of the proof consists in exploiting the exponential inequalities (4.77) and (4.79), and to transfer them to $\sup_{t \in [0, T]} \|z_t^n\|_{-1}^2$. For this purpose, let $S > 0$ be defined by

$$S := \left(\sum_{l \geq 0} \frac{1}{(1+l)^{4/3}} \right)^{-1}. \tag{4.80}$$

For some $\eta > 0$, it holds that

$$\begin{aligned}
\mathbf{P} \left(\sup_{t \in [0, T]} \|z_t^n\|_{-1, 1/q_\psi} > \eta \right) &\leq \mathbf{P} \left(\sum_{l \geq 0} \sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2 > \eta^2 \right) \leq \\
&\leq \sum_{l \geq 0} \mathbf{P} \left(\sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2 > \frac{S}{(1+l)^{4/3}} \eta^2 \right) \leq \\
&\leq c_0 \exp\{-c_0 S n \eta^2\} + C \sum_{l \geq 1} \exp \left\{ -\frac{S}{c^2} \frac{\lambda_l}{\log(2 + 2T\lambda_l)(1+l)^{4/3}} n \eta^2 \right\}.
\end{aligned} \tag{4.81}$$

Now we use the fact that $\lambda_l = \Theta(l^2)$ as l tends to infinity, see Proposition 4.B.3. In particular, there exists $L > 0$, depending on T , such that

$$\begin{aligned} \mathbf{P} \left(\sup_{t \in [0, T]} \|z_t^n\|_{-1, 1/q_\psi} > \eta \right) &\leq c_0 \exp\{-c_0 S n \eta^2\} + \\ &+ CL \exp \left\{ -\frac{S}{c^2 \log(2 + 2T\lambda_L)} n \eta^2 \right\} + C \sum_{l > L} \exp \left\{ -\frac{S}{c^2} \sqrt{l} n \eta^2 \right\}. \end{aligned} \quad (4.82)$$

Observing that $\int_1^\infty e^{-\sqrt{x}n} dx = \frac{2(n+1)}{n^2} e^{-n}$, taking n large enough and Z an suitable constant depending on c_0, S, L and C , the proof of (ii) is concluded.

Back to (i), observe that for $s, t \in [0, T]$ and for some $k \geq 1$

$$\|z_t^n - z_s^n\|_{-1}^2 \leq \sum_{l=0}^k \left| z_t^n(f_l^\psi) - z_s^n(f_l^\psi) \right|^2 + 2 \sum_{l > k} \sup_{t \in [0, T]} \left| z_t^n(f_l^\psi) \right|^2. \quad (4.83)$$

The first term can be made small by using the continuity of $z_t^n(e_l)$; for the second one, observe that we have just proven that $\mathbf{E} \left[\sum_{l \geq 1} \sup_{t \in [0, T]} |z_t^n(e_l)|^2 \right] < \infty$. This implies that there exists a subsequence $\{k_m\}_{m \in \mathbb{N}}$ such that $\sum_{l > k_m} \sup_{t \in [0, T]} |z_t^n(e_l)|^2$ tends to 0 almost surely as m tends to infinity. The almost sure continuity in (4.83) is then established by choosing s and t close enough and k large enough. \square

3.3. Proof of Theorem 2.3. Recall that μ_0^n converges in H_{-1} to $q_\psi \in M$. Next lemma assures that the projection of μ_0^n on M is well defined for n big enough.

LEMMA 3.5. [78, Lemma 2.8] *There exists $\sigma > 0$ such that for all $\mu \in H_{-1}$ such that $\text{dist}(\mu, M) \leq \sigma$, there exists a unique phase $\psi := \text{proj}_M(\mu) \in \mathbb{T}$ such that $P_\psi^0(\mu - q_\psi) = 0$ and the mapping $\mu \mapsto \text{proj}_M(\mu)$ is \mathcal{C}^∞ .*

Let $\psi_n = \text{proj}_M(\mu_0^n)$. Fix $\varepsilon > 0$, we place ourselves in

$$A_1^n = \{\|\mu_0^n - q_{\psi_n}\|_{-1} \leq \varepsilon/2\}. \quad (4.84)$$

Denote by d_t^n the distance between μ_t^n and M , i.e.

$$d_t^n := \text{dist}(\mu_t^n, M). \quad (4.85)$$

We want to prove that μ_t^n stays close to M for long times. Let $T > 0$, $N \in \mathbb{N}$ and define

$$T_i = iT, \quad \text{for } i = 0, \dots, N. \quad (4.86)$$

If, for $N = N_n = o(\exp(n))$, we show that

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T_{N_n}]} d_t^n \leq \varepsilon \right) = 1, \quad (4.87)$$

then we are done. For sake of notation, we just employ N .

For $0 \leq a < b < \infty$, define the events

$$E^n(a, b) = \left\{ \max \left\{ 2d_a^n, 2d_b^n, \sup_{t \in (a, b)} d_t^n \right\} \leq \varepsilon \right\}, \quad (4.88)$$

clearly

$$\mathbf{P} \left(\sup_{t \in [0, T_N]} d_t^n \leq \varepsilon \right) \geq \mathbf{P} (E^n(0, T_N)). \quad (4.89)$$

The Markov property of system (4.2) implies that

$$\begin{aligned} \mathbf{P} (E^n(0, T_N)) &\geq \mathbf{P} (E^n(0, T_N) | E^n(0, T_{N-1})) \mathbf{P} (E^n(0, T_{N-1})) = \\ &= \mathbf{P} (E^n(0, T)) \mathbf{P} (E^n(0, T_{N-1})) \geq \mathbf{P} (E^n(0, T))^N. \end{aligned} \quad (4.90)$$

Let's then focus on the bounded interval of time $[0, T]$ and consider $\nu^n = \mu^n - q_{\psi^n}$, which satisfies the stochastic partial differential equation (4.31). Taking the norm in H_{-1} on both sides of (4.31) and using the properties of the semigroup together with the fact that (e.g. [10, Lemma 7.3])

$$\|\partial_\theta(\mu(J * \nu))\|_{-2} \leq C \|\mu\|_{-1} \|\nu\|_{-1}, \quad \text{for all } \mu, \nu \in H_{-1}, \quad (4.91)$$

one is left with (with a new constant C)

$$\|\nu_t^n\|_{-1} \leq \|e^{tL_{\psi^n}} \nu_0^n\|_{-1} + \int_0^t \frac{C}{\sqrt{t-s}} \|\nu_s^n\|_{-1}^2 ds + \|g_t^n\|_{-1} + \|z_t^n\|_{-1}. \quad (4.92)$$

By taking ε small enough, one can apply a Gronwall-type inequality (similar to Lemma 4.B.4) that leads to (recall (4.84) and the fact that the semigroup is continuous)

$$\sup_{t \in [0, T]} \|\nu_t^n\|_{-1} \leq \frac{2}{3}\varepsilon + \sup_{t \in [0, T]} \|g_t^n\|_{-1} + \sup_{t \in [0, T]} \|z_t^n\|_{-1}. \quad (4.93)$$

For $\eta > 0$, define $A_2^n(\eta) = \{\sup_{t \in [0, T]} \|z_t^n\|_{-1} \leq \eta\}$. If n is large enough, Lemma 3.2 assures that $\sup_{t \in [0, T]} \|g_t^n\|_{-1}$ is arbitrarily small a.s., and, placing ourselves in $A_2^n(\varepsilon/10)$, one obtains

$$\sup_{t \in [0, T]} \|\nu_t^n\|_{-1} \leq \frac{2}{3}\varepsilon + \frac{1}{5}\varepsilon \leq \varepsilon. \quad (4.94)$$

Plugging this estimate in (4.92) for $t = T$, observing that $P_{\psi^n}^0 \nu_0^n = 0$ by construction so that $\|e^{tL_{\psi^n}} \nu_0^n\|_{-1} \leq C e^{-\lambda_1 t/2} \|\nu_0^n\|_{-1}$ (e.g. [78, Proposition B.6]), one obtains

$$\|\nu_T^n\|_{-1} \leq e^{-\lambda_1 T/2} \frac{\varepsilon}{2} + \varepsilon \left(\varepsilon \int_0^T \frac{C}{\sqrt{T-s}} ds \right) + \frac{\varepsilon}{5}, \quad (4.95)$$

choosing T and ε such that

$$\begin{aligned} T &\geq \frac{2}{\lambda_1} \log(5), \\ \varepsilon &\leq \frac{1}{20C\sqrt{T}}, \end{aligned} \quad (4.96)$$

one finally gets

$$\|\nu_T^n\|_{-1} \leq \frac{\varepsilon}{2}. \quad (4.97)$$

Since $d_t^n \leq \|\nu_t^n\|_{-1}$ for all $t \geq 0$, we have then proven that $A_2^n(\varepsilon/10) \subset E^n(0, T)$. In particular,

$$\mathbf{P} \left(\sup_{t \in [0, T_N]} d_t^n \leq \varepsilon \right) \geq \mathbf{P} (E^n(0, T))^N \geq \left(1 - \mathbf{P} \left(A_2^n(\varepsilon/10)^c \right) \right)^N. \quad (4.98)$$

We can then use the estimate (4.67) in Lemma 3.3 to get

$$\mathbf{P} \left(A_2^n(\varepsilon/10)^c \right) = \mathbf{P} \left(\sup_{t \in [0, T]} \|z_t^n\|_{-1} > \frac{\varepsilon}{10} \right) \leq \exp \left\{ -\frac{Z}{100} n \varepsilon^2 \right\}. \quad (4.99)$$

Putting all together, one is left with

$$\begin{aligned} \mathbf{P} \left(\sup_{t \in [0, T_N]} d_t^n \leq \varepsilon \right) &\geq \left(1 - \mathbf{P} \left(A_2^n(\varepsilon/10)^c \right) \right)^N \geq \left(1 - \exp \left\{ -\frac{Z}{100} n \varepsilon^2 \right\} \right)^N \\ &= \exp \left\{ N \log \left[1 - \exp \left(-\frac{Z}{100} n \varepsilon^2 \right) \right] \right\} \geq \exp \left\{ -\frac{3}{2} N \exp \left(-\frac{Z}{100} n \varepsilon^2 \right) \right\}, \end{aligned} \quad (4.100)$$

where we have used that $\log(1 - x) \geq -3/2x$ for $0 \leq x \leq 1/2$. But the right hand side of (4.100) tends to 1 for all $N = N_n = o(\exp(n))$ and the proof is concluded.

4. Longtime behavior around $1/2\pi$

In this section we will suppose that the finite time behavior is already known, so that for n large enough, μ_t^n is very close to μ_t ; thus, for a large T_0 , μ_{T_0} will be very close to $1/2\pi$ and so will be for $\mu_{T_0}^n$. At the end of the day, we may suppose that we are starting close to $1/2\pi$. Since we are not assuming any independence between initial conditions and graph, instead of proving Theorem 2.4, we rather prove the following Proposition.

PROPOSITION 4.1. *If for every $\varepsilon_0 > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left\| \mu_0^n - \frac{1}{2\pi} \right\|_{-1} \leq \varepsilon_0 \right) = 1. \quad (4.101)$$

Then, there exists $A > 0$ such that for every positive increasing sequence $\{T_n\}_{n \in \mathbb{N}}$ such that $T_n = \exp(o(n))$ and for all $0 < \varepsilon < A$, it holds

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T_n]} \left\| \mu_t^n - \frac{1}{2\pi} \right\|_{-1} \leq \varepsilon \right) = 1. \quad (4.102)$$

The end of the section is thus devoted to prove Proposition 4.1.

4.1. A mild formulation around $1/2\pi$. We place ourselves around the stationary solution $\frac{1}{2\pi}$. The system evolution is captured by the linear dynamics around $\frac{1}{2\pi}$ and the corresponding linear operator $L_{2\pi}$ is given by

$$L_{2\pi} u := \frac{1}{2} \partial_\theta^2 u - \frac{1}{2\pi} (\partial_\theta J) * u, \quad \text{for } u \in \mathcal{C}^2(\mathbb{T}), \quad \int_{\mathbb{T}} u(\theta) d\theta = 0. \quad (4.103)$$

The adjoint $L_{2\pi}^*$ of $L_{2\pi}$ in \mathcal{L}_0^2 has the following expression

$$L_{2\pi}^* u = \frac{1}{2} \partial_\theta^2 u - \frac{1}{2\pi} J * (\partial_\theta u), \quad (4.104)$$

and domain $D(L_{2\pi}^*) = D(L_{2\pi})$. These operators are diagonal in the Fourier basis $\{e_l\}_{l \geq 1}$, with eigenvalues denoted by $\{\lambda_l^{2\pi}\}_{l \geq 1}$. The spectrum is negative and bounded away from 0, let $\gamma_K = \lambda_1^{2\pi} = \frac{1-K}{2} > 0$ denote the spectral gap. The operator $L_{2\pi}$ (resp. $L_{2\pi}^*$) defines an analytic semigroup $e^{tL_{2\pi}}$ (resp. $e^{tL_{2\pi}^*}$) with the following property:

$$\|e^{tL_{2\pi}} u\|_{-1} \leq C \frac{e^{-\gamma t/2}}{\sqrt{t}} \|u\|_{-2}, \quad \text{for some } C > 0, \quad (4.105)$$

for all $\gamma \in [0, \gamma_K)$, all $t > 0$ and $u \in H_{-1}$. We will not prove (4.105) but refer to Appendix 4.B for similar estimates.

Define $\nu_t^n := \mu_t^n - \frac{1}{2\pi}$. As done in Section 3, we derive a mild formulation for ν^n . We omit the proof.

PROPOSITION 4.2. *The process $\nu_t^n \in H_{-1}$ satisfies the following stochastic partial differential equation in $C([0, T], H_{-1})$:*

$$\nu_t^n = e^{tL_{2\pi}} \nu_0^n - \int_0^t e^{(t-s)L_{2\pi}} \partial_\theta [\nu_s^n (J * \nu_s^n)] ds - g_t^n + z_t^n, \quad (4.106)$$

where

$$g_t^n = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_{2\pi}} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right] ds, \quad (4.107)$$

and $z_t^n \in H_{-1}$ is defined for $h \in H_1$ by

$$\langle z_t^n, h \rangle_{-1,1} = \frac{1}{n} \sum_{j=1}^n \int_0^t \partial_\theta e^{(t-s)L_{2\pi}^*} h(\theta_s^{j,n}) dB_s^j. \quad (4.108)$$

4.2. Control on the perturbations. Contrary to the supercritical case, the operator $L_{2\pi}$ is contracting along all direction or, in other words, all its eigenvalues are negative. This property gives a stronger control on g^n and z^n , as shown in the next Lemmas.

LEMMA 4.3 (Wellposedness and bounds on g_t^n). *For $n \in \mathbb{N}$ and $t \geq 0$, let g_t^n be given by*

$$g_t^n = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_{2\pi}} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right] ds. \quad (4.109)$$

Then

(1) $g^n \in C^0([0, \infty), H_{-1})$. In particular, for every $h \in H_1$ and $t \geq 0$

$$g_t^n(h) = -\frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) J(\theta_s^{i,n} - \theta_s^{j,n}) \partial_\theta e^{(t-s)L_{2\pi}^*} h(\theta_s^{i,n}) ds. \quad (4.110)$$

(2) There exists $D > 0$, independent of t , such that

$$\|g_t^n\|_{-1} \leq D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}, \quad \text{for all } t \geq 0. \quad (4.111)$$

PROOF. We only prove (ii). Observe that, as in (4.61),

$$\begin{aligned} & \left\langle \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)L_{2\pi}} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right], h \right\rangle_{-1,1} = \\ & = -\frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) \langle \delta_{\theta_s^{i,n}}, (J * \delta_{\theta_s^{j,n}}) \partial_\theta e^{(t-s)L_{2\pi}^*} h \rangle_{-1,1}. \end{aligned} \quad (4.112)$$

Applying Theorem 1.2, this time with

$$\begin{aligned} a_{ij} &= \left(\frac{\xi_{ij}}{p_n} - 1 \right), \\ S_i &= \delta_{\theta_s^{i,n}}, \\ T_j &= \frac{\sqrt{t-s}}{C e^{-\gamma(t-s)}} \left(J * \delta_{\theta_s^{j,n}} \right) \partial_\theta e^{(t-s)L_{2\pi}^*} \frac{h}{\|h\|_1}, \end{aligned} \quad (4.113)$$

allows us to bound the expression in (4.112) by

$$K_R \frac{C e^{-\gamma(t-s)}}{\sqrt{t-s}} \|h\|_1 \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}. \quad (4.114)$$

This shows that

$$\|g_t^n\|_{-1} \leq K_R C \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2} \int_0^t \frac{e^{-\gamma(t-s)}}{\sqrt{t-s}} ds \leq D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}, \quad (4.115)$$

where $D := K_R C \int_0^\infty \frac{e^{-\gamma s}}{\sqrt{s}} ds > 0$ since the integral converges. The proof is concluded. \square

We now turn to the stochastic term z_t^n in (4.106). Recall that $L_{2\pi}$ is diagonal in the Fourier basis $\{e_l\}_{l \geq 1}$ of H_{-1} , with eigenvalues denoted by $\lambda_l^{2\pi}$. Then

LEMMA 4.4 (Wellposedness and bounds on z_t^n). For $n \in \mathbb{N}$ and $t > 0$, let z_t^n be defined by

$$z_t^n = \sum_{l \geq 1} \langle z_t^n, e_l \rangle_{H_{-1}} e_l, \quad (4.116)$$

where

$$\langle z_t^n, e_l \rangle_{H_{-1}} = z_t^n \left(\frac{e^{il \cdot}}{l} \right) = \frac{i}{n} \sum_{j=1}^n \int_0^t e^{(t-s)\lambda_l^{2\pi}} e^{il\theta_s^{j,n}} dB_s^j. \quad (4.117)$$

Then

- (1) $z_t^n \in C^0([0, \infty), H_{-1})$ almost surely.
- (2) There exists $C > 0$ independent of n , such that for all $T > 0$

$$\mathbf{E} \left[\sup_{t \in [0, T]} \|z_t^n\|_{-1}^2 \right] \leq C \frac{\log(1 + 2\gamma_K T)}{n}. \quad (4.118)$$

(3) For every positive increasing sequence $\{T_n\}_{n \in \mathbb{N}}$ such that $T_n = \exp(o(n))$ and for all $\eta > 0$, it holds

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\sup_{t \in [0, T_n]} \|z_t^n\|_{-1} \leq \eta \right) = 1. \quad (4.119)$$

PROOF. We only prove (ii). For $l \geq 1$, let $x_t^l := \sqrt{2\lambda_l^{2\pi} n} e^{\lambda_l^{2\pi} t} |z_t^n(e^{i\cdot})|$. In particular

$$x_t^l = \left| \frac{\sqrt{2\lambda_l^{2\pi}}}{\sqrt{n}} \sum_{j=1}^n \int_0^t e^{s\lambda_l^{2\pi}} e^{i\theta_s^{j,n}} dB_s^j \right| = |a_t^l + i b_t^l|, \quad (4.120)$$

where a^l and b^l are two continuous real valued martingales. Let $\langle x^l \rangle_t = \langle a^l \rangle_t + \langle b^l \rangle_t$ where $\langle a^l \rangle_t$ and $\langle b^l \rangle_t$ are the quadratic variations of a_t^l and b_t^l respectively, then

$$\langle x^l \rangle_t = \frac{2\lambda_l^{2\pi}}{n} \sum_{j=1}^n \int_0^t e^{2s\lambda_l^{2\pi}} (\cos^2 + \sin^2)(l\theta_s^{j,n}) ds = e^{2\lambda_l^{2\pi} t} - 1. \quad (4.121)$$

We now use

LEMMA 4.5. *Let $Y_t = A_t + i B_t$, where A_t and B_t are continuous real valued martingales. Define $X_t = |Y_t|$ and $\langle X \rangle_t = \langle A \rangle_t + \langle B \rangle_t$, where $\langle A \rangle_t$ and $\langle B \rangle_t$ are the quadratic variations of A and B respectively. Then, there exists $C > 0$ such that, for all $T > 0$,*

$$\mathbf{E} \left[\sup_{t \in [0, T]} \frac{X_t^2}{1 + \langle X \rangle_t} \right] \leq C \log(1 + \log(1 + \langle X \rangle_t)). \quad (4.122)$$

The proof of Lemma 4.5 is presented at the end of the section. By choosing $X_t = x_t^l$, $A_t = a_t^l$ and $B_t = b_t^l$, one obtains that, for $T > 0$,

$$\mathbf{E} \left[\sup_{t \in [0, T]} |z_t^n(e_l)|^2 \right] = \frac{1}{2\lambda_l^{2\pi} n} \mathbf{E} \left[\sup_{t \in [0, T]} \frac{(x_t^l)^2}{1 + \langle x^l \rangle_t} \right] \leq \frac{C}{2\lambda_l^{2\pi} n} \log(1 + 2\lambda_l^{2\pi} T). \quad (4.123)$$

It remains to observe that

$$\mathbf{E} \left[\sup_{t \in [0, T]} \|z_t^n\|_{-1}^2 \right] \leq \mathbf{E} \left[\sum_{l \geq 1} \sup_{t \in [0, T]} |z_t^n(e_l)|^2 \right] \leq C \sum_{l \geq 1} \frac{1}{2\lambda_l^{2\pi} n} \log(1 + 2\lambda_l^{2\pi} T). \quad (4.124)$$

The conclusion holds by factorizing the first term of the sum and modifying the constant C accordingly: observe that $\sum_{l \geq 1} \sup_{T \geq 1} \frac{\log(1 + 2\lambda_l^{2\pi} T)}{\lambda_l^{2\pi} \log(1 + 2\lambda_l^{2\pi} T)} < \infty$.

The proof is concluded modulo Lemma 4.5, proven hereafter. \square

PROOF OF LEMMA 4.5. Recall that A_t is a martingale, in particular a slight variation of [52, Corollary 2.8] implies that there exists $D > 0$ such that

$$\mathbf{E} \left[\sup_{t \in [0, T]} \frac{A_t^2}{1 + \langle A \rangle_t} \right] \leq D \log(1 + \log(1 + \langle A \rangle_t)). \quad (4.125)$$

Thus, one can develop

$$\begin{aligned} \mathbf{E} \left[\sup_{t \in [0, T]} \frac{X_t^2}{1 + \langle X \rangle_t} \right] &\leq \mathbf{E} \left[\sup_{t \in [0, T]} \frac{A_t^2}{1 + \langle A \rangle_t} \right] + \mathbf{E} \left[\sup_{t \in [0, T]} \frac{B_t^2}{1 + \langle B \rangle_t} \right] \leq \\ &\leq D \log(1 + \log(1 + \langle A \rangle_t)) + D \log(1 + \log(1 + \langle B \rangle_t)) \leq \\ &\leq 2D \log(1 + \log(1 + \langle X \rangle_t)), \end{aligned} \quad (4.126)$$

and the proof is done by taking $C = 2D$. \square

4.3. Proof of Proposition 4.1. Fix $\varepsilon > 0$. From Proposition 4.2 we know that $\nu_t^n := \mu_t^n - \frac{1}{2\pi}$ satisfies

$$\nu_t^n = e^{tL_{2\pi}} \nu_0^n - \int_0^t e^{(t-s)L_{2\pi}} \partial_\theta [\nu_s^n (J * \nu_t^n)] ds - g_t^n + z_t^n. \quad (4.127)$$

Taking the norm and using the properties of $e^{tL_{2\pi}}$, together with the estimate (4.91), for all $0 < \gamma < \gamma_K$ one obtains

$$\|\nu_t^n\|_{-1} \leq \|\nu_0^n\|_{-1} + C \int_0^t \frac{e^{-\gamma(t-s)}}{\sqrt{t-s}} \|\nu_s^n\|_{-1}^2 ds + \|g_t^n\|_{-1} + \|z_t^n\|_{-1}. \quad (4.128)$$

Thanks to the contractive properties of $L_{2\pi}$, there exists $D > 0$ (Lemma 4.3) such that

$$\sup_{t \geq 0} \|g_t^n\|_{-1} < D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}. \quad (4.129)$$

Define now $B_1^n(\varepsilon_0) = \{\|\nu_0^n\| \leq \varepsilon_0\}$ and $B_2^n(\eta) = \{\sup_{t \in [0, T_n]} \|z_t^n\|_{-1} \leq \eta\}$. On $B_1^n(\varepsilon/3) \cap B_2^n(\varepsilon/4)$ and for n large enough, we can apply Lemma 4.B.4 with

$$\begin{aligned} \delta &= \frac{\varepsilon}{3}, \quad T = T_n, \\ f(t) &= \|\nu_t^n\|_{-1}, \\ g(t) &= \|g_t^n\|_{-1} + \|z_t^n\|_{-1}, \end{aligned} \quad (4.130)$$

and obtain

$$\sup_{t \in [0, T_n]} \|\nu_t^n\|_{-1} \leq \varepsilon. \quad (4.131)$$

The proof is concluded since by hypothesis $\mathbf{P}(B_1^n) \rightarrow 1$ and Lemma 4.4 implies that $\mathbf{P}(B_2^n) \rightarrow 1$ as n tends to infinity.

5. Finite time behavior

The aim of this section is to study the closeness of μ^n to μ . on bounded time interval.

PROOF OF THEOREM 2.1. Fix $\varepsilon > 0$ and $T > 0$. It is not difficult to see that $\mu^n - \mu$ satisfies again a mild equation in $\mathcal{C}^0([0, T], H_{-1})$, which is given by

$$\mu_t^n - \mu_t = e^{t\frac{\Delta}{2}} (\mu_0^n - \mu_0) - \int_0^t e^{(t-s)\frac{\Delta}{2}} \partial_\theta [\mu_s^n (J * \mu_s^n) - \mu_s (J * \mu_s)] ds - g_t^n + z_t^n, \quad (4.132)$$

where

$$g_t^n = \frac{1}{n^2} \sum_{i,j=1}^n \int_0^t \left(\frac{\xi_{ij}}{p_n} - 1 \right) e^{(t-s)\frac{\Delta}{2}} \partial_\theta \left[\delta_{\theta_s^{i,n}} (J * \delta_{\theta_s^{j,n}}) \right] ds, \quad (4.133)$$

and z_t^n is denoted for $h \in H_1$ by

$$z_t^n(h) = \frac{1}{n} \sum_{j=1}^n \int_0^t \partial_\theta e^{(t-s)\frac{\Delta}{2}} h(\theta_s^{j,n}) dB_s^j. \quad (4.134)$$

Observe that we are using the Laplacian operator which is very similar to $L_{2\pi}$ except for the first eigenvalue that is now given by $-(1-K)/2$. We will thus use all the results about $L_{2\pi}$ and its semigroup to control z_t^n and g_t^n .

Taking the H_{-1} norm in (4.132) and applying (4.91), one is left with

$$\|\mu_t^n - \mu_t\|_{-1} \leq \|\mu_0^n - \mu_0\|_{-1} + \int_0^t \frac{C}{\sqrt{t-s}} \|\mu_s^n - \mu_s\|_{-1} ds + \|g_t^n\|_{-1} + \|z_t^n\|_{-1}. \quad (4.135)$$

The term involving the graph g_t^n can be controlled again by $\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}$: minor modifications to Lemma 4.3 show that there exists $D > 0$ such that

$$\sup_{t \in [0, T]} \|g_t^n\|_{-1} \leq D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2}. \quad (4.136)$$

For the initial conditions and the stochastic part z_t^n , define the two sets:

$$\begin{aligned} C_1^n &= C_1^n(\varepsilon_0) = \{ \|\mu_0^n - \mu_0\|_{-1} \leq \varepsilon_0 \}; \\ C_2^n &= C_2^n(T, \eta) = \left\{ \sup_{t \in [0, T]} \|z_t^n\|_{-1} \leq \eta \right\}. \end{aligned} \quad (4.137)$$

On $C_1^n \cap C_2^n$, one obtains

$$\|\mu_t^n - \mu_t\|_{-1} \leq \varepsilon_0 + \int_0^t \frac{C}{\sqrt{t-s}} \|\mu_s^n - \mu_s\|_{-1} ds + D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2} + \eta. \quad (4.138)$$

Gronwall-Henry's inequality ([58, Lemma 7.1.1 and Exercice 1]) leads to

$$\sup_{t \in [0, T]} \|\mu_t^n - \mu_t\|_{-1} \leq 2 \left(\varepsilon_0 + D \frac{\|P^{(n)} - \mathbf{1}^{(n)}\|_{\infty \rightarrow 1}}{n^2} + \eta \right) e^{aT}, \quad (4.139)$$

where a is independent of n , ε_0 and η . Considering ε_0 and η small enough and n large enough, the proof is concluded modulo showing that

$$\lim_{n \rightarrow \infty} \mathbf{P}(C_1^n \cap C_2^n) = 1. \quad (4.140)$$

From the hypothesis on the initial condition (4.19), it is clear that for all ε_0 one has $\mathbf{P}(C_1^n(\varepsilon_0)) \rightarrow 1$ as n tends to infinity. The same conclusion holds for C_2^n by slightly modifying the proof of Lemma 4.4. The proof is concluded. \square

4.A. Graphs

4.A.1. General properties of the graphs under consideration. We observe that condition (4.16) implies a weak form of degree homogeneity (recall (4.10)):

LEMMA 4.A.1. *Suppose that (4.16) holds. Let $\delta > 0$, define*

$$I_n^\delta := \left\{ i \in \{1, \dots, n\} : \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{j=1}^n \frac{\xi_{ij}^{(n)}}{p_n} - 1 \right| \geq \delta \right\}. \quad (4.141)$$

Then $|I_n^\delta| = o(n)$.

PROOF. Suppose that $\lim_{n \rightarrow \infty} \frac{|I_n|}{n} = c$ for some $c > 0$. Then

$$\begin{aligned} \sup_{s_i, t_j \in \{\pm 1\}} \frac{1}{n^2} \sum_{i,j=1}^n \left(\frac{\xi_{ij}^{(n)}}{p_n} - 1 \right) s_i t_j &\geq \sup_{s_i \in \{\pm 1\}} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{ij}^{(n)}}{p_n} - 1 \right) \right] s_i \geq \\ &\geq \frac{1}{n} \sum_{i \in I_n} \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{ij}^{(n)}}{p_n} - 1 \right) \right| \geq \frac{|I_n|}{n} \inf_{i \in I_n} \left| \frac{1}{n} \sum_{j=1}^n \left(\frac{\xi_{ij}^{(n)}}{p_n} - 1 \right) \right|. \end{aligned} \quad (4.142)$$

This last term does not go to zero as n tends to infinity, against (4.16). \square

It also implies the existence of an unique giant component.

LEMMA 4.A.2. *Suppose that (4.16) holds. Then, there exists a unique sequence of connected components $\{\mathcal{C}^{(n)}\}$ in $\{\xi^{(n)}\}$ and $\lim_{n \rightarrow \infty} |\mathcal{C}^{(n)}|/n = 1$.*

PROOF. We prove the uniqueness first. Suppose that for every n there exist $\mathcal{C}_1^{(n)}$ and $\mathcal{C}_2^{(n)}$ distinct connected components of $\xi^{(n)}$ such that $|\mathcal{C}_i^{(n)}| = n_i = \Theta(n)$ for $i = 1, 2$. Without loss of generality, one can suppose $\mathcal{C}_1^{(n)}$ consisting in the first n_1 vertices of $\xi^{(n)}$ and $\mathcal{C}_2^{(n)}$ in the following n_2 .

Using the equivalence of $l_\infty \rightarrow l_1$ norm with the cut-norm (e.g. [3]), one obtains

$$\|P_n - \mathbf{1}_n\|_{\infty \rightarrow 1} \geq \sup_{x_i, y_j \in \{0,1\}} \left| \sum_{i,j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) x_i y_j \right| \geq \sum_{\substack{1 \leq i \leq n_1 \\ n_1 \leq j \leq n_2 - n_1}} 1 = n_1 n_2 = \Theta(n^2). \quad (4.143)$$

For the existence, suppose the connected components of $\xi^{(n)}$ are ordered from the biggest one in size (the first n_1 vertices) to the smallest one (the last vertices). Take the first m components such that $|\mathcal{C}_1 \cup \dots \cup \mathcal{C}_m| \geq n/4$. One easily sees that $|\mathcal{C}_1 \cup \dots \cup \mathcal{C}_m| \leq n/2$. Applying the same reasoning of before with $1 \leq i \leq n/4$ and $n/2 \leq j \leq n$, the proof is concluded. \square

4.A.2. Examples of graph sequences. We exhibit two classes of graphs, a random and a deterministic one, that satisfy assumption (4.16). The only hypothesis required on p_n is equivalent to asking that the mean degree per site diverges as n tends to infinity, i.e. $np_n \uparrow \infty$.

4.A.2.1. *Erdős-Rényi random graphs.* As mentioned in the introduction, $\|\cdot\|_{\infty \rightarrow 1}$ has been found very useful for random graph concentration and this is indeed the case of ER graphs (e.g. [55]). We recall the definition and give the result.

For every $n \in \mathbb{N}$, let $\{\xi_{ij}^{(n)}\}_{1 \leq i \neq j \leq n}$ be IID Bernoulli random variables with parameter p_n , \mathbb{P} denoting the associated probability. For every i , $\xi_{ii}^{(n)}$ is set equal to 0, i.e. self loop are not admitted.

LEMMA 4.A.3. *Assume that*

$$\lim_{n \rightarrow \infty} np_n = \infty. \quad (4.144)$$

There exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P} \left(\sup_{s_i, t_j} \frac{1}{n^2} \sum_{i, j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) s_i t_j \geq \frac{2}{\sqrt{np_n}} \right) \leq e^{-2n}, \quad \text{for all } n \geq n_0. \quad (4.145)$$

PROOF. The proof is just an union bound and an application of Bernstein's inequality. Indeed,

$$\mathbb{P} \left(\sup_{s_i, t_j} \frac{1}{n^2} \sum_{i, j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) s_i t_j \geq \frac{\delta}{\sqrt{np_n}} \right) \leq \sum_{s_i, t_j} \mathbb{P} \left(\frac{1}{n^2} \sum_{i, j=1}^n \left(\frac{\xi_{ij}}{p_n} - 1 \right) s_i t_j \geq \frac{\delta}{\sqrt{np_n}} \right). \quad (4.146)$$

Bernstein's inequality ([19, Corollary 2.11]) says that if X_1, \dots, X_n are independent zero-mean random variables such that $|X_j| \leq M$ a.s. for all j , then for all $t \geq 0$

$$\mathbb{P} \left(\sum_{j=1}^n X_j > t \right) \leq \exp \left\{ - \frac{t^2}{2 \sum_{j=1}^n \mathbb{E}[X_j^2] + \frac{2}{3} Mt} \right\}.$$

Let $X_{k(i,j)} = \frac{s_i t_j}{n^2 p_n} (\xi_{ij} - p_n)$ with k some bijection from $\{1, \dots, n\}^2$ to $\{1, \dots, n^2\}$. Then $|X_k| \leq \frac{1}{n^2 p_n}$ and $\mathbb{E}[X_k^2] \leq \frac{2}{n^4}$. For n large enough, we thus obtain

$$\mathbb{P} \left(\sum_{k=1}^{n^2} X_k \geq \frac{\delta}{\sqrt{np_n}} \right) \leq \exp \left\{ - \frac{n\delta^2}{4p_n + \frac{2}{3} \frac{\delta}{\sqrt{np_n}}} \right\} \leq \exp \{-n\delta^2\}. \quad (4.147)$$

The proof is concluded observing that the sum in (4.146) consists in 4^n elements and choosing $\delta = 2$. \square

We thus have

PROPOSITION 4.A.4. *Given (4.144), ER graphs satisfy condition (4.16) \mathbb{P} -almost surely.*

PROOF. It suffices to apply Borel-Cantelli lemma to (4.145). \square

Similarly one can prove that symmetric ER random graphs satisfy (4.16) a.s..

Ramanujan graphs. Let $d = 2, 3, \dots$, consider a d -regular graph, i.e. graph where each vertex has exactly d neighbors. We start recalling a well-known result

LEMMA 4.A.5 (Expander mixing lemma). *Let G be a d -regular random graph (G denoting the adjacency matrix itself), it holds*

$$\frac{1}{n^2} \left\| \frac{n}{d} G - \mathbf{1}^{(n)} \right\|_{\infty \rightarrow 1} \leq 4 \frac{\lambda(d)}{d}, \quad (4.148)$$

where $\lambda(d)$ is the second biggest eigenvalue (in absolute value) associated to G .

PROOF. The proof is classical but it is in general formulated in terms of the cut-norm (e.g. [59]). One easily sees that the cut-norm is equivalent (paying a factor 4, e.g. [3]) to the $l_\infty \rightarrow l_1$ norm. \square

Ramanujan graphs are d -regular graphs such that $\lambda(d) \leq 2\sqrt{d-1}$, they are very well known for their expander properties (e.g. [59]). Condition (4.16) holds whenever d_n diverges; indeed

PROPOSITION 4.A.6. *Let $d_n = np_n$. Suppose that (4.144) holds, i.e.*

$$\lim_{n \rightarrow \infty} d_n = \infty. \quad (4.149)$$

Then, every sequence of Ramanujan graphs satisfies condition (4.16).

PROOF. Rewriting (4.148) in terms of p_n , it becomes

$$\frac{1}{n^2} \left\| \frac{G}{p_n} - \mathbf{1}^{(n)} \right\|_{\infty \rightarrow 1} \leq \frac{8}{\sqrt{np_n}}. \quad (4.150)$$

The proof is concluded taking the limit for n which tends to infinity. \square

4.A.3. Links with graphons. The norm $\|\cdot\|_{\infty \rightarrow 1}$ is strictly related to the canonical distance $d_{\mathcal{W}}$ on the space of (sparse) graphons \mathcal{W} (e.g. [17]). In fact, whenever $\xi^{(n)}/p_n$ is (a realization of) a graphon $W^{(n)}$, condition (4.16) is implied by the convergence of $W^{(n)}$ to the constant graphon $W \equiv 1$ in \mathcal{W} . One can then consider system (4.2) on a sequence of (sparse) graphons and require, instead of condition (4.16), the convergence in \mathcal{W} to the constant graphon.

We have decided not to add another level of complexity in order to keep the results as clear as possible, but everything could be reformulated within this more general framework and the proofs would basically not change.

4.B. H_{-1} and Semigroups

4.B.1. On the relationship between H_{-1} and $\mathcal{P}(\mathbb{T})$. Consider $H_1 := H_{1,1}$, its dual space, denoted by H_{-1} , can be described through the Fourier orthonormal basis $\{e_l\}_{l \geq 1}$, where $e_l(\theta) = le^{il\theta}$. With this characterization one easily obtains that $\mathcal{P}(\mathbb{T}) - \frac{1}{2\pi} \subset H_{-1}$. Indeed, for $\mu \in \mathcal{P}(\mathbb{T})$,

$$\left\| \mu - \frac{1}{2\pi} \right\|_{-1} = \sqrt{\sum_{l \geq 1} |\langle \mu, le^{il\cdot} \rangle_{H_{-1}}|^2} = \sqrt{\sum_{l \geq 1} \frac{1}{l^2} |\langle \mu, e^{il\cdot} \rangle|^2} \leq \sqrt{\sum_{l \geq 1} \frac{1}{l^2}} < \infty. \quad (4.151)$$

In particular, the difference between two probability measures belongs to H_{-1} .

Observe now that H_{-1} induces a distance on $\mathcal{P}(\mathbb{T})$ which controls the bounded-Lipschitz distance d_{bL} , i.e. for all $\mu, \nu \in \mathcal{P}(\mathbb{T})$

$$\begin{aligned} d_{\text{bL}}(\mu, \nu) &= \sup_{\|f\|_{\text{bL}}=1} \int f(d\mu - d\nu) \leq \sup_{h \in \mathcal{C}_0^1, \|h\|_1=1} \int h(d\mu - d\nu) = \\ &= \sup_{h \in \mathcal{C}_0^1, \|h\|_1=1} \int h'(\mathcal{U} - \mathcal{V}) = \sup_{\|h\|_1=1} \langle \mu - \nu, h \rangle_{-1,1} = \\ &= \|\mu - \nu\|_{-1}. \end{aligned} \quad (4.152)$$

Where we have used the density of \mathcal{C}_0^1 in H_1 , and denoted by \mathcal{U} and \mathcal{V} the primitives of μ and ν respectively.

4.B.2. On the weighted Hilbert space $H_{-1,\omega}$. Recall that, one has this sequence of continuous and dense inclusions:

$$H_{1,1/\omega} \subset \mathcal{L}_0^2 = \mathcal{L}_0^{2*} \subset H_1^* =: H_{-1,\omega}, \quad (4.153)$$

where we have chosen the canonical identification for \mathcal{L}_0^2 . We can explicit the isometry between $H_{1,1/\omega}$ to $H_{-1,\omega}$. Consider the operator

$$\begin{aligned} A_\omega : \mathcal{C}^\infty(\mathbb{T}) &\rightarrow \mathcal{C}^\infty(\mathbb{T}) \\ f &\mapsto -\partial_\theta(\omega^{-1} \partial_\theta f) \end{aligned} \quad (4.154)$$

It is known [22, pag. 82] that $A_\omega(H_{1,1/\omega})$ is dense in $H_{-1,\omega}$ and the injection is continuous. This allows considering $H_{1,1/\omega}$ as a subset of $H_{-1,\omega}$ by identifying u and $A_\omega u$.

The inner product in $H_{-1,\omega}$, dual to the one in $H_{1,1/\omega}$, is given by

$$\langle u, v \rangle_{H_{-1,\omega}} = \int w \mathcal{U} \mathcal{V}, \quad (4.155)$$

where \mathcal{U} and \mathcal{V} are primitive of u and v respectively, such that $\int w \mathcal{U} = 0 = \int w \mathcal{V}$ (e.g. [9, Subsection 2.2]). Then, for $f, g \in \mathcal{C}^\infty$, it holds

$$\langle A_\omega f, A_\omega g \rangle_{-1,\omega} = \int \omega^{-1} f' g' = \langle f, g \rangle_{1,1/\omega}. \quad (4.156)$$

4.B.3. The linear operators L_ψ and L_ψ^* and their semigroups. This subsection recalls the known results on L_ψ , its dual L_ψ^* and the associated semigroups e^{tL_ψ} and $e^{tL_\psi^*}$.

We start with the spectral properties of L_ψ .

PROPOSITION 4.B.1. *The operator L_ψ (resp. L_ψ^*) is essentially self-adjoint with compact resolvent in $H_{-1,1/q}$ (resp. $H_{1,q}$). Its spectrum is pure point and lies in $(-\infty, -\lambda_1] \cup \{0\}$, where $\lambda_1 > 0$ and 0 is a simple eigenvalue of L_ψ with eigenvector $\partial_\theta q_\psi$.*

Moreover, both $L_{2\pi}$ and $L_{2\pi}^$ generate a \mathcal{C}^0 semigroup $t \mapsto e^{tL_\psi}$ (resp. $t \mapsto e^{tL_\psi^*}$) in \mathcal{L}_0^2 and $e^{tL_\psi^*} = (e^{tL_\psi})^*$.*

PROOF. The result about L_ψ is given in [9]. Observe that, due to the isometry (4.154) between $H_{-1,1/q_\psi}$ and H_{1,q_ψ} , $L_\psi^* = A_{1/q_\psi}^{-1} L_\psi A_{1/q_\psi}$ and it has thus the same spectral properties of L_ψ .

From the spectral properties of L_ψ and L_ψ^* , one deduces that the two operators are sectorial (and with dense domain in H_{-1}), standard techniques assure the existence of the analytic semigroup (e.g. [58]). \square

An accurate analysis of the semigroup has already been established in [10] by means of interpolating norms and Fourier decomposition. We recall here the most important properties. We will use the space H_{-2} , defined in an analogous way of H_{-1} .

PROPOSITION 4.B.2 ([10, Lemma 7.2]). *For all $t > 0$, the operator e^{tL_ψ} extends to a bounded operator from H_{-2} to H_{-1} and there exists $C > 0$ such that for all $u \in H_{-2}$*

$$\|e^{tL_\psi} u\|_{-1} \leq C \left(1 + \frac{1}{\sqrt{t}}\right) \|u\|_{-2}. \quad (4.157)$$

Moreover, for all $\epsilon \in (0, 1/2)$, $\delta \geq 0$ and all $u \in H_{-1}$

$$\|e^{(t+\delta)L_\psi} u - e^{tL_\psi} u\|_{-1} \leq C\delta^\epsilon \left(1 + \frac{1}{t^{1/2+\epsilon}}\right) \|u\|_{-2}. \quad (4.158)$$

By duality, observe that for all $h \in H_1$

$$\|e^{tL_\psi^*} h\|_2 \leq C \left(1 + \frac{1}{\sqrt{t}}\right) \|h\|_1. \quad (4.159)$$

We end this subsection with an useful result on the eigenvalues and eigenfunctions associated to L_ψ , recall (4.29).

PROPOSITION 4.B.3. *There exists $C > 1$ such that for all $l \in \mathbb{N}$*

$$\frac{l^2}{C} \leq \lambda_l \leq Cl^2. \quad (4.160)$$

Let $f_l^\psi = A_{1/q_\psi}^{-1} e_l^\psi$, then f_l^ψ is an eigenfunction of L_ψ^* associated to $-\lambda_l$ and

$$\sup_{l \in \mathbb{N}} \left\| \partial_\theta f_l^\psi \right\|_\infty < \infty. \quad (4.161)$$

PROOF. The first part is covered in [10, Remark 8.3] and the second one in [10, Corollary 8.6]. \square

4.B.4. Analytical estimate. A variation on Gronwall Lemma.

LEMMA 4.B.4. *Let $T > 0$, $\gamma \geq 0$. Let $f : [0, T] \rightarrow [0, \infty)$ be a continuous function and $g : [0, T] \rightarrow [0, \infty)$ be such that for all $0 \leq t \leq T$*

$$f(t) \leq f(0) + \int_0^t \frac{e^{-\gamma(t-s)}}{\sqrt{t-s}} f^2(s) ds + g(t). \quad (4.162)$$

There exists $A > 0$, which depends on T only if $\gamma = 0$, such that for all $0 < \delta < A$ and if $f(0) < \delta$, $\sup_{t \in [0, T]} g(t) < \delta$, then

$$\sup_{t \in [0, T]} f(t) \leq 3\delta. \quad (4.163)$$

PROOF. Consider the set $O = \{t : f(t) \leq 3\delta\} \subset [0, T]$. Since f is continuous and $f(0) \leq \delta$, O is a non-empty open set in $[0, T]$. Suppose that $\sup(O) = u < T$; we show that $u \in O$, which implies $O = [0, T]$.

Consider

$$\begin{aligned} f(u) &= f(0) + \int_0^u \frac{e^{-\gamma(u-s)}}{\sqrt{u-s}} f^2(s) ds + g(u) \leq \\ &\leq 2\delta + \delta \left(9\delta \int_0^u \frac{e^{-\gamma(u-s)}}{\sqrt{u-s}} ds \right) \leq 3\delta, \end{aligned} \quad (4.164)$$

where the last inequality holds for all $\delta \leq A := \left(9 \int_0^\infty \frac{e^{-\gamma s}}{\sqrt{s}} ds \right)^{-1}$ whenever $\gamma > 0$ or for all $\delta \leq \frac{1}{18\sqrt{T}}$ in case $\gamma = 0$. Thus $u \in O$ and the proof is concluded. \square

A Law of Large Numbers via a mild formulation

This chapter is based on the joint work with Florian Bechtold [8].

1. Introduction

The theory of weakly interacting particle systems has received great attention in the last fifty years. On the one hand, its mathematical tractability has allowed to obtain a deep understanding of the behavior of the empirical measure for such systems: law of large numbers [88, 29], fluctuations and central limit theorems [99, 45], large deviations [44, 56] and propagation of chaos properties [98] are by now established. On the other hand, the theory of weakly interacting particles enters in several areas of applied mathematics such as mean-field games or finance models [25], making it an area of active research.

Depending on the context of application, several results are available. The class of mean-field systems under the name of weakly interacting particles is rather large and models may substantially vary from one another depending on the regularity of the coefficients or the noise. This richness in models is reflected in a variety of different techniques implemented in their study (see e.g. [29, 88, 98] for three very different approaches).

If one focuses on models where the interaction function is regular enough, e.g. bounded and globally Lipschitz, one of the aspects that has not been completely investigated so far, concerns the initial condition. To the authors' knowledge, most of known results require a finite moment condition in order to prove tightness properties of the general sequence (e.g. [56]) or to apply a fixed-point argument in a suitable topological space (e.g. [29]). The only exceptions are given by [98, 99], although they require independent and identically distributed (IID) initial conditions. We want to point out that existence of a solution to the limiting system, a non-linear partial differential equation (PDE) known as Fokker-Planck or McKean-Vlasov equation, does not require any finite moment condition on the initial measure, see e.g. [98, Theorem 1.1]. Furthermore, whenever the particle system is deterministic, there is no need to assume independence (or any finite moment) for this same convergence, e.g. [41, 87].

We present a result in the spirit of the law of large numbers, without requiring any assumption on the initial conditions but the convergence of the associated empirical measure. Our main idea consists in exploiting a mild formulation associated to the stochastic partial differential equation satisfied by the empirical measure for a fixed (finite!) population. The main difficulty is giving a meaning to the noise term appearing in such formulation: exploiting the regularizing properties of the semigroup generated by the Laplacian in two different ways,

using rough paths theory and maximal inequalities for self-normalized processes respectively, we are able to adequately control it. By taking the limit for the size of the population which tends to infinity, the stochastic term vanishes and the limiting measure satisfies the well-known McKean-Vlasov equation.

1.1. Organization. The paper is organized as follows. In the rest of this section we present the model, known results and introduce the set-up in which the evolution of the empirical measure is studied along with notation used.

In Section 2 we give the definition of our notion of solution as well as a corresponding uniqueness statement. The law of large numbers, Theorem 2.3, is presented right after; the section ends with a discussion, the strategy of the proof and the existing literature.

The noise perturbation mentioned in the introduction is tackled in Section 3 where rough paths techniques and maximal inequalities for self-normalized processes are exploited. The proof of Theorem 2.3 is given at the end of this section.

Appendix A recalls general properties of analytic semigroups; Appendix B provides an extension of Gubinelli's theory for rough integration to our setting.

1.2. The model and known results. Consider $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ a filtered probability space, the filtration satisfying the usual conditions. Fix $d \in \mathbb{N}$, let $(B^i)_{i \in \mathbb{N}}$ be a sequence of IID \mathbb{R}^d -valued Brownian motions adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Fix $n \in \mathbb{N}$ and $T > 0$ a finite time horizon. Let $\Gamma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a bounded Lipschitz function, and $(x^{i,n})_{1 \leq i \leq n}$ the unique strong solution to

$$\begin{cases} dx_t^{i,n} &= \frac{1}{n} \sum_{j=1}^n \Gamma(x_t^{i,n}, x_t^{j,n}) dt + dB_t^i, & t \in (0, T), \\ x_0^{i,n} &= x_0^i, & \text{for } 1 \leq i \leq n. \end{cases} \quad (5.1)$$

The initial conditions are denoted by the sequence $(x_0^i)_{i \in \mathbb{N}} \subset \mathbb{R}^d$, whenever they are random they are taken independent of the Brownian motions. Existence and uniqueness for (5.1) is a classical result, e.g. [97].

The main quantity of interest in system (5.1) is the empirical measure $\nu^n = (\nu_t^n)_{t \in [0, T]}$, defined for $t \in [0, T]$ as the probability measure on \mathbb{R}^d such that

$$\nu_t^n := \frac{1}{n} \sum_{j=1}^n \delta_{x_t^{j,n}} \in \mathcal{P}(\mathbb{R}^d). \quad (5.2)$$

Observe that ν^n is a priori a probability measure on the continuous trajectories with values in \mathbb{R}^d , i.e. $\nu^n \in \mathcal{P}(C([0, T], \mathbb{R}^d))$, however in many instances we rather consider its projection $(\nu_t^n)_{t \in [0, T]} \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ as continuous function over the probability measures on \mathbb{R}^d . This last object does not carry the information of the time dependencies between time marginals, but is in our case more suitable when studying (5.1) in the limit for n which tends to infinity.

1.2.1. Known results. Fix a probability measure $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$. Whenever $(x_0^i)_{i \in \mathbb{N}}$ are taken either to be IID random variables sampled from ν_0 , or such that ν_0^n weakly converges to ν_0 with

some $p \geq 1$ finite moment, it is well known (e.g. [98, Theorem 1.4] and [29, Theorem 3.1]) that ν^n converges (in a precise sense depending on the setting) to the solution of the following PDE, known as non-linear Fokker-Planck (or McKean-Vlasov) equation

$$\begin{cases} \partial_t \nu_t = \frac{1}{2} \Delta \nu_t - \operatorname{div}[\nu_t(\Gamma * \nu_t)], & \text{for } t \in (0, T), \\ \nu|_{t=0} = \nu_0, \end{cases} \quad (5.3)$$

where $*$ denotes the integration with respect to the second argument, i.e. for $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$(\Gamma * \mu)(x) = \int_{\mathbb{R}^d} \Gamma(x, y) \mu(dy), \quad x \in \mathbb{R}^d.$$

REMARK 1.1. Observe that requiring IID initial conditions is not an innocent assumption as they are, in particular, exchangeable, see [98, §I.2] for more on this perspective. From an applied viewpoint, independence is often a hypothesis that we do not want to assume, see e.g. [36, Example II].

A solution to (5.3) is linked with the following non-linear process:

$$\begin{cases} x_t = x_0 + \int_0^t \int_{\mathbb{R}^d} \Gamma(x_s, y) \nu_s(dy) ds + B_t, \\ \nu_t = \operatorname{Law}(x_t), \end{cases} \quad (5.4)$$

where B is a Brownian motion independent of $(B^i)_{i \in \mathbb{N}}$ and x_0 . It is well-known that $\nu = (\nu_t)_{t \in [0, T]}$ is a solution to (5.3) if and only if the non-linear process $(x_t)_{t \in [0, T]}$ in (5.4) exists and is such that $\operatorname{Law}(x_t) = \nu_t$ for every $t \in [0, T]$.

In particular, we have to following theorem.

THEOREM 1.2 ([98, Theorem 1.1]). *Suppose Γ is bounded and Lipschitz and x_0 is a random variable with law $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$. Then, system (5.4) has a unique solution $(x_t)_{t \in [0, T]}$.*

Moreover, if $\nu = (\nu_t)_{t \in [0, T]}$ is the law of $(x_t)_{t \in [0, T]}$, then $\nu \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ and it solves the McKean-Vlasov equation (5.3) in the weak sense.

1.3. Set-up and notations. Let $W^{m,p} = W^{m,p}(\mathbb{R}^d)$ be the standard Sobolev space with $m \in \mathbb{N}$ and $p \in [1, \infty)$. Classical results as [2, Theorem 4.12] assure that

$$W_0^{m,p}(\mathbb{R}^d) = W^{m,p}(\mathbb{R}^d) \subset C_b(\mathbb{R}^d) \quad \text{whenever } mp > d, \quad (5.5)$$

where $C_b(\mathbb{R}^d)$ is the space of continuous bounded functions on \mathbb{R}^d . The space $W_0^{m,p}(\mathbb{R}^d)$ is the closure of $C_0^\infty(\mathbb{R}^d)$, i.e., the space of smooth functions with compact support, with respect to the norm

$$\|\varphi\|_{W^{m,p}} := \left(\sum_{0 \leq |\alpha| \leq m} \int_{\mathbb{R}^d} |\partial^\alpha \varphi(x)|^p dx \right)^{1/p}, \quad \varphi \in C_0^\infty(\mathbb{R}^d),$$

where $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = \alpha_1 + \dots + \alpha_d$ and $\partial^\alpha = (\partial_{x_1})^{\alpha_1} (\partial_{x_2})^{\alpha_2} \dots (\partial_{x_d})^{\alpha_d}$.

Fix $p = 2$ and $m > d/2$, we consider the Hilbert space $H^m := W^{m,2}(\mathbb{R}^d)$, with norm denoted by $\|\cdot\|_m$ and its dual space $H^{-m} := (H^m)^*$ with the standard dual norm defined by $\|\mu\|_{-m} := \sup_{\|h\|_m \leq 1} \langle \mu, h \rangle_{-m,m}$. The action the action of H^{-m} on H^m is denoted by $\langle \cdot, \cdot \rangle_{-m,m}$. By duality, it follows from (5.5) that

$$\mathcal{P}(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)^* \subset H^{-m}.$$

We denote by $(\cdot, \cdot)_m$ the scalar product in H^m and by $\langle \cdot, \cdot \rangle$ the natural action of a probability measure on test functions, i.e., for $\nu \in \mathcal{P}(\mathbb{R}^d)$ and a smooth function h , we write $\langle \nu, h \rangle = \int_{\mathbb{R}^d} h(x) \nu(dx)$. We often abuse of notation denoting the density of a probability measure by the probability measure itself.

Let $\nu \in \mathcal{P}(\mathbb{R}^d)$, and thus $\nu \in H^{-m}$, and let $\tilde{\nu} \in H^m$ be its Riesz representative, then we have for any $h \in H^m$

$$\langle \nu, h \rangle = \nu(h) = (\tilde{\nu}, h)_m = \langle \nu, h \rangle_{-m,m}$$

and therefore

$$|\langle \nu, h \rangle| \leq \|\nu\|_{-m} \|h\|_m.$$

In particular

$$\sup_{\|h\|_m \leq 1} \langle \nu, h \rangle = \|\nu\|_{-m}.$$

If $(\mu^n)_{n \in \mathbb{N}}$ is a sequence of probability measures which weakly converges to some $\mu \in \mathcal{P}(\mathbb{R}^d)$, we use the notation $\mu^n \rightharpoonup \mu$. For weak convergence and weak- $*$ -convergence of a sequence $(x_n)_n \subset X$ to some $x \in X$, X being a Banach space, we use the standard notations $x_n \rightharpoonup x$ and $x_n \xrightarrow{*} x$ respectively.

As introduced in [86], we will use $\|\cdot\|_{-m}$ as distance between probability measures and our results will be expressed with respect to this topology.

The various constants in the paper will always be denoted by C or C_α to emphasize the dependence on some parameter α . Their value may change from line to line.

2. Main result

Before stating the main result, we give the definition of weak-mild solution to (5.3) in the Hilbert space H^m . We denote by $S = (S_t)_{t \in [0, T]}$ the analytic semigroup generated by $\frac{\Delta}{2}$ on H^m , see Appendix A for general properties of S .

DEFINITION 2.1 (*m-weak-mild solutions to McKean-Vlasov PDEs*). Let ν_0 be an element in H^{-m} . We call $\nu \in L^\infty([0, T], H^{-m})$ an *m-weak-mild solution* to the problem (5.3), if for every $h \in H^m$ and $t \in [0, T]$, it holds

$$\langle \nu_t, h \rangle_{-m,m} = \langle \nu_0, S_t h \rangle_{-m,m} + \int_0^t \langle \nu_s, (\nabla S_{t-s} h)(\Gamma * \nu_s) \rangle_{-m,m} ds. \quad (5.6)$$

If Γ is sufficiently regular, uniqueness is a consequence of Gronwall's Lemma.

PROPOSITION 2.2 (Uniqueness). *Suppose that $\Gamma(\cdot_x, \cdot_y) \in H_y^m W_x^{m,\infty}$, i.e.*

$$\|\Gamma(\cdot_x, \cdot_y)\|_{H_y^m W_x^{m,\infty}} = \max_{|\beta| \leq m} \left\| \sum_{|\alpha| \leq m} \int_{\mathbb{R}^d} (\partial_x^\beta \partial_y^\alpha \Gamma(x, y))^2 dy \right\|_{L_x^\infty} < \infty. \quad (5.7)$$

Then, m -weak mild solutions to (5.3) are unique.

PROOF. Suppose $\nu, \rho \in L^\infty([0, T], H^{-m})$ are two m -weak mild solutions. Then, taking the difference between the two equations (5.6), one obtains that for every $h \in H^m$

$$\begin{aligned} \langle \nu_t - \rho_t, h \rangle_{-m, m} &= \int_0^t \langle \nu_s - \rho_s, (\nabla S_{t-s} h)(\Gamma * \nu_s) \rangle_{-m, m} ds + \\ &\quad + \int_0^t \langle \rho_s, (\nabla S_{t-s} h)(\Gamma * (\nu_s - \rho_s)) \rangle_{-m, m} ds. \end{aligned}$$

In particular,

$$\begin{aligned} \|\nu_t - \rho_t\|_{-m} &\leq \int_0^t \|\nu_s - \rho_s\|_{-m} \|(\nabla S_{t-s} h)(\Gamma * \nu_s)\|_m ds + \\ &\quad + \int_0^t \|\rho_s\|_{-m} \|(\nabla S_{t-s} h)(\Gamma * (\nu_s - \rho_s))\|_m ds. \end{aligned}$$

Observe that, for $\mu \in H^{-m}$ it holds that

$$\begin{aligned} \|(\nabla S_{t-s} h)(\Gamma * \mu)\|_m &\leq \|\nabla S_{t-s} h\|_m \|\Gamma * \mu\|_{W^{m,\infty}} \leq \\ &\leq \frac{C}{\sqrt{t-s}} \|h\|_m \|\mu\|_{-m} \|\Gamma(\cdot_x, \cdot_y)\|_{H_y^m W_x^{m,\infty}}, \end{aligned} \quad (5.8)$$

where we have used the properties of the semigroup. Using the continuous embedding of $\mathcal{P}(\mathbb{R}^d)$ into H^{-m} , we conclude that there exists a (new) constant $C > 0$:

$$\|\nu_t - \rho_t\|_{-m} \leq C \|\Gamma(\cdot_x, \cdot_y)\|_{H_y^m W_x^{m,\infty}} \int_0^t \frac{1}{\sqrt{t-s}} \|\nu_s - \rho_s\|_{-m} ds.$$

A Gronwall-like lemma yields the proof. \square

We are ready to state the main result.

THEOREM 2.3. *Assume $m > d/2 + 3$ and $\Gamma(\cdot_x, \cdot_y) \in H_y^m W_x^{m,\infty}$. If $\nu_0 \in H^{-m}$, then there exists $\nu \in L^\infty([0, T], H^{-m})$, unique m -weak-mild solution to (5.6). Suppose that the initial empirical measure associated to the particle system (5.1) is such that*

$$\nu_0^n \rightharpoonup \nu_0 \quad \text{in } H^{-m}$$

in probability. Then, the empirical measure ν^n of (5.1) satisfies

$$\nu^n \xrightarrow{*} \nu \quad \text{in } L^\infty([0, T], H^{-m})$$

in probability.

Moreover, if $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$, then ν is the unique weak solution of the McKean-Vlasov equation (5.3) and, in particular, $\nu \in C([0, T], \mathcal{P}(\mathbb{R}^d))$.

2.1. Discussion. Theorem 2.3 shows a law of large numbers in $L^\infty([0, T], H^{-m})$ by directly studying the evolution of the empirical measure. Contrary to most of the existing proofs in the literature, it does not establish any trajectorial estimates on system (5.1) and does not invoke propagation of chaos techniques, as, e.g., in [83, 98]. This allows to deal with very general initial data: the weak convergence of $(\nu_0^n)_{n \in \mathbb{N}}$ in H^{-m} – which is implied by the weak convergence in $\mathcal{P}(\mathbb{R}^d)$ – suffices.

Working in H^{-m} for $m > d/2$ assures a bound on $\|\nu\|_{-m}$ which is uniform in $\nu \in \mathcal{P}(\mathbb{R}^d)$ thanks to the continuous embedding of $\mathcal{P}(\mathbb{R}^d)$ in H^{-m} and the duality properties of probability measures, see Lemma 5.A.2. By exploiting the equation satisfied by ν^n , we are able to establish a compactness property for $(\nu^n)_{n \in \mathbb{N}}$, usually hard to obtain in $\mathcal{P}(\mathbb{R}^d)$, and which represents our main tool for obtaining the existence both of the limit solution and of a convergent subsequence.

Weak-mild solutions make sense for any $m > d/2$, yet we have to require the stronger condition $m > d/2 + 3$ in order to give a pathwise meaning to the stochastic term present in the dynamics. This implies that Γ is C^3 . In this last case, it is already known that a weak solution to the McKean-Vlasov equation (5.3) exists for any initial probability measure ν_0 . Since weak solutions are weak-mild solutions, as we will show in the sequel, a byproduct of our main result is the uniqueness of (weak) solutions to equation (5.3).

The particle system (5.1) represents an interaction setting where no transport is present in the dynamics. We have decided not to include other terms so as to keep the underlying ideas and techniques as clear as possible. However, all our arguments readily extend to the more general case of interacting particles given by

$$dx_t^{i,n} = F(x_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \Gamma(x_t^{i,n}, x_t^{j,n})dt + dB_t^i, \quad (5.9)$$

provided that $F \in H^m$.

Finally, we point out that the need of rather high regularity in Γ (and F) is an intrinsic requirement of rough paths theory and not of the particular class of models we are working with. In particular, proving Theorem 2.3 independently of rough paths arguments would likely yield less restrictive regularity constraints on Γ . On the other hand, rough paths theory allows to give a pathwise definition of the stochastic partial differential equation satisfied by the empirical measure. Such viewpoint appears to be new in the literature. Moreover, the proposed strategy represents an application to the algebraic integration with respect to semigroups, presented in [54], that can be interesting on its own.

2.2. Comparison with the existing literature. Proving a law of large numbers by directly studying the empirical measure and not the single trajectories is the classical approach in the deterministic setting [87, 41], i.e., when no Brownian motions are acting on system

(5.1). In the case of interacting diffusions, the idea of studying the equation satisfied by the empirical measure for a fixed n , comes from the two articles [10, 78] and the recent [30], where a weak-mild formulation is derived and carefully studied. Contrary to our case, in [10, 30, 78] the particles live in the one dimensional torus which considerably simplifies the analysis; we refer to Remark 3.5.

A Hilbertian approach for particle systems has already been discussed in [45], where it is used to study the fluctuations of the empirical measure around the McKean-Vlasov limit. However, [45] does not make use of the theory of semigroups but instead requires strong hypothesis on the initial conditions which have to be IID and with finite $(4d + 1)$ -moment (see [45, §3]). The evolution of the empirical measure (5.2) is then studied in weighted Hilbert spaces (or, more precisely, in spaces of Bessel potentials) so as to fully exploit the properties of mass concentration given by the condition on the moments. Observe that we are not able to present a fluctuation result, given the lack of a suitable uniform estimate on the noise term.

Studying the action of an analytic semigroup in the evolution of an interacting particle system has been recently proposed in similar settings; we refer to [46, 47] and references therein. This method is referred to as the semigroup approach. We want to stress that the cited works deal with smooth mollified empirical measures and work in a weaker topology (with respect to the time variable) than the one of Theorem 2.3.

The strategies developed in [84, 83, 65], and further applied in the case of mean-field games in [25], study the evolution of the joint law of system (5.1) and take a more abstract viewpoint. In particular, they study the system dynamics at the level of the flows and not directly addressing the empirical measure.

Finally, observe that under a suitable change of the time-scale, the n -dependent SPDE satisfied by the empirical measure (5.2) is the mild formulation of the Dean-Kawasaki equation [66, Theorem 1] and [67].

2.3. Strategy of the proof. Using Itô's formula, we derive an equation satisfied by ν^n for every fixed $n \in \mathbb{N}$, which turns out to be the McKean-Vlasov PDE perturbed by some noise w^n , see Lemma 3.1. This equation makes sense in $L^\infty([0, T], H^{-m})$ and in this space we study the convergence of $(\nu^n)_{n \in \mathbb{N}}$.

The main challenge towards the proof of Theorem 2.3 is giving a meaning to w^n and suitably controlling it. We first give a pathwise definition of such term through rough paths theory, see Lemma 3.2, referring to Appendix 5.B for a suitable theory of rough integration in our setting. This in turn will allow to show that $(\nu^n)_{n \in \mathbb{N}}$ is uniformly bounded in $L^\infty([0, T], H^{-m})$ and extract a weak-* converging subsequence, see Lemma 3.8.

To show that a converging subsequence satisfies the weak-mild solution (5.6) in the limit, as shown in Lemma 3.9, we need a further step: the pointwise estimate of $w^n(h)$, for a fixed $h \in H^m$. Using a suitable decomposition of the semigroup and a maximal inequality for self-normalized processes, we are able to prove that $w^n(h)$ converges to zero in probability as n

diverges, see Lemma 3.6. If on the one hand the rough paths bound cannot take advantage of the statistical independence of the Brownian motions and thus, cannot be improved in n , on the other hand the probability estimate does not suffice to define w^n as an element of $L^\infty([0, T], H^{-m})$. We refer to Subsection 3.2 and Remark 3.7 for more on this aspect.

The uniqueness of weak-mild solution, Proposition 2.2, is the last ingredient to obtain that any convergent subsequence of $(\nu^n)_{n \in \mathbb{N}}$ admits a further subsequence that converges \mathbb{P} -a.s. to the same ν satisfying (5.6). This is equivalent to the weak-* convergence in probability to the weak-mild solution ν .

3. Proofs

We start by giving the n -dependent stochastic equation satisfied by the empirical measure for each $n \in \mathbb{N}$. We then move to the control on the noise term and, finally, the proof of Theorem 2.3.

3.1. A weak-mild formulation satisfied by the empirical measure. Recall that $(S_t)_{t \in [0, T]}$ denotes the semigroup generated by $\frac{\Delta}{2}$ on H^m .

LEMMA 3.1. *Assume $m > d/2 + 2$. The empirical measure (5.2) associated to the particle systems (5.1) satisfies for every $h \in H^m$ and $t \in [0, T]$*

$$\langle \nu_t^n, h \rangle_{-m, m} = \langle \nu_0^n, S_t h \rangle_{-m, m} + \int_0^t \langle \nu_s^n, (\nabla S_{t-s} h)(\Gamma * \nu_s^n) \rangle_{-m, m} ds + w_t^n(h), \quad \mathbb{P}\text{-a.s.}, \quad (5.10)$$

where

$$w_t^n(h) = \frac{1}{n} \sum_{j=1}^n \int_0^t [\nabla S_{t-s} h](x_s^{j, n}) \cdot dB_s^j. \quad (5.11)$$

PROOF. Fix $t \in [0, T]$ and $h \in H^m$, by (5.5) h is $C^2(\mathbb{R}^d)$. For $s < t$, applying Itô's formula onto the test function $\varphi(x, s) = (S_{t-s} h)(x)$, we obtain

$$\begin{aligned} h(x_t^{i, n}) &= (S_t h)(x_0^{i, n}) + \frac{1}{n} \sum_{j=1}^n \int_0^t (\nabla S_{t-s} h)(x_s^{i, n}) \Gamma(x_s^{i, n}, x_s^{j, n}) ds \\ &\quad + \int_0^t (\nabla S_{t-s} h)(x_s^{j, n}) \cdot dB_s^j. \end{aligned}$$

Summing over all particles and dividing by $1/n$, the claim is proved modulo well-posedness of the noise term w^n which is presented in the following subsection. \square

3.2. Controlling the noise term. The aim of this subsection is to control the noise term w^n appearing in the weak-mild formulation (5.10) for the empirical measure. We start by giving a pathwise definition of the integral (5.11), i.e. for any $\omega \in A \subset \Omega$ where $\mathbb{P}(A) = 1$ and any $h \in H^m$ we define

$$w_t^n(h)(\omega) = \left(\frac{1}{n} \sum_{j=1}^n \int_0^t [\nabla S_{t-s} h](x_s^{j, n}) \cdot dB_s^j \right) (\omega),$$

which in turn allows to define w^n as an element of $L^\infty([0, T], H^{-m})$, via an inequality of the form

$$\sup_{\|h\|_m=1} |w_t^n(h)(\omega)| \leq C_T(\omega)$$

for $\omega \in A$, see Lemma 3.2. For this purpose, we extend Gubinelli's theory for rough integration (see [53] and [54, §3 and 4]) to our setting, see Appendix 5.B for notations and precise results on this extension.

A probabilistic estimate is then given, exploiting the independence of the Brownian motions; Lemma 3.6 shows that

$$\mathbb{E} \left[\sup_{t \in [0, T]} |w_t^n(h)|^2 \right] \leq \frac{C}{n} \|h\|_m^2, \quad h \in H^m.$$

This estimate will allow us to prove the convergence of (5.10) to (5.6) for every fixed $h \in H^m$, see Lemma 3.9.

3.2.1. Pathwise definition via rough paths theory for semigroup functionals. We start by observing that the noise term $w_t^n(h)$ in (5.11) is neither a stochastic convolution that could be treated using a maximal inequality in Hilbert spaces (e.g. [32, §6.4] and [7] in the context of an unbounded diffusion operator), nor a classical controlled rough path integral (e.g. [49]) as the integrand depends on the upper integration limit.

We combine the strategies in [54, 53] so to define $w_t^n(h)$ in a pathwise sense. Note that our setting is different from [54], where an infinite dimensional theory à la Da Prato-Zabczyk is constructed, while we are interested in finite dimensional stochastic integrals over functionals of such objects. Our construction is nonetheless similar to [54]: we fix the Itô-rough path lift associated to Brownian motion and extend the algebraic integration in [54] to our setting of semigroup functionals. This extension is presented in detail in Appendix B, where the main ingredient, the Sewing lemma, is proven. Before stating Lemma 3.2, we present in a heuristic fashion the main ideas towards a rough path construction of (5.11).

Note that it suffices to define integrals of the form

$$\int_s^t \nabla S_{t-u} f(x_u) \cdot dB_u \tag{5.12}$$

in a pathwise sense for a class of sufficiently regular functions f and where $(x_u)_u$ is an \mathbb{R}^d -valued process controlled by the Brownian motion $(B_u)_u$, such that

$$x_t - x_s = B_t - B_s + O(|t - s|), \quad \text{for } s, t \in [0, T], \mathbb{P}\text{-a.s.} \tag{5.13}$$

Recall that in the classical setting of rough paths theory, one has for $s \leq t$

$$\int_s^t f(x_u) dB_u = f(x_s) B_{ts} + (D_x f)(x_u) \mathbb{B}_{ts} + R_{ts}$$

where we have used the notation $B_{ts} := B_t - B_s$ as well as

$$\mathbb{B}_{ts} := \int_s^t B_{us} \otimes dB_u, \quad t \geq s \in [0, T].$$

In particular, $A_{ts} := f(x_s)B_{ts} + (D_x f)(x_u)\mathbb{B}_{ts}$ is a germ and, thanks to (5.13), $R_{ts} = o(|t - s|)$ is a remainder in the terminology of [53]. In the same spirit of [53], we rewrite the left hand side of this expression as

$$\int_s^t f(x_u)dB_u = [\delta I]_{ts} = I_t - I_s$$

where

$$I_t = \int_0^t f(x_u)dB_u.$$

We are thus left with

$$[\delta I]_{ts} = A_{ts} + R_{ts}. \quad (5.14)$$

Recall that Gubinelli's Sewing Lemma formulates precise conditions under which a given germ A gives rise to a unique remainder term $R_{ts} = o(|t - s|)$ and such that I can be obtained as

$$I_t := \lim_{|\mathcal{P}[0,t]| \downarrow 0} \sum_{[u,v] \in \mathcal{P}[0,t]} A_{vu}.$$

If one tries to follow a similar approach for the quantity of interest (5.12), a canonical candidate for local approximations to (5.12) would be

$$\int_s^t (\nabla S_{t-u} f)(x_u)dB_u = (\nabla S_{t-s} f)(x_s)B_{ts} + (D\nabla S_{t-s} f)(x_s)\mathbb{B}_{ts} + R_{ts}.$$

However, notice that if we were to set

$$I_t(f) := \int_0^t (\nabla S_{t-u} f)(x_u)dB_u$$

then, we would obtain

$$\begin{aligned} [\delta I(f)]_{ts} &= I_t(f) - I_s(f) = \int_0^t (\nabla S_{t-u} f)(x_u)dB_u + \int_0^s (\nabla S_{s-u} (S_{t-s} - \text{Id})f)(x_u)dB_u \\ &= \int_s^t (\nabla S_{t-u} f)(x_u)dB_u + I_s((S_{t-s} - \text{Id})f) \\ &\neq \int_s^t (\nabla S_{t-u} f)(x_u)dB_u, \end{aligned}$$

in contrast to the above setting, meaning the standard approach of [53] fails. If one defines, following Gubinelli and Tindel [54, p.16], the operator ϕ via

$$[\phi I(f)]_{ts} = I_s((S_{t-s} - \text{Id})f)$$

as well as the operator $\hat{\delta}$ via

$$[\hat{\delta} I(f)]_{ts} = [\delta I(f)]_{ts} - [\phi I(f)]_{ts},$$

the desired relationship is recovered, indeed

$$\begin{aligned} [\hat{\delta}I(f)]_{ts} &= \int_s^t (\nabla S_{t-u}f)(x_u)dB_u \\ &= (\nabla S_{t-s}f)(x_s)B_{ts} + (D\nabla S_{t-s}f)(x_s)\mathbb{B}_{ts} + R_{ts}. \end{aligned}$$

The idea is hence to change the cochain complex in [53] and to consider a perturbed version of it associated to the operator $\hat{\delta}$, this is done in Lemma 5.B.1. Lemma 5.B.2 proves a Sewing Lemma in this modified setting, which in turn allows to construct the above remainder R_{ts} . The germ will therefore be

$$[Af]_{ts} = (\nabla S_{t-s}f)(x_s)B_{ts} + (D\nabla S_{t-s}f)(x_s)\mathbb{B}_{ts}.$$

For $0 = t_0 < \dots < t_{n+1} = t$, note that due to

$$\begin{aligned} I_t(f) &= \int_0^t (\nabla S_{t-u}f)(x_u)dB_u \\ &= \sum_{k=0}^n \int_{t_k}^{t_{k+1}} (\nabla S_{t-u}f)(x_u)dB_u \\ &= \sum_{k=0}^n \int_{t_k}^{t_{k+1}} (\nabla S_{t_{k+1}-u}(S_{t-t_{k+1}}f))(x_u)dB_u \\ &= \sum_{k=0}^n [A(S_{t-t_{k+1}}f)]_{t_{k+1}t_k} + \sum_{k=0}^n R_{t_{k+1}t_k}, \end{aligned}$$

the correct way of sewing together the germs is given by

$$I_t(f) = \lim_{n \rightarrow \infty} \sum_{k=0}^n [A(S_{t-t_{k+1}}f)]_{t_{k+1}t_k},$$

which is reflected in equation (5.31) in Corollary 5.B.3. In particular, note that this Corollary comes with the stability estimate (5.30) which allows to eventually deduce the first crucial estimate (5.15) on the noise term, as shown in the next Lemma.

LEMMA 3.2. *Suppose $m > d/2 + 3$. For every $\alpha \in (1/3, 1/2)$, there exists a positive random constant $C = C_\alpha$ that is finite \mathbb{P} -a.s. (and of finite moments for all orders) such that \mathbb{P} -a.s.*

$$|w_t^n(h)| \leq C_\alpha (1+t)^{3\alpha} \|h\|_m \tag{5.15}$$

for any $t \geq 0$ and $h \in H^m$.

PROOF. We follow the notations of Appendix 5.B. Fix $\alpha \in (1/3, 1/2)$ and recall that (B, \mathbb{B}) is the Itô rough path lift, with

$$\mathbb{B}_{ts} := \int_s^t B_{us} \otimes dB_u, \quad t, s \in [0, T],$$

where $B_{us} := B_u - B_s$. Note that the above stochastic integral is understood in the Itô sense.

We use Lemma 5.B.3 to define the Itô integral (5.12). This in turn will imply the well-posedness of $w_t^n(h)$ with the choice $f = h$, $x = x^{i,n}$ and $B = B^i$ for $i = 1, \dots, n$. Indeed, $x^{i,n}$ is controlled by B^i (recall (5.1) and the fact that Γ is bounded), and thus

$$|w_t^n(h)| \leq \frac{1}{n} \sum_{i=1}^n \left| \int_0^t (\nabla S_{t-s} h)(x_s^{i,n}) dB_s^i \right| \leq C_\alpha (1+t)^{3\alpha} \|h\|_m.$$

Define the operator A acting on $f \in H^m$ into $C(\Delta_2, \mathbb{R})$ via

$$[Af]_{ts} := (\nabla S_{t-s} f)(x_s) \cdot B_{ts} + (D_x \nabla S_{t-s} f)(x_s) \cdot \mathbb{B}_{ts},$$

where D_x denotes the Jacobian in \mathbb{R}^d and \cdot the scalar product between tensors of the same dimension. In the sequel, we adopt the following shorter notation

$$[Af]_{ts} := \nabla S_{ts} f_s B_{ts} + D_x \nabla S_{ts} f_s \mathbb{B}_{ts}.$$

As in classical rough paths theory $[Af]_{ts}$ is not a 1-increment (i.e. a difference as B_{ts}) but a continuous function of the two variables s and t . In particular $A \in D_2$, i.e. A is a linear operator from the Banach space H^m to C_2 .

One can actually prove that $A \in D_2^\alpha$: for $0 \leq s \leq t \leq T$ and $f \in H^m$

$$|[Af]_{ts}| \leq \|\nabla S_{ts} f\|_\infty |B_{ts}| + \|D_x \nabla S_{ts} f\|_\infty |\mathbb{B}_{ts}| \leq C_\alpha \|f\|_m |t-s|^\alpha,$$

where $C_\alpha = C_\alpha(\omega)$ depends on the α -Hölder norm of $B(\omega)$ and $\mathbb{B}(\omega)$ and we have used the properties of S , see Lemma 5.A.1. Note in particular that $C_\alpha < \infty$, \mathbb{P} -a.s. and that C_α has finite moments of all orders.

Recall the definition of $\hat{\delta}$ (Lemma 5.B.1), in order to apply Lemma 5.B.2 and Corollary 5.B.3 we need to show that $\hat{\delta}A \in D_3^{1+}$. Let $f \in H^m$ and $s < u < t$, one has

$$\begin{aligned} [\hat{\delta}Af]_{tus} &= [\delta Af]_{tus} - [\phi Af]_{tus} = [Af]_{ts} - [Af]_{tu} - [Af]_{us} - [A(S_t \cdot - \text{Id})f]_{us} = \\ &= [Af]_{ts} - [Af]_{tu} - [AS_t \cdot f]_{us}. \end{aligned}$$

Observe that thanks to the properties of the semigroup

$$[AS_t \cdot f]_{us} = \nabla S_{us} S_{tu} f_s B_{us} + D_x \nabla S_{us} S_{tu} f_s \mathbb{B}_{us} = \nabla S_{ts} f_s B_{us} + D_x \nabla S_{ts} f_s \mathbb{B}_{us}.$$

In particular, using Chen's relation

$$\mathbb{B}_{ts} = \mathbb{B}_{us} + \mathbb{B}_{tu} + B_{tu} \otimes B_{us}$$

we obtain

$$\begin{aligned} [\hat{\delta}Af]_{tus} &= \nabla S_{ts} f_s B_{tu} - \nabla S_{tu} f_u B_{tu} + D_x \nabla S_{ts} f_s (\mathbb{B}_{ts} - \mathbb{B}_{us}) - D_x \nabla S_{tu} f_u \mathbb{B}_{tu} = \\ &= (\nabla S_{ts} f_s - \nabla S_{tu} f_u) B_{tu} + D_x (\nabla S_{ts} f_s - \nabla S_{tu} f_u) \mathbb{B}_{tu} + D_x S_{ts} f_s B_{tu} \otimes B_{us}. \end{aligned}$$

We rewrite everything as the sum of four terms

$$\begin{aligned} [\hat{\delta}Af]_{tus} &= \nabla(S_{ts} - S_{tu})f_u B_{tu} + D_x \nabla(S_{ts} - S_{tu})f_u \mathbb{B}_{tu} + \\ &\quad + (D_x \nabla S_{ts} f_s - D_x \nabla S_{ts} f_u) \mathbb{B}_{tu} + (\nabla S_{ts} f_s - \nabla S_{ts} f_u + D_x \nabla S_{ts} f_s B_{us}) B_{tu} \\ &=: A^1 + A^2 + A^3 + A^4. \end{aligned}$$

For A^1 we obtain

$$|\nabla(S_{ts} - S_{tu})f_u B_{tu}| \leq \|\nabla(S_{ts} - S_{tu})f_u\|_\infty |B_{tu}| \leq C_\alpha \|f\|_m |t - u|^\alpha |u - s|,$$

where $C_\alpha = C_\alpha(\omega)$ depends on the α -Hölder norm of $B(\omega)$ and we have used the properties of S , see Lemma 5.A.1. Note in particular that $C_\alpha < \infty$, \mathbb{P} -a.s. and that C_α has finite moments of all orders. Similarly, for A^2 (with a different C_α)

$$|D_x \nabla(S_{ts} - S_{tu})f_u \mathbb{B}_{tu}| \leq C_\alpha \|f\|_m |t - u|^{2\alpha} |u - s|^{1/2}.$$

Observe now that, since $f \in C_b^3$, the function $D_x \nabla S_{ts} f$ is Lipschitz uniformly in s and t , from which we extract that

$$\begin{aligned} |(D_x \nabla S_{ts} f_s - D_x \nabla S_{ts} f_u) \mathbb{B}_{tu}| &\leq C_\alpha \|f\|_m |x_s - x_u| |t - u|^{2\alpha} \\ &\leq C_\alpha \|f\|_m |t - u|^{2\alpha} |u - s|^\alpha. \end{aligned}$$

Using (5.13), we recognize in A^4 the Taylor expansion of $\nabla S_{ts} f$ around x_s , i.e.

$$\begin{aligned} |\nabla S_{ts} f_u - \nabla S_{ts} f_s - D_x \nabla S_{ts} f_s B_{us}| &\leq \\ &\leq |\nabla S_{ts} f_u - \nabla S_{ts} f_s - D_x \nabla S_{ts} f_s x_{us}| + c |D_x \nabla S_{ts} f_s| |u - s| \leq \\ &\leq c \|f\|_m |x_{us}|^2 + c \|f\|_m |u - s| \leq c \|f\|_m |u - s|^{2\alpha}. \end{aligned}$$

We conclude that

$$|A^4| \leq C_\alpha \|f\|_m |t - u|^\alpha |u - s|^{2\alpha}.$$

Putting the four estimates together, we have just shown $\hat{\delta}A \in D_3^{1+}$ and, in particular, that

$$\left\| \hat{\delta}A \right\|_{D_3^{3\alpha}} \leq C_\alpha$$

for some C_α which is finite \mathbb{P} -a.s. and admits moments of all orders. By Corollary 5.B.3, we know that there exists $I \in D_1$ such that

$$[\hat{\delta}If]_{ts} = \lim_{|\mathcal{P}^n[s,t]| \rightarrow 0} \sum_{[u,v] \in \mathcal{P}^n[s,t]} [AS_t f]_{vu}$$

is well defined. For $0 \leq s \leq t \leq T$, se set

$$\int_s^t \nabla S_{t-u} f(x_u) \cdot dB_u := [\hat{\delta}If]_{ts}.$$

Again Corollary 5.B.3 assures that there exists a (new) constant C_α , depending on the norm of A in D_2^α and the norm of $\hat{\delta}A$ in $D_3^{3\alpha}$, such that

$$\left| \int_0^t \nabla S_{t-u} f(x_u) \cdot dB_u \right| \leq C_\alpha \|f\|_m (1+t)^{3\alpha}.$$

The proof is concluded. \square

3.2.2. *Controlling $w_t^n(h)$ via a maximal inequality for self-normalized processes.* The aim of this subsection is to give a probabilistic bound on

$$w_t^n(h) = \frac{1}{n} \sum_{j=1}^n \int_0^t [\nabla S_{t-s} h](x_s^{j,n}) dB_s^j$$

by exploiting the independence of the Brownian motions (we have removed the product symbol \cdot for the sake of notation).

Observe that if $w_t^n(h)$ didn't involve a convolution with the semigroup S , $w_t^n(h)$ would be a standard martingale and classical estimates like the Burkholder-Davis-Gundy inequality could be used to establish the desired bound. While the convolution with the semigroup S destroys the martingale property, $w_t^n(h)$ is still closely related to maximal inequalities for self-normalized martingales for which the following fine estimate due to Graversen and Peskir [52] is available.

LEMMA 3.3 ([52, Corollary 2.8] and [62, Corollary 2.4]). *Let $(M_t)_{t \in [0, T]}$ be a continuous local martingale. There exists a universal constant C such that*

$$\mathbb{E} \left[\sup_{t \in [0, \tau]} \frac{|M_t|^2}{1 + \langle M \rangle_t} \right] \leq C \mathbb{E} [\log(1 + \log(1 + \langle M \rangle_\tau))]$$

for every stopping time $\tau \leq T$.

Observe that this result is a consequence of more general bounds on self-normalized processes of the form $X_t = A_t/B_t$ (e.g. [35]), where in this case $A_t = M_t$ is a martingale and $B_t^2 - 1 = \langle M \rangle_t$ its quadratic variation.

Let us illustrate in the following example how this interpretation can be used to directly obtain a bound on

$$v_t = \frac{1}{n} \sum_{j=1}^n \int_0^t e^{-a(t-s)} dB_s^j, \quad a > 0,$$

which could be seen as a most simple toy model for $w_t^n(h)$.

EXAMPLE 3.4. Let $(B^j)_{j \leq n}$ be independent Brownian motions on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$. For $a > 0$, let $(X^j)_{j \leq n}$ be the following associated family of Ornstein Uhlenbeck processes:

$$X_t^j := \int_0^t e^{-a(t-s)} dB_s^j, \quad t \in [0, T]$$

and consider the quantity

$$v_t := \frac{1}{n} \sum_{j=1}^n X_t^j.$$

We remark that we may rewrite

$$\sum_{j=1}^n X_t^j = \sqrt{\frac{n}{2a}} e^{-at} \left(\sum_{j=1}^n \sqrt{\frac{2a}{n}} \int_0^t e^{as} dB_s^j \right) =: \sqrt{\frac{n}{2a}} e^{-at} M_t.$$

Notice that M is a martingale of quadratic variation

$$\langle M \rangle_t = (e^{2at} - 1)$$

and therefore, by Lemma 3.3, we conclude that

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} |v_t|^2 \right] &= \frac{1}{2na} \mathbb{E} \left[\sup_{t \in [0, T]} |e^{-at} M_t|^2 \right] \\ &= \frac{1}{2na} \mathbb{E} \left[\sup_{t \in [0, T]} \frac{|M_t|^2}{1 + \langle M \rangle_t} \right] \\ &\leq C \frac{1}{2na} \log(1 + 2aT). \end{aligned}$$

Note that we crucially exploited the splitting $e^{-a(t-s)} = e^{-at} e^{as}$, which is not available in the semigroup setting we are concerned with. Intending to employ such a step suggests to pass by a functional calculus for the semigroup, which we briefly discuss next.

Recall that an analytic semigroup is a bounded linear operator that can be expressed by means of a Dunford integral (e.g. [58, 75] and Appendix 5.A). The integral representation of S is given for every $t \in [0, T]$ by

$$S_t = \frac{1}{2\pi i} \int_{\gamma_{r,\eta}} e^{t\lambda} R(\lambda, \frac{\Delta}{2}) d\lambda, \quad (5.16)$$

where $R(\lambda, \frac{\Delta}{2}) = (\lambda \text{Id} - \frac{\Delta}{2})^{-1}$ denotes the resolvent of $\frac{\Delta}{2}$ and where, for $r > 0$ and $\eta \in (\pi/2, \pi)$, $\gamma_{r,\eta}$ is the curve $\{\lambda \in \mathbb{C} : |\arg \lambda| = \eta, |\lambda| \geq r\} \cup \{\lambda \in \mathbb{C} : |\arg \lambda| \leq \eta, |\lambda| = r\}$, oriented counterclockwise.

Plugging (5.16) into the expression of $w_t^n(h)$ yields

$$w_t^n(h) = \frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \int_{\gamma_{r,\eta}} e^{(t-s)\lambda} [\nabla R(\lambda, \frac{\Delta}{2}) h] (x_s^{j,n}) d\lambda dB_s^j,$$

splitting the complex integral into three real integrals parametrizing $\gamma_{r,\eta}$, and then using stochastic Fubini, one is left with expressions similar to

$$\frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t e^{(t-s)\rho e^{i\eta}} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2}) h] (x_s^{j,n}) e^{i\eta} dB_s^j, \quad \rho > r,$$

which remind us of 1-dimensional self-normalized martingale for every ρ , similar to the process $(v_t)_t$ considered in Example 3.4.

It remains to establish a suitable bound on the expression

$$[\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h](x_s^{j,n})$$

and to ensure that this bound is integrable for $\rho \in (r, \infty)$, see Lemma 5.A.3.

Putting all the above considerations together with care, one obtains a maximal inequality for $w_t^n(h)$ that we present in Lemma 3.6.

REMARK 3.5. A similar control has already been used in [30, Lemma 3.3], see also [10, §3.1] and [78, §4] for an estimate using the Rodemich-Garsia-Rumsey lemma. However, in all these cases the particles are living in the one dimensional torus, making the (still highly technical) noise analysis considerably simpler due to the decomposition in Fourier series.

LEMMA 3.6. *Assume $m > d/2$. There exists a constant $C \geq 1$, independent of n and $h \in H^m$, such that for every $h \in H^m$*

$$\mathbb{E} \left[\sup_{t \in [0, T]} |w_t^n(h)|^2 \right] \leq \frac{C}{n} \|h\|_m^2. \quad (5.17)$$

PROOF. Let $h \in H^m$ and $\gamma_{r,\eta}$ be the curve in (5.16) with $\eta \in (\pi/2, \pi)$ and $r > 0$. Since the real values of η and r are not crucial for the proof, we may suppose $r > 1$. Using the decomposition of S we obtain:

$$\begin{aligned} w_t^n(h) &= \frac{1}{n} \sum_{j=1}^n \int_0^t [\nabla S_{t-s}h](x_s^{j,n}) dB_s^j = \\ &= \frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \left[\nabla \int_{\gamma_{r,\eta}} e^{(t-s)\lambda} R(\lambda, \frac{\Delta}{2}) h d\lambda \right] (x_s^{j,n}) dB_s^j = \\ &= \frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \int_{\gamma_{r,\eta}} e^{(t-s)\lambda} [\nabla R(\lambda, \frac{\Delta}{2})h](x_s^{j,n}) d\lambda dB_s^j = \\ &= Z_t^1(h) + Z_t^2(h) + Z_t^3(h), \end{aligned}$$

where in the third step we have used that ∇ is a closed linear operator on $D(\frac{\Delta}{2})$ and with

$$\begin{aligned} Z_t^1(h) &:= \frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \int_r^\infty e^{(t-s)\rho e^{i\eta}} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h](x_s^{j,n}) e^{i\eta} d\rho dB_s^j, \\ Z_t^2(h) &:= \frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \int_{-\eta}^\eta e^{(t-s)r e^{i\alpha}} [\nabla R(r e^{i\alpha}, \frac{\Delta}{2})h](x_s^{j,n}) i r e^{i\alpha} d\alpha dB_s^j, \\ Z_t^3(h) &:= -\frac{1}{2\pi i n} \sum_{j=1}^n \int_0^t \int_r^\infty e^{(t-s)\rho e^{-i\eta}} [\nabla R(\rho e^{-i\eta}, \frac{\Delta}{2})h](x_s^{j,n}) e^{-i\eta} d\rho dB_s^j. \end{aligned} \quad (5.18)$$

Using the classical estimate $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, it follows that

$$|w_t^n(h)|^2 \leq 3 \left[|Z_t^1(h)|^2 + |Z_t^2(h)|^2 + |Z_t^3(h)|^2 \right].$$

We focus on $Z_t^1(h)$, but similar estimates for $Z_t^2(h)$ and $Z_t^3(h)$ follow in exactly the same way.

Fix $\epsilon > 0$ small, the stochastic Fubini theorem (e.g. [32, §4.5]) and Cauchy-Schwartz inequality imply that

$$\begin{aligned} |Z_t^1(h)|^2 &= \left| \int_r^\infty \left[\int_0^t \frac{\rho^{\frac{1+\epsilon}{2}}}{2\pi i n} \sum_{j=1}^n e^{(t-s)\rho e^{i\eta}} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h] (x_s^{j,n}) e^{i\eta} dB_s^j \right] \frac{d\rho}{\rho^{\frac{1+\epsilon}{2}}} \right|^2 \leq \\ &\leq C \int_r^\infty \left| \int_0^t \frac{1}{n} \sum_{j=1}^n e^{(t-s)\rho e^{i\eta}} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h] (x_s^{j,n}) e^{i\eta} dB_s^j \right|^2 \rho^{1+\epsilon} d\rho \\ &= \frac{C}{n^2} \int_r^\infty e^{-2t\rho(-\cos \eta)} \underbrace{\left| \int_0^t \sum_{j=1}^n e^{-s\rho e^{i\eta}} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h] (x_s^{j,n}) dB_s^j \right|^2}_{=: M_t} \rho^{1+\epsilon} d\rho \end{aligned}$$

where $C = \frac{1}{4\pi^2} \int_r^\infty \frac{d\rho}{\rho^{1+\epsilon}}$.

We introduce the continuous martingale $X_t^{\epsilon, \rho}(h)$ defined for $t \geq 0$ by

$$X_t^{\epsilon, \rho}(h) := \rho^{1/2+\epsilon} \sqrt{\frac{-2\rho \cos \eta}{\|h\|_m^2 n}} M_t,$$

so to obtain

$$\begin{aligned} |Z_t^1(h)|^2 &\leq \frac{C}{-2n \cos \eta} \|h\|_m^2 \int_r^\infty e^{2t\rho \cos \eta} |X_t^{\epsilon, \rho}(h)|^2 \frac{d\rho}{\rho^{1+\epsilon}} \\ &\leq \frac{C}{n} \|h\|_m^2 \int_r^\infty e^{2t\rho \cos \eta} |X_t^{\epsilon, \rho}(h)|^2 \frac{d\rho}{\rho^{1+\epsilon}}. \end{aligned}$$

where we absorbed the factor $(-2 \cos \eta)^{-1}$ in the unessential constant C .

We compute the quadratic variation of $X_t^{\epsilon, \rho}(h)$:

$$\langle X^{\epsilon, \rho}(h) \rangle_t = \rho^{1+2\epsilon} \frac{(-2\rho \cos \eta)}{\|h\|_m^2 n} \sum_{j=1}^n \int_0^t e^{-2s\rho \cos \eta} [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h]^2 (x_s^{j,n}) ds.$$

Lemma 5.A.2 assures that for every ϵ such that $0 < 2\epsilon < (m - d/2) \wedge 1$, $\mathcal{P}(\mathbb{R}^d)$ is continuously embedded in $H^{-m+2\epsilon}$, in particular

$$\begin{aligned} \left| [\nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h] (x_s^{j,n}) \right| &= \left| \langle \delta_{x_s^{j,n}}, \nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h \rangle_{-m, m} \right| = \\ &= \left| \langle \delta_{x_s^{j,n}}, \nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h \rangle_{-m+2\epsilon, m-2\epsilon} \right| \leq \\ &\leq \left\| \delta_{x_s^{j,n}} \right\|_{-m+2\epsilon} \left\| \nabla R(\rho e^{i\eta}, \frac{\Delta}{2})h \right\|_{m-2\epsilon} \leq C \frac{\|h\|_m}{\rho^{1/2+\epsilon}}, \end{aligned} \tag{5.19}$$

where we have exploited the properties of the resolvent operator R , see Lemma 5.A.3.

Thus, the quadratic variation of $X_t^{\epsilon, \rho}(h)$ is bounded \mathbb{P} -a.s. by

$$\langle X^{\epsilon, \rho}(h) \rangle_t \leq C(-2\rho \cos \eta) \int_0^t e^{-2s\rho \cos \eta} ds = C(e^{-2t\rho \cos \eta} - 1). \quad (5.20)$$

Observe then

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} |Z_t^1(h)|^2 \right] &\leq \frac{C}{n} \|h\|_m^2 \int_r^\infty \mathbb{E} \left[\sup_{t \in [0, T]} e^{2t\rho \cos \eta} |X_t^{\epsilon, \rho}(h)|^2 \right] \frac{d\rho}{\rho^{1+\epsilon}} \leq \\ &\leq \frac{C}{n} \|h\|_m^2 \int_r^\infty \mathbb{E} \left[\sup_{t \in [0, T]} \frac{|X_t^{\epsilon, \rho}(h)|^2}{1 + \langle X^{\epsilon, \rho}(h) \rangle_t} \left(\sup_{t \in [0, T]} \frac{1 + \langle X^{\epsilon, \rho}(h) \rangle_t}{e^{-2t\rho \cos \eta}} \right) \right] \frac{d\rho}{\rho^{1+\epsilon}}. \end{aligned}$$

The term $\sup_{t \in [0, T]} \frac{1 + \langle X^{\epsilon, \rho}(h) \rangle_t}{e^{-2t\rho \cos \eta}}$ is bounded using (5.20) by a constant, wherefore we are left with

$$\mathbb{E} \left[\sup_{t \in [0, T]} |Z_t^1(h)|^2 \right] \leq \frac{C}{n} \|h\|_m^2 \int_r^\infty \mathbb{E} \left[\sup_{t \in [0, T]} \frac{|X_t^{\epsilon, \rho}(h)|^2}{1 + \langle X^{\epsilon, \rho}(h) \rangle_t} \right] \frac{d\rho}{\rho^{1+\epsilon}}.$$

We now invoke Lemma 3.3, which in conjunction with (5.20) allows to deduce that

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} \frac{|X_t^{\epsilon, \rho}(h)|^2}{1 + \langle X^{\epsilon, \rho}(h) \rangle_t} \right] &\leq C \mathbb{E} [\log(1 + \log(1 + \langle X^{\epsilon, \rho}(h) \rangle_T))] \\ &\leq C \mathbb{E} [\log(1 - 2T\rho \cos \eta + \log(C))] \end{aligned}$$

where in the last inequality, we have bounded the constant C appearing in (5.20) by $\max\{1, C\}$.

Further modifying C accordingly, we are thus left with

$$\mathbb{E} \left[\sup_{t \in [0, T]} |Z_t^1(h)|^2 \right] \leq \frac{C}{n} \|h\|_m^2 \int_r^\infty \log(1 - 2T\rho \cos \eta + \log(C)) \frac{d\rho}{\rho^{1+\epsilon}} \leq \frac{C}{n} \|h\|_m^2.$$

Concerning $Z_t^3(h)$, computations are the same if one replaces η by $-\eta$. Concerning $Z_t^2(h)$, computations are easier since there is no a priori diverging integral to deal with and we omit the proof. The overall bound on $w_t^n(h)$ is thus obtained by summing the three estimates and choosing the constant C accordingly. \square

REMARK 3.7. Note that Lemma 3.6 implies by Jensen's inequality the following bound

$$\mathbb{E} \left[\sup_{t \in [0, T]} |w_t^n(h)| \right] \leq \frac{C}{\sqrt{n}} \|h\|_m,$$

which is sharper in n with respect to (5.15), but in a weaker topology. One could ask if it is possible to establish a similar $O(1/\sqrt{n})$ bound for

$$\mathbb{E} \left[\sup_{t \in [0, T]} \|w_t^n\|_{-m} \right] = \mathbb{E} \left[\sup_{t \in [0, T]} \sup_{\|h\|_m \leq 1} |w_t^n(h)| \right].$$

Such a bound cannot be obtained by rough paths theory and a full probabilistic proof, which takes the independence between the Brownian motions into account, is desirable. To the authors' knowledge, this has been established only in the case of interacting oscillators; we refer to the noise term analysis in [10, 30, 78].

3.3. Proving Theorem 2.3. The proof of Theorem 2.3 consists in two steps: using the pathwise bound on w^n , Lemma 3.8 shows that we can extract from $(\nu^n)_{n \in \mathbb{N}}$ a weak- $*$ -convergence subsequence; then, by exploiting the probability bound on $w_t^n(h)$ for a fixed $h \in H^m$, we identify through Lemma 3.9 the limit with a solution to (5.6).

3.3.1. *Extraction of a weak- $*$ -convergent subsequence.* The main result of this subsection is given by the next lemma.

LEMMA 3.8. *The sequence $(\nu^n)_{n \in \mathbb{N}}$ is uniformly bounded in $L^\infty([0, T], H^{-m})$ and thus admits a subsequence that converges weak- $*$ to some $\nu \in H^{-m}$, \mathbb{P} -a.s..*

PROOF. It suffices to show that $(\nu^n)_{n \in \mathbb{N}}$ is uniformly bounded in $L^\infty([0, T], H^{-m})$ \mathbb{P} -a.s., an application of Banach-Alaoglu yields the existence of a convergent subsequence.

Exploiting the mild formulation in Lemma 3.1 and the bound on $w_t^n(h)$ in Lemma 3.2 for some $\alpha \in (1/3, 1/2)$, one obtains that

$$\begin{aligned} \|\nu_t^n\|_{-m} &\leq \|\nu_0^n\|_{-m} + \int_0^t \|\nu_s^n\|_{-m} \sup_{\|h\|_m \leq 1} \|(\nabla S_{t-s}h)(\Gamma * \nu_s^n)\|_m \, ds + \|w_t^n\|_{-m} \\ &\leq \|\nu_0^n\|_{-m} + \int_0^t \frac{C}{\sqrt{t-s}} \|\nu_s^n\|_{-m} \, ds + C_\alpha(1+t)^{3\alpha}, \end{aligned}$$

where we have exploited the properties of the semigroup and the bound already used in (5.8). A Gronwall-like argument implies the existence of a constant a independent of n and T such that

$$\sup_{t \in [0, T]} \|\nu_t^n\|_{-m} \leq 2 \left(\|\nu_0^n\|_{-m} + C_\alpha(1+T)^{3\alpha} \right) \sqrt{T} e^{a\sqrt{T}}.$$

In particular, using Lemma 5.A.2, we conclude

$$\sup_{n \in \mathbb{N}} \sup_{t \in [0, T]} \|\nu_t^n\|_{-m} \leq C_{\alpha, T}.$$

□

We move to the identification of the limit $\nu \in L^\infty([0, T], H^{-m})$.

3.3.2. *The limit coincides with an m -weak-mild solution.* We prove that any possible limit of $(\nu^n)_{n \in \mathbb{N}}$ is a weak-mild solution (5.6). Given the uniqueness of (5.6), this implies the weak- $*$ convergence in $L^\infty([0, T], H^{-m})$ of $(\nu^n)_{n \in \mathbb{N}}$ to the element ν given in Lemma 3.8.

LEMMA 3.9. *Let $(\nu^n)_{n \in \mathbb{N}}$ be converging weak- $*$ to some $\bar{\nu} \in L^\infty([0, T], H^{-m})$ \mathbb{P} -a.s. along a subsequence that we denote by $(\nu^{n_k})_{k \in \mathbb{N}}$. Then $\bar{\nu}$ satisfies (5.6), i.e.*

$$\langle \bar{\nu}_t, h \rangle_{-m, m} = \langle \bar{\nu}_0, S_t h \rangle_{-m, m} + \int_0^t \langle \bar{\nu}_s, (\nabla S_{t-s} h)(\Gamma * \bar{\nu}_s) \rangle_{-m, m} ds,$$

meaning $\bar{\nu}$ is an m -weak-mild solution to (5.3).

PROOF. Recall that for every n , ν^n solves the mild formulation (5.10), i.e. for $t \in [0, T]$

$$\langle \nu_t^n, h \rangle_{-m, m} = \langle \nu_0^n, S_t h \rangle_{-m, m} + \int_0^t \langle \nu_s^n, (\nabla S_{t-s} h)(\Gamma * \nu_s^n) \rangle_{-m, m} ds + w_t^n(h).$$

By hypothesis we have that for every $t \in [0, T]$ and $h \in H^m$

$$\lim_{k \rightarrow \infty} \langle \nu_t^{n_k}, h \rangle_{-m, m} = \langle \bar{\nu}_t, h \rangle_{-m, m}, \quad \mathbb{P}\text{-a.s.}$$

In particular, this is true for $(\nu_0^{n_k})_k$ since $S_t h \in H^m$. Furthermore, Lemma 3.6 implies that

$$\lim_{k \rightarrow \infty} w_t^{n_k}(h) = 0, \quad \text{in } \mathbb{P}\text{-probability}$$

and thus in particular the convergence holds \mathbb{P} -a.s. along a sub-subsequence $(n_{k_j})_j$. Thus, it remains to show that \mathbb{P} -a.s.

$$\lim_{j \rightarrow \infty} \int_0^t \langle \nu_s^{n_{k_j}}, (\nabla S_{t-s} h)(\Gamma * \nu_s^{n_{k_j}}) \rangle_{-m, m} ds = \int_0^t \langle \bar{\nu}_s, (\nabla S_{t-s} h)(\Gamma * \bar{\nu}_s) \rangle_{-m, m} ds. \quad (5.21)$$

For better readability and lighter notation, we will not distinguish between n and n_{k_j} in the following, understanding that we continue to work on the sub-subsequence. Consider then

$$\begin{aligned} & \langle \bar{\nu}_s^n, (\nabla S_{t-s} h)(\Gamma * \bar{\nu}_s^n) \rangle_{-m, m} - \langle \bar{\nu}_s, (\nabla S_{t-s} h)(\Gamma * \bar{\nu}_s) \rangle_{-m, m} = \\ & = \langle \bar{\nu}_s^n - \bar{\nu}_s, (\nabla S_{t-s} h)(\Gamma * \bar{\nu}_s) \rangle_{-m, m} + \langle \bar{\nu}_s^n, (\nabla S_{t-s} h)(\Gamma * (\bar{\nu}_s^n - \bar{\nu}_s)) \rangle_{-m, m}. \end{aligned}$$

Using again (5.8), it is easy to see that \mathbb{P} -a.s.

$$\mathbf{1}_{[0, t]}(s)(\nabla S_{t-s} h)(\Gamma * \nu_s) \in L^1([0, T], H^m)$$

wherefore it is indeed an element of the predual to $L^\infty([0, T], H^{-m})$ and thus by weak- $*$ convergence in this space, we have

$$\int_0^T \langle \nu_s^n - \nu_s, \mathbf{1}_{[0, t]}(s)(\nabla S_{t-s} h)(\Gamma * \nu_s) \rangle ds = \int_0^t \langle \nu_s^n - \nu_s, (\nabla S_{t-s} h)(\Gamma * \nu_s) \rangle ds \rightarrow 0.$$

For the second term, note that

$$\begin{aligned} & \langle \bar{\nu}_s^n, (\nabla S_{t-s} h)(\Gamma * (\bar{\nu}_s^n - \bar{\nu}_s)) \rangle_{-m, m} = \\ & = \langle \bar{\nu}_s^n - \bar{\nu}_s, \langle \bar{\nu}_s^n(dx), (\nabla S_{t-s} h)(x)(\Gamma(x, \cdot)) \rangle_{-m, m} \rangle_{-m, m} \end{aligned}$$

and that the function

$$\begin{aligned} y &\mapsto \langle \bar{\nu}_s^n(\mathrm{d}x), (\nabla S_{t-s}h)(x)(\Gamma(x, y)) \rangle_{-m, m} = \\ &= \frac{1}{n} \sum_{j=1}^n (\nabla S_{t-s}h(x_s^{j, n})) \cdot (\Gamma(x_s^{j, n}, y)) \end{aligned}$$

is in H^m for every $s \in [0, t]$, since $\Gamma(x, \cdot) \in H^m$ for a.e. $x \in \mathbb{R}^d$, recall (5.7). Namely,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{j=1}^n (\nabla S_{t-s}h(x_s^{j, n})) \cdot (\Gamma(x_s^{j, n}, \cdot)) \right\|_m &\leq \|\nabla S_{t-s}h\|_\infty \|\Gamma\|_{H_y^m L_x^\infty} \\ &\leq C \frac{\|h\|_m}{\sqrt{t-s}} \|\Gamma\|_{H_y^m W_x^{m, \infty}}. \end{aligned}$$

We conclude that

$$\mathbf{1}_{[0, t]} \langle \bar{\nu}_s^n(\mathrm{d}x), (\nabla S_{t-s}h)(x)(\Gamma(x, \cdot)) \rangle_{-m, m} \in L^1([0, T], H^m)$$

and in particular

$$\int_0^T \langle \bar{\nu}_s^n - \bar{\nu}_s, \langle \bar{\nu}_s^n(\mathrm{d}x), \mathbf{1}_{[0, t]}(s) (\nabla S_{t-s}h)(x) \cdot (\Gamma(x, \cdot)) \rangle_{-m, m} \rangle_{-m, m} \mathrm{d}s \rightarrow 0.$$

This establishes (5.21).

Overall, we have thus shown that any subsequence of $(\nu^n)_n$ converges along some further subsequence \mathbb{P} -a.s. weak* in $L^\infty([0, T], H^{-m})$, the limit $\bar{\nu}$ satisfying for every $h \in H^m$ the equation

$$\langle \bar{\nu}_t, h \rangle_{-m, m} = \langle \bar{\nu}_0, S_t h \rangle_{-m, m} + \int_0^t \langle \bar{\nu}_s, (\nabla S_{t-s}h)(\Gamma * \bar{\nu}_s) \rangle_{-m, m} \mathrm{d}s,$$

meaning that $\bar{\nu}$ is indeed an m -weak-mild solution. \square

3.3.3. Proof of Theorem 2.3. In order to show that $\nu^n \xrightarrow{*} \nu$ in $L^\infty([0, T], H^{-m})$ in probability, we show that any subsequence $(\nu^{n_k})_k$ admits a further subsequence that converges \mathbb{P} -a.s. in weak-* topology of $L^\infty([0, T], H^{-m})$ to ν .

Let $(\nu^{n_k})_k$ be hence a subsequence. By assumption of the Theorem, Lemmas 3.6 and 3.8, we find a further subsequence $(\nu^{n_{k_j}})_j$, along which

$$\begin{aligned} w_t^{n_{k_j}}(h) &\rightarrow 0 \quad \forall h \in H^m \\ \langle \nu_0^{n_{k_j}}, h \rangle &\rightarrow \langle \nu_0, h \rangle \quad \forall h \in H^m \\ \nu^{n_{k_j}} &\xrightarrow{*} \bar{\nu} \quad \text{in } L^\infty([0, T], H^{-m}) \end{aligned} \tag{5.22}$$

\mathbb{P} -a.s., where the limit $\bar{\nu} \in L^\infty([0, T], H^{-m})$ may a priori depend on the subsequence chosen. Notice however that due to Lemma 3.9, any such limit is a m -weak-mild solution to (5.3). By the uniqueness result of Proposition 2.2, we conclude that the limit $\bar{\nu} = \nu$ must be the same for any subsequence chosen.

The first part of the Theorem is proved. Note that a priori, our limit ν is only a distribution in H^{-m} at each fixed timepoint.

Suppose $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$. In order to show that ν_t is actually a probability measure for each $t \in [0, T]$, we observe that a weak solution $\mu \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ to (5.3) (which exists due to Theorem 1.2) is a weak-mild solution (5.6).

Indeed, let $\mu = (\mu_t)_{t \in [0, T]} \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ be a weak solution to (5.3). As done in Lemma 3.1, one can show that for every $f \in C_0^\infty$ and $t \in [0, T]$

$$\langle \mu_t, f \rangle_{-m, m} = \langle \mu_0, S_t f \rangle_{-m, m} + \int_0^t \langle \mu_s, (\nabla S_{t-s} f) \cdot (\Gamma * \mu_s) \rangle_{-m, m} ds \quad (5.23)$$

holds. Note that by standard approximation, (5.23) holds also for $f \in H^m \subset C_b^3$, meaning that μ is indeed a weak-mild solution. By the uniqueness statement of Proposition 2.2 we conclude $\mu = \nu$ and thus in particular $\nu \in C([0, T], \mathcal{P}(\mathbb{R}^d))$. This concludes the second part and thus the entire proof of the Theorem.

5.A. Hilbert spaces and semigroups

5.A.0.1. *The Laplacian semigroup.* The following definitions are taken from [58, 75]. For the sake of notation, we focus on Δ , the standard Laplacian on $L^2(\mathbb{R}^d)$, instead of $\frac{\Delta}{2}$. We can consider the part of Δ on (the complexification e.g. [75, Appendix A] of) H^m :

$$\Delta : \mathcal{D}(\Delta) \subset H^m \longrightarrow H^m.$$

It is not difficult to see that $H^{m+2} \subset \mathcal{D}(\Delta)$, where the inclusion is dense, and that Δ is a sectorial operator with spectrum given by $(-\infty, 0]$. In particular, it generates an analytic strongly continuous semigroup denoted for all $t \geq 0$ by S_t ; recall that $S_0 := \text{Id}$ is the identity operator.

We represent S for $t \in [0, T]$ as the following Dunford integral

$$S_t = \frac{1}{2\pi i} \int_{\gamma_{r, \eta}} e^{t\lambda} R(\lambda, \Delta) d\lambda,$$

where $R(\lambda, \Delta) = (\lambda \text{Id} - \Delta)^{-1}$ denotes the resolvent of Δ and where, for $r > 0$ and $\eta \in (\pi/2, \pi)$, $\gamma_{r, \eta}$ is the curve $\{\lambda \in \mathbb{C} : |\arg \lambda| = \eta, |\lambda| \geq r\} \cup \{\lambda \in \mathbb{C} : |\arg \lambda| \leq \eta, |\lambda| = r\}$, oriented counterclockwise.

Observe that $\gamma_{r, \eta}$ is contained in the resolvent set of Δ , i.e. $\gamma_{r, \eta} \subset \rho(\Delta)$, and that, for all regular values $\lambda \in \rho(\Delta)$, $R(\lambda, \Delta)$ is a bounded linear operator on H^m .

When computing the semigroup against a function h through (5.16), we use the following decomposition into three real integrals:

$$\begin{aligned} S_t h &= \frac{1}{2\pi i} \int_{\gamma_{r,\eta}} e^{t\lambda} R(\lambda, \Delta) h d\lambda = \frac{1}{2\pi i} \left[\int_r^\infty e^{t\rho e^{i\eta}} R(\rho e^{i\eta}, \Delta) e^{i\eta} d\rho + \right. \\ &\quad \left. + \int_{-\eta}^\eta e^{tr(\cos\alpha + i\sin\alpha)} R(re^{i\alpha}, \Delta) i r e^{i\alpha} d\alpha - \int_r^\infty e^{t\rho e^{-i\eta}} R(\rho e^{-i\eta}, \Delta) e^{-i\eta} d\rho \right]. \end{aligned} \quad (5.24)$$

The section ends with some estimates concerning the regularity of S .

LEMMA 5.A.1. *Assume $m > d/2 + 3$. Let $(S_t)_{t \leq T}$ be the heat semigroup acting on H^m . For $f \in H^m$, it holds that*

$$\begin{aligned} \|\nabla(S_t - \text{Id})f\|_\infty &\leq \sqrt{t} \|D^2 f\|_\infty \leq C \sqrt{t} \|f\|_m, \\ \|\nabla(S_t - \text{Id})f\|_\infty &\leq \frac{1}{2} t \|D^3 f\|_\infty \leq \frac{C}{2} t \|f\|_m, \end{aligned}$$

where D^2 is the Hessian and D^3 the tensor with third-order derivatives. In particular

$$\begin{aligned} \|\nabla(S_t - S_s)f\|_\infty &\leq \sqrt{|t-s|} \|D^2 f\|_\infty \leq C \sqrt{|t-s|} \|f\|_m, \\ \|\nabla(S_t - S_s)f\|_\infty &\leq \frac{1}{2} |t-s| \|D^3 f\|_\infty \leq \frac{C}{2} |t-s| \|f\|_m. \end{aligned}$$

PROOF. We calculate explicitly

$$\begin{aligned} |\nabla(S_t - \text{Id})f(x)| &= \left| \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} e^{-\frac{|y|^2}{2t}} (\nabla f(x-y) - \nabla f(x)) dy \right| \\ &= \left| \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} e^{-\frac{|y|^2}{2t}} \left(\nabla f(x-y) - \nabla f(x) + \frac{1}{2} (-y)^T (D^2 \nabla f)(x) \right) dy \right| \\ &\leq \frac{1}{2} \|D^3 f\|_\infty \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} e^{-\frac{|y|^2}{2t}} |y|^2 dy \\ &= \frac{1}{2} \|D^3 f\|_\infty t \end{aligned}$$

where we exploited the asymmetry of the first Taylor component. The first statement follows from a similar consideration, considering an order one Taylor expansion of $\nabla f(x-y)$ around x instead of an order two Taylor expansion. The proof follows by Sobolev's embeddings. \square

5.A.0.2. *The Hilbert space H^s .* It is useful to give an explicit definition of H^m through the Fourier transform (e.g. [2, 7.62]). Let $s > 0$, define $(H^s, \|\cdot\|_s)$ by

$$\begin{aligned} H^s &= \left\{ u \in L^2(\mathbb{R}^d) : \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\mathcal{F}(u)(\xi)|^2 d\xi < \infty \right\}, \\ \|u\|_s^2 &= \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\mathcal{F}(u)(\xi)|^2 d\xi. \end{aligned} \quad (5.25)$$

Whenever s is an integer, it is well known that this definition coincides with the standard definition of the Sobolev space $W^{s,2}(\mathbb{R}^d)$.

The next lemma extends the embedding (5.5) to H^s and its relationship with the space of probability measures.

LEMMA 5.A.2. *For all $s > d/2$, one has the following continuous embedding*

$$H^s \subset C_b(\mathbb{R}^d). \quad (5.26)$$

Moreover, there exists $C > 0$ (depending on s only) such that

$$\sup_{\mu \in \mathcal{P}(\mathbb{R}^d)} \|\mu\|_{-s} \leq C. \quad (5.27)$$

PROOF. The continuous embedding (5.26) is a consequence of the embedding of Besov spaces into the space of continuous bounded functions (e.g. [2, Theorem 7.34]) and the fact that they coincide with H_s for a particular choice of the indices.

Turning to (5.27), let $\mu \in \mathcal{P}(\mathbb{R}^d)$, then

$$\|\mu\|_{-s} = \sup_{h \in H^s} \frac{\langle \mu, h \rangle_{-s,s}}{\|h\|_s} = \sup_{h \in H^s} \frac{\langle \mu, h \rangle}{\|h\|_s} \leq \sup_{h \in H^s} \frac{\|h\|_\infty}{\|h\|_s} \leq C,$$

where C is the norm of the identity operator between H^s and $C_b(\mathbb{R}^d)$. \square

5.A.0.3. *Fractional operators on H^s .* We have the following lemma.

LEMMA 5.A.3. *Let $\lambda = \rho e^{i\eta} \in \rho(\Delta)$ and suppose $\rho > 1$. There exists a positive constant $C = C_\eta$ such that for every $\epsilon \in (0, 1/2)$*

$$\|\nabla R(\rho e^{i\eta}, \Delta)h\|_{m-2\epsilon}^2 \leq C_\eta \frac{\|h\|_m^2}{\rho^{1+2\epsilon}}, \quad h \in H^m. \quad (5.28)$$

PROOF. Exploiting the Fourier multipliers associated to ∇ and R , one obtains

$$\begin{aligned} \|\nabla R(\rho e^{i\eta}, \Delta)h\|_{m-2\epsilon}^2 &= \int_{\mathbb{R}^d} (1 + |\xi|^2)^{m-2\epsilon} |\mathcal{F}(\nabla R(\rho e^{i\eta}, \Delta)h)(\xi)|^2 d\xi \leq \\ &\leq \int_{\mathbb{R}^d} (1 + |\xi|^2)^{m-2\epsilon} |\mathcal{F}(h)(\xi)|^2 \left| \frac{\xi}{\rho e^{i\eta} + |\xi|^2} \right|^2 d\xi \leq \\ &\leq \int_{\mathbb{R}^d} (1 + |\xi|^2)^m |\mathcal{F}(h)(\xi)|^2 \frac{|\xi|^2}{|\rho e^{i\eta} + |\xi|^2|^2} \frac{1}{(1 + |\xi|^2)^{2\epsilon}} d\xi. \end{aligned}$$

Since we are assuming $\rho > 1$, we have that

$$\frac{|\xi|^2}{|\rho e^{i\eta} + |\xi|^2|^2} \frac{1}{(1 + |\xi|^2)^{2\epsilon}} \leq \frac{(\rho + |\xi|^2)^{1-2\epsilon}}{|\rho e^{i\eta} + |\xi|^2|^2} \leq C_\eta \frac{1}{(\rho + |\xi|^2)^{1+2\epsilon}} \leq C_\eta \frac{1}{\rho^{1+2\epsilon}},$$

with $C_\eta = (\sup_{x \geq 0} (1+x)/|e^{i\eta} + x|)^2 < \infty$ since $\eta \neq \pi$. \square

5.B. Rough integration associated to semigroup functionals

We mostly follow [54] and use very similar notations. Let $k \geq 1$ and Δ_k be the k -dimensional simplex given by

$$\Delta_k = \{t_1, \dots, t_k \in [0, T] : T \geq t_1 \geq t_2 \geq \dots \geq t_k \geq 0\}.$$

Let $C_k = C(\Delta_k; \mathbb{R})$ and W a Banach space with a strongly continuous semigroup $(S_t)_{t \in [0, T]}$ acting on it. Define D_k as the space of linear operators from W to C_k . Furthermore, let $D_* = \bigcup_{k \geq 1} D_k$ and define the following operators on D_* :

$$\delta : D_k \rightarrow D_{k+1}, \quad \phi : D_k \rightarrow D_{k+1}, \quad k \geq 1.$$

For $A \in D_k$ and $f \in W$, they are defined as

$$\begin{aligned} [\delta A f]_{t_1 \dots t_{k+1}} &= \sum_{i=1}^{k+1} (-1)^{i+1} [A f]_{t_1 \dots \cancel{t_i} \dots t_{k+1}}, \\ [\phi A f]_{t_1 \dots t_{k+1}} &= [A(S_{t_1 t_2} - \text{Id})f]_{t_2 \dots t_{k+1}}, \end{aligned}$$

where $\cancel{t_i}$ means that the argument t_i is omitted and $S_{t_1 t_2}$ stands for $S_{t_1 - t_2}$.

We are ready for the first lemma.

LEMMA 5.B.1. *Let $\hat{\delta} := \delta - \phi$. Then $(D_*, \hat{\delta})$ is an acyclic cochain complex. In particular*

$$\text{Ker } \hat{\delta}|_{D_{k+1}} = \text{Im } \hat{\delta}|_{D_k}, \quad \text{for any } k \geq 1.$$

PROOF. The proof mimics [54, Proposition 3.1]. We only mention that for proving $\text{Ker } \hat{\delta}|_{D_{k+1}} \subset \text{Im } \hat{\delta}|_{D_k}$, a possible choice for $A \in \text{Ker } \hat{\delta}|_{D_{k+1}}$ is given by $B \in D_k$ defined as

$$[B f]_{t_1 \dots t_k} = (-1)^{k+1} [A f]_{t_1 \dots t_k 0}, \quad f \in W.$$

□

We now introduce some analytical assumptions on the previous function spaces. We start with a Hölder-like norm on C_k for $k = 2, 3$. For $\mu > 0$ and $g \in C_2$, define

$$\|g\|_\mu := \sup_{t, s \in \Delta_2} \frac{|g_{ts}|}{|t - s|^\mu},$$

and consequently

$$C_2^\mu := \left\{ g \in C_2 ; \|g\|_\mu < \infty \right\}.$$

For $\gamma, \rho > 0$ and $g \in C_3$, define

$$\|g\|_{\gamma, \rho} := \sup_{t, u, s \in \Delta_3} \frac{|g_{tus}|}{|t - u|^\mu |u - s|^\rho}$$

and

$$\|g\|_\mu := \inf \left\{ \sum_i \|g_i\|_{\rho_i, \mu - \rho_i} ; g = \sum_i g_i, 0 < \rho_i < \mu \right\},$$

where the infimum is taken on all sequences $(g_i) \subset C_3$ such that $g = \sum_i g_i$ and $\rho_i \in (0, \mu)$. Again, $\|\cdot\|_\mu$ defines a norm on C_3 and we denote the induced subspace by

$$C_3^\mu := \left\{ g \in C_3; \|g\|_\mu < \infty \right\}.$$

With these definitions in mind, let

$$D_k^\mu := L(W, C_k^\mu), \quad D_k^{1+} := \bigcup_{\mu > 1} D_k^\mu, \quad k = 2, 3.$$

The space $L(W, C_k^\mu)$ is the space of linear bounded operators from W to C_k^μ equipped with its corresponding operator norm, i.e.

$$\|A\|_{D_k^\mu} := \sup_{\|f\|_W \leq 1} \|Af\|_\mu, \quad f \in D_k^\mu.$$

The main tool for constructing the pathwise integral associated to semigroup functionals is given by the next lemma and the following corollary. We use the notation $\hat{\delta}(D_k) := \text{Im } \hat{\delta}|_{D_k}$ for $k \geq 1$.

LEMMA 5.B.2 (Sewing). *There exists a unique linear operator*

$$\Lambda : D_3^{1+} \cap \hat{\delta}(D_2) \rightarrow D_2^{1+}$$

such that

$$\hat{\delta}\Lambda = \text{Id}|_{D_3^{1+} \cap \hat{\delta}(D_2)}.$$

Moreover, if $\eta > 1$ then Λ is a continuous operator from $D_3^\eta \cap \hat{\delta}(D_2)$ to D_2^η , i.e. there exists a constant $C = C_\eta > 0$ such that

$$\|\Lambda A\|_\eta \leq C_\eta \|A\|_\eta, \quad A \in D_3^\eta \cap \hat{\delta}(D_2). \quad (5.29)$$

PROOF. Concerning uniqueness, let $\tilde{\Lambda}$ be another map satisfying the conditions stated in the Lemma. Then for $A \in D_3^{1+} \cap \hat{\delta}D_2$ we have

$$\hat{\delta}(\tilde{\Lambda}A - \Lambda A) = A - A = 0$$

hence $Q := \tilde{\Lambda}A - \Lambda A \in \text{Ker}(\hat{\delta}) \cap D_2$. By Lemma 5.B.1 there exists $q \in D_1$ such that $Q = \hat{\delta}q$. Note that for any partition $\mathcal{P}^n([s, t]) = (t_i)_{0 \leq i \leq n+1}$ of the interval $[s, t] \subset [0, T]$ such that $t_0 = s$ and $t_{n+1} = t$, we have the following telescopic sum expansion

$$\begin{aligned} \sum_{i=0}^n [\hat{\delta}q S_{tt_{i+1}} f]_{t_{i+1}t_i} &= \sum_{i=0}^n [q S_{tt_{i+1}} f]_{t_{i+1}} - [q S_{tt_{i+1}}]_{t_i} - [q S_{tt_{i+1}} (S_{t_{i+1}t_i} - \text{Id}) f]_{t_i} \\ &= \sum_{i=0}^n [q S_{tt_{i+1}} f]_{t_{i+1}} - [q S_{tt_i} f]_{t_i} \\ &= [q f]_t - [q S_{ts} f]_s \\ &= [\hat{\delta}q f]_{ts} \end{aligned}$$

for any $f \in W$. We conclude

$$[Qf]_{ts} = [\hat{\delta}qf]_{ts} = \sum_{i=0}^n [\hat{\delta}qS_{tt_{i+1}}f]_{t_{i+1}t_i} = \sum_{i=0}^n [QS_{tt_{i+1}}f]_{t_{i+1}t_i}.$$

Letting $\mathcal{P}^n([s, t])$ be for example be the dyadic partition, one obtains for $Q \in D_2^\gamma$, with $\gamma > 1$, the estimate

$$|[QS_{tt_{i+1}}f]_{t_{i+1}t_i}| \leq 2^{-n\gamma} \|QS_{tt_{i+1}}f\|_\gamma |t - s|^\gamma \leq 2^{-n\gamma} \|Q\|_{D_2^\gamma} \|f\|_W |t - s|^\gamma$$

where we exploited that S is a contraction semigroup. Returning to the telescope sum, we obtain

$$|[Qf]_{ts}| \leq 2^{n(1-\gamma)} \|Q\|_{D_2^\gamma} \|f\|_W |t - s|^\gamma.$$

By passing to the limit for n which tends to infinity, we conclude that for any $f \in W$ and any $[s, t] \subset [0, T]$

$$[Qf]_{ts} = 0$$

yielding $Q = 0$, i.e. $\Lambda A = \tilde{\Lambda}A$ for any $A \in D_3^{1+} \cap \hat{\delta}(D_2)$, concluding uniqueness.

Towards existence, let $A \in D_3^{1+} \cap \hat{\delta}(D_2)$, i.e. there exist a $B \in D_2$ and $\eta > 1$ such that $\hat{\delta}B = A \in D_3^\eta$. Let $(r_k^n)_{0 \leq k \leq 2^n}$ be the dyadic partition of $[s, t]$. We set, following [54]

$$\begin{aligned} M^n : D_3^{1+} \cap \hat{\delta}(D_2) &\rightarrow D_2^{1+} \\ \hat{\delta}B &\mapsto M^n \hat{\delta}B \end{aligned}$$

where

$$[(M^n \hat{\delta}B)f]_{ts} := [B(f)]_{ts} - \sum_{k=0}^{2^n-1} [B(S_{tr_{k+1}^n}f)]_{r_{k+1}^n r_k^n}.$$

Note in particular that $[M^n \hat{\delta}f]_{ts} = 0$. We show that $(M^n \hat{\delta}B)_n$ is Cauchy in D_2^η . Note that

$$\begin{aligned} &[(M^n \hat{\delta}B)f]_{ts} - [(M^{n+1} \hat{\delta}B)f]_{ts} = \\ &= \sum_{k=0}^{2^n-1} [B(S_{tr_{2k+2}^n}f)]_{r_{2k+2}^n r_{2k}^n} - [B(S_{tr_{2k+2}^n}f)]_{r_{2k+2}^n r_{2k+1}^n} - [B(S_{tr_{2k+1}^n}f)]_{r_{2k+1}^n r_{2k}^n} \\ &= \sum_{k=0}^{2^n-1} [\delta B(S_{tr_{2k+2}^n}f)]_{r_{2k+2}^n r_{2k+1}^n r_{2k}^n} - \sum_{k=0}^{2^n-1} [\phi B(S_{tr_{2k+2}^n}f)]_{r_{2k+2}^n r_{2k+1}^n r_{2k}^n} \\ &= \sum_{k=0}^{2^n-1} [\hat{\delta}B(S_{tr_{2k+2}^n}f)]_{r_{2k+2}^n r_{2k+1}^n r_{2k}^n} \leq (t-s)^\eta \sum_{k=0}^{2^n-1} \left\| \hat{\delta}B(S_{tr_{2k+2}^n}f) \right\|_\eta 2^{-n\eta} \\ &\leq (t-s)^\eta \left\| \hat{\delta}B \right\|_{D_3^\eta} \|f\|_W 2^{-n(\eta-1)} \end{aligned}$$

From which we deduce that $(M^n \hat{\delta}B)_{n \in \mathbb{N}}$ is a Cauchy sequence in D_2^η , indeed

$$\left\| M^n \hat{\delta}B - M^{n+1} \hat{\delta}B \right\|_{D_2^\eta} \leq \left\| \hat{\delta}B \right\|_{D_3^\eta} 2^{-n(\eta-1)}.$$

Let $\Lambda \hat{\delta}B \in D_2^\eta$ be its limit. By a telescope argument

$$\left\| M^n \hat{\delta}B \right\|_{D_2^\eta} = \left\| \sum_{k=0}^{n-1} M^k \hat{\delta}B - M^{k+1} \hat{\delta}B \right\|_{D_2^\eta} \leq \sum_{k=0}^{n-1} 2^{-k(\eta-1)} \left\| \hat{\delta}B \right\|_{D_3^\eta} \leq C_\eta \left\| \hat{\delta}B \right\|_{D_3^\eta}$$

from which we obtain (5.29) using to weak-* lower semicontinuity of the norm. One can prove that the limit does not depend on the particular sequence, see [54, Proposition 2.3].

Finally, let $u = 2^m$ for some $m \in 0, \dots, n$ and note that

$$\begin{aligned} [(\hat{\delta}M^n \hat{\delta}B)f]_{tus} &= [(M^n \hat{\delta}B)f]_{ts} - [(M^n \hat{\delta}B)f]_{tu} - [(M^n \hat{\delta}B)f]_{us} - [(M^n \hat{\delta}B(S_{tu} - \text{Id}))f]_{us} \\ &= [Bf]_{ts} - [Bf]_{tu} - [BS_{tu}f]_{us} + \sum_{k=0}^{2^n-1} [B(S_{tr_{k+1}^n} f)]_{r_{k+1}^n r_k^n} \\ &\quad - \sum_{k=2^m}^{2^n-1} [B(S_{tr_{k+1}^n} f)]_{r_{k+1}^n r_k^n} - \sum_{k=0}^{2^m-1} [B(S_{ur_{k+1}^n} S_{tu} f)]_{r_{k+1}^n r_k^n} \\ &= [Bf]_{ts} - [Bf]_{tu} - [BS_{tu}f]_{us} \\ &= [\hat{\delta}Bf]_{sut} \end{aligned}$$

from which we recover, in the limit $n \rightarrow \infty$, that $\hat{\delta}\Lambda = \text{Id}|_{D_3^{1+} \cap \hat{\delta}(D_2)}$. \square

COROLLARY 5.B.3. *Suppose that $A \in D_2$ is such that $\hat{\delta}A \in D_3^{1+}$. Then there exists $I \in D_1$ such that*

$$\hat{\delta}I = (\text{Id} - \Lambda \hat{\delta}) A,$$

i.e. for every $f \in W$ and $(t, s) \in \Delta_2$, $[\hat{\delta}If]_{ts} = [Af]_{ts} - [\Lambda \hat{\delta}Af]_{ts}$. In particular, if $A \in D_2^\mu$ with $\mu > 0$ and $\hat{\delta}A \in D_3^\eta$ with $\eta > 1$, then for every $f \in W$

$$\left| [\hat{\delta}If]_{ts} \right| \leq \left(\|A\|_{D_2^\mu} (t-s)^\mu + \left\| \hat{\delta}A \right\|_{D_3^\eta} (t-s)^\eta \right) \|f\|_W. \quad (5.30)$$

Finally

$$\left[\hat{\delta}If \right]_{ts} = \lim_{|\mathcal{P}[s,t]| \downarrow 0} \sum_{[v,u] \in \mathcal{P}[s,t]} [AS_{tu}f]_{uv}, \quad (5.31)$$

where the limit is over any partition of $[s, t]$ whose mesh tends to zero.

PROOF. The proof is an easy application of the Sewing Lemma and the properties of $(D_*, \hat{\delta})$. Indeed, observe that $\hat{\delta}(Af - \Lambda \hat{\delta}Af) = 0$ for any $f \in W$, which means that $A - \Lambda \hat{\delta}A \in \text{Ker } \hat{\delta}|_{D_2}$, and thus there exists $I \in D_1$ such that $\hat{\delta}I = A - \Lambda \hat{\delta}A$.

The estimate (5.30) follows from (5.29). Concerning (5.31), observe that for a partition $|\mathcal{P}[s, t]|$, using the properties of $\hat{\delta}$, one obtains

$$\begin{aligned} \sum_{[v,u] \in \mathcal{P}[s,t]} [AS_{tuv}]_{uv} &= \sum_{[v,u] \in \mathcal{P}[s,t]} \left[\hat{\delta} I S_{tuv} \right]_{uv} + \left[\Lambda \hat{\delta} A S_{tuv} \right]_{uv} = \\ &= \left[\hat{\delta} I f \right]_{ts} + \sum_{[v,u] \in \mathcal{P}[s,t]} \left[\Lambda \hat{\delta} A S_{tuv} \right]_{uv}. \end{aligned}$$

By taking the limit for the mesh which tends to zero and using the fact that $\Lambda \hat{\delta} A \in D_2^{1+}$, the last sum converges to zero. \square

CHAPTER 6

Perspectives

We shortly list some consequences and open questions arising from the previous chapters.

1. From dense graph sequences to the sparse regime

In Chapters 2 and 4, the convergence to the classical mean-field behavior is shown not only for dense graph sequences, but also for the intermediate regimes, recall definition (1.7). Under a suitable normalization of the interaction – represented by the dilution parameter p_n – it appears that what really affects the behavior of the empirical measure is the degree inhomogeneity, as defined in Remark 2.3.

Moreover, most of the works addressing the graphon setting [6, 27, 77, 82, 90] deal with sequences in the intermediate regimes and not with the dense case only. Namely, the definition of W -random graph (1.12) can be easily extended to graphs with an arbitrary diverging average degree. Indeed, for every n one can take

$$\xi_{ij}^{(n)} \sim \text{Ber}(p_n W(U_i, U_j)) \quad \text{independently for } 1 \leq i < j \leq n,$$

where $\{p_n\}_{n \in \mathbb{N}}$ is a sequence of dilution parameters. It is not difficult to see that the sequence $\xi^{(n)}$ is in the intermediate regime and has mean average degree given by np_n . We refer to [16] where this was introduced for the first time, and where it is shown that $\xi^{(n)}$, suitably normalized, converges to W .

Even though a general result for particle systems on sequences in the intermediate regimes is missing, the cited results prove that whenever the graph limit is regular enough, the limit behavior is independent of the magnitude of np_n , provided that np_n diverges. In other words, the sparse regime appears to represent the real threshold with regards to the degree asymptotic.

Particle systems on sparse graph sequences show a macroscopic behavior not readily captured by the mean-field approximation, no matter the homogeneity of the underlying graph. This framework represents an interesting domain for future investigations; we refer to the only two works available [91, 70] so far.

2. Irregular graphons and random graphons

To the author's knowledge, the joint work with Gianmarco Bet and Francesca Nardi [11] represents the first work in the literature tackling interacting particle systems and random unlabeled graphons.

An unlabeled graphon can be a very irregular function given that the only requirements are measurability and symmetry. For instance, Chapter 3 leaves the following issues open:

- What is the limit behavior of the empirical measure on a strongly irregular graphon?
- Is there any key property related to the graph limit that allows to derive some macroscopic property for the particle system, e.g., existence and uniqueness of a stable stationary solution?
- Do random graphons – and random Fokker-Planck equations – provide an interesting framework with regards to applications?

The study of neural field equations may shed some light on these aspects; see [26], [77, §2.8] and references therein for some progress in this direction.

3. Unbounded graphons

The classical graphon theory, eventually extended to include graphs in the intermediate regimes [16], has proven to be the right framework to study the convergence of empirical measures in mean-field systems. However, it comes with a strong limitation with regards to applications: it describes graph sequences where each portion of vertices has roughly the same edge density – which, following the previous notation, is given by p_n . In other words, a graph sequence converging to a non-zero graphon has most of the degrees on the same scale as the size of the graph grows.

The graph limit framework does not cover many interesting cases where different densities appear within the same sequence as, e.g., in scale-free graphs. In this last example, a part of the vertices can have a density which scales as $n^{-\alpha}$, with $\alpha \geq 0$, and others with density which scales as $n^{-\beta}$ for $\beta > \alpha$. Such sequences are only trivially included in the graphon theory: if one renormalizes the degrees by the largest one, this usually yields a graph limit which is constantly equal to zero.

A substantial progress has been made by Borgs and coauthors in the two works [17, 18]. They put the basis for a theory of L^p -graphons – in contrast to the classical theory for which $p = \infty$ – and establish a notion of convergence for unbounded graphons in the intermediate regimes. In our notation, they establish a theory for kernels which are L^p -integrable on $[0, 1]^2$. The most common example is probably given by the kernel $W(x, y) = (xy)^{-\alpha}$, where $0 < \alpha < 1$, and which leads to a random graph sequence where the degrees follow a power law distribution.

A way to include L^p -graphons in the analysis of (1.13), is made by renormalizing each particle with a different dilution parameter. This means considering the following class of systems:

$$d\bar{x}_t^{i,n} = F(\bar{x}_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \frac{\xi_{ij}^{(n)}}{p_{i,n}} \Gamma(\bar{x}_t^{i,n}, \bar{x}_t^{j,n})dt + \sigma(\bar{x}_t^{i,n})dB_t^i, \quad (6.1)$$

where $p_{i,n}$ represents the dilution parameter for particle i , for $i = 1, \dots, n$. As already stressed in [77, §2.7] and in [82, §3], the right quantity to look at in (6.1) is not given by the graph

sequence $\xi^{(n)}$, but by the normalized variables

$$\bar{\xi}_{ij}^{(n)} = \frac{\xi_{ij}^{(n)}}{p_{i,n}}, \quad i, j = 1, \dots, n.$$

The entries $\bar{\xi}_{ij}^{(n)}$ no longer represent an undirected dense graph but form a matrix with real non-negative values. The underlying graph $\xi^{(n)}$ is now allowed to have vertices with out-degree¹ of different magnitudes, yet, by properly renormalizing the interaction, the limit is again given by a graphon. In other words, even if the starting graph sequence would lead to an L^p -kernel, the renormalization yields a graphon as n tends to infinity. In conclusion, although equation (6.1) allows to include a larger class of graphs, it does not necessarily provide any new asymptotic behavior of the empirical measure if the graph limit remains a graphon.

Extending the previous results to include kernels *in the limit* is challenging under several aspects. For instance, the analysis of the non-linear process (1.15) on kernels is already problematic, as stressed in [77]. A suitable framework for particle systems on unbounded graphons represents another direction towards new exciting results.

4. Long-time dynamics and uniform propagation of chaos

The approach given in Chapter 5 represents a complementary technique to the study of weakly interacting particle systems. Directly tackling the equation satisfied by the empirical measure, provides an useful tool for the study of the associated long-time dynamics, as perfectly illustrated in Chapter 4 and in the articles [10, 78].

We imagine that this approach can also be helpful for uniform in time propagation of chaos estimates, see, e.g., [83, 84]. Propagation of chaos has only been touched across Chapter 4 but can be easily derived under suitable assumptions on the initial data and the graph sequence.

Moreover, we believe that our method can be extended to systems as

$$dx_t^{i,n} = F(x_t^{i,n})dt + \frac{1}{n} \sum_{j=1}^n \Gamma(x_t^{i,n}, x_t^{j,n})dt + \sigma_1(x_t^{i,n})dB_t^i + \sigma_2(x_t^{i,n})dB_t \quad (6.2)$$

where σ_1 and σ_2 are suitable functions and B is a Brownian motion independent of $\{B^i\}_{i \in \mathbb{N}}$. In other words, we would like to include multiplicative and common noise.

¹In a directed graph ξ of n vertices, we can define the out-degree of vertex $i \in \{1, \dots, n\}$ as $\sum_{j=1}^n \xi_{ij}$. Observe that the normalization in $\bar{\xi}^{(n)}$ leads to a matrix which is no longer symmetric.

Bibliography

- [1] J. A. Acebron, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, 2005.
- [2] R. A. Adams and J. Fournier. *Sobolev spaces*. Number 65 in Pure and applied mathematics series. Acad. Press, 2 edition, 2003.
- [3] N. Alon and A. Naor. Approximating the Cut-Norm via Grothendieck’s Inequality. *SIAM Journal on Computing*, 35:787–803, 2006.
- [4] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, 2008.
- [5] A. Basak, S. Bhamidi, S. Chakraborty, and A. Nobel. Large subgraphs in pseudo-random graphs. *arXiv:1610.03762 [math]*, 2016.
- [6] E. Bayraktar, S. Chakraborty, and R. Wu. Graphon mean field systems. *arXiv:2003.13180 [math]*, 2020.
- [7] F. Bechtold. Strong solutions of semilinear SPDEs with unbounded diffusion. *arXiv:2001.08848 [math]*, 2020.
- [8] F. Bechtold and F. Coppini. A Law of Large Numbers for interacting diffusions via a mild formulation. *arXiv:2005.05624 [math]*, 2020.
- [9] L. Bertini, G. Giacomin, and K. Pakdaman. Dynamical aspects of mean field plane rotators and the Kuramoto model. *Journal of Statistical Physics*, 138(1-3):270–290, 2010.
- [10] L. Bertini, G. Giacomin, and C. Poquet. Synchronization and random long time dynamics for mean-field plane rotators. *Probability Theory and Related Fields*, 160(3-4):593–653, 2014.
- [11] G. Bet, F. Coppini, and F. R. Nardi. Weakly interacting oscillators on dense random graphs. *arXiv:2006.07670 [math]*, 2020.
- [12] G. Bet, R. van der Hofstad, and J. S. H. van Leeuwen. Big jobs arrive early: From critical queues to random graphs. *arXiv:1704.03406 [math]*, 2017.
- [13] S. Bhamidi, A. Budhiraja, and R. Wu. Weakly interacting particle systems on inhomogeneous random graphs. *Stochastic Processes and their Applications*, 129(6):2174–2206, 2019.
- [14] P. Billingsley. *Convergence of Probability Measures*. Wiley, 2 edition, 1999.
- [15] K. Bogerd, R. M. Castro, and R. van der Hofstad. Cliques in rank-1 random graphs: the role of inhomogeneity. *arXiv:1805.01688 [math]*, 2018.
- [16] B. Bollobas and O. Riordan. Metrics for sparse graphs. *Survey in combinatorics*, 365:211–287, 2009.
- [17] C. Borgs, J. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5):3019–3062, 2019.
- [18] 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. *The Annals of Probability*, 46(1):337–396, 2018.
- [19] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [20] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Communications in Mathematical Physics*, 56(2):101–113, 1977.

- [21] J. v. Brecht, T. Kolokolnikov, A. L. Bertozzi, and H. Sun. Swarming on Random Graphs. *Journal of Statistical Physics*, 151(1-2):150–173, 2013.
- [22] H. Brezis. *Analyse fonctionnelle : Théorie et applications*. Dunod, Paris, 1999.
- [23] A. Budhiraja, P. Dupuis, and M. Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *The Annals of Probability*, 40(1):74–102, 2012.
- [24] P. E. Caines and M. Huang. Graphon Mean Field Games and the GMFG Equations. *2018 IEEE Conference on Decision and Control (CDC)*, 2018.
- [25] R. Carmona, D. Cooney, C. Graves, and M. Lauriere. Stochastic Graphon Games: I. The Static Case. *arXiv:1911.10664 [math]*, 2019.
- [26] 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.
- [27] H. Chiba and G. S. Medvedev. The mean field analysis of the kuramoto model on graphs i. the mean field equation and transition point formulas. *Discrete & Continuous Dynamical Systems - A*, 39(1):131, 2019.
- [28] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.
- [29] M. Coghi, J.-D. Deuschel, P. Friz, and M. Maurelli. Pathwise McKean-Vlasov Theory with Additive Noise. *arXiv:1812.11773 [math]*, 2019.
- [30] F. Coppini. Long time dynamics for interacting oscillators on graphs. *arXiv:1908.01520 [math]*, 2019.
- [31] 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, 2020.
- [32] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Number 45 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.
- [33] P. Dai Pra and F. d. Hollander. McKean-Vlasov limit for interacting random processes in random media. *Journal of Statistical Physics*, 84(3-4):735–772, 1996.
- [34] D. A. Dawson and J. Gärtner. Large deviations from the mckean-vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.
- [35] V. H. de la Pena, M. J. Klass, and T. L. Lai. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *The Annals of Probability*, 32(3A):1902–1933, 2004.
- [36] 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, 2016.
- [37] A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *The Annals of Applied Probability*, 20(2):565–592, 2010.
- [38] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [39] F. den Hollander. *Large Deviations*. American Mathematical Society, 2000.
- [40] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28:33–61, 2008.
- [41] R. L. Dobrushin. Vlasov equations. *Functional Analysis and Its Applications*, 13(2):115–123, 1979.
- [42] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, 2011.
- [43] F. Dörfler and F. Bullo. Synchronization in complex networks of phase oscillators: A survey. *Automatica*, 50(6):1539–1564, 2014.
- [44] J. Feng and T. G. Kurtz. *Large Deviations for Stochastic Processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2006.

- [45] B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Processes and their Applications*, 71(1):33–53, 1997.
- [46] F. Flandoli, M. Leimbach, and C. Olivera. Uniform convergence of proliferating particles to the FKPP equation. *Journal of Mathematical Analysis and Applications*, 473(1):27–52, 2019.
- [47] F. Flandoli, C. Olivera, and M. Simon. Uniform approximation of $2d$ Navier-Stokes equation by stochastic interacting particle systems. *arXiv:2004.00458 [math]*, 2020.
- [48] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer US, New York, NY, 1984.
- [49] P. K. Friz and M. Hairer. *A Course on Rough Paths*. Universitext. Springer International Publishing, Cham, 2014.
- [50] T. Funaki. A certain class of diffusion processes associated with nonlinear parabolic equations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(3):331–348, 1984.
- [51] G. Giacomin, K. Pakdaman, and X. Pellegrin. Global attractor and asymptotic dynamics in the Kuramoto model for coupled noisy phase oscillators. *Nonlinearity*, 25(5):1247–1273, 2012.
- [52] S. E. Graversen and G. Peskir. Maximal inequalities for the Ornstein-Uhlenbeck process. *Proceedings of the American Mathematical Society*, 128(10):3035–3042, 2000.
- [53] M. Gubinelli. Controlling rough paths. *Journal of Functional Analysis*, 216(1):86–140, 2004.
- [54] M. Gubinelli and S. Tindel. Rough evolution equations. *The Annals of Probability*, 38(1):1–75, 2010.
- [55] O. Guédon and R. Vershynin. Community detection in sparse networks via Grothendieck’s inequality. *Probability Theory and Related Fields*, 165(3):1025–1049, 2016.
- [56] J. Gärtner. On the McKean-Vlasov Limit for Interacting Diffusions. *Mathematische Nachrichten*, 137(1):197–248, 1988.
- [57] E. Heiman, G. Schechtman, and A. Shraibman. Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 45(2):306–317, 2014.
- [58] D. Henry. *Geometric Theory of Semilinear Parabolic Equations*, volume 840 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1981.
- [59] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(04):439–562, 2006.
- [60] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes, Volume 24 - 2nd Edition*, volume 24 of *North-Holland Mathematical Library*. North Holland, 2nd edition, 1989.
- [61] P.-E. Jabin and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones mathematicae*, 214(1):523–591, 2018.
- [62] C. Jia and G. Zhao. Moderate maximal inequalities for the Ornstein-Uhlenbeck process. *arXiv:1711.00902 [math]*, 2017.
- [63] M. Kac. Foundations of Kinetic Theory. In *Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 1956. University of California Press.
- [64] Y. Katznelson. *An Introduction To Harmonic Analysis*. Cambridge Mathematical Library. Cambridge University Press, 3rd edition, 2004.
- [65] V. N. Kolokoltsov. *Markov processes, semigroups, and generators*. Number 38 in De Gruyter studies in mathematics. De Gruyter, Berlin ; New York, 2011.
- [66] V. Konarovskiy, T. Lehmann, and M.-K. v. Renesse. Dean-Kawasaki dynamics: ill-posedness vs. triviality. *Electronic Communications in Probability*, 24, 2019.
- [67] V. Konarovskiy, T. Lehmann, and M. von Renesse. On Dean-Kawasaki Dynamics with Smooth Drift Potential. *Journal of Statistical Physics*, 2019.

- [68] Y. Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics*, number 39 in Lecture Notes in Physics, pages 420–422. Springer Berlin Heidelberg, 1975.
- [69] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*, volume 19 of *Springer Series in Synergetics*. Springer-Verlag Berlin Heidelberg, 1984.
- [70] D. Lacker, K. Ramanan, and R. Wu. Large sparse networks of interacting diffusions. *arXiv:1904.02585 [math]*, 2019.
- [71] C. Lancellotti. On the Vlasov Limit for Systems of Nonlinearly Coupled Oscillators without Noise. *Transport Theory and Statistical Physics*, 34(7):523–535, 2005.
- [72] C. M. Le, E. Levina, and R. Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.
- [73] L. Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.
- [74] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.
- [75] A. Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*. Modern Birkhäuser classics. Birkhäuser, Basel, 2012.
- [76] E. Luçon. Quenched Large Deviations for Interacting Diffusions in Random Media. *Journal of Statistical Physics*, 166(6):1405–1440, 2017.
- [77] E. Luçon. Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *Stochastic Processes and their Applications*, 2020.
- [78] 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.
- [79] E. Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability*, 24(5):1946–1993, 2014.
- [80] C. Léonard. Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coefficients non bornés. *Annales de l’I.H.P. Probabilités et statistiques*, 22(2):237–262, 1986.
- [81] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907–1911, 1966.
- [82] G. S. Medvedev. The continuum limit of the kuramoto model on sparse random graphs. *arXiv:1802.03787 [nlin]*, 2018.
- [83] S. Mischler and C. Mouhot. Kac’s Program in Kinetic Theory. *Inventiones Mathematicae*, 193(1):1–147, 2013.
- [84] S. Mischler, C. Mouhot, and B. Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probability Theory and Related Fields*, 161(1-2):1–59, 2015.
- [85] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations*, number 1627 in Lecture Notes in Mathematics, pages 42–95. Springer Berlin Heidelberg, 1996.
- [86] M. Métivier. *Semimartingales, A Course on Stochastic Processes*. De Gruyter Studies in Mathematics 2, Berlin, Boston, 1982.
- [87] H. Neunzert. An introduction to the nonlinear boltzmann-vlasov equation. In *Kinetic Theories and the Boltzmann Equation*, Lecture Notes in Mathematics, pages 60–110, Berlin, Heidelberg, 1984. Springer.
- [88] K. Oelschläger. A Martingale Approach to the Law of Large Numbers for Weakly Interacting Stochastic Processes. *The Annals of Probability*, 12(2):458–479, 1984.

- [89] R. I. Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv:0911.0600 [math]*, 2009.
- [90] R. I. Oliveira and G. H. Reis. Interacting Diffusions on Random Graphs with Diverging Average Degrees: Hydrodynamics and Large Deviations. *Journal of Statistical Physics*, 176(5):1057–1087, 2019.
- [91] R. I. Oliveira, G. H. Reis, and L. M. Stolerman. Interacting diffusions on sparse graphs: hydrodynamics from local weak limits. *arXiv:1812.11924 [math]*, 2018.
- [92] E. Olivieri and M. E. Vares. *Large Deviations and Metastability*. Cambridge University Press, Cambridge, 2005.
- [93] G. Pisier. Grothendieck’s Theorem, past and present. *Bulletin of the American Mathematical Society*, 49(2):237–323, 2012.
- [94] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths. The Kuramoto model in complex networks. *Physics Reports*, 610:1–98, 2016.
- [95] T. Shiga and H. Tanaka. Central limit theorem for a system of Markovian particles with mean field interactions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69(3):439–459, 1985.
- [96] S. H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, 2000.
- [97] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Number 233 in Grundlehren der mathematischen Wissenschaften. Springer, Berlin; New York, 2006.
- [98] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d’Eté de Probabilités de Saint-Flour XIX - 1989*, volume 1464, pages 165–251. Springer Berlin Heidelberg, 1991.
- [99] H. Tanaka. Limit Theorems for Certain Diffusion Processes with Interaction. In *North-Holland Mathematical Library*, volume 32 of *Stochastic Analysis*, pages 469–488. Elsevier, 1984.
- [100] J. M. A. M. van Neerven, M. C. Veraar, and L. Weis. Stochastic integration in UMD Banach spaces. *The Annals of Probability*, 35(4):1438–1478, 2007.
- [101] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer-Verlag, Berlin Heidelberg, 6 edition, 1995.