WEAKLY INTERACTING OSCILLATORS ON DENSE RANDOM GRAPHS

GIANMARCO BET, FABIO COPPINI, AND FRANCESCA R. NARDI

To our friend and colleague Carlo Casolo

Abstract. We consider a class of weakly interacting particle systems of mean-field type. The interactions between the particles are encoded in a graph sequence, i.e., two particles are interacting if and only if they are connected in the underlying graph. We establish a Law of Large Numbers for the empirical measure of the system that holds whenever the graph sequence is convergent in the sense of graph limits theory. The limit is shown to be the solution to a non-linear Fokker-Planck equation weighted by the (possibly random) graph limit. No regularity assumptions are made on the graphon limit so that our analysis allows for very general graph sequences, such as exchangeable random graphs. For these, we also prove a propagation of chaos result. Finally, we fully characterize the graph sequences for which the associated empirical measure converges to the mean-field limit, i.e., to the solution of the classical McKean-Vlasov equation.

2020 MSC: 60K35, 05C80, 82B20, 60H20, 35Q84.

Keywords: Interacting oscillators, random graphons, mean-field systems, Fokker-Planck equation, exchangeable graphs, McKean-Vlasov.

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 (see, e.g., [22, 25]) 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 an underlying network, see, e.g., [2, 24].

The first mathematically rigorous results appeared only recently [7, 15]. 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 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 [13, 14, 19].

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. Our main object of study is the empirical measure associated to this system. We rely on the recent graphon theory for the notion of graph convergence and graph limit, see [18]. Graphons are a generalization of dense graph sequences, and have proven to be useful in a variety of contexts which nowadays include mean-field games as well, see [9, 10] and references therein.

Our main result 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 nor on the limiting graphon. Thus, our result holds for very general graph sequences such as exchangeable random graphs, see [16].

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. Notably, 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 [4, 19, 23] and the deterministic setting [11, 21]. 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 and they are not able to deal with general graph sequences as we do. Moreover, to our knowledge, the results presented here appear to be the first in the literature to tackle interacting particle systems on random graphons.

Our work stems from the fact that the empirical measure of a particle system is invariant under relabeling of the particles and thus 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. 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. We do this by establishing existence, uniqueness and convergence results without any regularity assumption on the graph structure. We also establish a propagation of chaos result, and propose a non-linear process that describes the behavior of a tagged particle sampled uniformly at random.

1.1. **A look at the literature.** Weakly interacting particle systems on graphs have first been studied in [7, 15], 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 [14] considers 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 [4, 19, 23] deal with more general sequences of graphs and take into account a few notions coming from graph limits theory. Namely, [23] establishes a Large Deviation Principle for the empirical measure of weakly interacting particles on W-random graphs, see (2.10) for the precise definition. The works [4, 19] 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 various graph sequences arising from labeled graphons, we refer to [11, 21] and references therein.

To the authors' knowledge, the only work addressing the long-time behavior of interacting particle systems on graphs is [13], where the Kuramoto model defined on pseudo-random graphs is shown to behave as the mean-field limit 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 [9, 10] 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). Finally, although the results in [4, 11, 19, 21, 23] 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, as well as the various distances between probability measures that we will consider.

In Section 2 we define the interacting particle system and the associated nonlinear 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 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 A we collect the most important results on graphons from the literature 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 $\{\mathcal{F}_\cdot\}$ is a filtration satisfying the usual conditions. All Brownian motions that we consider later on are adapted to $\{\mathcal{F}_t\}_{t\geqslant 0}$ 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

$$D_T(\bar{\mu},\bar{\nu}) := \inf_{m\in\gamma(\bar{\mu},\bar{\nu})} \left\{ \int \sup_{t\in[0,T]} |x_t - y_t|^2 \, m(\mathrm{d}x,\mathrm{d}y) \right\}^{1/2}, \tag{1.1}$$

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 (1.1) can be rewritten as

$$D_T(\bar{\mu},\bar{\nu}) = \inf_{X,Y} \left\{ \mathbf{E}\left[\sup_{t\in[0,T]} |X_t - Y_t|^2 \right] : \mathcal{L}(X) = \bar{\mu},\ \mathcal{L}(Y) = \bar{\nu} \right\}^{1/2} \tag{1.2}$$

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 (1.1) we obtain that for every $s \in [0,T]$

$$\sup_f \left| \int_{\mathbb{T}} f(\theta)\, \bar{\mu}_s(\mathrm{d}\theta) - \int_{\mathbb{T}} f(\theta)\, \bar{\nu}_s(\mathrm{d}\theta) \right| \leqslant D_s(\bar{\mu},\bar{\nu}), \tag{1.3}$$

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 A. We follow the notation of [18], the notions of labeled and unlabeled graphs are taken from [16], 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 *labeled graphon*, and the associated equivalence class, i.e., an unlabeled graphon in the notation of [18], 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 our two main models: a weakly interacting particle system (2.1) and a non-linear process (2.3).

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 $\xi_{ij}^{(n)} = 1$ whenever the vertices i and j are connected and $\xi_{ij}^{(n)} = 0$ otherwise. Let $\{\theta^{i,n}\}_{i=1,\dots,n}$ be the family of oscillators on $\mathbb{T}^n$ that satisfy

$$\begin{cases} \mathrm{d}\theta_t^{i,n} = F(\theta_t^{i,n})\mathrm{d}t + \frac{1}{n}\sum_{j=1}^n \xi_{ij}^{(n)}\,\Gamma(\theta_t^{i,n},\theta_t^{j,n})\mathrm{d}t + \mathrm{d}B_t^i, & 0 < t < T, \\ \theta_0^{i,n} = \theta_0^i, & i \in \{1,\dots,n\}, \end{cases} \tag{2.1}$$

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. Many interesting examples of interacting oscillators fit this framework, such as the Kuramoto model, the plane rotator model and other generalizations, see, e.g., [15, §1.2], [3] and also Subsection 2.3. We are interested in studying the empirical measure associated to (2.1). This is defined as the (random) probability measure on $\mathbb{T}$ such that

$$\mu_t^n := \frac{1}{n}\sum_{j=1}^n \delta_{\theta_t^{j,n}}, \tag{2.2}$$

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

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 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

$$\begin{cases} \theta_t = \theta_0 + \int_0^t F(\theta_s)\mathrm{d}s + \int_0^t \int_I W(U,y)\int_{\mathbb{T}} \Gamma(\theta_s,\theta)\mu_s^y(\mathrm{d}\theta)\mathrm{d}y\,\mathrm{d}s + B_t, \\ \mu_t^y = \mathcal{L}(\theta_t|U=y), \quad \text{for } y \in I,\ t\in[0,T], \end{cases} \tag{2.3}$$

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 and the uniqueness of the solution to equation (2.3). In Section 3, we prove that equation (2.3) is well-posed 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* (2.3)*. 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*

$$\partial_t\bar{\mu}_t(\theta) = \frac{1}{2}\partial_\theta^2\bar{\mu}_t(\theta) - \partial_\theta\left[\bar{\mu}_t(\theta)F(\theta)\right] - \partial_\theta\left[\int_{I\times I} W(x,y)\,\mu_t^x(\theta)\int_{\mathbb{T}}\Gamma(\theta,\widetilde{\theta})\,\mu_t^y(\mathrm{d}\widetilde{\theta})\mathrm{d}y\,\mathrm{d}x\right], \tag{2.4}$$

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

Recall that $\delta_\square$ defines a metric in the space of graphons $\widetilde{\mathcal{W}}_0$, see (A.8). 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* (2.3) *associated with graphons* W *and* V *respectively, then*

$$D_T(\bar\mu^W, \bar\mu^V) \leqslant C\, \delta_\square(W,V)^{1/2}. \tag{2.5}$$

The proof is postponed to Section 3.4. Note that taking the p-Wasserstein distance in (1.1) for $p \geqslant 1$ leads to a Hölder exponent as large as $1/p$. Propositions 2.1 and 2.2 imply that the following mapping is continuous:

$$\begin{aligned} \Psi : (\widetilde{\mathcal{W}}_0, \delta_\square) &\to (\mathcal{C}([0,T], \mathcal{P}(\mathbb{T})), D_T) \\ W &\mapsto \bar\mu^W, \end{aligned} \tag{2.6}$$

where $\bar\mu^W$ is the law of θ solving equation (2.3) with graphon W. In particular, to every random variable W in $\widetilde{\mathcal{W}}_0$ 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*

$$\lim_{n\to\infty} \mathbb{E}\left[\delta_\square\left(\xi^{(n)}, W\right)\right] = 0. \tag{2.7}$$

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

$$\mu^n \longrightarrow \bar\mu, \quad \text{in } \mathbf{P}\times\mathbb{P}\text{-probability, as } n\to\infty, \tag{2.8}$$

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* (2.4) *starting from* $\bar\mu_0$, *with graphon* $W(\omega)$.

Condition (2.7) 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 (2.7) and the convergence in probability for random graph sequences is proven in Lemma A.2. One may wonder if the convergence of μ^n holds under weaker conditions on the initial data. We conjecture that our results still holds if the independence assumption is replaced with exchangeability of the initial data. However, the exchangeability property seems to be necessary to be able 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 limiting 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 also holds if $\{\theta_0^i\}_{i\in\mathbb{N}}$ is 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 randomness in $\xi^{(n)}$ is further discussed in Subsection 2.3.

Applications to exchangeable graphs. Recall that an exchangeable random graph $\xi = \{\xi_{ij}\}_{i,j\in\mathbb{N}}$ (see [18]) is a infinite array of binary random variables, such that

$$\mathbb{P}\left(\xi_{ij} = e_{ij},\, 1 \leqslant i, j \leqslant n\right) = \mathbb{P}\left(\xi_{ij} = e_{\sigma(i)\sigma(j)},\, 1 \leqslant i, j \leqslant n\right) \tag{2.9}$$

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 [16].

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* (A.6) *and* [18, §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

$$\xi_{ij}^W = W(U_i, U_j), \tag{2.10}$$

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 [16] and references therein. We recall their main result here.

Theorem 2.5 ([16, Theorem 5.3] and [18, 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*

$$\xi^{(n)} \longrightarrow W \quad \mathbb{P}\text{-a.s. in } \widetilde{\mathcal{W}}_0, \tag{2.11}$$

as $n \to \infty$.

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*

$$\mu^n \longrightarrow \bar{\mu}, \quad \text{in } \mathbf{P} \times \mathbb{P}\text{-probability, as } n \to \infty, \tag{2.12}$$

where $\bar{\mu}$ is the solution to (2.4) *starting from $\bar{\mu}_0$ with graphon W.*

Propagation of Chaos. Whenever $\xi = \{\xi^{(n)}\}_{n\in\mathbb{N}}$ is a sequence of exchangeable graphs[1], 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 [25, 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 (2.8), we can thus deduce a propagation of chaos statement for the particle system (2.1). 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 $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \mathcal{L}(\theta^{1,n}, \dots, \theta^{k,n}) = \prod_{i=1}^{k} \mathcal{L}(\theta) = \prod_{i=1}^{k} \bar{\mu}. \tag{2.13}$$

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 randomness 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, Theorem 2.3 yields a precise characterization of the graph sequences for which the empirical measure converges to the mean-field limit. 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 (2.1) on a sequence of complete graphs, i.e., $\xi^{(n)}_{ij} \equiv 1$ for every i, j and n. It is well known [22, 25] 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:

$$\partial_t \rho_t(\theta) = \frac{1}{2}\partial^2_\theta \rho_t(\theta) - \partial_\theta \left[\rho_t(\theta) F(\theta)\right] - p\, \partial_\theta \left[\rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \widetilde{\theta})\, \rho_t(\mathrm{d}\widetilde{\theta})\right], \tag{2.14}$$

with initial condition $\bar{\mu}_0$ and $p = 1$. Existence and uniqueness for the solution to (2.14) hold under our assumptions on F, Γ and $\bar{\mu}_0$, see e.g., [22, 25].

Suppose that the graph sequence is converging to a deterministic limit; we discuss the case of a random limit in 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 (2.14) with corresponding p. Since the convergence of $\xi^{(n)}$ to a non-constant graphon gives rise to equation (2.4), which is – at least formally – different from (2.14), 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

[1] i.e. for each $n \in \mathbb{N}$ the random variables $\{\xi^{(n)}_{ij}\}_{i,j=1,\dots,n}$ are exchangeable. Observe that ξ is not necessarily an exchangeable random graph as in (2.9).

graphs with such asymptotic behavior are known in the literature as pseudo-random graphs, see [3, 12] and [18, §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.

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

$$\xi_{ij}^{(n)} \sim \mathrm{Ber}(g_i g_j), \quad \text{independently for each } 1 \leqslant i < j \leqslant n. \tag{2.15}$$

The graph $\xi^{(n)}$ is the dense analogue of the inhomogeneous random graph, also known as rank-1 model, see e.g., [6, 8]. 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 (2.2).

The construction made in (2.15) yields a binary array $\{\xi_{ij}^{(n)}\}_{i,j=1,\dots,n}$ of exchangeable random variables. In particular, they have the same expected value $\mathbb{E}[\xi_{ij}^{(n)}] = p$, for every $i \neq j$. We are interested in comparing the empirical measure of the system (2.1) defined on the graph (2.15) to the empirical measure of the corresponding *annealed* system that is obtained from (2.1) by replacing $\xi_{ij}^{(n)}$ with their expected values. More precisely, the annealed system is defined as the solution to

$$\mathrm{d}\theta_t^{i,n} = F(\theta_t^{i,n})\mathrm{d}t + \frac{p}{n}\sum_{j=1}^{n} \Gamma(\theta_t^{i,n}, \theta_t^{j,n})\mathrm{d}t + \mathrm{d}B_t^i, \tag{2.16}$$

for which the asymptotic behavior is known to be the mean-field limit (2.14).

Perhaps surprisingly, the behavior of system (2.1) on the graph sequence (2.15) is suitably described in the limit by (2.16) only when g is deterministic and $g = \sqrt{p}$. Recall the definition of W-random graph given in (2.10): we see that $\xi^{(n)}$ is a W_g-random graph with

$$W_g(x,y) = F_g^{-1}(x)\, F_g^{-1}(y), \quad \text{for } x, y \in I, \tag{2.17}$$

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 (2.4) 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 (2.16) if and only if W_g is arbitrarily close to the constant graphon p in the cut-distance, i.e., if and only if $\mathrm{Var}[g] \ll 1$. 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 [14].

Moreover, 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 converges[2] 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 (2.1) does not yield any *averaging property* on the empirical measure μ^n. Moreover, adding the extra randomness through Bernoulli random variables in (2.15) 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. □

Until now, we have focused on deterministic limits for the sequence $\xi^{(n)}$. Observe that a characterization of the exchangeable random graphs with deterministic limits is given in [16]; see also [18, §11.5]. We now consider the case where the limit W is random, and we address the relationship between the resulting system and the mean-field limit ρ given in (2.14). 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 bullet point (2) above.

Example II: random mean-field behavior. Consider the growing preferential attachment graph ξ_{pa} constructed iteratively as follows; see also [18, 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 ([18, Proposition 11.45])**.** *With probability 1, the sequence $\{\xi_{\text{pa}}^{(n)}\}_{n\in\mathbb{N}}$ converges to a random constant graphon.*

Consider a particle system defined on the graph sequence $\{\xi_{\text{pa}}^{(n)}\}_{n\in\mathbb{N}}$. The empirical measure converges to the solution of equation (2.14) with a random p. In other words, μ^n converges to a random mean-field limit. Integrating (2.14) 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

$$\partial_t \bar{\rho}_t(\theta) = \frac{1}{2}\partial_\theta^2 \bar{\rho}_t(\theta) - \partial_\theta \left[\bar{\rho}_t(\theta) F(\theta)\right] - \partial_\theta \left[\mathbb{E}\Big[p\, \rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \widetilde{\theta}) \rho_t(\mathrm{d}\widetilde{\theta})\Big]\right], \qquad (2.18)$$

for $t \in [0,T]$. Note that (2.18) 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

[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 [18, §11.4].

formally satisfies the mean-field limit, i.e., it is not a solution to (2.14) 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 [5, 13] defined on the sequence $\xi_{\mathrm{pa}}^{(n)}$. The model is defined as the solution to

$$\mathrm{d}\theta_t^{i,n} = \frac{K}{n} \sum_{j=1}^{n} \xi_{ij}^{(n)} \sin(\theta_t^{j,n} - \theta_t^{i,n}) \mathrm{d}t + \mathrm{d}B_t^i, \tag{2.19}$$

for $i = 1, \dots, n$ and $t \in [0,T]$. It corresponds to (2.1) 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 (2.19) converges to the solution of

$$\partial_t \rho_t(\theta) = \frac{1}{2} \partial_\theta^2 \rho_t(\theta) + pK \partial_\theta [\rho_t(\theta)(\sin * \rho_t)(\theta)], \tag{2.20}$$

where $*$ stands for the convolution operator. It is well-known that the system (2.20) undergoes a phase transition as the coupling strength pK crosses the critical threshold $pK = 1$. Hence, the phase transition for this specific 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 $\mathbb{T}$ ($0 \leqslant pK < 1$). The solution to equation (2.20) can be written down explicitly (see again [5, 13]) 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 (2.20) for some fixed $p \in [0,1]$. □

3. The non-linear process

We introduce a non-linear process (3.10) which has already been considered in the literature [4, 11, 19, 20, 23] as the natural candidate in case the particles in (2.1) are not exchangeable and their labels are fixed from the initial condition. This process is useful for studying the evolution of a tagged particle with a specific profile of connections, as stressed in [19].

Contrary to our setting, some regularity in the – now labeled – graphon is usually assumed to show the convergence of the empirical measure (2.2). We will exploit (3.10) to better understand (2.3) and to establish existence and uniqueness.

Before introducing (3.10), 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 (3.6). 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.,

$$\mathcal{M}_T := \{\mu \in \mathcal{P}(I \times \mathcal{C}([0,T], \mathbb{T})) : p_1 \circ \mu = \lambda\}, \tag{3.1}$$

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

$$\mu(\mathrm{d}x, \mathrm{d}\theta) = \mu^x(\mathrm{d}\theta)\lambda(\mathrm{d}x), \quad x \in I, \tag{3.2}$$

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 $\lambda(\mathrm{d}x)$ on I simply by $\mathrm{d}x$. For $\mu, \nu \in \mathcal{M}_T$, define their distance by

$$d_T(\mu, \nu) := \left(\int_I D_T^2(\mu^x, \nu^x) \mathrm{d}x \right)^{1/2}. \tag{3.3}$$

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.*

$$\mathcal{M}_0 = \{\mu_0 \in \mathcal{P}(I \times \mathbb{T}) : p_1 \circ \mu_0 = \lambda\}, \tag{3.4}$$

and

$$d_0(\mu_0, \nu_0) = \left(\int_I D_0^2(\mu_0^x, \nu_0^x) \, \mathrm{d}x \right)^{1/2}, \quad \text{for } \mu_0, \nu_0 \in \mathcal{M}_0. \tag{3.5}$$

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

$$\mu \sim \nu \quad \text{iff there exists } \varphi \in S_I \text{ such that } \mu^x = \nu^{\varphi(x)}, \ x\text{-a.s.}. \tag{3.6}$$

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

$$\widetilde{d}_T(\mu, \nu) := \inf_{\varphi \in S_I} d_T(\mu, \nu^{\varphi}), \tag{3.7}$$

where we have used the notation $\nu^{\varphi} = \{\nu^{\varphi(x)}\}_{x \in I}$. Observe that if $\mu \sim \nu$, then $\bar{\mu} = \int_I \mu^x \mathrm{d}x = \int_I \nu^{\varphi(x)} \mathrm{d}x = \int_I \nu^x \mathrm{d}x = \bar{\nu}$. In particular, for every $\varphi \in S_I$

$$D_T^2(\bar{\mu}, \bar{\nu}) = D_T^2(\bar{\mu}, \bar{\nu}^{\varphi}) \leqslant \int_I D_T^2(\mu^x, \nu^{\varphi(x)}) \mathrm{d}x = d_T^2(\mu, \nu^{\varphi}). \tag{3.8}$$

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

$$D_T(\bar{\mu}, \bar{\nu}) \leqslant \widetilde{d}_T(\mu, \nu). \tag{3.9}$$

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

$$\begin{cases} \theta_t^x = \theta_0^x + \int_0^t F(\theta_s^x) \mathrm{d}s + \int_0^t \int_I W(x,y) \int_{\mathbb{T}} \Gamma(\theta_s^x, \theta) \mu_s^y(\mathrm{d}\theta) \mathrm{d}y \, \mathrm{d}s + B_t^x, \\ \mu_t^x = \mathcal{L}(\theta_t^x), \quad \text{for } x \in I, \, t \in [0,T], \end{cases} \tag{3.10}$$

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 $\{B^x\}_{x \in I}$ 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 (3.10). The proof follows a classical argument by Sznitman [25] and is postponed to Section 3.3.

Proposition 3.2. *There exists a unique solution $\theta = \{\theta^x\}_{x\in I}$ to* (3.10)*. 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*

$$\partial_t \mu_t^x(\theta) = \frac{1}{2}\partial_\theta^2 \mu_t^x(\theta) - \partial_\theta\left[\mu_t^x(\theta)F(\theta)\right] - \partial_\theta\left[\mu_t^x(\theta)\int_I W(x,y)\int_{\mathbb{T}}\Gamma(\theta,\theta')\mu_t^y(\mathrm{d}\theta')\mathrm{d}y\right] \tag{3.11}$$

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 [4, 9, 19, 20, 23].

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 [4, 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* (3.10) *with $W \in \mathcal{W}_0$ and $V \in \mathcal{W}_0$ respectively, then*

$$d_T(\mu^W, \mu^V) \leqslant C\,\|W - V\|_\square^{1/2}. \tag{3.12}$$

The proof is again postponed to Subsection 3.3. As for Proposition 2.2, endowing the space of trajectories with with the p-Wasserstein metric for $p \geqslant 1$ yields a Hölder exponent as large as $1/p$.

Relationship with the non-linear process (2.3)*.* Consider a probability distribution $\mu_0 \in \mathcal{M}_0$ such that $\int_I \mu_0^x\,\mathrm{d}x = \bar{\mu}_0$. The solution to (2.4) is given by $\bar{\mu} = \int_I \mu^x\,\mathrm{d}x$, where μ^x is the law of θ^x solving (3.10) with initial condition μ_0^x and labeled graphon W. In other words, θ has the same law of θ^U solution to (3.10), 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* (3.10)*. By a change of variables, $\theta^{\varphi(x)}$ solves*

$$\theta_t^{\varphi(x)} = \theta_0^{\varphi(x)} + \int_0^t F(\theta_s^{\varphi(x)})\mathrm{d}s + \int_0^t \int_I W(\varphi(x),\varphi(y))\int_{\mathbb{T}}\Gamma(\theta_s^{\varphi(x)},\theta)\mu_s^{\varphi(y)}(\mathrm{d}\theta)\mathrm{d}y\,\mathrm{d}s + B_t^{\varphi(x)} \tag{3.13}$$

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

$$\psi_t^x = \theta_0^{\varphi(x)} + \int_0^t F(\psi_s^x)\mathrm{d}s + \int_0^t\int_I V(x,y)\int_{\mathbb{T}}\Gamma(\psi_s^x,\theta)\nu_s^y(\mathrm{d}\theta)\mathrm{d}y\,\mathrm{d}s + B_t^{\varphi(x)}, \tag{3.14}$$

which has the same law as (3.10) *with labeled graphon V and initial conditions $\{\theta^{\varphi(x)}\}_{x\in I}$.*

Observe that the laws ν and μ associated to (3.14) *and* (3.10) *respectively, differ only in the labeling of the vertices but their distance in $\mathcal{M}_T$ is not zero 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 $\bar\mu = \int_I \mu^x \,\mathrm{d}x$ and $\bar\nu = \int_I \nu^x \,\mathrm{d}x$, 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* (2.3) *is also equivalent to ψ^U, where ψ^x solves* (3.14)*, and U is uniformly distributed on I.*

3.3. **Proofs for the non-linear process** (3.10) **with fixed labels.**

Proof of Proposition 3.2. The proof follows a classical argument given in [25, Lemma 1.3]. Consider $\nu \in \mathcal{M}_T$ and $\{\theta^{x,\nu}\}_{x\in I}$ solving

$$\theta^{x,\nu}_t = \theta^x_0 + \int_0^t F(\theta^{x,\nu}_s)\,\mathrm{d}s + \int_0^t \int_I W(x,y) \int_{\mathbb{T}} \Gamma(\theta^{x,\nu}_s,\theta)\nu^y_s(\mathrm{d}\theta)\,\mathrm{d}y\,\mathrm{d}s + B^x_t, \tag{3.15}$$

where the initial conditions and the Brownian motions are the same of (3.10). Since F and Γ are bounded Lipschitz functions, there exists a unique solution to (3.15), which we denote by $\Phi(\nu)\in\mathcal{M}_T$. Thus, the map

$$\begin{aligned} \Phi : (\mathcal{M}_T, d_T) &\to (\mathcal{M}_T, d_T) \\ \nu &\to \Phi(\nu) \end{aligned} \tag{3.16}$$

is well defined. A solution to (3.10) is a fixed point of Φ and any fixed point of Φ is a solution to (3.10).

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

$$\begin{aligned} |\theta^{x,\mu}_t - \theta^{x,\nu}_t|^2 &\leqslant C\int_0^t |F(\theta^{x,\mu}_s) - F(\theta^{x,\nu}_s)|^2\,\mathrm{d}s \\ &+ C\int_0^t \left|\int_I W(x,y)\left(\int_{\mathbb{T}}\Gamma(\theta^{x,\mu}_s,\theta)\mu^y_s(\mathrm{d}\theta) - \int_{\mathbb{T}}\Gamma(\theta^{x,\nu}_s,\theta)\nu^y_s(\mathrm{d}\theta)\right)\mathrm{d}y\right|^2\mathrm{d}s \end{aligned}$$

Adding and subtracting in the second integral the quantity $\Gamma(\theta^{x,\mu}_s,\theta)\nu^y_s(\mathrm{d}\theta)$ and using that F and Γ are Lipschitz-continuous functions and that F, Γ and W are bounded, we get

$$\leqslant C\int_0^t |\theta^{x,\mu}_s - \theta^{x,\nu}_s|^2\,\mathrm{d}s + C\int_0^t\int_I\left|\int_{\mathbb{T}}\Gamma(\theta^{x,\mu}_s,\theta)\left[\mu^y_s - \nu^y_s\right](\mathrm{d}\theta)\right|^2\mathrm{d}y\,\mathrm{d}s, \tag{3.17}$$

From (1.3) we obtain

$$\left|\int_{\mathbb{T}}\Gamma(\theta^{x,\mu}_s,\theta)\left(\mu^y_s - \nu^y_s\right)(\mathrm{d}\theta)\right| \leqslant D_s(\mu^y,\nu^y) \tag{3.18}$$

from which, using (3.3), we deduce

$$|\theta^{x,\mu}_t - \theta^{x,\nu}_t|^2 \leqslant C\int_0^t |\theta^{x,\mu}_s - \theta^{x,\nu}_s|^2\,\mathrm{d}s + C\int_0^t d^2_s(\mu,\nu)\,\mathrm{d}s. \tag{3.19}$$

The definition of D_T (1.2) and an application of Gronwall's lemma lead to

$$d_T^2(\Phi(\mu),\Phi(\nu)) \leqslant \int_I \mathbf{E}\left[\sup_{t\in[0,T]} |\theta_t^{x,\mu}-\theta_t^{x,\nu}|^2\right] \mathrm{d}x \leqslant C\int_0^T d_s^2(\mu,\nu)\,\mathrm{d}s. \tag{3.20}$$

From the last relation we obtain the uniqueness of solutions to (3.10).

We prove that a solution exists by iterating (3.20). Indeed, for $k\geqslant 1$ and $\mu\in\mathcal{M}_T$, one gets

$$d_T^2(\Phi^{k+1}(\mu),\Phi^k(\mu)) \leqslant C^k\frac{T^k}{k!}\int_0^T d_t^2(\Phi(\mu),\mu)\,\mathrm{d}t. \tag{3.21}$$

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

$$\begin{aligned} f(\theta_t^x) \;&=\; f(\theta_0^x)+\frac{1}{2}\int_0^t \partial_\theta^2 f(\theta_s^x)\,\mathrm{d}s+\int_0^t \partial_\theta f(\theta_s^x)\,F(\theta_s^x)\,\mathrm{d}s \\ &+\int_0^t \partial_\theta f(\theta_s^x)\,\left[\int_I W(x,y)\int_{\mathbb{T}}\Gamma(\theta_s^x,\theta)\mu_s^y(\mathrm{d}\theta)\mathrm{d}y\right]\mathrm{d}s+\int_0^t \partial_\theta f(\theta_s^x)\,\mathrm{d}B_s^x. \end{aligned} \tag{3.22}$$

Integrating with respect to $\mathbf{P}$ yields the weak formulation of (3.11). □

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

$$\begin{aligned} \left|\theta_s^{x,W}-\theta_s^{x,V}\right|^2 \;&\leqslant C\int_0^s \left|F(\theta_r^{x,W})-F(\theta_r^{x,V})\right|^2\mathrm{d}r \\ &+C\int_0^s \left|\int_I W(x,y)\int_{\mathbb{T}}(\Gamma(\theta_r^{x,W},\theta)-\Gamma(\theta_r^{x,V},\theta))\mu_r^{y,W}(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r \\ &+C\int_0^s \left|\int_I W(x,y)\int_{\mathbb{T}}\Gamma(\theta_r^{x,V},\theta)(\mu_r^{y,W}-\mu_r^{y,V})(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r \\ &+C\int_0^s \left|\int_I (W(x,y)-V(x,y))\int_{\mathbb{T}}\Gamma(\theta_r^{x,V},\theta)\mu_r^{y,V}(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r. \end{aligned} \tag{3.23}$$

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

$$\begin{aligned} \left|\theta_s^{x,W}-\theta_s^{x,V}\right|^2 &\leqslant C\int_0^s \left|\theta_r^{x,W}-\theta_r^{x,V}\right|^2\mathrm{d}r + C\int_0^s d_r^2(\mu^W,\mu^V)\,\mathrm{d}r \\ &+\int_0^s \left|\int_I (W(x,y)-V(x,y))\left(\int_{\mathbb{T}}\Gamma(\theta_s^{x,V},\theta)\mu_r^{y,V}(\mathrm{d}\theta)\right)\mathrm{d}y\right|^2\mathrm{d}r. \end{aligned} \tag{3.24}$$

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 (3.20) to get

$$d_t^2(\mu^W, \mu^V) \leqslant \int_I \mathbf{E}\left[\sup_{s\in[0,t]} \left|\theta_s^{x,W} - \theta_s^{x,V}\right|^2\right] \mathrm{d}x \leqslant C\Big(\int_0^t d_s^2(\mu^W, \mu^V)\,\mathrm{d}s + G\Big), \quad (3.25)$$

where G is given by

$$G = \int_0^t \mathbf{E}\left[\int_I \left|\int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta)\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right|^2 \mathrm{d}x\right] \mathrm{d}s. \quad (3.26)$$

Applying Gronwall's inequality to (3.25) yields

$$d_t^2(\mu^W, \mu^V) \leqslant CG. \quad (3.27)$$

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

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

$$\Gamma(\theta, \psi) = \sum_{k,l\in\mathbb{Z}} \Gamma_{kl}\, e^{ik\theta} e^{il\psi}, \quad \theta, \psi \in \mathbb{T}, \quad (3.28)$$

where $\Gamma_{kl} = \int_{\mathbb{T}^2} \Gamma(\theta,\psi) e^{i(k\theta + l\psi)} \mathrm{d}\theta \mathrm{d}\psi$. Since $\Gamma \in \mathcal{C}^{1+\varepsilon}$, classical results on the asymptotic of Fourier series [17, pp. 24-26] imply that

$$C_\Gamma := \sum_{k,l\in\mathbb{Z}} (kl)^{1+\varepsilon} \left|\Gamma_{kl}\right|^2 < \infty. \quad (3.29)$$

Plugging this expression into (3.26), we obtain that

$$\begin{aligned}
&\int_I \left|\int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta)\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right|^2 \mathrm{d}x \\
&= \int_I \left|\sum_{kl} \Gamma_{kl}\, e^{ik\theta_s^{x,V}} \int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} e^{il\theta}\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right|^2 \mathrm{d}x.
\end{aligned} \quad (3.30)$$

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

$$\begin{aligned}
&\leqslant \int_I \Bigg|\sum_{kl} \left((kl)^{(1+\varepsilon)/2}\Gamma_{kl}\, e^{ik\theta_s^{x,V}}\right) \\
&\qquad \left((kl)^{-(1+\varepsilon)/2} \int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} e^{il\theta}\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right)\Bigg|^2 \mathrm{d}x. \quad (3.31) \\
&\leqslant C_\Gamma \sum_{kl} (kl)^{-1-\varepsilon} \int_I \left|\int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} e^{il\theta}\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right|^2 \mathrm{d}x
\end{aligned}$$

where in the second step we have applied Cauchy-Schwartz inequality and (3.29). Using that W and V are bounded, as well as the fact that

$$\left|\int_I (W(x,y) - V(x,y)) \left(\int_{\mathbb{T}} e^{il\theta}\mu_s^{y,V}(\mathrm{d}\theta)\right) \mathrm{d}y\right| \leqslant 1, \quad (3.32)$$

we conclude

$$\begin{aligned} G &\leqslant C \sup_{\|a\|_\infty, \|b\|_\infty \leqslant 1} \int_I \left| \int_I (W(x,y) - V(x,y)) (a(y) + ib(y)) \, \mathrm{d}y \right| \mathrm{d}x \\ &\leqslant C \|W - V\|_{\infty \to 1} . \end{aligned} \tag{3.33}$$

Since the norm $\|\cdot\|_{\infty\to 1}$ is equivalent to the cut-norm (A.4), the proof is concluded. □

3.4. **Proofs for the non-linear process** (2.3).

Proof of Proposition 2.1. The first part follows directly from Proposition 3.2 and Remark 3.4. The proof of (2.4) is similar to the proof of (3.11), but note that we are now integrating with respect to the randomness in U as well. □

Proof of Proposition 2.2. Let $\theta^{U,W}$ and $\theta^{U,V}$ be the two solutions to (2.3) 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 (2.3) 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

$$\begin{aligned} &\left|\theta_s^{U,W} - \theta_s^{\varphi(U),V}\right|^2 \leqslant C \int_0^s \left|F(\theta_r^{U,W}) - F(\theta_r^{\varphi(U),V})\right|^2 \mathrm{d}r \\ &+ C \int_0^s \left| \int_I W(U,y) \int_{\mathbb{T}} \left(\Gamma(\theta_r^{\varphi(U),W}, \theta) - \Gamma(\theta_r^{\varphi(U),V}, \theta)\right) \mu_r^{\varphi(y),W}(\mathrm{d}\theta) \, \mathrm{d}y \right|^2 \mathrm{d}r \\ &+ C \int_0^s \left| \int_I W(U,y) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta)(\mu_r^{y,W} - \mu_r^{\varphi(y),V})(\mathrm{d}\theta) \, \mathrm{d}y \right|^2 \mathrm{d}r \\ &+ C \int_0^s \left| \int_I (W(U,y) - V^\varphi(U,y)) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta) \mu_r^{\varphi(y),V}(\mathrm{d}\theta) \, \mathrm{d}y \right|^2 \mathrm{d}r. \end{aligned} \tag{3.34}$$

The first two integrals on the r.h.s. are bounded by $C \int_0^s \left|\theta_r^{U,W} - \theta_r^{\varphi(U),V}\right|^2 \mathrm{d}r$, using that F and Γ are Lipschitz-continuous. While the third integral in the r.h.s. can be estimated using (1.3) and the fact that $0 \leqslant W \leqslant 1$. Thus we get

$$\begin{aligned} &\left| \int_I W(U,y) \int_{\mathbb{T}} \Gamma(\theta_r^{\varphi(U),V}, \theta)(\mu_r^{y,W} - \mu_r^{\varphi(y),V})(\mathrm{d}\theta) \, \mathrm{d}y \right|^2 \\ &\leqslant \int_I D_r^2(\mu^{y,W}, \mu^{\varphi(y),V}) \, \mathrm{d}y = d_r^2\left(\mu^W, (\mu^V)^\varphi\right), \end{aligned} \tag{3.35}$$

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

$$\begin{aligned}\int_I \mathbf{E}\left[\sup_{s\in[0,t]} \left|\theta_s^{x,W} - \theta_s^{\varphi(x),V}\right|^2\right] \mathrm{d}x \leqslant C \int_0^t \int_I \mathbf{E}\left[\sup_{r\in[0,s]} \left|\theta_r^{x,W} - \theta_r^{\varphi(x),V}\right|^2\right] \mathrm{d}x\, \mathrm{d}s \\ + C\int_0^t d_s^2\left(\mu^W, (\mu^V)^\varphi\right) \mathrm{d}s + CG,\end{aligned} \tag{3.36}$$

where G is given by

$$G = \int_0^t \mathbf{E}\left[\int_I \left|\int_I (W(x,y) - V^\varphi(x,y)) \int_{\mathbb{T}} \Gamma(\theta_s^{\varphi(x),V}, \theta) \mu_s^{\varphi(y),V}(\mathrm{d}\theta)\mathrm{d}y\right|^2 \mathrm{d}x\right] \mathrm{d}s. \tag{3.37}$$

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

$$\begin{aligned} d_t^2\left(\mu^W, (\mu^V)^\varphi\right) &\leqslant \int_I \mathbf{E}\left[\sup_{s\in[0,t]} \left|\theta_s^{x,W} - \theta_s^{\varphi(x),V}\right|^2\right] \mathrm{d}x, \\ G &\leqslant C' \left\|W - V^\varphi\right\|_\square, \qquad \text{for some } C' > 0.\end{aligned} \tag{3.38}$$

Applying these bounds to (3.36) and using Gronwall's inequality twice as in the previous proof, yields

$$d_t^2\left(\mu^W, (\mu^V)^\varphi\right) \leqslant C \left\|W - V^\varphi\right\|_\square. \tag{3.39}$$

By taking the infimum with respect to $\varphi \in S_I$ and recalling the definition of the cut-distance (A.8) together with (3.9), we obtain

$$D_t(\bar\mu^W, \bar\mu^V) \leqslant \widetilde{d}_t\left(\mu^W, \mu^V\right) \leqslant C\, \delta_\square(W,V)^{1/2}. \tag{3.40}$$

The proof is concluded. □

4. Proof of Theorem 2.3

In order to prove Theorem 2.3, we couple the system (2.1) 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 (2.1).

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

$$\theta_t^i = \theta_0^i + \int_0^t F(\theta_s^i)\mathrm{d}s + \int_0^t \int_I W(U_i, y) \int_{\mathbb{T}} \Gamma(\theta_s^i, \theta)\mu_s^y(\mathrm{d}\theta)\mathrm{d}y\, \mathrm{d}s + B_t^i. \tag{4.1}$$

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*

$$\lim_{n\to\infty} \mathbb{E}\times\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,T]}\left|\theta_t^{i,n}-\theta_t^i\right|^2\right]=0. \tag{4.2}$$

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 (2.1) and (4.1). 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

$$\begin{aligned}\left|\theta_s^{i,n}-\theta_s^i\right|^2 \leqslant{}& C\int_0^s\left|F(\theta_r^{i,n})-F(\theta_r^i)\right|^2\mathrm{d}r\\&+C\int_0^s\left|\frac{1}{n}\sum_{j=1}^{n}\xi_{ij}^{(n)}\left(\Gamma(\theta_r^{i,n},\theta_r^{j,n})-\Gamma(\theta_r^i,\theta_r^j)\right)\right|^2\mathrm{d}r\\&+C\int_0^s\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)\Gamma(\theta_r^i,\theta_r^j)\right|^2\mathrm{d}r\\&+C\int_0^s\left|\frac{1}{n}\sum_{j=1}^{n}W(U_i,U_j)\Gamma(\theta_r^i,\theta_r^j)-\int_I W(I_i,y)\int_{\mathbb{T}}\Gamma(\theta_r^i,\theta)\mu_r^y(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r.\end{aligned} \tag{4.3}$$

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,

$$\begin{aligned}E\left[\frac{1}{n}\sum_{i=1}^{n}\sup_{s\in[0,t]}\left|\theta_s^{i,n}-\theta_s^i\right|^2\right]\leqslant{}& C\int_0^t E\left[\frac{1}{n}\sum_{i=1}^{n}\sup_{q\in[0,r]}\left|\theta_q^{i,n}-\theta_q^i\right|^2\right]\mathrm{d}r\\+C\int_0^t E\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)\Gamma(\theta_r^i,\theta_r^j)\right|^2\right]\mathrm{d}r&\\+C\int_0^t\frac{1}{n}\sum_{i=1}^{n}E\left[\left|\frac{1}{n}\sum_{j=1}^{n}W(U_i,U_j)\Gamma(\theta_r^i,\theta_r^j)-\int_I W(U_i,y)\int_{\mathbb{T}}\Gamma(\theta_r^i,\theta)\mu_r^y(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\right]\mathrm{d}r.&\end{aligned} \tag{4.4}$$

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

$$\mathbf{E}\left[W(U_i,U_j)\Gamma(\theta_s^i,\theta_s^j)\right]=\int_I W(U_i,y)\int_{\mathbb{T}}\Gamma(\theta_s^i,\theta)\mu_s^y(\mathrm{d}\theta)\,\mathrm{d}y \tag{4.5}$$

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

Turning to the second term, we will prove that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)\Gamma(\theta_s^i,\theta_s^j)\right|^2\right] \leqslant C\,\mathbb{E}\left[\delta_\square(\xi^{(n)},W^{(n)})\right]+o(1), \tag{4.6}$$

where $W^{(n)} := \{W(U_i,U_j)\}_{i,j=1,\dots,n}$ is a W-random graph with n vertices, see (2.10). This, together with a Gronwall argument implies that

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\sup_{s\in[0,t]}\left|\theta_t^{i,n}-\theta_t^i\right|^2\right] \leqslant C\,\mathbb{E}\left[\delta_\square(\xi^{(n)},W^{(n)})\right]+o(1) \tag{4.7}$$

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 (4.6), we use an argument similar to (3.26)–(3.29). Recall that since $\Gamma \in \mathcal{C}^{1+\varepsilon}$, it admits a Fourier series (3.28) 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 (4.6), multiplying and dividing by $(kl)^{(1+\varepsilon)/2}$, we get

$$\begin{aligned}
&E\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)\Gamma(\theta_s^i,\theta_s^j)\right|^2\right]\\
&=E\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)\sum_{k,l}\Gamma_{kl}e^{i\theta_s^i k}e^{i\theta_s^j l}\right|^2\right]\\
&\leqslant CE\left[\sum_{k,l}(kl)^{-1-\varepsilon}\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)e^{i\theta_s^i k}e^{i\theta_s^j l}\right|^2\right],
\end{aligned} \tag{4.8}$$

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

$$\mathbb{E}\left[\sup_{s_i,t_j\in\{\pm1\}}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\left(\xi_{ij}^{(n)}-W(U_i,U_j)\right)s_it_j\right|\right]. \tag{4.9}$$

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 $\hat{\delta}_\square$ (A.7), one can thus take the labeling of $\{U_i\}_{i=1,\dots,n}$ for every $n \in \mathbb{N}$, such that

$$\mathbb{E}\left[\sup_{s_i,t_j\in\{\pm1\}}\left|\frac{1}{n^2}\sum_{i,j=1}^{n}\left(\xi_{ij}^{(n)} - W(U_i,U_j)\right)s_it_j\right|\right] = \mathbb{E}\left[\hat{\delta}_\square(\xi^{(n)}, W^{(n)})\right]. \tag{4.10}$$

Using the asymptotic equivalence of $\hat{\delta}_\square$ with $\delta_\square$, see Remark 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 (2.11) is proven in Lemma A.2. We turn to the proof of the convergence of μ^n.

It is well known that the bounded Lipschitz distance, recall (1.3), 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

$$\lim_{n\to\infty}\mathbb{E}\times\mathbf{E}\left[\int f(\theta)\mu^n(\mathrm{d}\theta) - \int f(\theta)\bar{\mu}(\mathrm{d}\theta)\right] = 0, \tag{4.11}$$

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 (4.1)), it is enough to show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\times\mathbf{E}\left[\left|f(\theta^{j,n}) - f(\theta^j)\right|\right] = 0. \tag{4.12}$$

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

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\times\mathbf{E}\left[\left|f(\theta^{j,n}) - f(\theta^j)\right|\right] \leqslant \mathbb{E}\times\mathbf{E}\left[\frac{1}{n}\sum_{j=1}^{n}\sup_{t\in[0,T]}\left|\theta_t^{j,n} - \theta_t^j\right|^2\right]^{1/2}, \tag{4.13}$$

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

Appendix A. Graph convergence and random graphons

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 $1\leqslant i\leqslant j\leqslant n$.

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

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

It is well-known that this norm is equivalent to the $\ell_\infty\to\ell_1$ norm [1]

$$\|A\|_{\infty\to1} := \sup_{s_i,t_j\in\{\pm1\}}\sum_{i,j=1}^{n}A_{ij}s_it_j. \tag{A.2}$$

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

$$d_\square(\xi,\xi') := \|\xi-\xi'\|_\square. \tag{A.3}$$

A.2. **Labeled and unlabeled graphons.** Recall that $I=[0,1]$ and let $\mathcal{W} := \{W : I^2 \to \mathbb{R}$ 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 $0 \leqslant W \leqslant 1$. Let $\mathcal{W}_0$ denote the space of labeled graphons. The cut-norm of $W \in \mathcal{W}$ is defined as

$$\|W\|_\square := \max_{S,T\subset I} \left| \int_{S\times T} W(x,y)\mathrm{d}x\mathrm{d}y \right| \tag{A.4}$$

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 $L^\infty(I) \to L^1(I)$ [18, Theorem 8.11]. This is defined as

$$\|W\|_{\infty\to 1} := \sup_{\|g\|_\infty \leqslant 1} \|Wg\|_1, \tag{A.5}$$

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

The metric induced by $\|\cdot\|_\square$, or equivalently by $\|\cdot\|_{\infty\to1}$, in the space of labeled graphons $\mathcal{W}_0$ is again denoted by $d_\square(\cdot,\cdot)$. Definitions (A.1) and (A.4) 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

$$W_\xi(x,y) = \sum_{i,j=1}^n \xi_{ij}\, \mathbf{1}_{[\frac{i-1}{n},\frac{i}{n})\times[\frac{j-1}{n},\frac{j}{n})}(x,y), \qquad \text{for } x,y\in I. \tag{A.6}$$

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

$$\hat{\delta}_\square(\xi,\xi') := \min_{\hat{\xi}'} d_\square(\xi,\hat{\xi}'), \tag{A.7}$$

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

$$\delta_\square(W,V) := \min_{\varphi\in S_I} d_\square(W,V^\varphi), \tag{A.8}$$

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 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 $\delta_\square$ between W_ξ and $W_{\xi'}$. These turn out to be equivalent as the number of vertices tends to infinity* [18, Theorem 9.29]. *Formally, for every two graphs ξ,ξ' on n vertices, it holds that*

$$\delta_\square(W_\xi,W_{\xi'}) \leqslant \hat{\delta}_\square(\xi,\xi') \leqslant \delta_\square(W_\xi,W_{\xi'}) + \frac{17}{\sqrt{\log n}}. \tag{A.9}$$

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

Contrary to $d_\square$, the cut-distance $\delta_\square$ 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 [18, Theorem 9.23].

We are not going into the details of graph convergence for which we refer to the exhaustive reference [18]. 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 $\delta_\square(W_{\xi^{(n)}}, W) \to 0$ as $n \to \infty$ [18, Theorem 11.22]). We refer to the following subsection for a characterization of the convergence in probability.

A.3. **Convergence in probability.** The characterization of the convergence in distribution for a sequence of graphs has been originally given in [16]. We give here a useful notion of convergence in $\widetilde{\mathcal{W}}_0$ by means of the cut-distance $\delta_\square$, which is equivalent to the convergence in probability for graph sequences.

Lemma 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* (2.7) *holds, i.e., if and only if*

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

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

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(\delta_\square(\xi^{(n)}, W) > \varepsilon\right) = 0. \tag{A.10}$$

Observe that the sequence of positive real random variables $\{\delta_\square(\xi^{(n)}, W)\}_{n\in\mathbb{N}}$ is uniformly bounded by 1. Equation (A.10) is then equivalent to the convergence in L^1, i.e., equivalent to (2.7). □

Acknowledgements

F.C. is thankful to his supervisor Giambattista Giacomin for insightful discussions and advice. F.C. acknowledges the support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665850. F. R. N. was partially supported by the Netherlands Organisation for Scientific Research (NWO) [Gravitation Grant number 024.002.003–NETWORKS].

References

[1] N. Alon and A. Naor. Approximating the Cut-Norm via Grothendieck's Inequality. *SIAM J. Comput.*, 35:787–803, 2006.

[2] A. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, Dec. 2008. arXiv: 0805.2976.

[3] A. Basak, S. Bhamidi, S. Chakraborty, and A. Nobel. Large subgraphs in pseudo-random graphs. *arXiv:1610.03762 [math]*, Oct. 2016. arXiv: 1610.03762.

[4] E. Bayraktar, S. Chakraborty, and R. Wu. Graphon mean field systems. *arXiv:2003.13180 [math]*, Mar. 2020. arXiv: 2003.13180.

[5] 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, Feb. 2010. arXiv: 0911.1499.

[6] G. Bet, R. van der Hofstad, and J. S. H. van Leeuwaarden. Big jobs arrive early: From critical queues to random graphs. *arXiv:1704.03406 [math]*, Apr. 2017. arXiv: 1704.03406.

[7] S. Bhamidi, A. Budhiraja, and R. Wu. Weakly interacting particle systems on inhomogeneous random graphs. *Stochastic Processes and their Applications*, 129(6):2174–2206, June 2019.

[8] K. Bogerd, R. M. Castro, and R. van der Hofstad. Cliques in rank-1 random graphs: the role of inhomogeneity. *arXiv:1805.01688 [math]*, May 2018. arXiv: 1805.01688.

[9] P. E. Caines and M. Huang. Graphon Mean Field Games and the GMFG Equations. *2018 IEEE Conference on Decision and Control (CDC)*, 2018.

[10] R. Carmona, D. Cooney, C. Graves, and M. Lauriere. Stochastic Graphon Games: I. The Static Case. *arXiv:1911.10664 [math]*, Nov. 2019. arXiv: 1911.10664.

[11] 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.

[12] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, Dec. 1989.

[13] F. Coppini. Long time dynamics for interacting oscillators on graphs. *arXiv:1908.01520 [math]*, Dec. 2019. arXiv: 1908.01520.

[14] 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, Apr. 2020.

[15] S. Delattre, G. Giacomin, and E. Luçon. A Note on Dynamical Models on Random Graphs and Fokker-Planck Equations. *Journal of Statistical Physics*, 165(4):785–798, Nov. 2016.

[16] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, 28:33–61, 2008.

[17] Y. Katznelson. *An Introduction To Harmonic Analysis*. Cambridge Mathematical Library. Cambridge University Press, 3rd edition, 2004.

[18] L. Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American mathematical society edition, 2012.

[19] E. Luçon. Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *arXiv:1811.09229 [math, q-bio]*, Nov. 2018. arXiv: 1811.09229.

[20] E. Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability*, 24(5):1946–1993, Oct. 2014.

[21] G. S. Medvedev. The continuum limit of the Kuramoto model on sparse random graphs. *arXiv:1802.03787 [nlin]*, Feb. 2018. arXiv: 1802.03787.

[22] K. Oelschläger. A Martingale Approach to the Law of Large Numbers for Weakly Interacting Stochastic Processes. *The Annals of Probability*, 12(2):458–479, May 1984.

[23] G. H. Reis and R. I. Oliveira. Interacting diffusions on random graphs with diverging degrees: hydrodynamics and large deviations. *Journal of Statistical Physics*, 2019. arXiv: 1807.06898.

[24] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths. The Kuramoto model in complex networks. *Physics Reports*, 610:1–98, Jan. 2016.

[25] A.-S. Sznitman. Topics in propagation of chaos. In P.-L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX - 1989*, volume 1464, pages 165–251. Springer Berlin Heidelberg, 1991.

Dipartimento di Matematica e Informatica "Ulisse Dini", Università degli Studi di Firenze, Firenze, Italy

Laboratoire de Probabilités Statistique et Modélisation (UMR 8001), Université de Paris, Sorbonne Paris Cité, Paris, France

Dipartimento di Matematica e Informatica "Ulisse Dini", Università degli Studi di Firenze, Firenze, Italy, and Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, the Netherlands