WEAKLY INTERACTING OSCILLATORS ON DENSE RANDOM GRAPHS

GIANMARCO BET [iD],* ** AND

FABIO COPPINI,* *** *Università degli Studi di Firenze*

FRANCESCA ROMANA NARDI,* *Università degli Studi di Firenze and Eindhoven University of Technology*

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 to a graphon. The limit is the solution of a non-linear Fokker–Planck equation weighted by the (possibly random) graphon limit. In contrast with the existing literature, our analysis focuses on both deterministic and random graphons: no regularity assumptions are made on the graph limit and we are able to include general graph sequences such as exchangeable random graphs. Finally, we identify the sequences of graphs, both random and deterministic, for which the associated empirical measure converges to the classical McKean–Vlasov mean-field limit.

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

2020 Mathematics Subject Classification: Primary 60K35; 05C80
Secondary 82B20; 60H20; 35Q84

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. [23], [26]), in which all the particles are connected to 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. Applications of these models include mean-field games [9, 10], synchronization phenomena [2], neuroscience [25], and statistical mechanics [27], among others.

The first mathematically rigorous results on McKean–Vlasov particle systems and graphs appeared only recently [7, 16]: these consider certain graph sequences with diverging average degree. Under suitable homogeneity conditions on the degrees, the system is described by the

classical mean-field limit [26] as the number of particles tends to infinity. The cited 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 system dynamics to the degree inhomogeneity in the underlying graph? The existing literature (see e.g. [4], [15], [20]) focuses on a large class of particle systems and different graph regimes, i.e. from dense to almost sparse graphs, yet they are unable to fully include the well-known class of exchangeable random graphs [17] in their result.

Our aim is to prove that a law of large numbers for weakly interacting particle systems can be proved for any dense random graph sequence, including exchangeable random graphs. To keep the focus on the random graph limit viewpoint, we restrict the class of interacting particles to weakly interacting oscillators, a class of systems that shows a remarkably rich behavior [2, 25]. We note that, at the cost of more technicalities, our result could be extended to more general particle systems. The main result of this work is a law of large numbers for the empirical measure under the only assumption that the graph sequence converges in probability in the space of graphons; see [19] for a complete tour on graphons. Namely, we tackle the challenging case of a random graph limit, which includes pseudo-random graphs (see e.g. [3], [12]) and exchangeable random graphs (see e.g. [17]). To the authors' knowledge, this is the first result in the literature that explicitly links random graph limits to empirical measures.

As a byproduct, we are able to identify deterministic and random graph sequences for which the particle system behavior is approximately mean-field. We also provide an example of a particle system undergoing a phase transition, where the critical threshold depends on the randomness of the graph limit; see Proposition 3.2 and the example below.

1.1. Unlabeled graphons and empirical measures

To better illustrate our contribution, a clarification of the term graphon is in order. In the literature of McKean–Vlasov particle systems, the term *graphon* is usually employed for what the classical graph limits theory (we refer to [17], [19]) calls a *labeled graphon*, i.e. a symmetric, bounded, and measurable function $W: [0, 1] \times [0, 1] \to \mathbb{R}$. The space [0,1] is known as the label space and it is used as particle index: roughly speaking, every $x \in [0, 1]$ corresponds to a particle and the interaction strength between two particles x and y is given by the value of $W(x, y)$.

The mathematical space of graphons is not merely given by labeled graphons but constructed via certain equivalence classes, that is,

$$\widetilde{W} = \{W^\varphi \text{ with } \varphi: [0, 1] \to [0, 1] \text{ measure-preserving map}\},$$

where $W^\varphi(x, y) = W(\varphi(x), \varphi(y))$ for every x and y in [0,1]. The measure-preserving map φ corresponds to a label permutation in a finite graph (note that φ need not be invertible in the infinite setting; see [19, Proposition 7.10]). In other words, $\widetilde{W}$ contains the connection information of some labeled graphon W, while being independent of its labeling. Following [17], we refer to this object as an *unlabeled graphon*. The space of graphons is the space of unlabeled graphons with a suitable distance; see Section 2.3 for further details.

The notion of labeling is relevant when studying interacting particle systems. On the one hand, we are interested in understanding the correlation between two particles and need their labels to be fixed to suitably describe it. On the other hand, we want to describe the macroscopic behavior, e.g. by studying the empirical measure, and there the labels do not play any role. Hence, depending on the question we are interested in, the suitable modeling choice might be labeled or unlabeled graphs. From a graphon viewpoint, considering unlabeled graphons is not a technical recourse but in fact represents the cornerstone of the theory and the key ingredient

to many of its interesting properties. As will become clear later, when describing the dynamics of a particle population, the concept of labeled graphon is not satisfactory: different labeled graphons can lead to exactly the same dynamics!

Our contribution stems from the previous considerations. We refer to Section 3.3 for a comparison with the existing literature.

1.2. Organization

We now present the set-up and notation used, as well as the various distances between probability measures that will be considered below. In Section 2 we give a brief introduction to graphons and the main tools needed for this work.

In Section 3 we define the interacting particle system and the associated non-linear process. Existence, uniqueness, and stability results for the non-linear process are presented immediately afterwards; see Theorem 3.1. Our main result, Theorem 3.2, is given in Section 3.2. Exchangeable random graphs are then discussed together with a propagation of chaos result; see Section 3.4. Section 3.5 is devoted to the comparison with the classical mean-field behavior and to a few important consequences of Theorem 3.2; the discussion is supported by two explanatory examples.

In Section 4 we focus on the non-linear process. In particular, we discuss its relationship with other characterizations already known in the literature. The proofs of the main results for the non-linear process are given in Section 4.4.

Section 5 contains the proof of Theorem 3.2. Finally, in Appendix A we give a useful characterization of convergence in probability for random graph sequences.

2. 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 $\{\mathcal{F}.\}$ is a filtration satisfying the usual conditions.

We use two different notations for expressing conditional probabilities: the one referring to Brownian motions and initial conditions is denoted by **P** and its expectation by **E**; the one referring to the randomness in the graph sequences, and/or in its limit object, is denoted by $\mathbb{P}$ and its expectation by $\mathbb{E}$.

The interval $I := [0, 1]$ represents the space of (continuous) labels. We study functions with values in the one-dimensional torus denoted by $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$. The space of continuous functions from $[0, T]$ to $\mathbb{T}$ is denoted by $\mathcal{C}([0, T], \mathbb{T})$ and is endowed with the supremum norm. For $\alpha > 0$, the space of α-Hölder-continuous functions from $\mathbb{T}^2$ to $\mathbb{R}$ is denoted by $\mathcal{C}^\alpha(\mathbb{T}^2)$. Given a separable metric space E, the space of probability measures over E is denoted by $\mathcal{P}(E)$.

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.

2.1. Distance between probability measures

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

$$D_T(\bar{\mu}, \bar{\nu}) := \inf_{m \in \gamma(\bar{\mu}, \bar{\nu})} \left\{ \int \sup_{t \in [0,T]} |x_t - y_t|^2 m(\mathrm{d}x, \mathrm{d}y) \right\}^{1/2}, \tag{2.1}$$

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 (2.1) can be rewritten as

$$D_T(\bar{\mu}, \bar{\nu}) = \inf_{X,Y} \left\{ \mathbf{E} \left[\sup_{t \in [0,T]} |X_t - Y_t|^2 \right] : \mathcal{L}(X) = \bar{\mu}, \ \mathcal{L}(Y) = \bar{\nu} \right\}^{1/2}, \tag{2.2}$$

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

$$\sup_f \left| \int_{\mathbb{T}} f(\theta) \, \bar{\mu}_s(\mathrm{d}\theta) - \int_{\mathbb{T}} f(\theta) \, \bar{\nu}_s(\mathrm{d}\theta) \right| \leq D_s(\bar{\mu}, \bar{\nu}), \tag{2.3}$$

where the supremum is taken over all 1-Lipschitz functions from $\mathbb{T}$ to $\mathbb{R}$. Observe that these definitions also make sense with $T = 0$ and $\mathcal{C}([0, T], \mathbb{T})$ replaced by $\mathbb{T}$.

2.2. Distance between finite graphs

We denote $[n] := \{1, \ldots, n\}$ for $n \in \mathbb{N}$. Let ξ be a 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 an $n \times n$ real matrix. The *cut-norm* of A is a well-known norm for (not necessarily square) finite matrices (see e.g. [1], [19]); it is defined as

$$\|A\|_\square := \frac{1}{n^2} \max_{S,T \subset [n]} \left| \sum_{i \in S, j \in T} A_{ij} \right|. \tag{2.4}$$

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

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

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

$$d_\square(\xi, \xi') := \|\xi - \xi'\|_\square.$$

2.3. Labeled and unlabeled graphons

The following definitions come from the book by Lovász [19], whose notation we adopt. Recall that $I = [0, 1]$ and let

$$\mathcal{W} := \{W \colon I^2 \to \mathbb{R} \text{ bounded symmetric and measurable}\}$$

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 zero Lebesgue measure on I^2. A graphon is a kernel W such that $0 \leq W \leq 1$. Let $\mathcal{W}_0$ denote the space of graphons. The cut-norm of $W \in \mathcal{W}$ is defined as

$$\|W\|_\square := \max_{S,T \subset I} \left| \int_{S \times T} W(x, y) \, \mathrm{d}x \, \mathrm{d}y \right|, \tag{2.5}$$

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 $L^\infty(I) \to L^1(I)$ [19, Theorem 8.11]. This is defined as

$$\|W\|_{\infty \to 1} := \sup_{\|g\|_\infty \leq 1} \|Wg\|_1,$$

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 $\|\cdot\|_{\infty\to 1}$, in the space of graphons $\mathcal{W}_0$ is again denoted by $d_\square(\cdot, \cdot)$. Definitions (2.4) and (2.5) are consistent in the sense that to each graph ξ is associated a graphon $W_\xi \in \mathcal{W}_0$ such that $\|\xi\|_\square = \|W_\xi\|_\square$. The graphon W_ξ is usually defined almost everywhere as

$$W_\xi(x, y) = \sum_{i,j=1}^n \xi_{ij} \mathbf{1}_{[\frac{i-1}{n}, \frac{i}{n}) \times [\frac{i-1}{n}, \frac{i}{n})}(x, y) \quad \text{for } x, y \in I. \tag{2.6}$$

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

$$\hat{\delta}_\square(\xi, \xi') := \min_{\hat{\xi}'} d_\square(\xi, \hat{\xi}'), \tag{2.7}$$

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

$$\delta_\square(W, V) := \min_{\varphi \in S_I} d_\square(W, V^\varphi), \tag{2.8}$$

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 2.1. 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 [19, Theorem 9.29]. Formally, for every two graphs ξ, ξ' on n vertices, it holds that

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

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 graphons can be zero. This leads to the definition of the *unlabeled* graphon $\widetilde{W}$ associated to W. For a 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 unlabeled graphons. A celebrated result of graph limits theory is that $(\widetilde{\mathcal{W}}_0, \delta_\square)$ is a compact metric space [19, Theorem 9.23].

We will not go into the details of graph convergence, for which we refer to the exhaustive source [19]. 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 $n \to \infty$ [19, Theorem 11.22].

3. The models and main results

3.1. The models

We introduce our two main models: a weakly interacting particle system (3.1) and a non-linear process (3.3).

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

$$\begin{cases} d\theta_t^{i,n} = F(\theta_t^{i,n}) \, dt + \dfrac{1}{n} \sum_{j=1}^{n} \xi_{ij}^{(n)} \, \Gamma(\theta_t^{i,n}, \theta_t^{j,n}) \, dt + dB_t^i, & 0 < t < T, \\[2mm] \theta_0^{i,n} = \theta_0^i, & i \in \{1, \ldots, n\}, \end{cases} \tag{3.1}$$

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

Remark 3.1. At the cost of more technical details, one can relax the hypothesis on the initial conditions and require independent but not identically distributed $\{\theta_0^i\}_{i \in \mathbb{N}}$. Instead we keep the proofs simple, following the classical propagation of chaos arguments; see e.g. [26].

Many interesting examples of interacting oscillators fit this framework, such as the Kuramoto model, the plane rotator model, and other generalizations; see e.g. [16, § 1.2], [3], and also Section 3.5. The applications of interacting oscillators range from modeling synchronization phenomena (see e.g. [2]) to neuroscience (e.g. [25]), as well as for studying statistical mechanics properties of complex systems [27].

We are interested in studying the empirical measure associated to (3.1). This is defined as the (random) probability measure on $\mathbb{T}$ such that

$$\mu_t^n := \frac{1}{n} \sum_{j=1}^{n} \delta_{\theta_t^{j,n}} \tag{3.2}$$

for every $t \in [0, T]$. Alternatively, one can also consider the empirical measure μ^n of the trajectories, that is, on

$$\mu^n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\theta^{j,n}} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T})),$$

where $\theta^{j,n} = (\theta_t^{j,n})_{t \in [0,T]} \in \mathcal{C}([0, T], \mathbb{T})$ is the trajectory on $[0, T]$ of particle j for $j = 1, \ldots, n$. The two are equally considered in the literature; see e.g. [15, Remark 1.1]) for the relation between the two.

3.1.2. *The non-linear process.* 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:

$$\begin{cases} \theta_t = \theta_0 + \displaystyle\int_0^t F(\theta_s) \, ds + \int_0^t \int_I W(U, y) \int_{\mathbb{T}} \Gamma(\theta_s, \theta) \mu_s^y(d\theta) \, dy \, ds + B_t, \\[2mm] \mu_t^y = \mathcal{L}(\theta_t \mid U = y) \quad \text{for } y \in I, \ t \in [0, T], \end{cases} \tag{3.3}$$

where the initial condition θ_0 has law $\mathcal{L}(\theta_0) = \bar{\mu}_0$ and is taken to be independent of U. The Brownian motion B is independent of the previous sequence $\{B^i\}_{i \in \mathbb{N}}$, of U, of the initial condition θ_0, and also independent of the sequence $\xi^{(n)}$.

The next theorem on existence and uniqueness of the solution to equation (3.3) is proved in Section 4, together with well-posedness with respect to W. For the latter, see Remark 4.4.

Theorem 3.1. *Suppose that F and Γ are bounded, uniformly Lipschitz functions. For every uniform random variable U on I, there exists a unique pathwise solution to* (3.3).

If $\bar{\mu} \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$ denotes the law of the solution $\theta = \{\theta_t\}_{t \in [0, T]}$ 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:

$$\partial_t \bar{\mu}_t(\theta) = \frac{1}{2} \partial_\theta^2 \bar{\mu}_t(\theta) - \partial_\theta [\bar{\mu}_t(\theta) F(\theta)] - \partial_\theta \left[\int_{I \times I} W(x, y) \, \mu_t^x(\theta) \int_{\mathbb{T}} \Gamma(\theta, \widetilde{\theta}) \, \mu_t^y(\mathrm{d}\widetilde{\theta}) \, \mathrm{d}y \, \mathrm{d}x \right],$$
(3.4)

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

3.2. Main result

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 3.2. *Assume the hypothesis of Theorem 3.1 and further suppose that $\Gamma \in \mathcal{C}^{1+\varepsilon}(\mathbb{T}^2)$ for some $\varepsilon > 0$. 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*

$$\lim_{n \to \infty} \mathbb{E} \big[\delta_\square \big(\xi^{(n)}, W \big) \big] = 0.$$
(3.5)

Suppose that the initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ are i.i.d. random variables with law $\bar{\mu}_0$ and are independent of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$. Then

$$\mu^n \longrightarrow \bar{\mu} \quad \text{in } \mathbf{P} \times \mathbb{P}\text{-probability, as } n \to \infty,$$
(3.6)

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, that is, for almost every $\omega \in \Omega$, $\bar{\mu}(\omega)$ solves equation (3.4) starting from $\bar{\mu}_0$, with graphon $W(\omega)$.

The hypothesis (3.5) on the graph convergence is very general. Indeed, it holds for any $\xi^{(n)}$ in $\mathcal{W}$, meaning that $\xi^{(n)}$ can take values in $\mathbb{R}$ rather than $\{0, 1\}$; recall the definition of kernels in Section 2.3. Moreover, the sequence $\xi^{(n)}$ can be deterministic or random; if it is random, we also cover the convergence in probability to a random graphon limit W. In this last case, well-posedness and measurability of equation (3.4) are granted by Theorem 4.1.

Looking at the proof of Theorem 3.2, 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 3.2 also holds if $\{\theta_0^i\}_{i \in \mathbb{N}}$ is independent of the randomness in W but not necessarily 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 Section 3.5.

Requiring that $\Gamma \in \mathcal{C}^{1+\varepsilon}(\mathbb{T}^2)$ for $\varepsilon > 0$ has no specific physical meaning. It is a technical assumption to ensure the convergence of estimates involving the Fourier coefficients of Γ; we refer to the proof of Theorem 3.2 in Section 5.

3.3. Comparison with the existing literature

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, 20, 24] and

in the deterministic one [11, 22]. Most of these results vary in terms of the class of particle systems considered, the notion of convergence of the graph sequence, or the hypothesis on the initial condition. To our knowledge, no result in the literature considers graph sequences issued from random graphons. In addition, we are not aware of any result directly using the cut-distance as we do.

We focus on dense graph sequences where, roughly speaking, the average degree in the graph is proportional to the number of vertices. All the cited results are established in more generality with respect to the graph density, i.e. they consider random graph sequences in different regimes, from dense to almost sparse. There is no common agreement on the terminology used for dense/sparse graphs; we refer to [13] for a presentation of the subject.

In the deterministic setting, Medvedev and co-authors (see [11], [22], and references therein) were the first to consider random graph sequences arising from a graphon. They focus on a restricted class of deterministic models, as the (deterministic) Kuramoto model. In the stochastic setting, Luçon [20] consider particle systems defined on $\mathbb{R}^d$ with additive noise and a transport term in the dynamics; the initial conditions are assumed to be independent but not identically distributed. The hypothesis on the convergence of the graph sequence is somewhat involved as it is not expressed in terms of the cut-norm: the labeled graph limit must satisfy some extra regularity assumption, such as continuity or integrability. Oliveira and Reis [24] focus on a system of interacting particles on $\mathbb{R}$ where the dynamics is defined by means of a Hamiltonian, and the initial conditions are taken to be i.i.d. They establish a large deviation principle for the empirical measure by explicitly using the cut-norm for the convergence of graph sequences. The proofs are based on large deviation techniques. The labeled graph limit must be Lipschitz.

Bayraktar *et al.* [4] consider interacting particle systems on $\mathbb{R}^d$ with multiplicative noise. The graph sequences arise from deterministic graphons under regularity assumptions that are comparable to ours (namely, only measurability for establishing the law of large numbers). They establish a law of large numbers and a propagation of chaos property. Their proof is based on careful trajectorial estimates as well as a polynomial representation of the interaction, suitable for the use of the cut-norm. They do not consider random graph limits.

3.4. Applications to exchangeable graphs

Recall that an exchangeable random graph $\xi = \{\xi_{ij}\}_{i,j \in \mathbb{N}}$ (see [17]) is an infinite array of jointly exchangeable binary random variables, that is, it satisfies

$$\mathbb{P}(\xi_{ij} = e_{ij}, \ 1 \leq i, j \leq n) = \mathbb{P}(\xi_{ij} = e_{\sigma(i)\sigma(j)}, \ 1 \leq i, j \leq n) \tag{3.7}$$

for all $n \in \mathbb{N}$, all permutations σ on $\{1, \ldots, n\}$, and all $e_{ij} \in \{0, 1\}$.

Remark 3.2. Any finite deterministic graph ξ leads to an exchangeable random graph by performing a uniform random sampling on its associated graphon W_ξ; see (2.6) and [19, §10].

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

$$\xi_{ij}^W = W(U_i, U_j), \tag{3.8}$$

where $\{U_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. uniform random variables on I. As recalled by the next theorem, the converse 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 [17] and references therein. We recall their main result here.

Theorem 3.3. ([17, *Theorem 5.3*], [19, *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 $W \in \mathcal{W}_0$. Moreover, let $\xi^{(n)} := \{\xi_{ij}\}_{i,j=1,\dots,n}$ for every $n \in \mathbb{N}$. It holds that*

$$\xi^{(n)} \longrightarrow W \quad \mathbb{P}\text{-a.s. in } \widetilde{\mathcal{W}}_0, \tag{3.9}$$

as $n \to \infty$.

We are now ready to state the main corollary of Theorem 3.2, which deals with exchangeable random graphs.

Corollary 3.1. *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 3.3. Assume that the initial conditions $\{\theta_0^i\}_{i \in \mathbb{N}}$ are i.i.d. and independent of $\{\xi^{(n)}\}_{n \in \mathbb{N}}$. Then*

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

where $\bar{\mu}$ is the solution to (3.4) starting from $\bar{\mu}_0$ with graphon W.

We say that $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of exchangeable graphs when the random variables $\{\xi_{ij}^{(n)}\}_{i,j=1,\dots,n}$ are exchangeable for each $n \in \mathbb{N}$. Observe that ξ is not necessarily an exchangeable random graph as in (3.7). Whenever $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of exchangeable graphs, the particles $\{\theta^{i,n}\}_{i=1,\dots,n}$ are exchangeable as well (recall that the initial conditions are assumed i.i.d.) and, in particular, their joint distribution is symmetric, i.e. invariant under permutation of the labels. Observe that this is not true when the graph is not exchangeable. A classical result by Sznitman [26, 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 (3.6), we can thus deduce a propagation of chaos statement for the particle system (3.1). This is illustrated in the next proposition.

Proposition 3.1. *Assume the hypothesis of Theorem 3.2. If $\xi = \{\xi^{(n)}\}_{n \in \mathbb{N}}$ is a sequence of exchangeable graphs, then for every $k \in \mathbb{N}$,*

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

where $\mathcal{L}$ stands for the law in $\mathbf{P} \times \mathbb{P}$-probability and $\bar{\mu}$ is the solution to (3.4).

We omit the proof of Proposition 3.1.

3.5. Mean-field behavior and two explanatory examples

Theorem 3.2 allows for a better understanding of the relationship between random graph sequences and the behavior of the empirical measure.

(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 3.2 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 (3.1) on a sequence of complete graphs, i.e. $\xi_{ij}^{(n)} \equiv 1$ for every i, j, and n. It is well known [23, 26] 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:

$$\partial_t \rho_t(\theta) = \frac{1}{2} \partial_\theta^2 \rho_t(\theta) - \partial_\theta[\rho_t(\theta) F(\theta)] - p\, \partial_\theta \left[\rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \widetilde{\theta})\, \rho_t(\mathrm{d}\widetilde{\theta}) \right], \qquad (3.10)$$

with initial condition $\bar{\mu}_0$ and $p = 1$. Existence and uniqueness for the solution to (3.10) hold under our assumptions on F, Γ, and $\bar{\mu}_0$; see again [23] and [26].

Suppose that the graph sequence is converging to a deterministic limit; we discuss the case of a random limit after the first example. Theorem 3.2 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 in the limit satisfies equation (3.10) with corresponding p. Since the convergence of $\xi^{(n)}$ to a non-constant graphon gives rise to equation (3.4), which is – at least formally – different from (3.10), we conclude that if the sequence $\xi^{(n)}$ converges to a constant graphon, then the limit of μ^n is formally mean-field. The graphs with such asymptotic behavior are known in the literature as pseudo-random graphs; see [3], [12], and [19, §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 next example.

3.5.1. *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 i.i.d. copies of g. Conditionally on $\{g_i\}_{i \in \mathbb{N}}$, $\xi_{ij}^{(n)}$ is defined as

$$\xi_{ij}^{(n)} \sim \mathrm{Ber}(g_i g_j), \quad \text{independently for each } 1 \le i < j \le n. \qquad (3.11)$$

The graph $\xi^{(n)}$ is the dense analogue of the inhomogeneous random graph, also known as a rank-1 model; see e.g. [6] and [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 are formed by particle i. We expect that assigning different distributions to g leads to different behaviors for the empirical measure (3.2).

The construction made in (3.11) yields a binary array $\{\xi_{ij}^{(n)}\}_{i,j=1,\ldots,n}$ of exchangeable random variables. In particular, the edges have the same expected value $\mathbb{E}[\xi_{ij}^{(n)}] = p$, for every $i \ne j$. We are interested in comparing the empirical measure of the system (3.1) defined on the graph (3.11) to the empirical measure of the corresponding *annealed* system that is obtained from (3.1) by replacing $\xi_{ij}^{(n)}$ with their expected value. More precisely, the annealed system is defined as the solution to

$$\mathrm{d}\theta_t^{i,n} = F(\theta_t^{i,n})\, \mathrm{d}t + \frac{p}{n} \sum_{j=1}^{n} \Gamma(\theta_t^{i,n}, \theta_t^{j,n})\, \mathrm{d}t + \mathrm{d}B_t^i, \qquad (3.12)$$

for which the asymptotic behavior is known to be the mean-field limit (3.10).

The behavior of system (3.1) on the graph sequence (3.11) is described in the limit by (3.12) only when g is deterministic and equal to $\sqrt{p}$. Recall the definition of a W-random graph given in (3.8): we see that $\xi^{(n)}$ is a W_g-random graph with

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

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 (3.4) with $W = W_g$. Theorems 3.2 and 4.1 imply that if W_g is arbitrarily close to the constant graphon p in the cut-distance, i.e. if $\mathrm{Var}[g] \ll 1$, then the empirical measure of the system associated to $\xi^{(n)}$ is arbitrarily close to the mean-field limit of the annealed system (3.12). In this case, $\{\xi^{(n)}\}_{n\in\mathbb{N}}$ is close to an Erdös–Rényi graph sequence, for which the mean-field behavior is already known; see [15]. Finally, observe that 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 to the same limit W_g, $\mathbb{P}$-a.s. in the realization of the Bernoulli random variables and possibly at the cost of requiring some regularity on W_g; see [19, §11.4].

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 (3.1) does not yield any *averaging property* on the empirical measure μ^n. Moreover, adding the extra randomness through Bernoulli random variables in (3.11) 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 $\xi^{(n)}$. 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 (3.10). One might be led to conjecture that it is possible to recover the mean-field behavior by, for example, 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 in their own right, as pointed out in (2) above.

3.5.2. *Example II: random mean-field behavior.* Consider the growing preferential attachment graph ξ_{pa} constructed iteratively as follows; see also [19, Example 11.44]. Begin with a single node and, assuming that at the nth step there are already n nodes, create a new node with label $n + 1$ and connect it to each node $i \in \{1, \ldots, n\}$ 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 $\xi_{\mathrm{pa}}^{(n+1)}$.

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 3.2. ([19, Proposition 11.45].) *With probability* 1, *the sequence* $\{\xi_{\mathrm{pa}}^{(n)}\}_{n\in\mathbb{N}}$ *converges to a random constant graphon.*

Consider a particle system defined on the graph sequence $\{\xi_{\mathrm{pa}}^{(n)}\}_{n\in\mathbb{N}}$. The empirical measure converges to the solution of equation (3.10) with a random p. In other words, μ^n converges to a random mean-field limit. Integrating (3.10) 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

$$\partial_t \bar{\rho}_t(\theta) = \frac{1}{2} \partial_\theta^2 \bar{\rho}_t(\theta) - \partial_\theta[\bar{\rho}_t(\theta) F(\theta)] - \partial_\theta\left[\mathbb{E}\left[p \, \rho_t(\theta) \int_{\mathbb{T}} \Gamma(\theta, \widetilde{\theta}) \rho_t(\mathrm{d}\widetilde{\theta})\right]\right] \qquad (3.13)$$

for $t \in [0, T]$. Note that (3.13) 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 satisfy the mean-field limit, i.e. it is not a solution to (3.10) with some deterministic $p \in [0, 1]$. By definition, $\bar{\rho}$ is a mixture of mean-field limits, weighted by the distribution of p.

To have an intuitive understanding of what $\bar{\rho}$ may look like, consider the stochastic Kuramoto model without natural frequencies [5], [14] defined on the sequence $\xi_{\mathrm{pa}}^{(n)}$. The model is defined as the solution to

$$\mathrm{d}\theta_t^{i,n} = \frac{K}{n} \sum_{j=1}^n \xi_{ij}^{(n)} \sin(\theta_t^{j,n} - \theta_t^{i,n}) \, \mathrm{d}t + \mathrm{d}B_t^i \qquad (3.14)$$

for $i = 1, \ldots, n$ and $t \in [0, T]$. It corresponds to (3.1) with the choices $F \equiv 0$ and $\Gamma(\theta, \psi) = -K \sin(\theta - \psi)$. An application of Theorem 3.2 and Proposition 3.2 implies that the empirical measure of (3.14) converges to the solution of

$$\partial_t \rho_t(\theta) = \frac{1}{2} \partial_\theta^2 \rho_t(\theta) + pK \partial_\theta[\rho_t(\theta)(\sin * \rho_t)(\theta)], \qquad (3.15)$$

where $*$ stands for the convolution operator. It is well known that the system (3.15) 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 $\mathbb{T}$ ($0 \leq pK < 1$). The solution to equation (3.15) can be written down explicitly (see again [5], [14]) 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 (3.15) for some fixed $p \in [0, 1]$.

4. The non-linear process

We introduce a non-linear process (4.4) which has already been considered in the literature [4, 11, 20, 21, 24] as the natural candidate if the particles in (3.1) 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 [20].

Contrary to our setting, some regularity in the – now labeled – graphon is usually assumed to show the convergence of the empirical measure (3.2). We will exploit the non-linear process with fixed labels (4.4) to better understand (3.3) and to establish existence and uniqueness.

Before introducing (4.4), 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 (4.2). This will allow us to prove Theorem 4.1, where we establish that the empirical measure is Hölder-continuous with respect to the underlying graphon.

4.1. 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, that is,

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

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

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

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

$$d_T(\mu, \nu) := \left(\int_I D_T^2(\mu^x, \nu^x) \, \mathrm{d}x\right)^{1/2}. \tag{4.1}$$

Remark 4.1. Observe that the previous expression makes sense, as

$$I \ni x \mapsto \mu^x \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$$

is a λ-measurable function. This is because we consider $\mu \in \mathcal{P}(I \times \mathcal{C}([0, T], \mathbb{T}))$. If one considers $\mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))^I$, as done in [4], then an extra assumption is required, namely that every $\eta = \{\eta^x\}_{x \in I} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))^I$ is λ-measurable in $x \in I$.

Remark 4.2. 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, that is,

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

and

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

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

$$\mu \sim \nu \text{ if and only if there exists } \varphi \in S_I \text{ such that } \mu^x = \nu^{\varphi(x)}, x\text{-a.s.} \tag{4.2}$$

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

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

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 $\varphi \in S_I$,

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

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

$$D_T(\bar{\mu}, \bar{\nu}) \leq \widetilde{d}_T(\mu, \nu). \tag{4.3}$$

4.2. 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

$$
\begin{cases}
\theta_t^x = \theta_0^x + \int_0^t F(\theta_s^x)\, \mathrm{d}s + \int_0^t \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta_s^x, \theta) \mu_s^y(\mathrm{d}\theta)\, \mathrm{d}y\, \mathrm{d}s + B_t^x, \\
\mu_t^x = \mathcal{L}(\theta_t^x) \quad \text{for } x \in I,\ t \in [0, T],
\end{cases}
\tag{4.4}
$$

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 i.i.d. Brownian motions independent of $\{\theta_0^x\}_{x \in I}$. The following proposition shows existence and uniqueness for the solution of (4.4). The proof follows a classical argument by Sznitman [26] and is postponed to Section 4.3.

Proposition 4.1. *There exists a unique pathwise solution $\theta = \{\theta^x\}_{x \in I}$ to (4.4). If $\nu^x \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{T}))$ denotes the law of θ^x for $x \in I$, then $I \ni x \mapsto \nu^x \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$ is Lebesgue-measurable. For every $x \in I$, ν^x satisfies the following non-linear Fokker–Planck equation in the weak sense:*

$$
\partial_t \mu_t^x(\theta) = \frac{1}{2} \partial_\theta^2 \mu_t^x(\theta) - \partial_\theta [\mu_t^x(\theta) F(\theta)] - \partial_\theta \left[\mu_t^x(\theta) \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta, \theta') \mu_t^y(\mathrm{d}\theta')\, \mathrm{d}y \right]
\tag{4.5}
$$

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], [20], [21], and [24].

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 systems much more general than those considered 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 4.2. *Assume the hypothesis of Theorem 3.2. There exists a positive constant C, depending only on Γ and T, such that if μ^W and μ^V denote the laws of the solutions to (4.4) with $W \in \mathcal{W}_0$ and $V \in \mathcal{W}_0$ respectively, then*

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

The proof is based on classical trajectorial estimates and Fourier analysis. As Γ is a function on the torus, it is possible to apply Fourier-type arguments to factorize it into its two components; this is the key point to make the graphon norm $\|\cdot\|$ appear and prove Proposition 4.2. The full proof is postponed to Section 4.3.

The above proposition can be translated in terms of the cut-distance $\delta_\square$ (recall (2.8)) and the space of graphons $\widetilde{\mathcal{W}}_0$.

Theorem 4.1. *Assume the hypothesis of Theorem 3.2. Then there exists a positive constant C, depending only on Γ and T, such that if $\bar{\mu}^W$ and $\bar{\mu}^V$ denote the laws of the solutions to equation (3.3) associated with graphons W and V respectively, it holds that*

$$
D_T\big(\bar{\mu}^W, \bar{\mu}^V\big) \le C\, \delta_\square(W, V)^{1/2}.
\tag{4.6}
$$

The proof of Theorem 4.1 is a straightforward consequence of Proposition 4.2; it is given in Section 4.4.

The 2-Wasserstein distance D_T (recall the definition (2.1)) could be replaced by the p-Wasserstein distance in (2.1) for $p \geq 1$. This would lead to a Hölder exponent in (4.6) as large as $1/p$. We stick to $p = 2$ but any choice could be possible, modulo a different constant C.

Observe that Theorems 3.1 and 4.1 imply that the following mapping is continuous:

$$\Psi \colon (\widetilde{\mathcal{W}}_0, \delta_\square) \to (\mathcal{C}([0, T], \mathcal{P}(\mathbb{T})), D_T),$$
$$W \mapsto \bar{\mu}^W,$$

where $\bar{\mu}^W$ is the law of θ solving equation (3.3) 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 $\omega \in \Omega$, $\bar{\mu}^W(\omega) = \bar{\mu}^{W(\omega)}$.

Remark 4.3. We point out that Theorem 4.1 allows us to conclude that two solutions to equation (3.4) are close as probability measures if the corresponding graphons are close in $\widetilde{\mathcal{W}}_0$. However, whether two different graphons can lead to similar behaviors in the particle system is still not clear. In other words, we are not able to provide any lower bound complementary to the upper bound given in equation (4.6). To our knowledge, this aspect may be model-dependent and needs further investigation.

4.2.1. *Relation between label and unlabeled non-linear processes.* 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 (3.4) is given by $\bar{\mu} = \int_I \mu^x \, \mathrm{d}x$, where μ^x is the law of θ^x solving (4.4) with initial condition μ_0^x and labeled graphon W. In other words, θ has the same law as θ^U solution to (4.4), 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 depends on neither the representative W nor μ_0.

Remark 4.4. Let $\varphi \in S_I$, i.e. φ is an invertible measure-preserving map from I to itself, and let $\nu = \{\nu^x\}_{x \in I}$ be the law of $\{\theta^{\varphi(x)}\}_{x \in I}$ solving (4.4). By a change of variables, $\theta^{\varphi(x)}$ solves

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

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

$$\psi_t^x = \theta_0^{\varphi(x)} + \int_0^t F\big(\psi_s^x\big) \, \mathrm{d}s + \int_0^t \int_I V(x, y) \int_{\mathbb{T}} \Gamma\big(\psi_s^x, \theta\big) \nu_s^y(\mathrm{d}\theta) \, \mathrm{d}y \, \mathrm{d}s + B_t^{\varphi(x)}, \qquad (4.7)$$

which has the same law as (4.4) with labeled graphon V and initial conditions $\{\theta^{\varphi(x)}\}_{x \in I}$.

Observe that the laws ν and μ associated to (4.7) and (4.4), 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 (3.3) is also equivalent to ψ^U, where ψ^x solves (4.7), and U is uniformly distributed on I.

4.3. Proofs for the non-linear process (4.4) with fixed labels

Proof of Proposition 4.1. The proof follows a classical argument given in [26, Lemma 1.3]. Consider $\nu \in \mathcal{M}_T$ and $\{\theta^{x,\nu}\}_{x \in I}$ solving

$$\theta_t^{x,\nu} = \theta_0^x + \int_0^t F(\theta_s^{x,\nu}) \, \mathrm{d}s + \int_0^t \int_I W(x, y) \int_{\mathbb{T}} \Gamma(\theta_s^{x,\nu}, \theta) \nu_s^y(\mathrm{d}\theta) \, \mathrm{d}y \, \mathrm{d}s + B_t^x, \tag{4.8}$$

where the initial conditions and the Brownian motions are the same as (4.4). Observe that by Remark 4.1, the mapping $I \ni x \mapsto \nu_s^x \in \mathcal{P}(\mathbb{T})$ is Lebesgue-measurable for every $s \in [0, T]$. Since F and Γ are bounded Lipschitz functions, there exists a unique solution to (4.8), which we denote by $\Phi(\nu) \in \mathcal{M}_T$. The solution of (4.8) is constructed as the limit of a Cauchy sequence of elements of $\mathcal{M}_T$. Since these are measurable as functions of $x \in I$, the mapping $I \ni x \mapsto \mathrm{Law}(\theta^{x,\nu}) \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{T}))$ is also Lebesgue-measurable. Thus the map

$$\Phi : (\mathcal{M}_T, d_T) \to (\mathcal{M}_T, d_T),$$
$$\nu \to \Phi(\nu)$$

is well-defined. A solution to (4.4) is a fixed point of Φ and any fixed point of Φ is a solution to (4.4).

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

$$\left| \theta_t^{x,\mu} - \theta_t^{x,\nu} \right|^2 \leq C \int_0^t \left| F(\theta_s^{x,\mu}) - F(\theta_s^{x,\nu}) \right|^2 \mathrm{d}s$$
$$+ C \int_0^t \left| \int_I W(x, y) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,\mu}, \theta) \mu_s^y(\mathrm{d}\theta) - \int_{\mathbb{T}} \Gamma(\theta_s^{x,\nu}, \theta) \nu_s^y(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}s.$$

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

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

From (2.3) we obtain

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

from which, using (4.1), we deduce

$$\left| \theta_t^{x,\mu} - \theta_t^{x,\nu} \right|^2 \leq C \int_0^t |\theta_s^{x,\mu} - \theta_s^{x,\nu}|^2 \, \mathrm{d}s + C \int_0^t d_s^2(\mu, \nu) \, \mathrm{d}s.$$

The definition of D_T (2.2) and an application of Gronwall's lemma lead to

$$d_T^2(\Phi(\mu), \Phi(\nu)) \leq \int_I \mathbf{E} \left[\sup_{t \in [0,T]} |\theta_t^{x,\mu} - \theta_t^{x,\nu}|^2 \right] \mathrm{d}x \leq C \int_0^T d_s^2(\mu, \nu) \, \mathrm{d}s. \tag{4.9}$$

From the last relation we obtain the uniqueness of solutions to (4.4).

We prove that a solution exists by iterating (4.9). Indeed, for $k \geq 1$ and $\mu \in \mathcal{M}_T$, we get

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

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 4.1, apply Itô's formula to $f(\theta_t^x)$ with $f\in\mathcal{C}_0^\infty$ to get

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

Integrating with respect to **P** yields the weak formulation of (4.5). □

Next we move to the proof of Proposition 4.2.

Proof of Proposition 4.2. 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 in the proof of Proposition 4.1, by adding and subtracting the term

$$W(x,y)\Gamma\big(\theta_r^{x,V},\theta\big)\big(\mu_r^{y,W}-\mu_r^{y,V}\big)$$

in the integrals, we get

$$\big|\theta_s^{x,W}-\theta_s^{x,V}\big|^2\le C\int_0^s\big|F\big(\theta_r^{x,W}\big)-F\big(\theta_r^{x,V}\big)\big|^2\,\mathrm{d}r$$
$$+C\int_0^s\left|\int_I W(x,y)\int_\mathbb{T}\big(\Gamma\big(\theta_r^{x,W},\theta\big)-\Gamma\big(\theta_r^{x,V},\theta\big)\big)\mu_r^{y,W}(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r$$
$$+C\int_0^s\left|\int_I W(x,y)\int_\mathbb{T}\Gamma\big(\theta_r^{x,V},\theta\big)\big(\mu_r^{y,W}-\mu_r^{y,V}\big)(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r$$
$$+C\int_0^s\left|\int_I(W(x,y)-V(x,y))\int_\mathbb{T}\Gamma\big(\theta_r^{x,V},\theta\big)\mu_r^{y,V}(\mathrm{d}\theta)\,\mathrm{d}y\right|^2\mathrm{d}r.$$

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

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

After taking the supremum over $s\in[0,t]$, the expectation **E** and integrating with respect to $x\in I$, we are able to apply Gronwall's lemma as in (4.9) to get

$$d_t^2\big(\mu^W,\mu^V\big)\le\int_I\mathbf{E}\left[\sup_{s\in[0,t]}\big|\theta_s^{x,W}-\theta_s^{x,V}\big|^2\right]\mathrm{d}x\le C\left(\int_0^t d_s^2\big(\mu^W,\mu^V\big)\,\mathrm{d}s+G\right),\qquad(4.10)$$

where G is given by

$$G=\int_0^t\mathbf{E}\left[\int_I\left|\int_I(W(x,y)-V(x,y))\left(\int_\mathbb{T}\Gamma\big(\theta_s^{x,V},\theta\big)\mu_s^{y,V}(\mathrm{d}\theta)\right)\mathrm{d}y\right|^2\mathrm{d}x\right]\mathrm{d}s.\qquad(4.11)$$

Applying Gronwall's inequality to (4.10) yields

$$d_t^2(\mu^W, \mu^V) \leq CG.$$

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

We take advantage of the Fourier series representation of Γ. As we are working on the torus, we can factorize Γ into its two components, that is,

$$\Gamma(\theta, \psi) = \sum_{k,l \in \mathbb{Z}} \Gamma_{kl} \, \mathrm{e}^{\mathrm{i}k\theta} \mathrm{e}^{\mathrm{i}l\psi}, \quad \theta, \psi \in \mathbb{T}, \tag{4.12}$$

where

$$\Gamma_{kl} = \int_{\mathbb{T}^2} \Gamma(\theta, \psi) \, \mathrm{e}^{\mathrm{i}(k\theta + l\psi)} \, \mathrm{d}\theta \, \mathrm{d}\psi.$$

The condition $\Gamma \in \mathcal{C}^{1+\varepsilon}$ allows us to have a decay estimate on the Fourier coefficients Γ_{kl}. Indeed, classical results on the asymptotic of Fourier series [18, pp. 24–26] imply that

$$C_\Gamma := \sum_{k,l \in \mathbb{Z}} (kl)^{1+\varepsilon} |\Gamma_{kl}|^2 < \infty. \tag{4.13}$$

Plugging this expression into (4.11), we obtain that

$$\int_I \left| \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} \Gamma(\theta_s^{x,V}, \theta) \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x$$

$$= \int_I \left| \sum_{kl} \Gamma_{kl} \, \mathrm{e}^{\mathrm{i}k\theta_s^{x,V}} \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}l\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x.$$

Multiplying and dividing by $(kl)^{(1+\varepsilon)/2}$, we are left with

$$\leq \int_I \left| \sum_{kl} \left((kl)^{(1+\varepsilon)/2} \Gamma_{kl} \, \mathrm{e}^{\mathrm{i}k\theta_s^{x,V}} \right) \right.$$

$$\left. \left((kl)^{-(1+\varepsilon)/2} \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}l\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right) \right|^2 \mathrm{d}x$$

$$\leq C_\Gamma \sum_{kl} (kl)^{-1-\varepsilon} \int_I \left| \int_I (W(x, y) - V(x, y)) \left(\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}l\theta} \mu_s^{y,V}(\mathrm{d}\theta) \right) \mathrm{d}y \right|^2 \mathrm{d}x,$$

where in the second step we applied the Cauchy–Schwarz inequality and (4.13). Using that W and V are bounded, as well as the fact that

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

we conclude

$$G \leq C \sup_{\|a\|_\infty, \|b\|_\infty \leq 1} \int_I \left| \int_I (W(x, y) - V(x, y))(a(y) + \mathrm{i}b(y)) \, \mathrm{d}y \right| \mathrm{d}x$$

$$\leq C\|W - V\|_{\infty \to 1}.$$

Since the norm $\|\cdot\|_{\infty \to 1}$ is equivalent to the cut-norm (2.5), the proof is concluded. $\qquad\square$

4.4. Proofs for the non-linear process (3.3)

Proof of Theorem 3.1. The first part follows directly from Proposition 4.1 and Remark 4.4. The proof of (3.4) is similar to the proof of (4.5), but note that we are now integrating with respect to the randomness in U as well. $\qquad\square$

Proof of Theorem 4.1. Let $\theta^{U,W}$ and $\theta^{U,V}$ be the two solutions to (3.3) 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 (3.3) with V^φ; see Remark 4.4. 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

$$W(U, y)\Gamma\big(\theta_r^{\varphi(U),V}, \theta\big)\big(\mu_r^{y,W} - \mu_r^{\varphi(y),V}\big)$$

to obtain that

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

The first two integrals on the right-hand side are bounded by $C \int_0^s \big|\theta_r^{U,W} - \theta_r^{\varphi(U),V}\big|^2 \, dr$, using that F and Γ are Lipschitz-continuous, while the third integral on the right-hand side can be estimated using (2.3) and the fact that $0 \leq W \leq 1$. Thus we get

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

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

$$
\begin{aligned}
\int_I \mathbf{E}\left[\sup_{s \in [0,t]} \big|\theta_s^{x,W} - \theta_s^{\varphi(x),V}\big|^2 \right] dx &\leq C \int_0^t \int_I \mathbf{E}\left[\sup_{r \in [0,s]} \big|\theta_r^{x,W} - \theta_r^{\varphi(x),V}\big|^2 \right] dx \, ds \\
&\quad + C \int_0^t d_s^2\big(\mu^W, (\mu^V)^\varphi\big) \, ds + CG, \qquad (4.14)
\end{aligned}
$$

where G is given by

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

In the proof of Proposition 4.2 we proved the following estimates:

$$d_t^2\big(\mu^W, (\mu^V)^\varphi\big) \le \int_I \mathbf{E}\bigg[\sup_{s\in[0,t]} \big|\theta_s^{x,W} - \theta_s^{\varphi(x),V}\big|^2\bigg]\,\mathrm{d}x,$$

$$G \le C'\|W - V^\varphi\|_\square \quad \text{for some } C' > 0.$$

Applying these bounds to (4.14) and using Gronwall's inequality twice as in the previous proof, yields

$$d_t^2\big(\mu^W, (\mu^V)^\varphi\big) \le C\|W - V^\varphi\|_\square.$$

By taking the infimum with respect to $\varphi \in S_I$ and recalling the definition of the cut-distance (2.8) together with (4.3), we obtain

$$D_t\big(\bar\mu^W, \bar\mu^V\big) \le \widetilde{d}_t\big(\mu^W, \mu^V\big) \le C\,\delta_\square(W, V)^{1/2}.$$

The proof is concluded. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5. Proof of Theorem 3.2

In order to prove Theorem 3.2, we couple the system (3.1) to a sequence of identically distributed copies of the non-linear process θ, which is obtained by sampling $\{U_i\}_{i\in\mathbb{N}}$ i.i.d. uniform random variables and choosing the same initial conditions and Brownian motions of (3.1).

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

$$\theta_t^i = \theta_0^i + \int_0^t F(\theta_s^i)\,\mathrm{d}s + \int_0^t \int_I W(U_i, y) \int_\mathbb{T} \Gamma(\theta_s^i, \theta)\mu_s^y(\mathrm{d}\theta)\,\mathrm{d}y\,\mathrm{d}s + B_t^i. \qquad (5.1)$$

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 3.2, we give a trajectorial estimate.

Lemma 5.1. *Under the hypothesis of Theorem 3.2, it holds that*

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

Proof. As before, we compare the trajectories $\theta^{i,n}$ and θ^i, by studying the equation satisfied by $\big|\theta_s^{i,n} - \theta_s^i\big|^2$; recall (3.1) and (5.1). Add and subtract the term

$$\big(\xi_{ij}^{(n)} - W(U_i, U_j)\big)\Gamma\big(\theta_r^i, \theta_r^j\big)$$

in the integrals, to get

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

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 $\mathbb{E} \times \mathbf{E}$, for which we just write E for simplicity:

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

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, we obtain for $i \neq j$

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

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

Turning to the second term, we will prove that

$$
E\left[\frac{1}{n} \sum_{i=1}^n \left|\frac{1}{n} \sum_{j=1}^n \left(\xi_{ij}^{(n)} - W(U_i, U_j)\right)\Gamma\left(\theta_s^i, \theta_s^j\right)\right|^2\right] \leq C\,\mathbb{E}[\delta_\square(\xi^{(n)}, W^{(n)})] + \mathrm{o}(1), \tag{5.2}
$$

where $W^{(n)} := \{W(U_i, U_j)\}_{i,j=1,\ldots,n}$ is a W-random graph with n vertices; see (3.8). This together with a Gronwall argument implies that

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

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 3.3.

Turning to (5.2), we use an argument similar to (4.11)–(4.13). Recall that since $\Gamma \in \mathcal{C}^{1+\varepsilon}$, it admits a Fourier series (4.12) 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 (5.2), multiplying and dividing by $(kl)^{(1+\varepsilon)/2}$, we get

$$E\left[\frac{1}{n}\sum_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \big(\xi_{ij}^{(n)} - W(U_i, U_j)\big)\Gamma\big(\theta_s^i, \theta_s^j\big)\right|^2\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \big(\xi_{ij}^{(n)} - W(U_i, U_j)\big)\sum_{k,l}\Gamma_{kl}\mathrm{e}^{\mathrm{i}\theta_s^i k}\mathrm{e}^{\mathrm{i}\theta_s^j l}\right|^2\right]$$

$$\leq CE\left[\sum_{k,l}(kl)^{-1-\varepsilon}\frac{1}{n}\sum_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \big(\xi_{ij}^{(n)} - W(U_i, U_j)\big)\mathrm{e}^{\mathrm{i}\theta_s^i k}\mathrm{e}^{\mathrm{i}\theta_s^j l}\right|^2\right],$$

where we used the Cauchy–Schwarz inequality as in the proof of Theorem 4.1. Observe that $\sum_{kl} (kl)^{-1-\varepsilon}$ is convergent and that $|\mathrm{e}^{\mathrm{i}\theta_s^i k}| \leq 1$ for all k and s: we can thus bound the previous term **P**-a.s. by

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

Recall that $W^{(n)} = \{W(U_i, U_j)\}_{i,j=1,\ldots,n}$ 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 for any order of $\{\theta^i\}_{i=1,\ldots,n}$ and, in particular, of $\{U^i\}_{i=1,\ldots,n}$. In particular, the last inequality holds for every relabeling of $W^{(n)}$.

From the definition of $\hat{\delta}_\square$ (2.7), one can thus take the labeling of $\{U_i\}_{i=1,\ldots,n}$ for every $n \in \mathbb{N}$, such that

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

Using the asymptotic equivalence of $\hat{\delta}_\square$ with $\delta_\square$ (see Remark 2.1), the claim is proved and the proof is concluded. $\qquad\square$

Proof of Theorem 3.2. The equivalence between the convergence in $\mathbb{P}$-probability of $\xi^{(n)}$ and equation (3.9) is proved in Lemma A.1. We turn to the proof of the convergence of μ^n.

It is well known that the bounded Lipschitz distance (recall (2.3)) metrizes 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

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

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 (5.1)), it is enough to show that

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

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

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

which goes to zero as $n\to\infty$ by Lemma 5.1. □

Appendix A. Graph convergence and random graphons

A.1. Convergence in probability

The characterization of the convergence in distribution for a sequence of graphs was originally given in [17]. Here we give 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.1. *Assume that* $\{\xi^{(n)}\}_{n\in\mathbb{N}}$ *is a sequence of random graphs and W is a random graphon in* $\widetilde{\mathcal{W}}_0$. *Then $\xi^{(n)}$ converges in $\mathbb{P}$-probability to W if and only if (3.5) holds, i.e. if and only if*

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

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

$$\lim_{n\to\infty}\mathbb{P}(\delta_\square(\xi^{(n)},W)>\varepsilon)=0\quad\text{for all }\varepsilon>0.\tag{A.1.}$$

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.1) is then equivalent to the convergence in L^1, i.e. equivalent to (3.5). □

Acknowledgements

Funding information

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALON, N. AND NAOR, A. (2006). Approximating the Cut-Norm via Grothendieck's inequality. *SIAM J. Comput.* **35**, 787–803.

[2] ARENAS, A., DIAZ-GUILERA, A., KURTHS, J., MORENO, Y. AND ZHOU, C. (2008). Synchronization in complex networks. *Phys. Rep.* **469**, 93–153.

[3] Basak, A., Bhamidi, S., Chakraborty, S. and Nobel, A. (2016). Large subgraphs in pseudo-random graphs. Available at arXiv:1610.03762.

[4] Bayraktar, E., Chakraborty, S. and Wu, R. (2020). Graphon mean field systems. Available at arXiv:2003.13180.

[5] Bertini, L., Giacomin, G. and Pakdaman, K. (2010). Dynamical aspects of mean field plane rotators and the Kuramoto model. *J. Statist. Phys.* **138**, 270–290.

[6] Bet, G., van der Hofstad, R. and van Leeuwaarden, J. S. (2020). Big jobs arrive early: from critical queues to random graphs. *Stoch. Syst.* **10**, 310–334.

[7] Bhamidi, S., Budhiraja, A. and Wu, R. (2019). Weakly interacting particle systems on inhomogeneous random graphs. *Stoch. Process. Appl.* **129**, 2174–2206.

[8] Bogerd, K., Castro, R. M. and van der Hofstad, R. (2020). Cliques in rank-1 random graphs: the role of inhomogeneity. *Bernoulli* **26**, 253–285.

[9] Caines, P. E. and Huang, M. (2018). Graphon mean field games and the GMFG equations. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4129–4134.

[10] Carmona, R., Cooney, D. B., Graves, C. V. and Lauriere, M. (2022). Stochastic graphon games I: The static case. *Math. Operat. Res.* **47**, 750–778.

[11] Chiba, H. and Medvedev, G. S. (2019). The mean field analysis of the Kuramoto model on graphs I: The mean field equation and transition point formulas. *Discrete Contin. Dyn. Syst. A* **39**, 131.

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

[13] Coppini, F. (2020). Weakly interacting diffusions on graphs. Doctoral thesis, Université de Paris.

[14] Coppini, F. (2022). Long time dynamics for interacting oscillators on graphs. *Ann. Appl. Prob.* **32**, 360–391.

[15] Coppini, F., Dietert, H. and Giacomin, G. (2020). A law of large numbers and large deviations for interacting diffusions on Erdős–Rényi graphs. *Stoch. Dyn.* **20**, 2050010.

[16] Delattre, S., Giacomin, G. and Luçon, E. (2016). A note on dynamical models on random graphs and Fokker–Planck equations. *J. Statist. Phys.* **165**, 785–798.

[17] Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat.* **28**, 33–61.

[18] Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd edn (Cambridge Mathematical Library). Cambridge University Press.

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

[20] Luçon, E. (2020). Quenched asymptotics for interacting diffusions on inhomogeneous random graphs. *Stoch. Process. Appl.* **130**, 6783–6842.

[21] Luçon, E. and Stannat, W. (2014). Mean field limit for disordered diffusions with singular interactions. *Ann. Appl. Prob.* **24**, 1946–1993.

[22] Medvedev, G. S. (2019). The continuum limit of the Kuramoto model on sparse random graphs. *Commun. Math. Sci.* **17**, 883–898.

[23] Oelschläger, K. (1984). A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Prob.* **12**, 458–479.

[24] Oliveira, R. I. and Reis, G. H. (2019). Interacting diffusions on random graphs with diverging average degrees: hydrodynamics and large deviations. *J. Statist. Phys.* **176**, 1057–1087.

[25] Rodrigues, F. A., Peron, T. K. D., Ji, P. and Kurths, J. (2016). The Kuramoto model in complex networks. *Phys. Rep.* **610**, 1–98.

[26] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX - 1989* (Lecture Notes Math. **1464**), ed. P.-L. Hennequin, pp. 165–251. Springer, Berlin and Heidelberg.

[27] Van Enter, A. C. D. and Ruszel, W. M. (2009). Gibbsianness versus non-Gibbsianness of time-evolved planar rotor models. *Stoch. Process. Appl.* **119**, 1866–1888.