GRAPHON MEAN FIELD SYSTEMS

ERHAN BAYRAKTAR, SUMAN CHAKRABORTY, AND RUOYU WU

Abstract. We consider heterogeneously interacting diffusive particle systems and their large population limit. The interaction is of mean field type with weights characterized by an underlying graphon. A law of large numbers result is established as the system size increases and the underlying graphons converge. The limit is given by a graphon mean field system consisting of independent but heterogeneous nonlinear diffusions whose probability distributions are fully coupled. Well-posedness, continuity and stability of such systems are provided. We also consider a not-so-dense analogue of the finite particle system, obtained by percolation with vanishing rates and suitable scaling of interactions. A law of large numbers result is proved for the convergence of such systems to the corresponding graphon mean field system.

Contents

Date: October 5, 2022.

2020 *Mathematics Subject Classification.* 05C80 60J60 60K35.

Key words and phrases. graphons, graphon particle systems, mean field interaction, heterogeneous interaction, networks, percolation.

E. Bayraktar is partially supported by the National Science Foundation under grant DMS2106556 and by the Susan M. Smith chair.

1. Introduction

In this work we study mean field diffusive particle systems with heterogeneous interaction and their large population limit. The interaction is of mean field type and is characterized through a step graphon. More precisely, denoting by X_i^n the state of the i-th particle,

$$X_i^n(t) = X_{\frac{i}{n}}(0) + \int_0^t \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s))\, ds + \int_0^t \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \sigma(X_i^n(s), X_j^n(s))\, dB_{\frac{i}{n}}(s), \quad i \in \{1, \dots, n\}, \tag{1.1}$$

where b and σ are some Lipschitz functions, $\{B_u : u \in [0,1]\}$ are i.i.d. d-dimensional Brownian motions, and $X_u(0)$ is a collection of independent $\mathbb{R}^d$-valued random variables, with probability distribution $\mu_u(0)$ for each $u \in [0,1]$, and independent of $\{B_u : u \in [0,1]\}$. Here $\xi_{ij}^n \in [0,1]$ determines the interaction between particles i and j, and depends on some step graphon G_n converging to a limiting graphon in the cut metric.

The classic mean field system with homogeneous interaction, which corresponds to $\xi_{ij}^n \equiv 1$ in (1.1), dates back to works of Boltzmann, Vlasov, McKean and others (see [30, 34, 43] and references therein). While the original motivation for the study came from statistical physics, similar models have arisen in many different application areas, including economics, chemical and biological systems, communication networks and social sciences (see e.g. [12] for an extensive list of references). Systems with inhomogeneity described by multi-type populations, where the interaction between two particles depends on their types, have been proposed in social sciences [20], statistical mechanics [19], neurosciences [1], and others [13,37]. In recent years, there have been an increasing attention on the study of mean field systems on large networks, including [3, 5, 6, 14, 21–24, 26, 28, 33, 35, 36, 38], where the majority of focus is on Erdős-Rényi type random graphs. Among these, [38] allows the edge probability between two nodes to depend on independent random media variables associated with these two nodes, and [23] analyzes the mean field game on Erdős-Rényi random graphs.

We extend the study of mean field models to a much larger class of graphs. To put our work in context, let us describe the class of graphs that we are going to consider. We consider sequences of dense graphs that converge to a limit in an appropriate sense (see [32] and references therein). Roughly speaking, this limit theory treats a graph $\mathbb{G}_n$ on n vertices as a function $G_n \colon I \times I \to \mathbb{R}$, where $I := [0,1]$. This function is what we call a graphon. Then $\mathbb{G}_n$ is said to converge to the function G if and only if G_n converges to G in "cut metric" (see Section 2 for the definition). We consider mean field models on such converging graph sequences. The motivation for considering such graph sequences is that aside from its theoretical implications, many important graph models (both random and deterministic) have been shown to converge to a limit. See [2, 7, 18, 31] for many such examples. Unfortunately, the graph limit theory only works for dense graphs (graphs with order of n^2 many edges). To extend our study to the not-so-dense settings we also consider bond percolated models of graphs; see [8], and also see [10, 11] and references therein for more interests in and analysis on not-so-dense graphs.

Among the aforementioned works on mean field systems on large networks, the ones closest to our setup are [33] and [5]. The work [33] considers large population of diffusions interacting on a random directed inhomogeneous graph, and obtains quenched convergence of particles, empirical measures, and also spatial fields (empirical measures of particle states and positions/indices). Compared with our models and results, there are two main differences. Firstly, [33] considers independent Bernoulli edges scaled by a certain dilution parameter and

assumes the collection of scaled success probabilities, viewed as graphons/kernels, converges in a sense stronger than the usual cut metric of graphons (Assumption 2.9), with Lipschitz assumptions on initial states (Assumption 2.6) and suitable regularity assumptions on the graphons (Assumptions 2.11 and 2.14). While we consider both Bernoulli edges and $[0,1]$-valued weighted edges, converging in the cut metric, with continuity assumptions on initial states. Secondly, [33] obtains quenched convergence, conditioning on the realization of random graphs; while we are mostly concerned with the annealed setup and some of our results (in Theorems 3.1 and 4.1) could be applied to the quenched setup (see Condition 3.1(b) and Remark 3.3(b)).

After the initial submission of this work, another paper on graphon particle systems ([5]) was posted on arXiv a few months later. [5] considers weakly interacting oscillators on dense random graphs and obtains an annealed law of large numbers for the empirical measure. The limiting system is given by describing the evolution of a particle uniformly chosen from I. However, [5] does not consider not-so-dense graphs as in our Section 4, and their results do not imply our results in Section 3 (and ours do not imply theirs either). In particular, [5] assumes $\{0,1\}$-valued random adjacency matrix ξ^n for n-particle systems converging to a possibly random graphon in the cut distance (and the argument is expected to work for $[0,1]$-valued weighted edges as well), while we consider both such random ξ^n and deterministic $[0,1]$-valued ξ^n. Also, [5] assumes homogeneous initial state distribution and constant diffusion coefficient, while we allow for heterogeneous initial state distribution and state-and-interaction-dependent diffusion coefficient. Lastly, particles' state space is bounded (one dimensional torus) in [5], and hence the system coefficients are bounded, while our work allows coefficients to have linear growth.

We should point out that the use of graphons to analyze heterogeneous interaction in game theory emerged recently (see e.g. [15, 17, 39]). Among these, [39] analyzes static graphon games and the convergence of the n-player game with interaction network sampled from a given graphon. Static graphon games are considered in [17] and the convergence of the n-player game with general interaction network that converges to a given graphon is obtained. The diffusive dynamics for the states of the particles, with constant diffusion coefficients, is considered in [15] for continuum graphon mean field games. However, [15] does not address the convergence problem of the finite particle system to the limiting problem they analyze.

The goal of this work is to study the asymptotics of the diffusive particle system (1.1) with heterogeneous interaction, and their not-so-dense analogue in (1.2) below. Our first main result is the existence, uniqueness, continuity, and stability property (Proposition 2.1 and Theorem 2.1) for the limiting graphon particle system (2.1), consisting of a continuum of independent but non-identical nonlinear diffusions. Among these, the stability property (Theorem 2.1(c)) in particular says that the system solution converges in a suitable sense provided that the underlying graphon converges in the cut metric. The proof makes use of a coupling argument but challenge is two fold: First, the cut metric is in general very weak, in that the convergence $G_n \to G$ does not necessarily imply the L^2-convergence of G_n as operators on $I \times I$, namely one may not have $\int_{I\times I}[G_n(x,y) - G(x,y)]^2\,dx\,dy \to 0$. However, one could alternatively view G_n as operators from $L^\infty(I)$ to $L^1(I)$ that are continuous with respect to the cut metric (see Remark 2.1). This observation is actually an important building block of many proofs in this work. Second, the interaction in the graphon particle system (2.1) does not match with such a choice of operator, unless the coefficients $b(x,y)$ and $\sigma(x,y)$ could be decomposed as the product of functions of each variable. For this, a truncation and approximation argument is applied to these coefficients, and the associated errors are carefully analyzed (see Section 5.2).

The second main result is the convergence of the n particle system (1.1) to the graphon particle system (2.1), for a sequence of convergent underlying step graphons (graphons with blockwise constant values; see (3.2)). A law of large numbers (LLN) is established in Theorem 3.1, which says that the empirical measure of n particles in (1.1) converges in probability to the averaged distribution of a continuum particles in (2.1). This is first proved in Lemma 6.1 under certain regularity assumptions of the graphon (Condition 2.2(b)), by applying again a truncation and approximation argument to the system coefficients. Then the stability property (Theorem 2.1(c)) is used in Section 6.2 to show that the result also holds for general graphons. In Theorem 3.2, we also obtain a precise particle-wise uniform convergence rate, when the underlying step graphons are sampled from a given graphon with a certain continuity property.

Our third main result is the analysis of the not-so-dense analogue of (1.1):

$$X_i^n(t) = X_{\frac{i}{n}}(0) + \int_0^t \frac{1}{n\beta_n} \sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s))\, ds + \int_0^t \sigma(X_i^n(s))\, dB_{\frac{i}{n}}(s), \quad i \in \{1, \dots, n\}, \tag{1.2}$$

where $\beta_n \in (0, 1]$ is some sequence of numbers that may converge to 0, and $\{\xi_{ij}^n\}$ are independent Bernoulli random variables with possibly vanishing probabilities (of order β_n) associated with the underlying step graphon. Similar to mean field systems on Erdős-Rényi random graphs [6, 23, 38], the strength of interaction here is scaled by the order of the number of neighbors (see Remark 4.1 for more explanations). In Theorem 4.1, we show a LLN that the limit of such systems is again given by a graphon particle system, provided that the underlying step graphons converge and

$$\lim_{n\to\infty} n\beta_n = \infty.$$

The main challenge lies in the heterogeneity of the system and the average of interactions of order ξ_{ij}^n/β_n that is unbounded in n. The unbounded interaction ξ_{ij}^n/β_n is taken care of in [6, 23, 38] using exchangeability. However, due to the lack of exchangeability here, a new approach is needed. Indeed, besides the application of coupling, truncation and approximation arguments, the key ingredient in the proof of Theorem 4.1 is (7.15) in Lemma 7.2, which shows that the expected effect of unbounded interactions ξ_{ij}^n/β_n on the coupled difference $|X_j^n - X_{\frac{j}{n}}|$ is roughly the same as $\mathbb{E}[\xi_{ij}^n/\beta_n]\mathbb{E}|X_j^n - X_{\frac{j}{n}}|$, up to some constant multiples and negligible errors. The proof of Lemma 7.2 applies a collection of change of measure arguments separately to each pair (resp. triplet) of certain auxiliary particles and the edge (resp. edges) connecting them. For each change of measure, the Radon-Nikodym derivative and the difference among pre-limit, limiting and auxiliary particles are carefully analyzed. Due to the technical application of the Girsanov's Theorem, the diffusion coefficient in (1.2) is taken to be state-dependent only. Lastly, we also obtain a precise rate of convergence in Theorem 4.2, when the underlying step graphons are sampled from a given graphon with a certain continuity property.

1.1. **Organization.** The paper is organized as follows. In Section 2 we analyze the graphon particle system (2.1). The existence and uniqueness is proved in Proposition 2.1. The continuity and stability of the system is presented in Theorem 2.1 in Section 2.1. Concrete examples are given in Section 2.2. In Section 3 we study the convergence of the n particle system (1.1). The LLN is given in Theorem 3.1, and a precise rate of convergence is given in Theorem 3.2 under conditions. In Section 4 we study the convergence of the n particle system with not-so-dense interaction (1.2). The LLN is given in Theorem 4.1, and a precise rate of

convergence is given in Theorem 4.2 under conditions. Sections 5, 6 and 7 are devoted to the proofs of results in Section 2, 3 and 4, respectively.

We close this section by introducing some frequently used notation.

1.2. **Notation.** Given a Polish space $\mathbb{S}$, denote by $\mathcal{P}(\mathbb{S})$ the space of probability measures on $\mathbb{S}$ endowed with the topology of weak convergence. For $\mu \in \mathcal{P}(\mathbb{S})$ and a μ-integrable function $f\colon \mathbb{S} \to \mathbb{R}$, let $\langle f, \mu\rangle := \int_{\mathbb{S}} f(x)\,\mu(dx)$. For $f\colon \mathbb{S} \to \mathbb{R}$, let $\|f\|_\infty := \sup_{x\in\mathbb{S}} |f(x)|$. The probability law of a random variable X will be denoted by $\mathcal{L}(X)$. Fix $T \in (0,\infty)$ and all processes will be considered over the time horizon $[0,T]$. Denote by $\mathbb{C}([0,T] : \mathbb{S})$ the space of continuous functions from $[0,T]$ to $\mathbb{S}$, endowed with the topology of uniform convergence. Let $\mathcal{C}_d := \mathbb{C}([0,T] : \mathbb{R}^d)$ and $\|x\|_{*,t} := \sup_{0\le s\le t} |x_s|$ for $x \in \mathcal{C}_d$ and $t \in [0,T]$. We will use κ to denote various constants in the paper and $\kappa(m)$ to emphasize the dependence on some parameter m. Their values may change from line to line. Expectations under $\mathbb{P}$ will be denoted by $\mathbb{E}$. To simplify the notation, we will usually write $\mathbb{E}[X^2]$ as $\mathbb{E}X^2$.

2. Graphon particle systems

We follow the notation used in [32, Chapters 7 and 8]. Let $I := [0,1]$. Denote by $\mathcal{G}$ the space of all bounded symmetric measurable functions $G\colon I \times I \to \mathbb{R}$. A graphon G is an element of $\mathcal{G}$ with $0 \le G \le 1$. The cut norm on $\mathcal{G}$ is defined by

$$\|G\|_\square := \sup_{S,T\in\mathcal{B}(I)} \left| \int_{S\times T} G(u,v)\,du\,dv \right|,$$

and the corresponding cut metric is defined by

$$d_\square(G_1, G_2) := \|G_1 - G_2\|_\square.$$

Remark 2.1. *We will also view a graphon G as an operator from $L^\infty(I)$ to $L^1(I)$ with the operator norm*

$$\|G\| := \|G\|_{\infty\to 1} := \sup_{\|g\|_\infty\le 1} \|Gg\|_1 = \sup_{\|g\|_\infty\le 1} \int_I \left| \int_I G(u,v)g(v)\,dv \right| du.$$

From [32, Lemma 8.11] *it follows that if $\|G_n - G\|_\square \to 0$ for a sequence of graphons G_n, then $\|G_n - G\| \to 0$.*

Given a graphon G and an initial distribution $\mu(0) := (\mu_u(0) \in \mathcal{P}(\mathbb{R}^d) : u \in I)$, consider the following graphon particle system:

$$\begin{aligned} X_u(t) = X_u(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(X_u(s), x) G(u,v)\,\mu_{v,s}(dx)\,dv\,ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(X_u(s), x) G(u,v)\,\mu_{v,s}(dx)\,dv\,dB_u(s), \;\; \mu_{u,t} = \mathcal{L}(X_u(t)), \;\; u \in I. \end{aligned} \tag{2.1}$$

As introduced in Section 1, here $\{B_u : u \in I\}$ are i.i.d. d-dimensional Brownian motions, and $X_u(0)$ is a collection of independent $\mathbb{R}^d$-valued random variables, with law $\mu_u(0)$ for each $u \in I$, and independent of $\{B_u : u \in I\}$, defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$. We make the following assumptions on the initial states and system coefficients.

Condition 2.1. *(a) The map $I \ni u \mapsto \mu_u(0) = \mathcal{L}(X_u(0)) \in \mathcal{P}(\mathbb{R}^d)$ is measurable. There exists some $\varepsilon \in (0,\infty)$ such that*

$$\sup_{u\in I} \mathbb{E}|X_u(0)|^{2+\varepsilon} < \infty. \tag{2.2}$$

(b) The coefficients b and σ are Lipschitz functions, namely there exists some $K \in [0,\infty)$ such that

$$|b(x,y)-b(x',y')|+|\sigma(x,y)-\sigma(x',y')| \le K(|x-x'|+|y-y'|), \quad \forall x,x',y,y' \in \mathbb{R}^d.$$

Remark 2.2. *The assumption that $\varepsilon > 0$ is used for some technical arguments treating the unboundedness of b and σ. If b and σ are bounded and Lipschitz functions, then one could take $\varepsilon = 0$ (see Remarks 5.2 and 6.1).*

The following proposition gives well-posedness for the system (2.1). For graphon mean field games with diffusive dynamics, under certain regularity conditions, existence and uniqueness were shown in [15, Theorem 2] (and the journal version [16, Theorem 3.12]). But we provide a proof in Section 5.1 as our graphon G may not be continuous.

Proposition 2.1. *Suppose Condition 2.1 holds. Then there exists a unique pathwise solution $\{X_u\}$ to the graphon particle system (2.1). Moreover, $\sup_{u\in I}\mathbb{E}\|X_u\|^{2+\varepsilon}_{*,T} < \infty$ and the map $I \ni u \mapsto \mu_u \in \mathcal{P}(\mathcal{C}_d)$ is measurable.*

Remark 2.3. *(a) We note that processes $\{X_u\}$ in (2.1) are independent but not identically distributed nonlinear diffusions. In particular, the Fokker–Planck equations for the probability distributions of $\{X_u\}$ are nonlinear and fully coupled. In general, each X_u may not be a McKean–Vlasov process, as the probability law μ_u plays a negligible role in its evolution.*

*(b) We also note that we are not assuming $u \mapsto B_u$ is measurable or claiming $u \mapsto X_u$ is measurable. Proposition 2.1 gives a measurable dependence of the **law** μ_u on u, which is sufficient for later use since (2.1) involves integrals of μ_u, instead of X_u or B_u, with respect to $u \in [0,1]$.*

In order to analyze the collection of probability laws $\mu = (\mu_u : u \in I)$, consider the following space of probability measures

$$\mathcal{M} := \{\nu = (\nu_u : u \in I) \in [\mathcal{P}(\mathcal{C}_d)]^I \mid I \ni u \mapsto \nu_u \in \mathcal{P}(\mathcal{C}_d) \text{ is measurable and } \sup_{u\in I}\int_{\mathcal{C}_d}\|x\|^2_{*,T}\,\nu_u(dx) < \infty\}.$$

For the convenience of analysis (see e.g. Remark 2.4), we make use of the following Wasserstein-2 metrics

$$W_2(\mu,\nu) := \Big(\inf\Big\{\mathbb{E}|X-\tilde{X}|^2 : \mathcal{L}(X)=\mu, \mathcal{L}(\tilde{X})=\nu\Big\}\Big)^{1/2},\ \mu,\nu \in \mathcal{P}(\mathbb{R}^d), \tag{2.3}$$

$$W_{2,t}(\mu,\nu) := \Big(\inf\Big\{\mathbb{E}\|X-\tilde{X}\|^2_{*,t} : \mathcal{L}(X)=\mu, \mathcal{L}(\tilde{X})=\nu\Big\}\Big)^{1/2},\ t\in[0,T],\ \mu,\nu \in \mathcal{P}(\mathcal{C}_d), \tag{2.4}$$

$$W^{\mathcal{M}}_{2,t}(\mu,\nu) := \sup_{u\in I} W_{2,t}(\mu_u,\nu_u),\ t\in[0,T],\ \mu,\nu\in\mathcal{M}. \tag{2.5}$$

Remark 2.4. *From (2.3), (2.4) and (2.5) clearly we have*

$$W_2(\mu,\nu) \ge \sup_f \left|\int_{\mathbb{R}^d} f(x)\,\mu(dx) - \int_{\mathbb{R}^d} f(x)\,\nu(dx)\right|, \quad \mu,\nu\in\mathcal{P}(\mathbb{R}^d),$$

$$W_{2,t}(\mu,\nu) \ge \sup_f \left|\int_{\mathbb{R}^d} f(x)\,\mu_t(dx) - \int_{\mathbb{R}^d} f(x)\,\nu_t(dx)\right|, \quad \mu,\nu\in\mathcal{P}(\mathcal{C}_d),$$

$$W^{\mathcal{M}}_{2,t}(\mu,\nu) \ge \sup_{u\in I}\sup_f \left|\int_{\mathbb{R}^d} f(x)\,\mu_{u,t}(dx) - \int_{\mathbb{R}^d} f(x)\,\nu_{u,t}(dx)\right|, \quad \mu,\nu\in\mathcal{M},$$

for each $t \in [0,T]$, where the supremum is taken over all $f\colon \mathbb{R}^d \to \mathbb{R}$ such that the integrals exist and $|f(x)-f(\tilde{x})| \le |x-\tilde{x}|$ for $x, \tilde{x} \in \mathbb{R}^d$.

2.1. **Continuity and stability of the system.** In this section we are interested in establishing the continuity and stability properties for the graphon particle system (2.1). We usually make the following assumption on the initial distribution $\mu(0)$ and the graphon G.

Condition 2.2. *There exists a finite collection of intervals $\{I_i : i = 1,\dots,N\}$ for some $N \in \mathbb{N}$, such that $\cup_{i=1}^N I_i = I$ and for each $i \in \{1,\dots,N\}$:*

(a) The map $I_i \ni u \mapsto \mu_u(0) \in \mathcal{P}(\mathbb{R}^d)$ is continuous with respect to the W_2 metric.
(b) For each interior point $u \in I_i$, there exists a subset $A_u \subset I$ such that $\lambda_I(A_u) = 0$ and $G(u,v)$ is continuous at $(u,v) \in I \times I$ for each $v \in I \setminus A_u$, where λ_I denotes the Lebesgue measure on I.

Remark 2.5. *Condition 2.2(b) is not necessary for, but will help strengthen, convergence results in Sections 3 and 4. It holds naturally if G is continuous, or if G is continuous when restricted to each block $I_i \times I_j$. For example, it holds for graphons $G(u,v) = \mathbf{1}_{[0,\frac{1}{2}]^2}(u,v)$ and $G(u,v) = \mathbf{1}_{\{|u-v|\le\frac{1}{4}\}}(u,v)$.*

Sometimes we may work with a special class of $\mu(0)$ and G having certain Lipschitz properties.

Condition 2.3. *There exists some $\kappa \in (0,\infty)$ and a finite collection of intervals $\{I_i : i = 1,\dots,N\}$ for some $N \in \mathbb{N}$, such that $\cup_{i=1}^N I_i = I$ and*

$$W_2(\mu_{u_1}(0), \mu_{u_2}(0)) \le \kappa|u_1-u_2|, \quad u_1,u_2 \in I_i, \quad i \in \{1,\dots,N\},$$
$$|G(u_1,v_1)-G(u_2,v_2)| \le \kappa(|u_1-u_2|+|v_1-v_2|),\ (u_1,v_1),(u_2,v_2) \in I_i \times I_j,\ i,j \in \{1,\dots,N\}.$$

The following theorem gives continuity and stability of the graphon particle system (2.1). The proof is given in Section 5.2.

Theorem 2.1. *Suppose Condition 2.1 holds.*

(a) (Continuity) Suppose Condition 2.2 holds. Then for each $i \in \{1,\dots,N\}$, the map $I_i \ni u \mapsto \mu_u \in \mathcal{P}(\mathcal{C}_d)$ is continuous with respect to the $W_{2,T}$ metric.
(b) (Lipschitz continuity) Suppose Condition 2.3 holds. Then there exists some $\kappa \in (0,\infty)$ such that $W_{2,T}(\mu_u,\mu_v) \le \kappa|u-v|$ whenever $u,v \in I_i$ for some $i \in \{1,\dots,N\}$.
(c) (Stability) Let μ^G be the probability law of (2.1) associated with G. The map $G \mapsto \mu^G$ is continuous in the sense that $\int_I [W_{2,T}(\mu_u^{G_n}, \mu_u^G)]^2\,du \to 0$ if a sequence of graphons $G_n \to G$ in the cut metric.

Remark 2.6. *(a) Theorem 2.1 (a,b) will be needed in Sections 3 and 4 to analyze the convergence of n-particle systems with graphon mean field interactions.*
(b) Theorem 2.1(c) implies that the solution law to (2.1) depends on the underlying graphon G in a continuous manner. This and Proposition 2.1 together guarantees that the analysis of (2.1) is "well-posed" according to Hadamard's principle (cf. [4, Page 368] and [27, Page 38]). It will also be used in Sections 3 and 4 to analyze the convergence of n-particle systems when G is not necessarily continuous.

2.2. **Some special graphon particle systems.** In this section we introduce two special scenarios under which the system (2.1) is more tractable.

Example 2.1. *Suppose G is blockwise constant (which arises as a limit of the stochastic block model), that is, there exists a finite collection of intervals $\{I_i : i = 1, \dots, N\}$ and constants $\{p_{ij} = p_{ji} \in [0,1] : i, j = 1, \dots, N\}$ for some $N \in \mathbb{N}$, such that $\cup_{i=1}^N I_i = I$ and*

$$G(u,v) = p_{ij}, \quad (u,v) \in I_i \times I_j, \quad i,j \in \{1,\dots,N\}.$$

Also suppose the initial distribution is the same within each interval:

$$\mu_u(0) = \mu_v(0), \quad u,v \in I_i, \quad i \in \{1,\dots,N\}.$$

Due to the homogeneity in this case, the system (2.1) could be written in terms of just N representatives $u_i \in I_i$, $i \in \{1,\dots,N\}$:

$$\begin{aligned} X_{u_i}(t) = X_{u_i}(0) &+ \int_0^t \sum_{j=1}^N |I_j| p_{ij} \left(\int_{\mathbb{R}^d} b(X_{u_i}(s), x)\, \mu_{u_j,s}(dx) \right) ds \\ &+ \int_0^t \sum_{j=1}^N |I_j| p_{ij} \left(\int_{\mathbb{R}^d} \sigma(X_{u_i}(s), x)\, \mu_{u_j,s}(dx) \right) dB_{u_i}(s), \quad \mu_{u_i,t} = \mathcal{L}(X_{u_i}(t)), \end{aligned}$$

where $|A|$ denotes the Lebesgue measure of $A \subset I$. Note that this is simply a finite collection of multi-type McKean–Vlasov processes.

Example 2.2. *Suppose $b(x,y) = c_1 + c_2 x + c_3 y$ is linear, σ is constant, and the initial distributions $\{\mu_u(0) : u \in I\}$ are Gaussian. Then the system (2.1) is just consisting of a collection of Gaussian processes. Letting $m_u(t) := \mathbb{E}[X_u(t)]$ and $M_u(t) := \mathbb{E}[X_u^2(t)]$, we have*

$$\begin{aligned} m_u(t) &= m_u(0) + \int_0^t \int_0^1 (c_1 + c_2 m_u(s) + c_3 m_v(s)) G(u,v)\, dv\, ds, \\ M_u(t) &= M_u(0) + \mathbb{E}\left[\int_0^t 2X_u(s)\, dX_u(s) \right] + \sigma^2 \left(\int_0^1 G(u,v)\, dv \right)^2 t \\ &= M_u(0) + 2\int_0^t \int_0^1 (c_1 m_u(s) + c_2 M_u(s) + c_3 m_u(s) m_v(s)) G(u,v)\, dv\, ds \\ &\quad + \left(\sigma \int_0^1 G(u,v)\, dv \right)^2 t. \end{aligned}$$

This is a system of coupled ordinary differential equations.

Remark 2.7. *We note that even in the above two examples, it may not be an easy work to obtain explicit forms of solutions $\mathcal{L}(X_u(t))$. In the next two sections we will show the convergence of finite particle systems (1.1) and (1.2) to the graphon particle system (2.1). On one hand, such results could be used to approximate large finite particle systems with heterogeneous interactions by a graphon particle system. On the other hand, for graphon particle systems that are not tractable, one may choose a suitable finite particle system (and even its Euler discretizations) to approximate the former.*

3. Mean-field particle systems on dense graphs

In this section, we consider a sequence of n interacting diffusions (1.1) with the strength of interaction governed by ξ_{ij}^n associated with some kernel G_n:

$$X_i^n(t) = X_{\frac{i}{n}}(0) + \int_0^t \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s))\, ds + \int_0^t \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \sigma(X_i^n(s), X_j^n(s))\, dB_{\frac{i}{n}}(s), \quad i \in \{1,\dots,n\}. \tag{3.1}$$

Here the pathwise existence and uniqueness of the solution is guaranteed by the Lipschitz property of b and σ.

We would like to consider the natural correspondence between the adjacency matrix $\{\xi_{ij}^n\}$ and a function on $I \times I$ with constant value ξ_{ij}^n on each block $(\frac{i-1}{n}, \frac{i}{n}] \times (\frac{j-1}{n}, \frac{j}{n}]$ of side length $\frac{1}{n}$. So we make the following assumption on the strength of interaction ξ_{ij}^n and the associated kernel G_n.

Condition 3.1. *(a) The kernel G_n is a step graphon, that is, $0 \le G_n \le 1$ and*

$$G_n(u,v) = G_n\left(\frac{\lceil nu\rceil}{n}, \frac{\lceil nv\rceil}{n}\right), \quad \text{for } (u,v) \in I \times I. \tag{3.2}$$

(b) The strength of interaction ξ_{ij}^n is generated by the kernel G_n as follows. Either one of the following two cases holds for all $\{\xi_{ij}^n : i,j \in \{1,\dots,n\}, n \in \mathbb{N}\}$:

(b.1) $\xi_{ij}^n = G_n(\frac{i}{n}, \frac{j}{n})$;

(b.2) $\xi_{ij}^n = \xi_{ji}^n =$ Bernoulli$(G_n(\frac{i}{n}, \frac{j}{n}))$ independently for $1 \le i \le j \le n$, and independent of $\{B_u, X_u(0) : u \in I\}$.

(c) $G_n \to G$ in the cut metric as $n \to \infty$.

Remark 3.1. *Although in this work we consider $G_n \to G$ in the cut metric, it is also possible to analyze the convergence to the limiting graphon particle system (2.1) when $G_n \to G$ in the cut distance, with suitable additional assumptions, illustrated as follows.*

The cut distance on $\mathcal{G}$ is defined by

$$\delta_\square(G_1, G_2) := \inf_{\varphi \in S_I} \|G_1 - G_2^\varphi\|_\square,$$

where S_I denotes the set of all invertible measure preserving maps $I \to I$, and $G^\varphi(u,v) := G(\varphi(u), \varphi(v))$. If $\delta_\square(G_n, G) \to 0$ for a sequence of step graphons G_n, then it follows from [32, Theorem 11.59] *that $\|\widetilde{G}_n - G\|_\square \to 0$, where $\widetilde{G}_n$ is a suitable relabeling of G_n. That is, there exists some permutation $\tilde{\varphi}_n$ on $\{1,\dots,n\}$ such that, letting $\varphi_n \in S_I$ be*

$$\varphi_n(\frac{i}{n} - u) = \frac{\tilde{\varphi}_n(i)}{n} - u, \quad 0 \le u < \frac{1}{n}, \quad i \in \{1,\dots,n\},$$

and $\widetilde{G}_n = G_n^{\varphi_n}$, one has $d_\square(\widetilde{G}_n, G) \to 0$.

Therefore Condition 3.1(c) could be replaced by the following one, in terms of the cut distance with an additional assumption to guarantee that the initial conditions are matching (which holds naturally if $\{X_i^n(0)\}$ are i.i.d.):

(c*) *$\delta_\square(G_n, G) \to 0$ as $n \to \infty$, so that there exists some invertible measure preserving map $\varphi_n \colon I \to I$, interpreted as a relabeling of the graph G_n, such that $d_\square(G_n^{\varphi_n}, G) \to 0$. Also suppose $X_i^n(0) = X_{\varphi_n^{-1}(\frac{i}{n})}(0)$.*

One could then apply results in this work to $Y_i^n := X^n_{n\varphi_n(\frac{i}{n})}$ *with the kernel* $\widetilde{G}_n := G_n^{\varphi_n}$*, on noting that* $Y_i^n(0) = X^n_{n\varphi_n(\frac{i}{n})}(0) = X_{\frac{i}{n}}(0)$*. This would also imply, as one would naturally expect, the convergence of empirical measures is independent of the labeling of the particles; see e.g.* [5] *for more discussions on this.*

The following convergence holds for the system (3.1). The proof is given in Section 6.2.

Theorem 3.1. *Suppose Conditions* 2.1*,* 2.2*(a) and* 3.1 *hold. Then*

$$\mu^n \to \bar{\mu} \text{ in } \mathcal{P}(\mathcal{C}_d) \text{ in probability} \tag{3.3}$$

as $n \to \infty$*, where*

$$\mu^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}, \quad \bar{\mu} := \int_I \mu_u \, du.$$

If in addition Condition 2.2*(b) holds, then*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|^2_{*,T} \to 0 \tag{3.4}$$

as $n \to \infty$*.*

Remark 3.2. *We note that the regularity assumption (Condition* 2.2*(b)) of* G *is not needed to obtain the convergence in* (3.3)*, but it helps obtain the* L^2 *convergence in* (3.4)*.*

Remark 3.3. *We also note that the interaction* $\{\xi_{ij}^n\}$ *in Condition* 3.1*(b) could be either weighted or random.*

(a) When Condition 3.1*(b.2) holds so that* $\{\xi_{ij}^n\}$ *are random, the convergence in Theorem* 3.1 *(and Theorem* 4.1*) is understood in the annealed sense, where the probability and expectation are taken with respect to all the randomness in the Brownian motions, initial states, and random* $\{\xi_{ij}^n\}$*.*

(b) When Condition 3.1*(b.1) holds so that* $\{\xi_{ij}^n\}$ *are deterministic, the convergence* (3.3) *in Theorem* 3.1 *(and* (4.3) *in Theorem* 4.1*) could be interpreted in the quenched sense under suitable assumptions, given the realization of* $\{\xi_{ij}^n\}$ *possibly generated in a random manner. For example, suppose* $\widetilde{G}_n$ *is sampled from* G *as follows: sample* n *points* $x_1^n, \dots, x_n^n$ *i.i.d. from the uniform distribution on* I *and relabel the indices so that* $x_1^n \le x_2^n \le \cdots \le x_n^n$*. Define* $\widetilde{G}_n$ *as the step graphon with* $\widetilde{G}_n(\frac{i}{n}, \frac{j}{n}) = Bernoulli(G(x_i^n, x_j^n))$ *independently for* $1 \le i \le j \le n$*, and independent of* $\{B_u, X_u(0) : u \in I\}$*. It then follows from* [32, Lemma 10.16] *and a Borel–Cantelli lemma argument (see e.g.* [17, Remark 6] *and* [39, Theorem 5]*) that one has the following almost sure convergence in the cut distance*

$$\delta_\square(\widetilde{G}_n, G) \to 0, \qquad \text{as } n \to \infty.$$

As illustrated in Remark 3.1*, one can then find a relabeling* G_n *of* $\widetilde{G}_n$ *such that* $d_\square(G_n, G) \to 0$ *almost surely, namely Condition* 3.1*(b.1) holds. Now suppose the initial states are i.i.d. so that the empirical measure is not affected by relabeling. Then one can apply* (3.3) *to almost every graph realization* G_n *to get a quenched law of large numbers, that is,*

$$\mu^n \to \bar{\mu} \text{ in } \mathcal{P}(\mathcal{C}_d) \text{ in } \mathbb{P}[\cdot \mid \{G_m\}_{m\in\mathbb{N}}] \text{ as } n \to \infty,$$

for almost every graph realization G_n*.*

If the interaction ξ^n_{ij} is sampled from a common graphon G that is Lipschitz, one could get a uniform rate of convergence.

Condition 3.2. *Suppose the strength of interaction ξ^n_{ij} is generated as follows. Either one of the following two cases holds for all $\{\xi^n_{ij} : i, j \in \{1, \dots, n\}, n \in \mathbb{N}\}$:*

(i) $\xi^n_{ij} = G(\frac{i}{n}, \frac{j}{n})$;
(ii) $\xi^n_{ij} = \xi^n_{ji} =$*Bernoulli*$(G(\frac{i}{n}, \frac{j}{n}))$ *independently for* $1 \le i \le j \le n$, *and independent of* $\{B_u, X_u(0) : u \in I\}$.

The proof of the following rate of convergence is given in Section 6.3. We note that the rate is consistent with that in the classic mean-field setup (in e.g. [43]).

Theorem 3.2. *Suppose Conditions* 2.1, 2.3 *and* 3.2 *hold. Then there exists some $\kappa \in (0, \infty)$ such that*

$$\max_{i=1,\dots,n} \mathbb{E}\|X^n_i - X_{\frac{i}{n}}\|^2_{*,T} \le \frac{\kappa}{n}, \quad \forall\, n \in \mathbb{N}. \tag{3.5}$$

4. Mean-field particle systems on not-so-dense graphs

In this section we consider a sequence of n interacting diffusions (1.2) with the strength of interaction governed by ξ^n_{ij} associated with some kernel G_n in a not-so-dense manner:

$$X^n_i(t) = X_{\frac{i}{n}}(0) + \int_0^t \frac{1}{n\beta_n} \sum_{j=1}^n \xi^n_{ij} b(X^n_i(s), X^n_j(s))\, ds + \int_0^t \sigma(X^n_i(s))\, dB_{\frac{i}{n}}(s), \quad i \in \{1, \dots, n\}. \tag{4.1}$$

We make the following assumptions on the initial states and system coefficients. The forms and assumptions on b and σ are taken for technical requirements, in order to apply certain Girsanov's arguments.

Condition 4.1. *(a) The map $I \ni u \mapsto \mu_u(0) = \mathcal{L}(X_u(0)) \in \mathcal{P}(\mathbb{R}^d)$ is measurable. Moreover,*

$$\sup_{u \in I} \mathbb{E}|X_u(0)|^2 < \infty.$$

(b) The function b is bounded and Lipschitz.
(c) The function σ is bounded, Lipschitz and invertible with bounded inverse $1/\sigma$.
(d) $\beta_n \in (0, 1]$ and $\lim_{n\to\infty} n\beta_n = \infty$.

We make the following assumptions on the strength of interaction ξ^n_{ij} and the associated kernel G_n.

Condition 4.2. *(a) G_n is a step graphon, that is,* (3.2) *holds: $G_n(u, v) = G_n(\frac{\lceil nu \rceil}{n}, \frac{\lceil nv \rceil}{n})$ for $(u, v) \in I \times I$.*
(b) $\xi^n_{ij} = \xi^n_{ji} =$ Bernoulli$(\beta_n G_n(\frac{i}{n}, \frac{j}{n}))$ independently for $1 \le i \le j \le n$, and independent of $\{B_u, X_u(0) : u \in I\}$.
(c) $G_n \to G$ in the cut metric as $n \to \infty$.

Remark 4.1. *(a) The interaction in* (4.1) *is locally mean-field, where the strength of interaction between particles i and j is ξ^n_{ij} scaled down by $n\beta_n$, the order of number of neighbors of i or j.*

(b) Condition 4.2(b) is also known as bond percolation, which has been studied for converging graph sequences in [8]. *If we take* $\lim_{n\to\infty}\beta_n = 0$ *then graphs that we obtain are not-so-dense, in that for a graph* $\mathbb{G}_n$ *with an order of* n^2 *edges, the percolated graph will have approximately an order of* $n^2\beta_n$ *edges. Therefore* β_n *can be interpreted as the global sparsity parameter.*

The limiting graphon particle system is given by

$$X_u(t) = X_u(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(X_u(s), x)G(u, v)\,\mu_{v,s}(dx)\,dv\,ds + \int_0^t \sigma(X_u(s))\,dB_u(s), \quad \mu_{u,t} = \mathcal{L}(X_u(t)), \quad u \in I. \tag{4.2}$$

This is a special case of (2.1) and hence Proposition 2.1 and Theorem 2.1 (with Remark 5.2) still hold.

The following theorem gives a LLN for the system (4.1). The proof is given in Section 7.2.

Theorem 4.1. *Suppose Conditions 2.2(a), 4.1 and 4.2 hold. Then*

$$\mu^n \to \bar{\mu} \text{ in } \mathcal{P}(\mathcal{C}_d) \text{ in probability} \tag{4.3}$$

as $n \to \infty$, *where*

$$\mu^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}, \quad \bar{\mu} := \int_I \mu_u\,du.$$

If in addition Condition 2.2(b) holds, then

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|^2_{*,T} \to 0 \tag{4.4}$$

as $n \to \infty$.

If the interaction ξ_{ij}^n is sampled from a common graphon G that is Lipschitz, one could get a precise rate of convergence.

Condition 4.3. $\xi_{ij}^n = \xi_{ji}^n$ =*Bernoulli*$(\beta_n G(\frac{i}{n}, \frac{j}{n}))$ *independently for* $1 \le i \le j \le n$, *and independent of* $\{B_u, X_u(0) : u \in I\}$.

The proof of the following rate of convergence is given in Section 7.2.

Theorem 4.2. *Suppose Conditions 2.3, 4.1 and 4.3 hold. Then for each* $q \in (1, \infty)$ *there exists some* $\kappa(q) \in (0, \infty)$ *such that*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|^2_{*,T} \le \frac{\kappa(q)}{(n\beta_n)^{1/q}}, \quad \forall\, n \in \mathbb{N}. \tag{4.5}$$

5. Proofs for Section 2

In this section we prove Proposition 2.1 and Theorem 2.1.

5.1. **Proof of Proposition 2.1.** Define the map $\mathcal{M} \ni \mu \mapsto \Phi(\mu) \in [\mathcal{P}(\mathcal{C}_d)]^I$ by $\Phi(\mu) := (\mathcal{L}(X_u^\mu) : u \in I)$, where X_u^μ is the solution of

$$\begin{aligned} X_u^\mu(t) = X_u(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx)\, dv\, dB_u(s). \end{aligned} \tag{5.1}$$

Note that the pathwise uniqueness of $\{X_u^\mu : u \in I\}$ is guaranteed by the Lipschitz properties of b and σ (see e.g. [29, Theorem 5.2.5]). We claim that

$$\text{the pathwise existence of } \{X_u^\mu : u \in I\} \text{ holds and } \Phi(\mu) \in \mathcal{M} \text{ for } \mu \in \mathcal{M}. \tag{5.2}$$

The proof of (5.2) is deferred to the end.

Next we claim that

$$W_{2,t}^{\mathcal{M}}(\Phi(\mu), \Phi(\nu)) \le \kappa \int_0^t W_{2,s}^{\mathcal{M}}(\mu, \nu)\, ds, \quad \mu, \nu \in \mathcal{M}. \tag{5.3}$$

Note that the right hand side is well-defined since the integrand $W_{2,s}^{\mathcal{M}}(\mu,\nu)$ is increasing and hence measurable in $s \in [0,T]$. To show (5.3), consider the coupling

$$\begin{aligned} X_u^\mu(t) = X_u(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx)\, dv\, dB_u(s), \\ X_u^\nu(t) = X_u(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(X_u^\nu(s), x) G(u,v)\, \nu_{v,s}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(X_u^\nu(s), x) G(u,v)\, \nu_{v,s}(dx)\, dv\, dB_u(s). \end{aligned}$$

It then follows from Holder's inequality and the Burkholder-Davis-Gundy inequality that

$$\begin{aligned} &\mathbb{E}\|X_u^\mu - X_u^\nu\|_{*,t}^2 \\ &\le \kappa \mathbb{E} \int_0^t \int_I \left| \int_{\mathbb{R}^d} b(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx) - \int_{\mathbb{R}^d} b(X_u^\nu(s), x) G(u,v)\, \nu_{v,s}(dx) \right|^2 dv\, ds \\ &\quad + \kappa \mathbb{E} \int_0^t \int_I \left| \int_{\mathbb{R}^d} \sigma(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx) - \int_{\mathbb{R}^d} \sigma(X_u^\nu(s), x) G(u,v)\, \nu_{v,s}(dx) \right|^2 dv\, ds. \end{aligned}$$

By adding and subtracting terms, we have

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} b(X_u^\mu(s), x) G(u,v)\, \mu_{v,s}(dx) - \int_{\mathbb{R}^d} b(X_u^\nu(s), x) G(u,v)\, \nu_{v,s}(dx) \right|^2 \\ &\le 2 \left| \int_{\mathbb{R}^d} [b(X_u^\mu(s), x) - b(X_u^\nu(s), x)] G(u,v)\, \mu_{v,s}(dx) \right|^2 \\ &\quad + 2 \left| \int_{\mathbb{R}^d} b(X_u^\nu(s), x) G(u,v)\, [\mu_{v,s} - \nu_{v,s}](dx) \right|^2 \\ &\le \kappa |X_u^\mu(s) - X_u^\nu(s)|^2 + \kappa [W_{2,s}^{\mathcal{M}}(\mu,\nu)]^2, \end{aligned}$$

where the last line uses the Lipschitz property of b and Remark 2.4. The same estimate holds when b is replaced by σ in the last display. It then follows from Gronwall's inequality that

$$\mathbb{E}\|X_u^\mu - X_u^\nu\|_{*,t}^2 \le \kappa \int_0^t [W_{2,s}^{\mathcal{M}}(\mu,\nu)]^2\, ds.$$

Therefore the claim (5.3) holds.

Using the claim (5.3), we can immediately get pathwise uniqueness for the solution of (2.1). The pathwise existence also follows from (5.3) and a standard contraction argument (see e.g. [43, Section I.1]). To be precise, taking $\nu = (\mathcal{L}(Y_u) : u \in I)$ where $Y_u(t) \equiv X_u(0)$ for $u \in I$ and $t \in [0,T]$, iterating (5.3) gives

$$W_{2,T}^{\mathcal{M}}(\Phi^{k+1}(\nu), \Phi^k(\nu)) \le \kappa^k \frac{T^k}{k!} W_{2,T}^{\mathcal{M}}(\Phi(\nu),\nu), \quad k \in \mathbb{N}.$$

Using (2.2), the Lipschitz properties of b and σ, and the fact that $\Phi(\nu) \in \mathcal{M}$, one clearly has $W_{2,T}^{\mathcal{M}}(\Phi(\nu),\nu) < \infty$. Therefore $\Phi^k(\nu)$ is a Cauchy sequence, and hence there exists some $\mu = (\mu_u)_{u\in I} \in [\mathcal{P}(\mathcal{C}_d)]^I$ such that $\lim_{k\to\infty} W_{2,T}^{\mathcal{M}}(\Phi^k(\nu),\mu) = 0$ and $\sup_{u\in I}\int_{\mathcal{C}_d} \|x\|_{*,T}^2\, \mu_u(dx) < \infty$. This gives the existence in law of the solution of (2.1), which together with the pathwise uniqueness gives the pathwise existence of the solution of (2.1). Using the claim (5.2), we have $\Phi^k(\nu) \in \mathcal{M}$ and hence $u \mapsto [\Phi^k(\nu)]_u$ is measurable by the definition of $\mathcal{M}$. Since $\mathcal{P}(\mathcal{C}_d)$ is Polish and $\lim_{k\to\infty} W_{2,T}^{\mathcal{M}}(\Phi^k(\nu),\mu) = 0$, we have the measurability of $I \ni u \mapsto \mu_u \in \mathcal{P}(\mathcal{C}_d)$ (cf. [25, Theorem 4.2.2]). Using the Lipschitz properties of b and σ, one has $\sup_{u\in I}\mathbb{E}\|X_u\|_{*,T}^{2+\varepsilon} < \infty$.

Finally we verify the claim (5.2). Fix $\mu \in \mathcal{M}$. Let $\tilde{X}_u^0(t) \equiv X_u(0)$ for $t \in [0,T]$ and $u \in I$. For $n \in \mathbb{N}$, let

$$\begin{aligned}\tilde{X}_u^n(t) = \tilde{X}_u^{n-1}(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_u^{n-1}(s),x) G(u,v)\, \mu_{v,s}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(\tilde{X}_u^{n-1}(s),x) G(u,v)\, \mu_{v,s}(dx)\, dv\, dB_u(s). \end{aligned} \tag{5.4}$$

Since $\mu \in \mathcal{M}$, it follows that $\tilde{X}_u^n$ and integrals in (5.4) are well-defined for all $u \in I$ and $n \in \mathbb{N}$. We will prove by induction that

$$I \ni u \mapsto \mathcal{L}(\tilde{X}_u^n, B_u) \in \mathcal{P}(\mathcal{C}_d \times \mathcal{C}_d) \text{ is measurable for each } n = 0,1,\dots \tag{5.5}$$

By construction and Condition 2.1(a), (5.5) holds for $n = 0$. Next suppose (5.5) holds up to $k-1$ for some $k \in \mathbb{N}$. To complete the proof of (5.5), it suffices to show that

$$I \ni u \mapsto \mathcal{L}(\tilde{X}_u^k(t_1), B_u(t_1), \dots, \tilde{X}_u^k(t_m), B_u(t_m)) \in \mathcal{P}((\mathbb{R}^d \times \mathbb{R}^d)^m)$$

is measurable for all $0 \le t_1 \le \cdots \le t_m \le T$ and $m \in \mathbb{N}$. It then suffices to show that

$$I \ni u \mapsto \mathbb{E}\left[\prod_{i=1}^m \Big(f_i(\tilde{X}_u^k(t_i)) g_i(B_u(t_i))\Big)\right] \in \mathbb{R}$$

is measurable for all $0 \le t_1 \le \cdots \le t_m \le T$, $m \in \mathbb{N}$ and bounded and continuous functions $\{f_i, g_i : i = 1,\dots,m\}$ on $\mathbb{R}^d$. Now consider the following auxiliary processes

$$\begin{aligned}\tilde{X}_u^{k,\delta}(t) = \tilde{X}_u^{k-1}(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_u^{k-1}(\lfloor \tfrac{s}{\delta} \rfloor \delta),x) G(u,v)\, \mu_{v,\lfloor \frac{s}{\delta}\rfloor\delta}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(\tilde{X}_u^{k-1}(\lfloor \tfrac{s}{\delta} \rfloor \delta),x) G(u,v)\, \mu_{v,\lfloor \frac{s}{\delta}\rfloor\delta}(dx)\, dv\, dB_u(s), \end{aligned}$$

where $\delta \in (0,1)$. Clearly, $\tilde{X}^{k,\delta}_u(t)$ converges to $\tilde{X}^k_u(t)$ in probability as $\delta \to 0$ for each $u \in I$. So it suffices to prove that

$$I \ni u \mapsto \mathbb{E}\left[\prod_{i=1}^m \Big(f_i(\tilde{X}^{k,\delta}_u(t_i))g_i(B_u(t_i))\Big)\right] \in \mathbb{R}$$

is measurable for all $0 \le t_1 \le \cdots \le t_m \le T$, $m \in \mathbb{N}$ and bounded and continuous functions $\{f_i, g_i : i = 1, \dots, m\}$ on $\mathbb{R}^d$. Fix $t \in [0,T]$. Since (5.5) holds for $k-1$, it further suffices to show that

$$\tilde{X}^{k,\delta}_u(t) = h(u, \tilde{X}^{k-1}_u, B_u)$$

for some measurable function $h\colon I \times \mathcal{C}_d \times \mathcal{C}_d \to \mathbb{R}$. Noting that $\tilde{X}^{k,\delta}_u$ is in fact a finite sum of terms depending on $\{\tilde{X}^{k-1}_u(0), \tilde{X}^{k-1}_u(\delta), \dots, \tilde{X}^{k-1}_u(\lfloor \frac{t}{\delta} \rfloor \delta)\}$ and $\{B_u(0), B_u(\delta), \dots, B_u(\lfloor \frac{t}{\delta} \rfloor \delta)\}$ continuously, we have that $h(u,\cdot,\cdot)$ is continuous on $\mathcal{C}_d \times \mathcal{C}_d$ for each $u \in I$, and that $h(\cdot, x, w)$ is measurable on I for each $(x,w) \in \mathcal{C}_d \times \mathcal{C}_d$. Therefore h is measurable and this verifies (5.5) by induction. Using the Lipschitz Properties of b and σ, from (5.4) we can get

$$\sup_{u\in I} \mathbb{E}\|\tilde{X}^{n+1}_u - \tilde{X}^n_u\|^2_{*,t} \le \kappa \int_0^t \sup_{u\in I} \mathbb{E}\|\tilde{X}^n_u - \tilde{X}^{n-1}_u\|^2_{*,s}\, ds.$$

Therefore $\{\tilde{X}^n_u : n \in \mathbb{N}\}$ is Cauchy and converges uniformly in $u \in I$ in probability to some X^μ_u that satisfies (5.1). Since $\mathcal{P}(\mathcal{C}_d)$ is Polish, the measurability of $u \mapsto \mathcal{L}(\tilde{X}^n_u)$ then guarantees that $I \ni u \mapsto \mathcal{L}(X^\mu_u) \in \mathcal{P}(\mathcal{C}_d)$ is measurable (cf. [25, Theorem 4.2.2]). Also noting that $\sup_{n\in\mathbb{N}} \sup_{u\in I} \mathbb{E}\|\tilde{X}^n_u\|^2_{*,T} < \infty$, we see that Φ is actually a well-defined map from $\mathcal{M}$ to $\mathcal{M}$. This verifies (5.2) and completes the proof of Proposition 2.1. ∎

From the proof of the claim (5.2) we can also get the following joint measurability, which will be used in the proof of Theorem 2.1.

Remark 5.1. *Given graphons G_i and measures μ^{G_i}, $i = 1,2$, let $\tilde{X}^{G_i,0}_u(t) \equiv X_u(0)$ for $t \in [0,T]$ and $u \in I$. For $n \in \mathbb{N}$, let*

$$\begin{aligned}\tilde{X}^{G_i,n}_u(t) = \tilde{X}^{G_i,n-1}_u(0) &+ \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}^{G_i,n-1}_u(s), x) G_i(u,v)\, \mu^{G_i}_{v,s}(dx)\, dv\, ds \\ &+ \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(\tilde{X}^{G_i,n-1}_u(s), x) G_i(u,v)\, \mu^{G_i}_{v,s}(dx)\, dv\, dB_u(s).\end{aligned}$$

A similar proof to that of (5.5) shows that $I \ni u \mapsto \mathcal{L}(\tilde{X}^{G_1,n}_u, \tilde{X}^{G_2,n}_u, B_u) \in \mathcal{P}(\mathcal{C}_d \times \mathcal{C}_d \times \mathcal{C}_d)$ is measurable for each $n = 0, 1, \dots$. It then follows that $\{(\tilde{X}^{G_1,n}_u, \tilde{X}^{G_2,n}_u) : n \in \mathbb{N}\}$ is Cauchy and converges uniformly in $u \in I$ in probability to some $(X^{G_1,\mu^{G_1}}_u, X^{G_2,\mu^{G_2}}_u)$, where $X^{G_i,\mu^{G_i}}_u$ satisfies (5.1) with G and μ replaced by G_i and μ^{G_i}. Therefore we have the joint measurability of $I \ni u \mapsto \mathcal{L}(X^{G_1,\mu^{G_1}}_u, X^{G_2,\mu^{G_2}}_u) \in \mathcal{P}(\mathcal{C}_d \times \mathcal{C}_d)$.

5.2. **Proof of Theorem 2.1.** (a) (b) Fix $u_1, u_2 \in I$. Consider the following diffusions:

$$\begin{aligned}
\tilde{X}_{u_1}(t) &= \tilde{X}_{u_1}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{u_1}(s), x) G(u_1, v)\, \mu_{v,s}(dx)\, dv\, ds \\
&\quad + \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(\tilde{X}_{u_1}(s), x) G(u_1, v)\, \mu_{v,s}(dx)\, dv\, dB(s), \\
\tilde{X}_{u_2}(t) &= \tilde{X}_{u_2}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{u_2}(s), x) G(u_2, v)\, \mu_{v,s}(dx)\, dv\, ds \\
&\quad + \int_0^t \int_I \int_{\mathbb{R}^d} \sigma(\tilde{X}_{u_2}(s), x) G(u_2, v)\, \mu_{v,s}(dx)\, dv\, dB(s).
\end{aligned}$$

Here B is a d-dimensional Brownian motion independent of $\{\tilde{X}_{u_1}(0), \tilde{X}_{u_2}(0)\}$, $\mathcal{L}(\tilde{X}_{u_1}(0)) = \mu_{u_1}(0)$, $\mathcal{L}(\tilde{X}_{u_2}(0)) = \mu_{u_2}(0)$, but $\tilde{X}_{u_1}(0)$ and $\tilde{X}_{u_2}(0)$ may not be independent. From Proposition 2.1 we have $\mathcal{L}(\tilde{X}_{u_1}) = \mu_{u_1}$ and $\mathcal{L}(\tilde{X}_{u_2}) = \mu_{u_2}$. Also note that

$$\begin{aligned}
&\mathbb{E}\|\tilde{X}_{u_1} - \tilde{X}_{u_2}\|^2_{*,t} \\
&\le \kappa \mathbb{E}|\tilde{X}_{u_1}(0) - \tilde{X}_{u_2}(0)|^2 \\
&\quad + \kappa \mathbb{E} \int_0^t \int_I \int_{\mathbb{R}^d} \left| b(\tilde{X}_{u_1}(s), x) G(u_1, v) - b(\tilde{X}_{u_2}(s), x) G(u_2, v) \right|^2 \mu_{v,s}(dx)\, dv\, ds \\
&\quad + \kappa \mathbb{E} \int_0^t \int_I \int_{\mathbb{R}^d} \left| \sigma(\tilde{X}_{u_1}(s), x) G(u_1, v) - \sigma(\tilde{X}_{u_2}(s), x) G(u_2, v) \right|^2 \mu_{v,s}(dx)\, dv\, ds \\
&\le \kappa \mathbb{E}|\tilde{X}_{u_1}(0) - \tilde{X}_{u_2}(0)|^2 + \kappa \mathbb{E} \int_0^t |\tilde{X}_{u_1}(s) - \tilde{X}_{u_2}(s)|^2\, ds + \kappa \int_I |G(u_1, v) - G(u_2, v)|^2\, dv,
\end{aligned}$$

where the last line follows on adding and subtracting terms, using the Lipschitz properties of b and σ, and uniformly finite second moment of μ_u. It then follows from Gronwall's inequality that

$$[W_{2,T}(\mu_{u_1}, \mu_{u_2})]^2 \le \mathbb{E}\|\tilde{X}_{u_1} - \tilde{X}_{u_2}\|^2_{*,T} \le \kappa \mathbb{E}|\tilde{X}_{u_1}(0) - \tilde{X}_{u_2}(0)|^2 + \kappa \int_I |G(u_1, v) - G(u_2, v)|^2\, dv.$$

Taking the infimum over random variables $\tilde{X}_{u_1}(0)$ and $\tilde{X}_{u_2}(0)$ such that $\mathcal{L}(\tilde{X}_{u_1}(0)) = \mu_{u_1}(0)$ and $\mathcal{L}(\tilde{X}_{u_2}(0)) = \mu_{u_2}(0)$, we have

$$[W_{2,T}(\mu_{u_1}, \mu_{u_2})]^2 \le \kappa [W_2(\mu_{u_1}(0), \mu_{u_2}(0))]^2 + \kappa \int_I |G(u_1, v) - G(u_2, v)|^2\, dv.$$

Part (a) and Part (b) then follow from Condition 2.2 and Condition 2.3, respectively.

(c) Fix $G_n \to G$ in the cut metric as $n \to \infty$. Let X^{G_n}, μ^{G_n} (resp. X^G, μ^G) be the solution of (2.1) associated with the graphon G_n (resp. G). Fix $t \in [0, T]$. Using Remark 5.1, the left

hand side below is well-defined and we have

$$\begin{aligned}\int_I \mathbb{E}\|X_u^{G_n} - X_u^G\|_{*,t}^2\,du \le \kappa \int_0^t \int_I \mathbb{E}\left|\int_I \int_{\mathbb{R}^d} b(X_u^{G_n}(s),x) G_n(u,v)\,\mu_{v,s}^{G_n}(dx)\,dv\right. \\ \left. - \int_I \int_{\mathbb{R}^d} b(X_u^G(s),x) G(u,v)\,\mu_{v,s}^{G}(dx)\,dv\right|^2 du\,ds \\ + \kappa \int_0^t \int_I \mathbb{E}\left|\int_I \int_{\mathbb{R}^d} \sigma(X_u^{G_n}(s),x) G_n(u,v)\,\mu_{v,s}^{G_n}(dx)\,dv\right. \\ \left. - \int_I \int_{\mathbb{R}^d} \sigma(X_u^{G}(s),x) G(u,v)\,\mu_{v,s}^{G}(dx)\,dv\right|^2 du\,ds. \end{aligned} \tag{5.6}$$

We will analyze the first integrand above for fixed $s \in [0,t]$, and the analysis for σ is similar. By adding and subtracting terms, we have

$$\begin{aligned}&\int_I \mathbb{E}\left|\int_I \int_{\mathbb{R}^d} b(X_u^{G_n}(s),x) G_n(u,v)\,\mu_{v,s}^{G_n}(dx)\,dv - \int_I \int_{\mathbb{R}^d} b(X_u^{G}(s),x) G(u,v)\,\mu_{v,s}^{G}(dx)\,dv\right|^2 du \\ &\le \kappa \int_I \int_I \mathbb{E}\left|\int_{\mathbb{R}^d} [b(X_u^{G_n}(s),x) - b(X_u^G(s),x)] G_n(u,v)\,\mu_{v,s}^{G_n}(dx)\right|^2 dv\,du \\ &\quad + \kappa \int_I \int_I \mathbb{E}\left|\int_{\mathbb{R}^d} b(X_u^{G}(s),x) G_n(u,v)\,[\mu_{v,s}^{G_n} - \mu_{v,s}^{G}](dx)\right|^2 dv\,du \\ &\quad + \kappa \int_I \mathbb{E}\left|\int_I \int_{\mathbb{R}^d} b(X_u^{G}(s),x) [G_n(u,v) - G(u,v)]\,\mu_{v,s}^{G}(dx)\,dv\right|^2 du \\ &=: \kappa\left(\mathcal{J}_s^{n,1} + \mathcal{J}_s^{n,2} + \mathcal{J}_s^{n,3}\right). \end{aligned} \tag{5.7}$$

Now we analyze each term $\mathcal{J}_s^{n,k}$, $k = 1,2,3$. Using the Lipschitz property of b we have

$$\mathcal{J}_s^{n,1} \le \kappa \int_I \mathbb{E}|X_u^{G_n}(s) - X_u^G(s)|^2\,du. \tag{5.8}$$

Using Remark 2.4 and the Lipschitz property of b we have

$$\mathcal{J}_s^{n,2} \le \kappa \int_I [W_{2,s}(\mu_v^{G_n}, \mu_v^G)]^2\,dv. \tag{5.9}$$

For the last term $\mathcal{J}_s^{n,3}$, fix $M \in (0,\infty)$ and write

$$b_M(x,y) := b(x,y)\mathbf{1}_{\{|x|\le M, |y|\le M\}}. \tag{5.10}$$

Since b is Lipschitz continuous, it follows from [42, Corollary 2 of Theorem 3.1] that there exist some $m = m(M) \in \mathbb{N}$ and polynomials

$$\tilde{b}_m(x,y) := \sum_{k=1}^m a_k(x) c_k(y) \mathbf{1}_{\{|x|\le M, |y|\le M\}}, \tag{5.11}$$

where $a_k(x)$ and $c_k(y)$ are powers of x and y respectively, for each $k = 1,\dots,m$, such that

$$|b_M(x,y) - \tilde{b}_m(x,y)| \le 1/M, \quad x,y \in \mathbb{R}^d. \tag{5.12}$$

By adding and subtracting terms, we have

$$\begin{aligned}\mathcal{J}_s^{n,3} &\le \kappa \int_I \int_I \mathbb{E}\left| \int_{\mathbb{R}^d} [b(X_u^G(s),x) - b_M(X_u^G(s),x)][G_n(u,v)-G(u,v)]\, \mu_{v,s}^G(dx)\right|^2 dv\, du \\ &+ \kappa \int_I \int_I \mathbb{E}\left| \int_{\mathbb{R}^d} [b_M(X_u^G(s),x) - \tilde{b}_m(X_u^G(s),x)][G_n(u,v)-G(u,v)]\, \mu_{v,s}^G(dx)\right|^2 dv\, du \\ &+ \kappa \int_I \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_u^G(s),x)[G_n(u,v)-G(u,v)]\, \mu_{v,s}^G(dx)\, dv\right|^2 du \\ &=: \kappa \sum_{k=1}^3 \mathcal{J}_s^{n,3,k}. \end{aligned} \tag{5.13}$$

Next we analyze each term $\mathcal{J}_s^{n,3,k}$, $k=1,2,3$. For $\mathcal{J}_s^{n,3,1}$, using the Lipschitz property of b, (5.10) and Proposition 2.1 we have

$$\mathcal{J}_s^{n,3,1} \le \kappa \int_I \int_I \mathbb{E}\left| \int_{\mathbb{R}^d} (1+|X_u^G(s)|+|x|)(\mathbf{1}_{\{|X_u^G(s)|>M\}} + \mathbf{1}_{\{|x|>M\}})\, \mu_{v,s}^G(dx)\right|^2 dv\, du \le \frac{\kappa}{M^\varepsilon}. \tag{5.14}$$

For $\mathcal{J}_s^{n,3,2}$, using (5.12) we have

$$\mathcal{J}_s^{n,3,2} \le \frac{\kappa}{M^2}. \tag{5.15}$$

For $\mathcal{J}_s^{n,3,3}$, using the definition of the bounded function $\tilde{b}_m$ in (5.11) we have

$$\begin{aligned}\mathcal{J}_s^{n,3,3} &\le \kappa(M) \int_I \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_u^G(s),x)[G_n(u,v)-G(u,v)]\, \mu_{v,s}^G(dx)\, dv\right| du \\ &\le \kappa(M) \sum_{k=1}^m \int_I \mathbb{E}\left[|a_k(X_u^G(s))| \mathbf{1}_{\{|X_u^G(s)|\le M\}}\right] \\ &\qquad \cdot \left| \int_I [G_n(u,v)-G(u,v)] \left[\int_{\mathbb{R}^d} c_k(x) \mathbf{1}_{\{|x|\le M\}} \mu_{v,s}^G(dx)\right] dv\right| du \\ &\le \kappa(M) \|G_n - G\|, \end{aligned} \tag{5.16}$$

where $\kappa(M)$ is some constant that depends on M but not on n. Combining (5.6)–(5.9) and (5.13)–(5.16) with Gronwall's inequality, we have

$$\begin{aligned}\int_I [W_{2,t}(\mu_u^{G_n}, \mu_u^G)]^2\, du &\le \int_I \mathbb{E}\|X_u^{G_n} - X_u^G\|_{*,t}^2\, du \\ &\le \kappa \left(\int_0^t \int_I [W_{2,s}(\mu_u^{G_n}, \mu_u^G)]^2\, du\, ds + \frac{1}{M^\varepsilon} + \kappa(M)\|G_n - G\| \right). \end{aligned}$$

It then follows from the Gronwall's inequality again that

$$\int_I [W_{2,t}(\mu_u^{G_n}, \mu_u^G)]^2\, du \le \kappa \left(\frac{1}{M^\varepsilon} + \kappa(M)\|G_n - G\| \right).$$

Since $G_n \to G$ in the cut metric, from Remark 2.1 we have $\|G_n - G\| \to 0$ as $n \to \infty$. Therefore, by taking $\limsup_{n\to\infty}$ and then $\limsup_{M\to\infty}$ in the last display, we have the desired result. This gives part (c) and completes the proof of Theorem 2.1. ■

Remark 5.2. *The assumption that $\varepsilon > 0$ is used in* (5.14)*. If b and σ are bounded functions, then one can allow $\varepsilon = 0$ and replace the estimate in* (5.14) *by*

$$\mathcal{J}_s^{n,3,1} \le \kappa \int_I \int_I \mathbb{E}\left| \int_{\mathbb{R}^d} (\mathbf{1}_{\{|X_u^G(s)|>M\}} + \mathbf{1}_{\{|x|>M\}})\, \mu_{v,s}^G(dx) \right|^2 dv\, du \le \frac{\kappa}{M^2}.$$

6. Proofs for Section 3

In this section we prove Theorems 3.1 and 3.2.

6.1. **Preliminary Estimates.**

We first provide some preliminary estimates under Condition 2.2.

Lemma 6.1. *Suppose Conditions* 2.1, 2.2 *and* 3.1*(a,b) hold. Then there exist some $\kappa, \kappa(M) \in (0,\infty)$ for each $M \in (0,\infty)$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,T}^2 \le \kappa(M) \limsup_{n\to\infty} \|G_n - G\| + \frac{\kappa}{M^\varepsilon}.$$

Proof. Fix $t \in [0,T]$.

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,t}^2 \\
&\le \kappa \int_0^t \left[\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \right] ds \\
&\quad + \kappa \int_0^t \left[\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \sigma(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} \sigma(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \right] ds.
\end{aligned} \tag{6.1}$$

We will analyze the first integrand above for fixed $s \in [0,t]$, and the analysis for σ is similar. By adding and subtracting terms, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \\
&\le \frac{3}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \left(b(X_i^n(s), X_j^n(s)) - b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) \right) \right|^2 \\
&\quad + \frac{3}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \left(\xi_{ij}^n b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx) \right) \right|^2 \\
&\quad + \frac{3}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \\
&=: 3\left(\mathcal{T}_s^{n,1} + \mathcal{T}_s^{n,2} + \mathcal{T}_s^{n,3} \right).
\end{aligned} \tag{6.2}$$

Fix $s \in [0,T]$. For $\mathcal{T}_s^{n,1}$, using the Lipschitz property of b we have

$$\begin{aligned}\mathcal{T}_s^{n,1} &\le \kappa \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \left(|X_i^n(s) - X_{\frac{i}{n}}(s)|^2 + |X_j^n(s) - X_{\frac{j}{n}}(s)|^2\right)\right] \\ &\le \kappa \frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i^n(s) - X_{\frac{i}{n}}(s)|^2. \end{aligned} \tag{6.3}$$

For $\mathcal{T}_s^{n,2}$, using a weak LLN type argument, we have

$$\begin{aligned}\mathcal{T}_s^{n,2} &= \frac{1}{n}\sum_{i=1}^n \frac{1}{n^2}\sum_{j=1}^n\sum_{k=1}^n \mathbb{E}\left[\left(\xi_{ij}^n b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx)\right)\right. \\ &\qquad \left.\cdot \left(\xi_{ik}^n b(X_{\frac{i}{n}}(s), X_{\frac{k}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{k}{n})\, \mu_{\frac{k}{n},s}(dx)\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^n \frac{1}{n^2}\sum_{j=1}^n \sum_{k\in\{i,j\}} \mathbb{E}\left[\left(\xi_{ij}^n b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx)\right)\right. \\ &\qquad \left.\cdot \left(\xi_{ik}^n b(X_{\frac{i}{n}}(s), X_{\frac{k}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G_n(\frac{i}{n}, \frac{k}{n})\, \mu_{\frac{k}{n},s}(dx)\right)\right] \\ &\le \frac{\kappa}{n}, \end{aligned} \tag{6.4}$$

where the second equality follows from the observation that the expectation is zero whenever $k \notin \{i,j\}$ by Condition 3.1(b) and the independence of $\{X_{\frac{i}{n}}\}$ and $\{\xi_{ij}^n\}$, and the inequality uses the boundedness of ξ_{ij}^n, G_n, Lipschitz property of b, and the uniformly finite second moment of X_u.

The analysis of $\mathcal{T}_s^{n,3}$ is similar to that of $\mathcal{J}_s^{n,3}$ in the proof of Theorem 2.1(c) but is more involved. Fix $M \in (0,\infty)$ and let b_M and $\tilde{b}_m$ be defined as in (5.10) and (5.11), such that

(5.12) holds. By adding and subtracting terms, we have

$$\begin{aligned}
\mathcal{T}_s^{n,3} \le & \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d}[b(X_{\frac{i}{n}}(s),x)-b_M(X_{\frac{i}{n}}(s),x)]G_n(\frac{i}{n},\frac{j}{n})\,\mu_{\frac{j}{n},s}(dx)\right|^2 \\
& + \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\int_I\int_{\mathbb{R}^d}[b(X_{\frac{i}{n}}(s),x)-b_M(X_{\frac{i}{n}}(s),x)]G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2 \\
& + \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d}[b_M(X_{\frac{i}{n}}(s),x)-\tilde{b}_m(X_{\frac{i}{n}}(s),x)]G_n(\frac{i}{n},\frac{j}{n})\,\mu_{\frac{j}{n},s}(dx)\right|^2 \\
& + \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\int_I\int_{\mathbb{R}^d}[b_M(X_{\frac{i}{n}}(s),x)-\tilde{b}_m(X_{\frac{i}{n}}(s),x)]G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2 \\
& + \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d}\tilde{b}_m(X_{\frac{i}{n}}(s),x)G_n(\frac{i}{n},\frac{j}{n})\,\mu_{\frac{j}{n},s}(dx)\right. \\
& \qquad \left. - \int_I\int_{\mathbb{R}^d}\tilde{b}_m(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2 \\
=: & \ \kappa\sum_{k=1}^5 \mathcal{T}_s^{n,3,k}. \qquad (6.5)
\end{aligned}$$

Next we analyze each term. For $\mathcal{T}_s^{n,3,1}$ and $\mathcal{T}_s^{n,3,2}$, using Proposition 2.1 and (5.10) we have

$$\mathcal{T}_s^{n,3,1} \le \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d}(1+|X_{\frac{i}{n}}(s)|+|x|)[\mathbf{1}_{\{|X_{\frac{i}{n}}(s)|>M\}}+\mathbf{1}_{\{|x|>M\}}]\,\mu_{\frac{j}{n},s}(dx)\right|^2 \le \frac{\kappa}{M^\varepsilon}, \qquad (6.6)$$

$$\mathcal{T}_s^{n,3,2} \le \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left|\int_I\int_{\mathbb{R}^d}(1+|X_{\frac{i}{n}}(s)|+|x|)[\mathbf{1}_{\{|X_{\frac{i}{n}}(s)|>M\}}+\mathbf{1}_{\{|x|>M\}}]\,\mu_{v,s}(dx)\,dv\right|^2 \le \frac{\kappa}{M^\varepsilon}. \qquad (6.7)$$

For $\mathcal{T}_s^{n,3,3}$ and $\mathcal{T}_s^{n,3,4}$, using (5.12) we have

$$\mathcal{T}_s^{n,3,3} \le \frac{\kappa}{M^2}, \quad \mathcal{T}_s^{n,3,4} \le \frac{\kappa}{M^2}. \qquad (6.8)$$

For $\mathcal{T}_s^{n,3,5}$, using the step graphon structure (3.2) of G_n we have

$$\begin{aligned}
\mathcal{T}_s^{n,3,5} &= \int_I \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G_n(u,v)\, \mu_{\frac{\lceil nv\rceil}{n},s}(dx)\, dv \right. \\
&\qquad \left. - \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G(\frac{\lceil nu\rceil}{n},v)\, \mu_{v,s}(dx)\, dv \right|^2 du \\
&\le 2\int_I \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) \left[G_n(u,v)-G(u,v)\right] \mu_{\frac{\lceil nv\rceil}{n},s}(dx)\, dv \right|^2 du \\
&\quad + 2\int_I \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G(u,v)\, \mu_{\frac{\lceil nv\rceil}{n},s}(dx)\, dv \right. \\
&\qquad \left. - \int_I \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G(\frac{\lceil nu\rceil}{n},v)\, \mu_{v,s}(dx)\, dv \right|^2 du \\
&=: \mathcal{T}_s^{n,3,6} + \mathcal{T}_s^{n,3,7}. \qquad (6.9)
\end{aligned}$$

For $\mathcal{T}_s^{n,3,6}$, using the definition of the bounded function $\tilde{b}_m$ in (5.11) we have

$$\begin{aligned}
\mathcal{T}_s^{n,3,6} &\le \kappa(M) \sum_{k=1}^m \int_I \left| \int_I [G_n(u,v)-G(u,v)] \left[\int_{\mathbb{R}^d} c_k(x) \mathbf{1}_{\{\|x\|\le M\}} \mu_{\frac{\lceil nv\rceil}{n},s}(dx) \right] dv \right| du \\
&\le \kappa(M) \|G_n - G\|,
\end{aligned}$$

where $\kappa(M)$ depends on M but not on n. For $\mathcal{T}_s^{n,3,7}$, using Condition 2.2 and Theorem 2.1(a) we see that

$$\int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G(u,v)\, \mu_{\frac{\lceil nv\rceil}{n},s}(dx) - \int_{\mathbb{R}^d} \tilde{b}_m(X_{\frac{\lceil nu\rceil}{n}}(s),x) G(\frac{\lceil nu\rceil}{n},v)\, \mu_{v,s}(dx) \to 0$$

as $n\to\infty$, for each $u\in I$ and $v\in I\setminus A_u$. Since the set A_u is assumed to have Lebesgue measure zero in Condition 2.2(b), we have

$$\lim_{n\to\infty} \int_0^T \mathcal{T}_s^{n,3,7}\, ds = 0. \qquad (6.10)$$

Combining (6.5)–(6.10) gives

$$\limsup_{n\to\infty} \int_0^T \mathcal{T}_s^{n,3}\, ds \le \kappa(M) \limsup_{n\to\infty} \|G_n - G\| + \frac{\kappa}{M^\varepsilon}. \qquad (6.11)$$

Combining (6.1)–(6.4) gives

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,t}^2 \le \kappa \int_0^t \frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,s}^2\, ds + \frac{\kappa}{n} + \kappa \int_0^t \mathcal{T}_s^{n,3}\, ds.$$

Using the Gronwall's inequality and (6.11) completes the proof. ∎

Remark 6.1. *The assumption that $\varepsilon>0$ is used in* (6.6) *and* (6.7)*. If b and σ are bounded functions, then one can allow $\varepsilon=0$ and replace the estimate in* (6.6) *and* (6.7) *by*

$$\begin{aligned}
\mathcal{T}_s^{n,3,1} &\le \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d} \left[\mathbf{1}_{\{|X_{\frac{i}{n}}(s)|>M\}} + \mathbf{1}_{\{|x|>M\}}\right] \mu_{\frac{j}{n},s}(dx) \right|^2 \le \frac{\kappa}{M^2}, \\
\mathcal{T}_s^{n,3,2} &\le \frac{\kappa}{n}\sum_{i=1}^n \mathbb{E}\left| \int_I \int_{\mathbb{R}^d} \left[\mathbf{1}_{\{|X_{\frac{i}{n}}(s)|>M\}} + \mathbf{1}_{\{|x|>M\}}\right] \mu_{v,s}(dx)\, dv \right|^2 \le \frac{\kappa}{M^2}.
\end{aligned}$$

As a consequence, the result in Lemma 6.1 *is rewritten as*

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,T}^2 \le \kappa(M)\limsup_{n\to\infty}\|G_n - G\| + \frac{\kappa}{M^2}.$$

Lemma 6.2. *Suppose Conditions* 2.1 *and* 2.2 *hold. Let* $\bar\mu^n := \frac{1}{n}\sum_{i=1}^n \delta_{X_{\frac{i}{n}}}$. *Then* $\bar\mu^n \to \bar\mu$ *in* $\mathcal{P}(\mathcal{C}_d)$ *in probability as* $n\to\infty$.

Proof. Let $\tilde\mu^n := \frac{1}{n}\sum_{i=1}^n \mu_{\frac{i}{n}}$. For each bounded and continuous function f on $\mathcal{C}_d$, using the independence of $\{X_{\frac{i}{n}}\}$, we have

$$\mathbb{E}\left(\langle f,\bar\mu^n\rangle - \langle f,\tilde\mu^n\rangle\right)^2 = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n\Big(f(X_{\frac{i}{n}}) - \mathbb{E}f(X_{\frac{i}{n}})\Big)\right)^2 \le \frac{\|f\|_\infty^2}{n}\to 0.$$

It then follows from Condition 2.2 and Theorem 2.1(a) that

$$\langle f,\tilde\mu^n\rangle - \langle f,\bar\mu\rangle = \frac{1}{n}\sum_{i=1}^n\langle f,\mu_{\frac{i}{n}}\rangle - \int_I\langle f,\mu_u\rangle\,du\to 0.$$

Combining these two estimates completes the proof. ∎

6.2. **Proof of Theorem 3.1.** If Condition 2.2(b) holds, then (3.4) follows on a direct application of Lemma 6.1 by sending $M\to\infty$.

Next we prove (3.3) by an approximation argument that is commonly used to analyze graphons (see e.g. [9]). Fix $\eta\in(0,1)$. Since $0\le G\le 1$, using Lusin's theorem ([41, Theorem 2.24]), we can approximate G by continuous functions $\tilde G$ with $0\le\tilde G\le 1$. Namely, we can find some continuous graphon $\tilde G = \tilde G_\eta$ such that

$$\|\tilde G - G\| < \eta.$$

It then follows from Theorem 2.1(c) that

$$W_{2,T}(\bar\mu^{\tilde G},\bar\mu)\le\kappa\eta, \tag{6.12}$$

where we denote by $X^{\tilde G} = \{X_u^{\tilde G} : u\in I\}$ the solution of (2.1) corresponding to the graphon $\tilde G$ and let $\bar\mu^{\tilde G} := \int_I \mathcal{L}(X_u^{\tilde G})\,du$. Since Condition 2.2 holds for $\tilde G$, from Lemma 6.2 we have

$$\frac{1}{n}\sum_{i=1}^n\delta_{X_{\frac{i}{n}}^{\tilde G}}\to\bar\mu^{\tilde G} \tag{6.13}$$

in $\mathcal{P}(\mathcal{C}_d)$ in probability as $n\to\infty$. Note that

$$[W_{2,T}(\mu^n,\frac{1}{n}\sum_{i=1}^n\delta_{X_{\frac{i}{n}}^{\tilde G}})]^2 \le \frac{1}{n}\sum_{i=1}^n\|X_i^n - X_{\frac{i}{n}}^{\tilde G}\|_{*,T}^2.$$

Since Condition 2.2 holds for $\tilde G$, it follows from Lemma 6.1 and Remark 2.1 that

$$\limsup_{n\to\infty}\mathbb{E}W_{2,T}(\mu^n,\frac{1}{n}\sum_{i=1}^n\delta_{X_{\frac{i}{n}}^{\tilde G}}) \le \sqrt{\kappa(M)\limsup_{n\to\infty}\|G_n-\tilde G\| + \frac{\kappa}{M^\varepsilon}} \le \sqrt{\kappa(M)\eta + \frac{\kappa}{M^\varepsilon}}.$$

Taking $\eta\to 0$ and then $M\to\infty$ gives

$$\limsup_{\eta\to 0}\limsup_{n\to\infty}\mathbb{E}W_{2,T}(\mu^n,\frac{1}{n}\sum_{i=1}^n\delta_{X_{\frac{i}{n}}^{\tilde G}}) = 0. \tag{6.14}$$

Combining (6.12)–(6.14), taking $n \to \infty$ and then sending $\eta \to 0$, we have (3.3). ■

6.3. **Proof of Theorem 3.2.** Fix $t \in [0,T]$ and $i \in \{1,\dots,n\}$. Similar to the proof of Lemma 6.1, we have

$$\begin{aligned}
&\mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,t}^2 \\
&\le \kappa\mathbb{E}\int_0^t \left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 ds \\
&\quad + \kappa\mathbb{E}\int_0^t \left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \sigma(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} \sigma(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 ds. \quad (6.15)
\end{aligned}$$

We will analyze the first integrand above for fixed $s \in [0,t]$, and the analysis for σ is similar. By adding and subtracting terms, we have

$$\begin{aligned}
&\mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n b(X_i^n(s), X_j^n(s)) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \\
&\le 3\mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \xi_{ij}^n \left(b(X_i^n(s), X_j^n(s)) - b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) \right) \right|^2 \\
&\quad + 3\mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \left(\xi_{ij}^n b(X_{\frac{i}{n}}(s), X_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx) \right) \right|^2 \\
&\quad + 3\mathbb{E}\left| \frac{1}{n}\sum_{j=1}^n \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx) - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \right|^2 \\
&=: 3\left(\tilde{\mathcal{T}}_s^{n,1} + \tilde{\mathcal{T}}_s^{n,2} + \tilde{\mathcal{T}}_s^{n,3} \right). \quad (6.16)
\end{aligned}$$

For $\tilde{\mathcal{T}}_s^{n,1}$, using the Lipschitz property of b we have

$$\tilde{\mathcal{T}}_s^{n,1} \le \kappa\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \left(|X_i^n(s) - X_{\frac{i}{n}}(s)|^2 + |X_j^n(s) - X_{\frac{j}{n}}(s)|^2 \right) \right] \le 2\kappa \max_{i=1,\dots,n} \mathbb{E}|X_i^n(s) - X_{\frac{i}{n}}(s)|^2. \quad (6.17)$$

For $\tilde{\mathcal{T}}_s^{n,2}$, using Condition 3.2, the independence of $\{X_{\frac{i}{n}}\}$ and $\{\xi_{ij}^n\}$, the Lipschitz property of b, Proposition 2.1, and a weak LLN type argument, we have

$$\tilde{\mathcal{T}}_s^{n,2} \le \frac{\kappa}{n}. \quad (6.18)$$

For $\tilde{\mathcal{T}}_s^{n,3}$, we have

$$\begin{aligned}\tilde{\mathcal{T}}_s^{n,3} &= \mathbb{E}\left|\int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},\frac{\lceil nv\rceil}{n})\,\mu_{\frac{\lceil nv\rceil}{n},s}(dx)\,dv - \int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2\\ &\le 2\mathbb{E}\left|\int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)\left[G(\frac{i}{n},\frac{\lceil nv\rceil}{n}) - G(\frac{i}{n},v)\right]\mu_{\frac{\lceil nv\rceil}{n},s}(dx)\,dv\right|^2\\ &\quad + 2\mathbb{E}\left|\int_I\left[\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)\,\mu_{\frac{\lceil nv\rceil}{n},s}(dx)\,dv - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)\,\mu_{v,s}(dx)\right]G(\frac{i}{n},v)\,dv\right|^2\\ &\le \frac{\kappa}{n^2}, \end{aligned} \tag{6.19}$$

where the last inequality uses Condition 2.3, Theorem 2.1(b) and Remark 2.4. Combining (6.15)–(6.19) with Gronwall's inequality gives (3.5) and completes the proof of Theorem 3.2. ■

7. Proofs for Section 4

In this section we prove Theorems 4.1 and 4.2.

7.1. **Preliminary Estimates.** We first provide some preliminary estimates.

Lemma 7.1. *Suppose Condition* 4.1 *holds. Suppose either Condition* 4.2 *or Condition* 4.3 *holds. Then*

$$\sup_{n\in\mathbb{N}}\max_{i=1,\dots,n}\mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,T}^k < \infty, \quad \forall k\in\mathbb{N}.$$

Proof of Lemma 7.1. Fix $k\in\mathbb{N}$ and $i\in\{1,\dots,n\}$. Since b and σ are bounded, we have

$$\begin{aligned}\mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,T}^k &\le \kappa\mathbb{E}\left|\frac{1}{n\beta_n}\sum_{j=1}^n \xi_{ij}^n\right|^k + \kappa\\ &\le \kappa\frac{\sum_{j=1}^n \mathbb{E}[(\xi_{ij}^n)^k] + \left(\sum_{j=1}^n E[(\xi_{ij}^n)^2]\right)^{k/2}}{(n\beta_n)^k} + \kappa \le \kappa,\end{aligned}$$

where the second inequality follows from the Rosenthal's inequality [40, Theorem 3] and the last inequality uses the assumption $\lim_{n\to\infty} n\beta_n = \infty$. ■

Lemma 7.2. *Suppose Conditions* 2.2, 4.1 *and* 4.2*(a,b) hold. Then there exist some* $\kappa, \kappa(M)\in(0,\infty)$ *for each* $M\in(0,\infty)$ *such that*

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,T}^2 \le \kappa(M)\limsup_{n\to\infty}\|G_n - G\| + \frac{\kappa}{M^2}.$$

Proof. Fix $t \in [0,T]$.

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\|X_i^n - X_{\frac{i}{n}}\|_{*,t}^2 \\
&\leq \kappa\int_0^t\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n\beta_n}\sum_{j=1}^{n}\xi_{ij}^n b(X_i^n(s),X_j^n(s)) - \int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2\right]ds \\
&\quad + \kappa\int_0^t\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|X_i^n(s) - X_{\frac{i}{n}}(s)\right|^2\right]ds. \qquad (7.1)
\end{aligned}$$

We will analyze the first integrand above for fixed $s \in [0,t]$. By adding and subtracting terms, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n\beta_n}\sum_{j=1}^{n}\xi_{ij}^n b(X_i^n(s),X_j^n(s)) - \int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2 \\
&\leq \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_{ij}^n}{\beta_n}\left(b(X_i^n(s),X_j^n(s)) - b(X_{\frac{i}{n}}(s),X_j^n(s))\right)\right|^2 \\
&\quad + \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_{ij}^n}{\beta_n}\left(b(X_{\frac{i}{n}}(s),X_j^n(s)) - b(X_{\frac{i}{n}}(s),X_{\frac{j}{n}}(s))\right)\right|^2 \\
&\quad + \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\xi_{ij}^n}{\beta_n}b(X_{\frac{i}{n}}(s),X_{\frac{j}{n}}(s)) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G_n(\frac{i}{n},\frac{j}{n})\,\mu_{\frac{j}{n},s}(dx)\right)\right|^2 \\
&\quad + \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G_n(\frac{i}{n},\frac{j}{n})\,\mu_{\frac{j}{n},s}(dx) - \int_I\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s),x)G(\frac{i}{n},v)\,\mu_{v,s}(dx)\,dv\right|^2 \\
&=: 4\left(\mathcal{R}_s^{n,1} + \mathcal{R}_s^{n,2} + \mathcal{R}_s^{n,3} + \mathcal{R}_s^{n,4}\right). \qquad (7.2)
\end{aligned}$$

Fix $s \in [0,T]$. For $\mathcal{R}_s^{n,1}$, using the Lipschitz property of b we have

$$\begin{aligned}
\mathcal{R}_s^{n,1} &\leq \kappa\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_{ij}^n}{\beta_n}|X_i^n(s) - X_{\frac{i}{n}}(s)|\right)^2 \\
&\leq \kappa\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\xi_{ij}^n - \beta_n G_n(\frac{i}{n},\frac{j}{n})}{\beta_n}|X_i^n(s) - X_{\frac{i}{n}}(s)|\right)^2 \\
&\quad + \kappa\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}G_n(\frac{i}{n},\frac{j}{n})|X_i^n(s) - X_{\frac{i}{n}}(s)|\right)^2.
\end{aligned}$$

Since $\{\xi^n_{ij} - \beta_n G_n(\frac{i}{n}, \frac{j}{n})\}$ are centered and independent, we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\xi^n_{ij} - \beta_n G_n(\frac{i}{n}, \frac{j}{n})}{\beta_n}\right)^4 \\
&= \frac{1}{(n\beta_n)^4}\sum_{j=1}^{n}\mathbb{E}\left(\xi^n_{ij} - \beta_n G_n(\frac{i}{n}, \frac{j}{n})\right)^4 \\
&\quad + \frac{3}{(n\beta_n)^4}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\mathbb{E}\left(\xi^n_{ij} - \beta_n G_n(\frac{i}{n}, \frac{j}{n})\right)^2\mathbb{E}\left(\xi^n_{ik} - \beta_n G_n(\frac{i}{n}, \frac{k}{n})\right)^2 \\
&\leq \frac{1}{(n\beta_n)^3} + \frac{3}{(n\beta_n)^2}.
\end{aligned}$$

From these two estimates, Cauchy–Schwarz inequality and Lemma 7.1 we have

$$\mathcal{R}^{n,1}_s \leq \frac{\kappa}{n\beta_n} + \frac{\kappa}{n}\sum_{i=1}^{n}\mathbb{E}|X^n_i(s) - X_{\frac{i}{n}}(s)|^2. \tag{7.3}$$

For $\mathcal{R}^{n,3}_s$, using Condition 4.2, the independence of $\{X_{\frac{i}{n}}\}$ and $\{\xi^n_{ij}\}$, the boundedness of b, and a weak LLN type argument, we have

$$\mathcal{R}^{n,3}_s \leq \frac{\kappa}{n\beta_n}. \tag{7.4}$$

For $\mathcal{R}^{n,4}_s$, similar to the proof of (6.11) in Lemma 6.1 and Remark 6.1, using Conditions 2.2, 4.1 and 4.2(a,b) we could get

$$\limsup_{n\to\infty}\int_0^T \mathcal{R}^{n,4}_s\,ds \leq \kappa(M)\limsup_{n\to\infty}\|G_n - G\| + \frac{\kappa}{M^2}. \tag{7.5}$$

The analysis of $\mathcal{R}^{n,2}_s$ is based on a collection of change of measure arguments. First note that

$$\begin{aligned}
\mathcal{R}^{n,2}_s &\leq \kappa\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\xi^n_{ij}}{\beta_n}|X^n_j(s) - X_{\frac{j}{n}}(s)|\right)^2 \\
&= \kappa\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(n\beta_n)^2}\left(\sum_{j=1}^{n}\mathbb{E}\left[\xi^n_{ij}|X^n_j(s) - X_{\frac{j}{n}}(s)|^2\right]\right. \\
&\quad \left. + \sum_{j=1}^{n}\sum_{k\neq j}^{n}\mathbb{E}\left[\xi^n_{ij}\xi^n_{ik}|X^n_j(s) - X_{\frac{j}{n}}(s)||X^n_k(s) - X_{\frac{k}{n}}(s)|\right]\right).
\end{aligned} \tag{7.6}$$

Fix $i, j \in \{1, \ldots, n\}$. We want to show that $\mathbb{E}\left[\xi^n_{ij}(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right]$ is roughly $\mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right]\beta_n G_n(\frac{i}{n}, \frac{j}{n})$ up to some errors, namely ξ^n_{ij} is kind of independent of $X^n_j(s) - X_{\frac{j}{n}}(s)$. For this, we want to compare $\mathcal{L}(X^n_j, X_{\frac{j}{n}}\,|\,\xi^n_{ij} = 1)$ and $\mathcal{L}(X^n_j, X_{\frac{j}{n}})$ using

Girsanov's theorem. This heuristic suggests considering the following auxiliary processes

$$\begin{aligned}\tilde{X}_{\frac{i}{n}}(t) &= X_{\frac{i}{n}}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv\, ds + \int_0^t \sigma(\tilde{X}_{\frac{i}{n}}(s))\, dB_{\frac{i}{n}}(s) \\ &\quad + \int_0^t \frac{\sigma(\tilde{X}_{\frac{i}{n}}(s))}{\sigma(X_i^n(s))} \frac{1}{n\beta_n} \left(\xi_{ij}^n - 1\right) b(X_i^n(s), X_j^n(s))\, ds, \\ \tilde{X}_{\frac{j}{n}}(t) &= X_{\frac{j}{n}}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{\frac{j}{n}}(s), x) G(\frac{j}{n}, v)\, \mu_{v,s}(dx)\, dv\, ds + \int_0^t \sigma(\tilde{X}_{\frac{j}{n}}(s))\, dB_{\frac{j}{n}}(s) \\ &\quad + \int_0^t \frac{\sigma(\tilde{X}_{\frac{j}{n}}(s))}{\sigma(X_j^n(s))} \frac{1}{n\beta_n} \left(\xi_{ji}^n - 1\right) b(X_j^n(s), X_i^n(s))\, ds.\end{aligned}$$

Note that the existence and uniqueness of such processes are guaranteed by the bounded and Lipschitz properties of b, σ and σ^{-1}. Also using these properties and Gronwall's inequality we can show that

$$\mathbb{E}\|X_{\frac{j}{n}} - \tilde{X}_{\frac{j}{n}}\|_{*,T}^m \leq \frac{\kappa(m)}{(n\beta_n)^m}, \quad m \geq 0. \tag{7.7}$$

Define $Q^{i,j,n}$ by

$$\begin{aligned}\frac{dQ^{i,j,n}}{d\mathbb{P}} &= \mathcal{E}_T \left(\int_0^\cdot \frac{1}{\sigma(X_i^n(s))} \frac{1}{n\beta_n} \left(1 - \xi_{ij}^n\right) b(X_i^n(s), X_j^n(s))\, dB_{\frac{i}{n}}(s) \right. \\ &\quad \left. + \int_0^\cdot \frac{1}{\sigma(X_j^n(s))} \frac{1}{n\beta_n} \left(1 - \xi_{ji}^n\right) b(X_j^n(s), X_i^n(s))\, dB_{\frac{j}{n}}(s) \right),\end{aligned}$$

where

$$\mathcal{E}_t(M) := \exp\left\{ M_t - \frac{1}{2}[M]_t \right\}$$

is the Doleans exponential for a semi-martingale M_t. Since b and σ^{-1} are bounded, we have that $B_{\frac{i}{n}}(\cdot) - \int_0^\cdot \frac{1}{\sigma(X_i^n(s))} \frac{1}{n\beta_n}(1 - \xi_{ij}^n) b(X_i^n(s), X_j^n(s))\, ds$, $B_{\frac{j}{n}}(\cdot) - \int_0^\cdot \frac{1}{\sigma(X_j^n(s))} \frac{1}{n\beta_n}(1 - \xi_{ji}^n) b(X_j^n(s), X_i^n(s))\, ds$ and $B_{\frac{k}{n}}$, $k \neq i, j$, are independent Brownian motion under $Q^{i,j,n}$, and

$$\mathbb{P}\left(\left(X_i^n, X_j^n, X_{\frac{i}{n}}, X_{\frac{j}{n}} \right) \in \cdot \,|\, \xi_{ij}^n = 1 \right) = Q^{i,j,n}\left(\left(X_i^n, X_j^n, \tilde{X}_{\frac{i}{n}}, \tilde{X}_{\frac{j}{n}} \right) \in \cdot \right) \tag{7.8}$$

by Girsanov's theorem, and

$$\mathbb{E}\left[\left(\frac{dQ^{i,j,n}}{d\mathbb{P}} \right)^m \right] \leq \exp\left\{ \frac{m|m-1|\kappa}{(n\beta_n)^2} \right\}, \quad m \geq 0.$$

From this it then follows that

$$\begin{aligned}\mathbb{E}\left[\left| \frac{dQ^{i,j,n}}{d\mathbb{P}} - 1 \right|^m \right] &\leq \sqrt{ \mathbb{E}\left[\left| \frac{dQ^{i,j,n}}{d\mathbb{P}} - 1 \right|^2 \right] \mathbb{E}\left[\left| \frac{dQ^{i,j,n}}{d\mathbb{P}} - 1 \right|^{2m-2} \right] } \\ &\leq \sqrt{ \mathbb{E}\left[\left(\frac{dQ^{i,j,n}}{d\mathbb{P}} \right)^2 - 1 \right] \kappa(m) \mathbb{E}\left[\left| \frac{dQ^{i,j,n}}{d\mathbb{P}} \right|^{2m-2} + 1 \right] } \\ &\leq \sqrt{ \left[1 + \frac{\kappa(m)}{(n\beta_n)^2} - 1 \right] \kappa(m) } \leq \frac{\kappa(m)}{n\beta_n}, \quad m \geq 1,\end{aligned} \tag{7.9}$$

where the third inequality uses the assumption that $\lim_{n\to\infty} n\beta_n = \infty$. From (7.8) we have

$$\begin{aligned}\mathbb{E}\left[\xi^n_{ij}(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right] &= \mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2 \,|\, \xi^n_{ij} = 1\right]\beta_n G_n(\frac{i}{n},\frac{j}{n}) \\ &= \mathbb{E}_{Q^{i,j,n}}\left[(X^n_j(s) - \tilde{X}_{\frac{j}{n}}(s))^2\right]\beta_n G_n(\frac{i}{n},\frac{j}{n}).\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}_{Q^{i,j,n}}\left[(X^n_j(s) - \tilde{X}_{\frac{j}{n}}(s))^2\right] &= \mathbb{E}\left[(X^n_j(s) - \tilde{X}_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,n}}{d\mathbb{P}}\right] \\ &\le 2\mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,n}}{d\mathbb{P}}\right] + 2\mathbb{E}\left[(X_{\frac{j}{n}}(s) - \tilde{X}_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,n}}{d\mathbb{P}}\right].\end{aligned}$$

For the first term, using Holder's inequality, Lemma 7.1 and (7.9) we have

$$\begin{aligned}&\mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,n}}{d\mathbb{P}}\right] - \mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right] \\ &\le \left(\mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^{2p}\right]\right)^{1/p}\left(\mathbb{E}\left[\left|\frac{dQ^{i,j,n}}{d\mathbb{P}} - 1\right|^q\right]\right)^{1/q} \\ &\le \frac{\kappa(q)}{(n\beta_n)^{1/q}}, \quad \forall q > 1.\end{aligned}$$

For the second term, using Holder's inequality and (7.7) we have

$$\mathbb{E}\left[\left(X_{\frac{j}{n}}(s) - \tilde{X}_{\frac{j}{n}}(s)\right)^2\frac{dQ^{i,j,n}}{d\mathbb{P}}\right] \le \sqrt{\mathbb{E}\left[\left(X_{\frac{j}{n}}(s) - \tilde{X}_{\frac{j}{n}}(s)\right)^4\right]\mathbb{E}\left[\left(\frac{dQ^{i,j,n}}{d\mathbb{P}}\right)^2\right]} \le \frac{\kappa}{(n\beta_n)^2}.$$

Combining these three estimates gives

$$\mathbb{E}\left[\xi^n_{ij}(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right] \le \left(2\mathbb{E}\left[(X^n_j(s) - X_{\frac{j}{n}}(s))^2\right] + \frac{\kappa(q)}{(n\beta_n)^{1/q}}\right)\beta_n G_n(\frac{i}{n},\frac{j}{n}). \tag{7.10}$$

Now fix $i,j,k \in \{1,\dots,n\}$ with $j \neq k$. The following argument is similar to the above change of measure, but we provide the proof for completeness. The intuition is again that $\mathbb{E}\left[\xi^n_{ij}\xi^n_{ik}|X^n_j(s) - X_{\frac{j}{n}}(s)||X^n_k(s) - X_{\frac{k}{n}}(s)|\right]$ is roughly $\mathbb{E}\left[|X^n_j(s) - X_{\frac{j}{n}}(s)||X^n_k(s) - X_{\frac{k}{n}}(s)|\right]\beta^2_n G_n(\frac{i}{n},\frac{j}{n})G_n(\frac{i}{n},\frac{k}{n})$ up to some errors, namely (ξ^n_{ij},ξ^n_{ik}) is kind of independent of $(X^n_j(s) - X_{\frac{j}{n}}(s))(X^n_k(s) - X_{\frac{k}{n}}(s))$. For this, we want to compare $\mathcal{L}(X^n_j, X_{\frac{j}{n}}, X^n_k, X_{\frac{k}{n}} \,|\, \xi^n_{ij} = 1, \xi^n_{ik} = 1)$ and $\mathcal{L}(X^n_j, X_{\frac{j}{n}}, X^n_k, X_{\frac{k}{n}})$ using Girsanov's theorem. This

heuristic suggests considering the following auxiliary processes

$$\begin{aligned}
\tilde{X}_{\frac{i}{n}}(t) &= X_{\frac{i}{n}}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv\, ds + \int_0^t \sigma(\tilde{X}_{\frac{i}{n}}(s))\, dB_{\frac{i}{n}}(s) \\
&\quad + \sum_{l=j,k} \int_0^t \frac{\sigma(\tilde{X}_{\frac{i}{n}}(s))}{\sigma(X_i^n(s))} \frac{1}{n\beta_n} (\xi_{il}^n - 1)\, b(X_i^n(s), X_l^n(s))\, ds, \\
\tilde{X}_{\frac{j}{n}}(t) &= X_{\frac{j}{n}}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{\frac{j}{n}}(s), x) G(\frac{j}{n}, v)\, \mu_{v,s}(dx)\, dv\, ds + \int_0^t \sigma(\tilde{X}_{\frac{j}{n}}(s))\, dB_{\frac{j}{n}}(s) \\
&\quad + \int_0^t \frac{\sigma(\tilde{X}_{\frac{j}{n}}(s))}{\sigma(X_j^n(s))} \frac{1}{n\beta_n} \left(\xi_{ji}^n - 1\right) b(X_j^n(s), X_i^n(s))\, ds \\
\tilde{X}_{\frac{k}{n}}(t) &= X_{\frac{k}{n}}(0) + \int_0^t \int_I \int_{\mathbb{R}^d} b(\tilde{X}_{\frac{k}{n}}(s), x) G(\frac{k}{n}, v)\, \mu_{v,s}(dx)\, dv\, ds + \int_0^t \sigma(\tilde{X}_{\frac{k}{n}}(s))\, dB_{\frac{k}{n}}(s) \\
&\quad + \int_0^t \frac{\sigma(\tilde{X}_{\frac{k}{n}}(s))}{\sigma(X_k^n(s))} \frac{1}{n\beta_n} (\xi_{ki}^n - 1)\, b(X_k^n(s), X_i^n(s))\, ds.
\end{aligned}$$

Note that the existence and uniqueness of such processes are again guaranteed by the bounded and Lipschitz properties of b, σ and σ^{-1}. Also using these properties and Gronwall's inequality we can show that

$$\mathbb{E}\|X_{\frac{j}{n}} - \tilde{X}_{\frac{j}{n}}\|_{*,T}^m + \mathbb{E}\|X_{\frac{k}{n}} - \tilde{X}_{\frac{k}{n}}\|_{*,T}^m \leq \frac{\kappa(m)}{(n\beta_n)^m}, \quad m \geq 0. \tag{7.11}$$

Define $Q^{i,j,k,n}$ by

$$\begin{aligned}
\frac{dQ^{i,j,k,n}}{d\mathbb{P}} = \mathcal{E}_T \Bigg(&\sum_{l=j,k} \int_0^{\cdot} \frac{1}{\sigma(X_i^n(s))} \frac{1}{n\beta_n} (1 - \xi_{il}^n)\, b(X_i^n(s), X_l^n(s))\, dB_{\frac{i}{n}} \\
&+ \sum_{l=j,k} \int_0^{\cdot} \frac{1}{\sigma(X_l^n(s))} \frac{1}{n\beta_n} (1 - \xi_{li}^n)\, b(X_l^n(s), X_i^n(s))\, dB_{\frac{l}{n}} \Bigg).
\end{aligned}$$

Since b and σ^{-1} are bounded, we have

$$\begin{aligned}
&\mathbb{P}\left(\left(X_i^n, X_j^n, X_k^n, X_{\frac{i}{n}}, X_{\frac{j}{n}}, X_{\frac{k}{n}}\right) \in \cdot \,|\, \xi_{ij}^n = 1, \xi_{ik}^n = 1\right) \\
&\quad = Q^{i,j,k,n}\left(\left(X_i^n, X_j^n, X_k^n, \tilde{X}_{\frac{i}{n}}, \tilde{X}_{\frac{j}{n}}, \tilde{X}_{\frac{k}{n}}\right) \in \cdot\right)
\end{aligned} \tag{7.12}$$

by Girsanov's theorem, and

$$\mathbb{E}\left[\left(\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right)^m\right] \leq \exp\left\{\frac{m|m-1|\kappa}{(n\beta_n)^2}\right\}, m \geq 0.$$

Using this and the assumption that $\lim_{n\to\infty} n\beta_n = \infty$, we have

$$\mathbb{E}\left[\left|\frac{dQ^{i,j,n}}{d\mathbb{P}} - 1\right|^m\right] \leq \sqrt{\mathbb{E}\left[\left|\frac{dQ^{i,j,n}}{d\mathbb{P}} - 1\right|^2\right] \mathbb{E}\left[\left|\frac{dQ^{i,j,n}}{d\mathbb{P}} - 1\right|^{2m-2}\right]} \leq \frac{\kappa(m)}{n\beta_n}, m \geq 1. \tag{7.13}$$

From (7.12) we have

$$\begin{aligned}
&\mathbb{E}\left[\xi^n_{ij}\xi^n_{ik}|X^n_j(s)-X_{\frac{j}{n}}(s)||X^n_k(s)-X_{\frac{k}{n}}(s)|\right]\\
&=\mathbb{E}\left[|X^n_j(s)-X_{\frac{j}{n}}(s)||X^n_k(s)-X_{\frac{k}{n}}(s)|\,|\,\xi^n_{ij}=1,\xi^n_{ik}=1\right]\beta_n^2G_n(\frac{i}{n},\frac{j}{n})G_n(\frac{i}{n},\frac{k}{n})\\
&=\mathbb{E}_{Q^{i,j,k,n}}\left[|X^n_j(s)-\tilde{X}_{\frac{j}{n}}(s)||X^n_k(s)-\tilde{X}_{\frac{k}{n}}(s)|\right]\beta_n^2G_n(\frac{i}{n},\frac{j}{n})G_n(\frac{i}{n},\frac{k}{n}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E}_{Q^{i,j,k,n}}\left[|X^n_j(s)-\tilde{X}_{\frac{j}{n}}(s)||X^n_k(s)-\tilde{X}_{\frac{k}{n}}(s)|\right]\\
&\le\frac{1}{2}\mathbb{E}_{Q^{i,j,k,n}}\left[|X^n_j(s)-\tilde{X}_{\frac{j}{n}}(s)|^2\right]+\frac{1}{2}\mathbb{E}_{Q^{i,j,k,n}}\left[|X^n_k(s)-\tilde{X}_{\frac{k}{n}}(s)|^2\right]\\
&=\frac{1}{2}\mathbb{E}\left[(X^n_j(s)-\tilde{X}_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right]+\frac{1}{2}\mathbb{E}\left[(X^n_k(s)-\tilde{X}_{\frac{k}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right]\\
&\le\mathbb{E}\left[(X^n_j(s)-X_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right]+\mathbb{E}\left[(X_{\frac{j}{n}}(s)-\tilde{X}_{\frac{j}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right]\\
&\quad+\mathbb{E}\left[(X^n_k(s)-X_{\frac{k}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right]+\mathbb{E}\left[(X_{\frac{k}{n}}(s)-\tilde{X}_{\frac{k}{n}}(s))^2\frac{dQ^{i,j,k,n}}{d\mathbb{P}}\right].
\end{aligned}$$

Similar to the derivation of (7.10), using Holder's inequality, Lemma 7.1, (7.13), and (7.11) we have

$$\begin{aligned}
&\mathbb{E}\left[\xi^n_{ij}\xi^n_{ik}|X^n_j(s)-X_{\frac{j}{n}}(s)||X^n_k(s)-X_{\frac{k}{n}}(s)|\right] \qquad (7.14)\\
&\le\left(\mathbb{E}\left[(X^n_j(s)-X_{\frac{j}{n}}(s))^2\right]+\mathbb{E}\left[(X^n_k(s)-X_{\frac{k}{n}}(s))^2\right]+\frac{\kappa_q}{(n\beta_n)^{1/q}}\right)\beta_n^2G_n(\frac{i}{n},\frac{j}{n})G_n(\frac{i}{n},\frac{k}{n}).
\end{aligned}$$

Applying (7.10) and (7.14) to (7.6) gives

$$\mathcal{R}^{n,2}_s\le\frac{\kappa}{n}\sum_{j=1}^n\mathbb{E}|X^n_j(s)-X_{\frac{j}{n}}(s)|^2+\frac{\kappa(q)}{(n\beta_n)^{1/q}}. \qquad (7.15)$$

Finally, combining (7.1)–(7.4) and (7.15) gives

$$\frac{1}{n}\sum_{i=1}^n\mathbb{E}\|X^n_i-X_{\frac{i}{n}}\|^2_{*,t}\le\kappa\int_0^t\frac{1}{n}\sum_{i=1}^n\mathbb{E}\|X^n_i-X_{\frac{i}{n}}\|^2_{*,s}+\frac{\kappa(q)}{(n\beta_n)^{1/q}}+\kappa\int_0^t\mathcal{R}^{n,4}_s\,ds.$$

Using the Gronwall's inequality and (7.5) completes the proof. ■

7.2. **Proof of Theorem 4.1.** The proof is similar to that of Theorem 3.1. If Condition 2.2(b) holds, then (4.4) follows on a direct application of Lemma 7.2 by sending $M\to\infty$.

Next we prove (4.3). Fix $\eta\in(0,1)$. Since $0\le G\le 1$, using Lusin's theorem ([41, Theorem 2.24]), we can approximate G by continuous functions $\tilde{G}$ with $0\le\tilde{G}\le 1$. Namely, we can find some continuous graphon $\tilde{G}=\tilde{G}_\eta$ such that

$$\|\tilde{G}-G\|<\eta.$$

It then follows from Theorem 2.1(c) that

$$W_{2,T}(\bar{\mu}^{\tilde{G}},\bar{\mu})\le\kappa\eta \qquad (7.16)$$

where we denote by $X^{\tilde{G}} = \{X_u^{\tilde{G}} : u \in I\}$ the solution of (4.2) corresponding to the graphon $\tilde{G}$ and let $\bar{\mu}^{\tilde{G}} := \int_I \mathcal{L}(X_u^{\tilde{G}})\, du$. Since Condition 2.2 holds for $\tilde{G}$, from Lemma 6.2 we have

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X^{\tilde{G}}_{\frac{i}{n}}} \to \bar{\mu}^{\tilde{G}} \tag{7.17}$$

in $\mathcal{P}(\mathcal{C}_d)$ in probability as $n \to \infty$. Note that

$$[W_{2,T}(\mu^n, \frac{1}{n}\sum_{i=1}^{n} \delta_{X^{\tilde{G}}_{\frac{i}{n}}})]^2 \le \frac{1}{n}\sum_{i=1}^{n} \|X_i^n - X^{\tilde{G}}_{\frac{i}{n}}\|^2_{*,T}.$$

Since Condition 2.2 holds for $\tilde{G}$, it follows from Lemma 7.2 and Remark 2.1 that

$$\limsup_{n\to\infty} \mathbb{E} W_{2,T}(\mu^n, \frac{1}{n}\sum_{i=1}^{n} \delta_{X^{\tilde{G}}_{\frac{i}{n}}}) \le \sqrt{\kappa(M) \limsup_{n\to\infty} \|G_n - \tilde{G}\| + \frac{\kappa}{M^\varepsilon}} \le \sqrt{\kappa(M)\eta + \frac{\kappa}{M^\varepsilon}}.$$

Taking $\eta \to 0$ and then $M \to \infty$ gives

$$\limsup_{\eta\to 0} \limsup_{n\to\infty} \mathbb{E} W_{2,T}(\mu^n, \frac{1}{n}\sum_{i=1}^{n} \delta_{X^{\tilde{G}}_{\frac{i}{n}}}) = 0. \tag{7.18}$$

Combining (7.16)–(7.18), taking $n \to \infty$ and then sending $\eta \to 0$, we have (4.3). ■

7.3. **Proof of Theorem 4.2.** Note that the estimates in (7.1)–(7.4) and (7.15) still hold under Conditions 2.3, 4.1 and 4.3. In order to show (4.5), it suffices to argue

$$\mathcal{R}_s^{n,4} \le \frac{\kappa}{n}.$$

For this, using Condition 4.3 we have

$$\begin{aligned}
\mathcal{R}_s^{n,4} &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\bigg| \frac{1}{n}\sum_{j=1}^{n} \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, \frac{j}{n})\, \mu_{\frac{j}{n},s}(dx) \\
&\qquad - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \bigg|^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\bigg| \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, \frac{\lceil nv \rceil}{n})\, \mu_{\frac{\lceil nv \rceil}{n},s}(dx)\, dv \\
&\qquad - \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) G(\frac{i}{n}, v)\, \mu_{v,s}(dx)\, dv \bigg|^2 \\
&\le 2\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\bigg| \int_I \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x) \left[G(\frac{i}{n}, \frac{\lceil nv \rceil}{n}) - G(\frac{i}{n}, v) \right] \mu_{\frac{\lceil nv \rceil}{n},s}(dx)\, dv \bigg|^2 \\
&\quad + 2\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\bigg| \int_I \left[\int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x)\, \mu_{\frac{\lceil nv \rceil}{n},s}(dx) - \int_{\mathbb{R}^d} b(X_{\frac{i}{n}}(s), x)\, \mu_{v,s}(dx) \right] G(\frac{i}{n}, v)\, dv \bigg|^2 \\
&\le \frac{\kappa}{n^2},
\end{aligned}$$

where the last inequality uses Condition 2.3, Theorem 2.1(b) and Remark 2.4. This completes the proof of Theorem 4.2. ■

8. Acknowledgment

We would like to thank the anonymous referees for their careful reading and many valuable suggestions on the paper.

References

[1] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul, *Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons*, The Journal of Mathematical Neuroscience **2** (2012), no. 1, 10.

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

[3] E. Bayraktar and R. Wu, *Mean field interaction on random graphs with dynamically changing multi-color edges*, Stochastic Processes and their Applications **141** (2021), 197–244.

[4] M. J Beeson, *Foundations of constructive mathematics: Metamathematical studies*, Vol. 6, Springer Science & Business Media, 2012.

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

[6] S. Bhamidi, A. Budhiraja, and R. Wu, *Weakly interacting particle systems on inhomogeneous random graphs*, Stochastic Processes and their Applications **129** (2019), no. 6, 2174–2206.

[7] S. Bhamidi, S. Chakraborty, S. Cranmer, and B. Desmarais, *Weighted exponential random graph models: Scope and large network limits*, Journal of Statistical Physics **173** (2018), no. 3-4, 704–735.

[8] B. Bollobás, C. Borgs, J. Chayes, O. Riordan, et al., *Percolation on dense graph sequences*, The Annals of Probability **38** (2010), no. 1, 150–183.

[9] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Structures & Algorithms **31** (2007), no. 1, 3–122.

[10] C. Borgs, J. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Transactions of the American Mathematical Society **372** (2019), no. 5, 3019–3062.

[11] C. Borgs, J. T Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence*, The Annals of Probability **46** (2018), no. 1, 337–396.

[12] A. Budhiraja, P. Dupuis, M. Fischer, and K. Ramanan, *Limits of relative entropies associated with weakly interacting particle systems*, Electronic Journal of Probability **20** (2015), no. 80, 1–22.

[13] A. Budhiraja and R. Wu, *Some fluctuation results for weakly interacting multi-type particle systems*, Stochastic Processes and their Applications **126** (2016), no. 8, 2253–2296.

[14] A. Budhiraja, D. Mukherjee, and R. Wu, *Supermarket model on graphs*, Ann. Appl. Probab. **29** (201906), no. 3, 1740–1777.

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

[16] P. E Caines and M. Huang, *Graphon mean field games and their equations*, SIAM Journal on Control and Optimization **59** (2021), no. 6, 4373–4399.

[17] R. Carmona, D. B Cooney, C. V Graves, and M. Lauriere, *Stochastic graphon games: I. the static case*, Mathematics of Operations Research **47** (2022), no. 1, 750–778.

[18] S. Chatterjee, P. Diaconis, et al., *Estimating and understanding exponential random graph models*, The Annals of Statistics **41** (2013), no. 5, 2428–2461.

[19] F. Collet, *Macroscopic limit of a bipartite Curie–Weiss model: A dynamical approach*, Journal of Statistical Physics **157** (2014), no. 6, 1301–1319.

[20] P. Contucci, I. Gallo, and G. Menconi, *Phase transitions in social sciences: Two-population mean field theory*, International Journal of Modern Physics B **22** (2008), no. 14, 2199–2212.

[21] F. Coppini, *Long time dynamics for interacting oscillators on graphs*, The Annals of Applied Probability **32** (2022), no. 1, 360 –391.

[22] 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 **0** (2019), no. 0, 2050010, available at `https://doi.org/10.1142/S0219493720500100`.

[23] F. Delarue, *Mean field games: A toy model on an Erdös-Renyi graph.*, ESAIM: Proceedings and Surveys **60** (2017), 1–26.

[24] S. Delattre, G. Giacomin, and E. Luçon, *A note on dynamical models on random graphs and Fokker–Planck equations*, Journal of Statistical Physics **165** (2016), no. 4, 785–798.
[25] R. M Dudley, *Real analysis and probability*, CRC Press, 2018.
[26] P. Dupuis and G. S Medvedev, *The large deviation principle for interacting dynamical systems on random graphs*, Communications in Mathematical Physics **390** (2022), no. 2, 545–575.
[27] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations*, Courier Corporation, 2003.
[28] D. Kaliuzhnyi-Verbovetskyi and G. S Medvedev, *The mean field equation for the kuramoto model on graph sequences with non-lipschitz limit*, SIAM Journal on Mathematical Analysis **50** (2018), no. 3, 2441–2465.
[29] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, vol. 113, Springer New York, 1991.
[30] V. N. Kolokoltsov, *Nonlinear Markov Processes and Kinetic Equations*, Cambridge Tracts in Mathematics, vol. 182, Cambridge University Press, 2010.
[31] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, 2006, pp. 199–262.
[32] L. Lovász, *Large networks and graph limits*, Vol. 60, American Mathematical Soc., 2012.
[33] E. Luçon, *Quenched asymptotics for interacting diffusions on inhomogeneous random graphs*, Stochastic Processes and their Applications **130** (2020), no. 11, 6783–6842.
[34] H. P. McKean, *Propagation of chaos for a class of non-linear parabolic equations*, Stochastic differential equations (Lecture Series in Differential Equations, Session 7, Catholic University, 1967), 1967, pp. 41–57.
[35] G. S Medvedev, *The nonlinear heat equation on dense graphs and graph limits*, SIAM Journal on Mathematical Analysis **46** (2014), no. 4, 2743–2766.
[36] G. S Medvedev, *The nonlinear heat equation on w-random graphs*, Archive for Rational Mechanics and Analysis **212** (2014), no. 3, 781–803.
[37] S. Nadtochiy and M. Shkolnikov, *Mean field systems on networks, with singular interaction through hitting times*, The Annals of Probability **48** (2020), no. 3, 1520–1556.
[38] R. I. Oliveira and G. H. Reis, *Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations*, Journal of Statistical Physics (2019Jul).
[39] F. Parise and A. E Ozdaglar, *Graphon games: A statistical framework for network games and interventions*, Available at SSRN: https://ssrn.com/abstract=3437293 (2019).
[40] H. P Rosenthal, *On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables*, Israel Journal of Mathematics **8** (1970), no. 3, 273–303.
[41] W. Rudin, *Real and complex analysis*, Mathematics series, McGraw-Hill, 1987.
[42] M. H. Schultz, *L^∞-multivariate approximation theory*, SIAM Journal on Numerical Analysis **6** (1969), no. 2, 161–183.
[43] A-S. Sznitman, *Topics in propagation of chaos*, Ecole d'Eté de Probabilités de Saint-Flour XIX—1989, 1991, pp. 165–251.

Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109

Simons Institute of Theory of Computing, UC Berkeley, CA 94720

Department of Mathematics, Iowa State University, 411 Morrill Road, Ames, IA 50011
Email address: erhan@umich.edu, sumanc@berkeley.edu, ruoyu@iastate.edu