



HAL
open science

Mean-field limit of non-exchangeable multi-agent systems over hypergraphs with unbounded rank

Nathalie Ayi, Nastassia Pouradier Duteil, David Poyato

► **To cite this version:**

Nathalie Ayi, Nastassia Pouradier Duteil, David Poyato. Mean-field limit of non-exchangeable multi-agent systems over hypergraphs with unbounded rank. 2024. hal-04600947v2

HAL Id: hal-04600947

<https://hal.science/hal-04600947v2>

Preprint submitted on 9 Oct 2024

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

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

MEAN-FIELD LIMIT OF NON-EXCHANGEABLE MULTI-AGENT SYSTEMS OVER HYPERGRAPHS WITH UNBOUNDED RANK

NATHALIE AYI, NASTASSIA POURADIER DUTEIL, AND DAVID POYATO

ABSTRACT. Interacting particle systems are known for their ability to generate large-scale self-organized structures from simple local interaction rules between each agent and its neighbors. In addition to studying their emergent behavior, a main focus of the mathematical community has been concentrated on deriving their large-population limit. In particular, the mean-field limit consists of describing the limit system by its population density in the product space of positions and labels. The strategy to derive such limits is often based on a careful combination of methods ranging from analysis of PDEs and stochastic analysis, to kinetic equations and graph theory. In this article, we focus on a generalization of multi-agent systems that includes higher-order interactions, which has largely captured the attention of the applied community in the last years. In such models, interactions between individuals are no longer assumed to be binary (*i.e.* between a pair of particles). Instead, individuals are allowed to interact by groups so that a full group jointly generates a non-linear force on any given individual. The underlying graph of connections is then replaced by a hypergraph, which we assume to be dense, but possibly non-uniform and of unbounded rank. For the first time in the literature, we show that when the interaction kernels are regular enough, then the mean-field limit is determined by a limiting Vlasov-type equation, where the hypergraph limit is encoded by a so-called UR-hypergraphon (unbounded-rank hypergraphon), and where the resulting mean-field force admits infinitely-many orders of interactions.

CONTENTS

| | |
|--|----|
| 1. Introduction | 2 |
| 2. Preliminaries | 6 |
| 2.1. Graphs and hypergraph limit theories | 6 |
| 2.2. Hypergraphons of unbounded rank | 13 |
| 2.3. Some examples of multi-agent systems over hypergraphs | 16 |
| 2.4. Functional setting | 19 |
| 3. Propagation of independence over hypergraphs | 23 |
| 4. Well-posedness of the Vlasov equation over UR-hypergraphons | 31 |
| 4.1. Properties of the force and well-posedness of characteristics | 31 |
| 4.2. Notion of distributional solution | 33 |
| 4.3. Fixed point argument | 35 |
| 5. Stability estimate of the Vlasov equation over UR-hypergraphons | 38 |
| 6. Proof of main Theorem 1.1 | 45 |
| 7. Numerical simulations | 48 |

Date: October 11, 2024.

2020 Mathematics Subject Classification. 05C65; 34C15; 35Q70; 35Q83; 35Q92; 35R02; 60G09; 92B20.

Key words and phrases. Higher-order interactions, Hypergraphs, Mean-field limit, Non-exchangeable multi-agent systems, Propagation of chaos, Unbounded rank, Vlasov equations.

Acknowledgment. NA has received funding from the Emergence grant of Sorbonne University and partially from the French National Research Agency for the project CONVIVIALITY (ANR-23-CE40-0003). NPD has received funding from the Emergence grant of Sorbonne University. DP has received funding from the European Union’s Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement No 101064402, and partially from Grant C-EXP-265-UGR23 funded by Consejería de Universidad, Investigación e Innovación & ERDF/EU Andalusia Program, from Grant PID2022-137228OB-I00 funded by the Spanish Ministerio de Ciencia, Innovación y Universidades, MICIU/AEI/10.13039/501100011033 & “ERDF/EU A way of making Europe”, and from Modeling Nature Research Unit, project QUAL21-011. We also thank the creators of the HGX Python library [52], which we have used in some of our plots.

| | |
|---|----|
| 7.1. Hypergraph for a homogeneous group of rank 3 | 49 |
| 7.2. Hypergraph for a balanced group of rank 3 | 50 |
| 8. Conclusions and perspectives | 51 |
| Appendix A. Convergence of hypergraphs for homogeneous groups | 52 |
| References | 53 |

1. INTRODUCTION

The study of collective behavior has attracted much attention in the last twenty years, including within the mathematical community. Its numerous applications include among many others the study of animal group coordination, opinion formation and cell organization in tissues. Both theoretical and experimental communities have tackled the question of self-organization, in the aim of elucidating how global collective patterns emerge in a group driven by only local interactions. To address this question, mathematical models have been developed, relying mainly on two important hypotheses:

Hypothesis 1 (Exchangeability). *The particles are assumed to be indistinguishable (or **exchangeable**), that is that exchanging any two particles does not modify the overall dynamics of the group. In other words, the population is composed of identical particles.*

Hypothesis 2 (Pairwise interaction). *The particles are assumed to interact **pairwise**. More precisely, this implies that the influence of the group on each given particle is viewed as the superposition of the pairwise interactions between this particle and all remaining ones.*

Hypotheses 1 and 2 can be used to write a simple general model giving the evolution of each particle's position (or opinion, or velocity, etc.) $X_i^N \in \mathbb{R}^d$ as the result of pairwise interactions with all remaining particles:

$$\frac{dX_i^N(t)}{dt} = \frac{1}{N} \sum_{j=1}^N K(X_i^N(t), X_j^N(t)), \quad i \in \{1, \dots, N\}. \quad (1.1)$$

Here, because of Hypothesis 1, the interaction weights are all equal (normalized to $\frac{1}{N}$), and due to Hypothesis 2, the total effect of the group on the evolution of X_i^N is modeled by the sum of the pairwise interactions.

Although this model provides a good description of many multi-agent systems with indistinguishable particles, in other applications, Hypotheses 1 and 2 need to be revised. One way to model the non-exchangeable nature of the particles (*i.e.* to remove Hypothesis 1) is to describe their pairwise interaction via an underlying weighted graph. This leads to the system

$$\frac{dX_i^N(t)}{dt} = \sum_{j=1}^N w_{ij}^N K(X_i^N(t), X_j^N(t)), \quad i \in \{1, \dots, N\}, \quad (1.2)$$

in which the interaction between two particles i and j no longer depends only on their positions X_i^N and X_j^N in the state space (via the interaction kernel K) but also on their individual labels i and j , via the weight w_{ij}^N . Agents' labels play the role of the vertices V_N of the underlying graph $G_N = (V_N, E_N, W_N)$, by setting $E_N := \{1, \dots, N\}$, while directed edges $V_N := \{(i, j) \in V_N^2 : w_{ij}^N \neq 0\}$ are entirely determined by their adjacency matrix $W_N := (w_{ij}^N)_{1 \leq i, j \leq N}$.

A growing body of works has subsequently introduced a generalization of this second model in order to remove Hypothesis 2, relying on the theory of *hypergraphs* [7, 8, 14]. These approaches are based on the idea that the basic interaction unit of many dynamical systems includes more than two nodes [55, 67, 59, 69, 73, 74, 75, 76]. Applications are numerous and involve the modeling of opinion dynamics [17, 25, 60, 68], contagion propagation [7, 57], synchronization of oscillators [75, 74, 73, 55, 76], animal communication [38] and evolutionary game dynamics

[7, 57]. This leads to the following multi-agent system with high-order interactions, which will be the focus of the present article:

$$\begin{cases} \frac{dX_i^N(t)}{dt} = \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{ij_1 \dots j_\ell}^{\ell, N} K_\ell(X_i^N(t), X_{j_1}^N(t), \dots, X_{j_\ell}^N(t)), \\ X_i^N(0) = X_{i,0}^N, \quad i \in \{1, \dots, N\}. \end{cases} \quad (1.3)$$

The above system (1.3) describes the evolution of N agents with states $X_i^N = X_i^N(t) \in \mathbb{R}^d$. Interactions among agents are given as the superposition of all possible $(\ell+1)$ -body interactions. The functions $K_\ell = K_\ell(x, x_1, \dots, x_\ell)$ represent the $(\ell+1)$ -body interaction kernels, and each weight $w_{ij_1 \dots j_\ell}^{\ell, N}$ describes the underlying $(\ell+1)$ -body couplings or connections among agents. More precisely, the tensor $(w_{ij_1 \dots j_\ell}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ can be seen as the adjacency tensor of a $(\ell+1)$ -uniform hypergraph describing the underlying communication architecture among agents. In particular, we note that the above sum moves over all possible types of multi-body interactions ranging from $\ell = 1$ (*i.e.*, pairwise interactions as in Equation (1.2)) to $\ell = N - 1$ (*i.e.*, full group interactions). For modeling purposes, we neglect summands with $\ell \geq N$ which would necessarily lead to undesired self-interaction. For similar reasons, we shall always assume that $w_{ij_1 \dots j_\ell}^{\ell, N} = 0$ whenever $i \in \{j_1, \dots, j_\ell\}$.

The goal of the present article is to derive the mean-field limit of the multi-agent system (1.3) as N tends to infinity. Our result studies a form of ‘‘propagation of chaos’’ for this highly non-exchangeable multi-agent system and it provides quantitative convergence rates as N tends to infinity towards the associated Vlasov equation:

$$\begin{cases} \partial_t \mu_t^\xi + \operatorname{div}_x (F_w[\mu_t](\cdot, \xi) \mu_t^\xi) = 0, \quad t \geq 0, x \in \mathbb{R}^d, \xi \in [0, 1], \\ \mu_{t=0}^\xi = \mu_0^\xi. \end{cases} \quad (1.4)$$

Here, $(\mu_t^\xi)_{\xi \in [0, 1]} \subset \mathcal{P}(\mathbb{R}^d)$ is a measurable family of probability measures parametrized by a continuous label $\xi \in [0, 1]$ for any $t \geq 0$. Note that all existing individuals contribute to jointly generate a mean-field force onto individuals with label ξ of the form:

$$\begin{aligned} F_w[\mu_t](x, \xi) &:= \sum_{\ell=1}^{\infty} \int_{[0, 1]^\ell} w_\ell(\xi, \xi_1, \dots, \xi_\ell) \\ &\quad \times \left(\int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\mu_{i_1}^{\xi_1}(x_1) \cdots d\mu_{i_\ell}^{\xi_\ell}(x_\ell) \right) d\xi_1, \dots, d\xi_\ell. \end{aligned} \quad (1.5)$$

Additionally, the above object $w = (w_\ell)_{\ell \in \mathbb{N}}$ consists in a sequence of uniformly bounded functions $w_\ell \in L^\infty([0, 1]^{\ell+1})$. It describes, at the macroscopic scale, the connection among agents via groups of fixed size $\ell+1$ for every $\ell \in \mathbb{N}$.

Inspired by the classical limit theory of dense graphs (*i.e.*, graphons [54]), and their subsequent extensions to sparse graphs (*i.e.*, L^p graphons [19, 20], graphops [5] and s-graphons [48]), some recent works have also proposed limit theories that are valid for the broader class of dense hypergraphs [30, 66, 77]. On the one hand, the case of hypergraphs with uniform cardinality $k \in \mathbb{N}$ was treated in [30, 77] using the notion of k -uniform hypergraphons, but a new doubling of variables was introduced compared to the treatment of graphons, which complicates the study of mean-field limits operating over these structures. On the other hand, the limit of non-uniform hypergraphs with bounded ranks was recently considered in [78], where a multilinear extension of graphops [5] was obtained. Lastly, in the more general setting of non-uniform hypergraphs with unbounded rank, a limit theory of dense simplicial complexes was obtained in [66], which exploits the concept of complexons. Simplicial complexes are a subclass of hypergraphs whose hyperedges are closed under restrictions, that is, all subgroups of nodes forming a hyperedge also form a smaller hyperedge. Their limiting objects were named complexon and consist of a sequence $w = (w_\ell)_{\ell \in \mathbb{N}}$ of equivalence classes (under a suitable quotient map) of uniformly bounded functions $w_\ell \in L^\infty([0, 1]^{\ell+1})$ with $\ell \in \mathbb{N}$.

In this paper, we shall not restrict ourselves to coupling weights determined by complexons, which are defined as the above-mentioned quotient objects, but we will rather exploit the primitive non-quotient objects endowed with a suitable topology, which, for distinction with the terminology in [66], we will call *UR-hypergraphons*, that is, hypergraphons of unbounded rank (*cf.* Section 2.2). This framework allows considering hypergraphs and hypergraphons with no *a priori* bound on the maximum hyperedge size.

Importantly, several studies have reported the frequent occurrence of large-size hyperedges in hypergraphs constructed from real-world datasets. For instance, in [26], the authors studied the hypergraph structure of fMRI data acquired among 77 subjects performing cognitively demanding tasks. It was found that the majority of subjects have a hypergraph composed of one large hyperedge and many small hyperedges, with a rough power law for the distribution of the smaller sizes. More recently, a detailed analysis of hypergraphs coming from six different fields (such as high-school contacts, email recipients, drug substances, or publication co-authors) have revealed that hyperedge sizes follow a heavy-tail (power law) distribution [44, 45]. These studies of real-world data highlight the importance of proposing and analyzing models for hypergraphs of unbounded rank, *i.e.* hypergraphs whose largest hyperedge can potentially contain all the hypergraph nodes.

While in most steps of this article (but not all), the coupling weights encoded by the hypergraphs can be assumed to be quite general (supposing of course some natural summability conditions), some more particular assumptions will be considered in the statement of our main result. For readability, we thus present a list of sufficient assumptions for our main result to hold, and expose Theorem 1.1 in this setting. As will be recalled throughout the article, not all assumptions are in fact necessary, and the specific requirements for each result will be specified in the corresponding statements. The following assumptions concern the interaction kernels, the coupling weights and the initial data of the multi-agent system (1.3).

Assumption 1 (On the interaction kernels).

(i) **(Bounded-Lipschitz interaction kernels):**

Assume that $K_\ell \in \text{BL}(\mathbb{R}^{d(\ell+1)})$, *i.e.*, there exist $B_\ell, L_\ell > 0$ such that

$$|K_\ell(x, x_1, \dots, x_\ell)| \leq B_\ell. \quad (1.6)$$

$$|K_\ell(x, x_1, \dots, x_\ell) - K_\ell(y, y_1, \dots, y_\ell)| \leq L_\ell \left(|x - y| + \sum_{k=1}^{\ell} |x_k - y_k| \right), \quad (1.7)$$

and denote the bounded-Lipschitz norm of K_ℓ by

$$BL_\ell := \max\{B_\ell, L_\ell\}, \quad \ell \in \mathbb{N}. \quad (1.8)$$

Additionally, assume that there exists $\eta \geq 0$ such that

$$\sum_{\ell=1}^{\infty} \frac{\sqrt{\ell!}}{\eta^\ell} B_\ell < \infty, \quad \sum_{\ell=1}^{\infty} \ell L_\ell < \infty. \quad (1.9)$$

(ii) **(Higher regularity of interaction kernels):**

Assume that $K_\ell \in H^{\frac{d(\ell+1)}{2} + \varepsilon}$ for some $\varepsilon > 0$, and additionally

$$\sum_{\ell=1}^{\infty} \frac{4^\ell \pi^{\frac{d\ell}{4}}}{\sqrt{\Gamma(\frac{d\ell}{2})}} \|K_\ell\|_{H^{\frac{d(\ell+1)}{2} + \varepsilon}} < \infty, \quad (1.10)$$

where Γ represents the Gamma function $\Gamma(a) := \int_0^\infty s^{a-1} e^{-s} ds$ for $a > 0$.

Assumption 2 (On the coupling weights).

(iii) **(Absence of loops):**

Assume that $w_{i_{j_1 \dots j_\ell}}^{\ell, N}$ verify

$$w_{i_{j_1 \dots j_\ell}}^{\ell, N} = 0, \quad \text{if } i \in \{j_1, \dots, j_\ell\}. \quad (1.11)$$

(iv) **(Scaling of the coupling weights):**

Assume that there exists $W > 0$ such that

$$\sup_{N \in \mathbb{N}} \max_{1 \leq \ell \leq N-1} \max_{1 \leq i, j_1, \dots, j_\ell \leq N} N^\ell w_{ij_1 \dots j_\ell}^{\ell, N} \leq W, \quad (1.12)$$

Assumption 3 (On the symmetry of kernels and weights).

(v) **(Symmetry of coupling weights):**

Assume that $w_{ij_1 \dots j_\ell}^{\ell, N}$ verify

$$w_{ij_1 \dots j_\ell}^{\ell, N} = w_{\sigma(i)\sigma(j_1) \dots \sigma(j_\ell)}^{\ell, N}, \quad \forall \sigma \in \mathcal{S}_{\ell+1}, \quad (1.13)$$

which in particular implies

$$w_{ij_1 \dots j_\ell}^{\ell, N} = w_{i\sigma(j_1) \dots \sigma(j_\ell)}^{\ell, N}, \quad \forall \sigma \in \mathcal{S}_\ell. \quad (1.14)$$

(vi) **(Symmetry of interaction kernels):**

Assume that K_ℓ verify

$$K_\ell(x, x_1, \dots, x_\ell) = K_\ell(x, x_{\sigma(1)}, \dots, x_{\sigma(\ell)}), \quad \forall \sigma \in \mathcal{S}_\ell. \quad (1.15)$$

Assumption 4 (On the initial data).

(vii) **(Scaling of the initial data):**

Assume that there exists $p \in [1, 2]$ such that $X_{i,0}^N$ satisfies

$$\sup_{N \in \mathbb{N}} \max_{1 \leq i \leq N} \mathbb{E} |X_{i,0}^N|^p < \infty. \quad (1.16)$$

Given this list of assumptions, the main contribution of the article can be stated in an informal way as follows:

Theorem 1.1. *Assume that the kernels K_ℓ and the weights $w_{ij_1 \dots j_\ell}^{\ell, N}$ satisfy Assumptions 1, 2 and 3. For any $(X_{1,0}^N, \dots, X_{N,0}^N)$ with *i.i.d.* $X_{i,0}^N$ (but N dependent law) satisfying Assumption 4, consider the unique solutions (X_1^N, \dots, X_N^N) to (1.3). Then, there is a subsequence $N_k \rightarrow \infty$ such that the mean-field limit of the multi-agent system (1.3) is characterized in a suitable sense by a solution to the Vlasov-type equation (1.4)-(1.5) for some $(\mu_t^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$ and some $w = (w_\ell)_{\ell \in \mathbb{N}}$ such that $\sup_{\ell \in \mathbb{N}} \|w_\ell\|_{L^\infty} \leq W$.*

A more rigorous statement is given in Theorem 6.1. We refer to Remarks 5.3 and 5.3 for some relaxation of the regularity assumption (1.10) under more particular shapes on the interaction kernels K_ℓ , and to Remark 6.2 for some comments about the need for the *i.i.d.* assumption on the initial data. Finally, we also refer to Remark 6.5 for an alternative reformulation of the Vlasov equation (1.4)-(1.5) in terms of a hierarchy of observables indexed by directed hypertrees (which extends the hierarchy indexed by trees and derived in [39] in the binary case).

The paper is organized as follows. Section 2 is devoted to recalling the definitions of various types of hypergraphs, as well as of their various limit objects, which will allow us to situate our work in the vast literature of graph and hypergraph limit theories. We also provide a few examples inspired from the existing literature to illustrate the notions of hypergraphs and hypergraphons, establishing rigorously the derivation from the former to the latter in some specific cases. As remarked in [39], the classical notion of propagation of chaos cannot be expected to hold for non-exchangeable particle systems, which are by definition not identically distributed. Instead, we show the weaker notion of propagation of *independence*, generalizing the approach of [39] to hypergraphs. Section 3 is devoted to showing that propagation of independence holds for an auxiliary particle system which approximates the original particle system (1.3) as N tends to infinity. In Section 4, we prove the well-posedness of the limit Vlasov equation (1.4)-(1.5) by a fixed-point argument relying on the Lipschitz regularity of the force F in (1.5) with respect to the measure μ_t , for an L^p -Bounded Lipschitz fibered distance. Section 5 is dedicated to proving the stability of the Vlasov equation with respect to both the initial datum μ_0 and the UR-hypergraphon (w_ℓ) . In Section 6, we state our main result in full detail and prove convergence of the solution to the particle system (1.3) toward the solution to the

Vlasov equation (1.4)-(1.5). Lastly, in Section 7 we illustrate the convergence result of Theorem 1.1 via numerical simulations based on some examples of hypergraphs and multi-agent dynamics presented in Section 2.3. We end with some conclusions and future perspectives in Section 8.

Notation:

Along the paper, we will alleviate the notation by introducing the following shorthand. For any $N \in \mathbb{N}$ we denote $\llbracket 1, N \rrbracket = \{1, \dots, N\}$. Given $\ell \in \mathbb{N}$ and $1 \leq k \leq \ell$, we define the multi-indices

$$\mathbf{j}_\ell = (j_1, \dots, j_\ell) \in \llbracket 1, N \rrbracket^\ell, \quad \hat{\mathbf{j}}_{\ell,k} = (j_1, \dots, j_{k-1}, j_{k+1}, \dots, j_\ell) \in \llbracket 1, N \rrbracket^{\ell-1}.$$

Additionally, given $\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell$ and $\sigma \in \mathcal{S}_\ell$ (a permutation of the finite set $\llbracket 1, \ell \rrbracket$), we define the new multi-index $\sigma(\mathbf{j}_\ell)$ by rearranging their components according to σ , that is,

$$\sigma(\mathbf{j}_\ell) = (j_{\sigma(1)}, \dots, j_{\sigma(\ell)}) \in \llbracket 1, N \rrbracket^\ell.$$

Similarly, we shall denote the multi-variables

$$\mathbf{x}_\ell = (x_1, \dots, x_\ell) \in \mathbb{R}^{d\ell}, \quad \boldsymbol{\xi}_\ell = (\xi_1, \dots, \xi_\ell) \in [0, 1]^\ell, \quad \hat{\boldsymbol{\xi}}_{\ell,k} = (\xi_1, \dots, \xi_{k-1}, \xi_{k+1}, \dots, \xi_\ell) \in [0, 1]^{\ell-1}.$$

Let us introduce here the different spaces which will appear through the article. We denote:

- for $p \in [1, \infty]$, $L_+^p(X)$ the space of functions belonging to $L^p(X)$ and taking nonnegative values,
- $\mathcal{M}(X)$ the space of finite Borel measures on X and $\mathcal{M}_+(X)$ the space of finite Borel nonnegative measures on X ,
- $\mathcal{P}(X)$ the space of probability measure on X ,
- $\text{BL}(X)$ the space of bounded-Lipschitz functions ϕ such that the bounded-Lipschitz norm is finite *i.e.*

$$\|\phi\|_{\text{BL}} := \max\{\|\phi\|_{L^\infty}, [\phi]_{\text{Lip}}\} < +\infty$$

where $[\phi]_{\text{Lip}}$ stands for the Lipschitz constant of ϕ . $\text{BL}(X)$ is equipped with the following distance: for $\mu_1, \mu_2 \in \text{BL}(X)$,

$$d_{\text{BL}}(\mu_1, \mu_2) = \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_X \phi(x) (d\mu_1(x) - d\mu_2(x)),$$

- $\text{BL}_1(X)$ the subspace of bounded-Lipschitz functions $\phi \in \text{BL}(X)$ such that $\|\phi\|_{\text{BL}} \leq 1$.

2. PRELIMINARIES

2.1. Graphs and hypergraph limit theories. In this section, we briefly introduce some useful notation and the necessary terminology to study hypergraphs, and their graph limit theories as proposed in previous literature [30, 66, 77, 78].

2.1.1. Basics on hypergraphs. We first recall notation for (finite) hypergraphs.

Definition 2.1 (Hypergraph). *A weighted hypergraph (or simply hypergraph) consists in a triple $H = (V, E, W)$, where V is a finite set, $E \subset 2^V$ is a finite collection of subsets of V , and $W : 2^V \rightarrow \mathbb{R}_+$ satisfies $W(e) > 0$ if, and only if, $e \in E$. Elements in V are called nodes (or vertices), and without loss of generality we shall assume that $V = \llbracket 1, N \rrbracket$ for some $N \in \mathbb{N}$, elements in E are called hyperedges, and W is called the weight function. We also define:*

- (i) (Cardinality) *The cardinality of a hyperedge edge $e \in E$ is defined by $\#e$.*
- (ii) (Rank) *The rank of the hypergraph H is the largest cardinality of any hyperedge.*
- (iii) (Adjacency tensors) *For each $\ell \in \llbracket 1, N-1 \rrbracket$, we define the $(\ell+1)$ -order adjacency tensor*

$$(w_{i_1 \dots i_\ell}^{\ell, N})_{1 \leq i_1, \dots, i_\ell \leq N} : \quad w_{i_1 \dots i_\ell}^{\ell, N} := W(i, j_1, \dots, j_\ell), \quad i, j_1, \dots, j_\ell \in \llbracket 1, N \rrbracket.$$

We remark that hypergraphs are generalizations of graphs. More specifically, hypergraphs whose hyperedges have cardinality 2 amount to usual graphs.

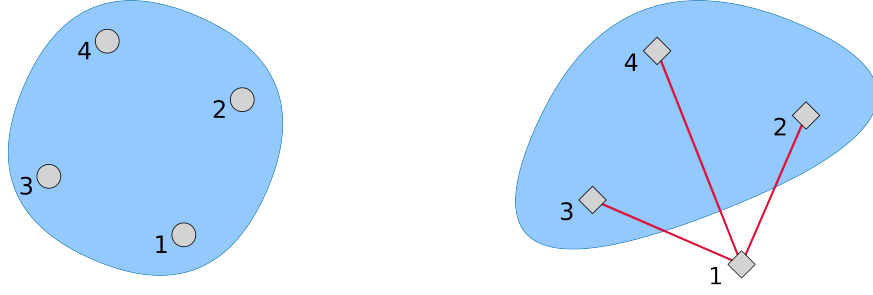


FIGURE 1. Undirected hyperedge (left) and directed hyperedge (right).

Remark 2.2 (Directed hypergraphs). *By definition the above hypergraphs have been considered undirected. Specifically, hyperedges are regarded as non-ordered subsets of nodes and therefore weights are assumed to be symmetric under permutations, that is, $w_{ij_1 \dots j_\ell}^{\ell, N} = w_{\sigma(i)\sigma(j_1) \dots \sigma(j_\ell)}^{\ell, N}$ for all hyperedge $e = \{i, j_1, \dots, j_\ell\}$, all $\sigma \in \mathcal{S}_{\ell+1}$ and all $\ell \in \llbracket 1, N-1 \rrbracket$.*

However, as for graphs, directed weighted hypergraphs (or simply directed hypergraphs) could have also been considered by replacing hyperedges by directed hyperedges. A directed hyperedge (or hyperarch) is a pair $e = \{i; h\}$ consisting of a tail $i \in V$ and a head $h \in 2^V$ with $i \notin h$. Since the head is a non-ordered subset of nodes, the corresponding weights are assumed to be symmetric under permutations on the nodes of the head, but not the tail, that is, $w_{ij_1 \dots j_\ell}^{\ell, N} = w_{i\sigma(j_1) \dots \sigma(j_\ell)}^{\ell, N}$ for all directed hyperedge $e = \{i; j_1, \dots, j_\ell\}$, all $\sigma \in \mathcal{S}_\ell$ and all $\ell \in \llbracket 1, N-1 \rrbracket$.

See Figure 1 for a graphical representation of the undirected edge $e = \{1, 2, 3, 4\}$ and the directed edge $e = \{1; 2, 3, 4\}$ with same nodes 1, 2, 3, 4.

So defined, directed hypergraphs are well suited to formulate dynamical multi-agent system under the form (3.1), since each term $w_{ij_1 \dots j_\ell}^{\ell, N} K_\ell(X_i^N, X_{j_1}^N, \dots, X_{j_\ell}^N)$ represents the strength of the force generated jointly by the agents $X_{j_1}^N, \dots, X_{j_\ell}^N$ (acting as the head) and felt by the agent X_i^N (acting as the tail). However, more general hypergraphs could have been considered by admitting that tails of the directed hyperedges may contain more than one element, or further considering directed hyperedges simply as ordered lists of nodes. The more general the directed hyperedges is, the less symmetries we assume on the weights.

Definition 2.3 (Types of hypergraphs).

- (i) (Unweighted hypergraph) *An unweighted hypergraph is a weighted hypergraph with weights valued in $\{0, 1\}$. In such case, given $i, j_1, \dots, j_\ell \in \llbracket 1, N \rrbracket$, the set $e = \{i, j_1, \dots, j_\ell\}$ is a hyperedge if, and only if, $w_{ij_1 \dots j_\ell}^{\ell, N} = 1$.*
- (ii) (Uniform hypergraphs) *A k -uniform hypergraph is a weighted hypergraph whose hyperedges all have the same cardinality k . In other words, $w_{ij_1 \dots j_\ell}^{\ell, N} = 0$ when $\ell \neq k-1$.*
- (iii) (Simplicial complexes) *A simplicial complex is an unweighted hypergraph whose hyperedges are closed under taking subsets, that is, if $e' \subset e \subset 2^V$ and $e \in E$ then $e' \in E$.*

2.1.2. *Graph limit theories.* Before introducing the recent theories on hypergraph limits, we briefly review some of the main results in the theory of graph limits, and its use to derive mean-field limits for multi-agent systems with binary interactions like (1.2).

Definition 2.4 (Graphons). *Given $W > 0$, we define the set of graphons (graph functions) as*

$$\mathcal{G}_W := \{w \in L_+^\infty([0, 1]^2) : \|w\|_{L^\infty} \leq W, \text{ and } w \text{ is symmetric}\}.$$

For any two graphons $w, \bar{w} \in \mathcal{G}_W$, we define the labeled cut distance

$$d_\square(w, \bar{w}) := \sup_{S, T \subset [0, 1]} \left| \iint_{S \times T} (w(\xi, \zeta) - \bar{w}(\xi, \zeta)) d\xi d\zeta \right|,$$

as well as the (unlabeled) cut distance

$$\delta_{\square}(w, \bar{w}) = \inf_{\Phi} d_{\square}(w, \bar{w}^{\Phi}),$$

where Φ ranges over all bijective measure-preserving maps $\Phi : [0, 1] \rightarrow [0, 1]$, and $\bar{w}^{\Phi} \in \mathcal{G}_W$ represents the rearranged graphon $\bar{w}^{\Phi}(\xi, \zeta) = \bar{w}(\Phi(\xi), \Phi(\zeta))$.

The space of graphons $(\mathcal{G}_W, \delta_{\square})$ was first introduced in [54] as one of the first limit theories for a sequence of graphs. Specifically, the authors proved that δ_{\square} is a pseudodistance on the space of graphons \mathcal{G}_W (as it is invariant under rearrangements of graphons) which, induced on the quotient space \mathcal{G}_W / \sim that identifies graphons identical modulo rearrangements, becomes a compact metric space. Note that for any sequence of graphs $(G_N)_{N \in \mathbb{N}}$ with an increasing number of nodes N we can associate a graphon $w^{G_N} \in \mathcal{G}_1$ via the piecewise function

$$w^{G_N}(\xi, \zeta) = N \sum_{i,j=1}^N w_{ij}^N \mathbf{1}_{I_i^N \times I_j^N}(\xi, \zeta), \quad \xi, \zeta \in [0, 1],$$

where $(w_{ij}^N)_{1 \leq i, j \leq N}$ is the adjacency matrix of G_N and $I_i^N := [\frac{i-1}{N}, \frac{i}{N})$, $I_j^N := [\frac{j-1}{N}, \frac{j}{N})$ for all $1 \leq i, j \leq N$. In view of the scaling $w_{ij}^N \leq \frac{W}{N}$ natural for mean-field systems (see also (1.12) for hypergraphs), we have intentionally rescaled the coupling weights w_{ij}^N by N in the above piecewise definition. By compactness of the space $(\mathcal{G}_W, \delta_{\square})$ such sequence must have a converging subsequence to a limiting graphon $w \in \mathcal{G}_W$ with respect to the cut distance δ_{\square} . Additionally, in [54] it was also proved that any graphon $w \in \mathcal{G}_W$ can be approximated in the cut distance by a sequence of finite graphs as above.

We remark that the labeled cut distance is strictly weaker than the L^1 norm by triangle inequality, that is,

$$d_{\square}(w, \bar{w}) \leq \|w - \bar{w}\|_{L^1},$$

for all $w, \bar{w} \in \mathcal{G}_W$. Approximating any bounded Borel-measurable function by characteristic functions of Borel sets, the labeled cut distance d_{\square} can be reformulated as follows

$$d_{\square}(w, \bar{w}) = \sup_{\phi, \psi: [0,1] \rightarrow [0,1]} \left| \iint_{[0,1]^2} \phi(\xi) \psi(\zeta) (w(\xi, \zeta) - \bar{w}(\xi, \zeta)) d\xi d\zeta \right|,$$

and therefore we obtain the following useful equivalence with operator norms

$$d_{\square}(w, \bar{w}) \leq \|T^w - T^{\bar{w}}\|_{\infty \rightarrow 1} \leq 4d_{\square}(w, \bar{w}),$$

where above $T^w : L^{\infty}([0, 1]) \rightarrow L^1([0, 1])$ stands for the adjacency operator of w , which consists in the bounded linear operator defined by

$$T^w[\psi](\xi) := \int_0^1 w(\xi, \zeta) \psi(\zeta) d\zeta, \quad \xi \in [0, 1],$$

for each $\psi \in L^{\infty}([0, 1])$, and similarly we define $T^{\bar{w}}$.

An additional important feature of the cut distance is that the associated notion of convergence is characterized by an object intimately related to graph theory, namely, the homomorphism density. Specifically, $\delta_{\square}(w_n, w) \rightarrow 0$ if, and only if, $\tau(F, w_n) \rightarrow \tau(F, w)$ for all finite graphs F . Here, $\tau(F, w)$ is the graphon extension of the homomorphism density of a finite graph F into a second finite graph G , which counts the proportion of graph homomorphisms from $V(F)$ to $V(G)$ among all maps from $V(F)$ to $V(G)$, that is,

$$\tau(F, w) := \int_{[0,1]^{\#V(F)}} \prod_{(i,j) \in E(F)} w(\xi_i, \xi_j) \prod_{i \in V(F)} d\xi_i.$$

We refer to the monograph [53] for further details.

For general sequences of unweighted graphs $(G_N)_{N \in \mathbb{N}}$ the obtained limit $w \in \mathcal{G}_1$ is the trivial graphon $w \equiv 0$ unless the sequence of graphs is dense, meaning that the number of edges is quadratic on the number of vertices, *i.e.*,

$$\#E(G_N) \approx \#V(G_N)^2.$$

The above suggests that the space $(\mathcal{G}_W, \delta_\square)$ is too small to capture the limit of sequences of sparse graphs, and a more general theory is needed. This is the origin of other recent graph limit theories operating over sequences of graphs with intermediate density/sparseness:

$$\#V(G_N) \lesssim \#E(G_N) \lesssim \#V(G_N)^2,$$

which we briefly review below.

◇ **(Graphings)** In [12] a first graph limit theory appeared to study limits of very sparse graph sequences $(G_N)_{N \in \mathbb{N}}$ having a number of edges linear on the number of vertices, that is, $\#E(G_N) \approx \#V(G_N)$. More specifically, this limit theory is known as the local or Benjamini-Schram convergence, it applies to sequences of graphs with bounded degree and limiting objects are probability measures over the set of rooted, connected (possibly infinite) graphs with bounded degree. The theory was later extended in [18, 35], where a more general notion of convergence of a sequence of graphs with bounded degree was proposed and was called the local-global convergence in contrast to the local or Benjamini-Schram convergence. In the local-global convergence, limiting objects are called graphings and are defined as (possibly infinite) unweighted graphs $G = (V(G), E(G))$ with bounded degree, where nodes $V(G) = \mathcal{X}$ form a Polish space, edges $E(G) \subset \mathcal{X} \times \mathcal{X}$ are Borel-measurable, and there is a probability measure $\nu \in \mathcal{P}(\mathcal{X})$ such that

$$\int_A e(\xi, B) d\nu(\xi) = \int_B e(\xi, A) d\nu(\xi),$$

for all $A, B \subset \mathcal{X}$ Borel-measurable, where $e(\xi, A)$ is the number of edges joining the node $\xi \in \mathcal{X}$ to any node in $A \subset \mathcal{X}$.

◇ **(L^p graphons)** The theory of L^p graphons was proposed in [20, 19] as an extension of the theory of L^∞ graphons and proved to be useful for sequences of graphs $(G_N)_{N \in \mathbb{N}}$ with intermediate density/sparsity. More particularly, L^p graphons are defined as

$$\mathcal{G}_W^p := \{w \in L_+^p([0, 1]^2) : \|w\|_{L^p} \leq W, \text{ and } w \text{ is symmetric}\},$$

for any $p \in (1, \infty]$. The novelty of this setting is that, as proved in [20, Theorem 2.8], given a sparse (unweighted) graph sequence $(G_N)_{N \in \mathbb{N}}$ which are uniformly L^p upper regular (see [20] for a precise definition), one can define the renormalized L^p graphons w^{G_N} given by

$$w^{G_N}(\xi, \zeta) = \frac{1}{D(G_N)} \sum_{i,j=1}^N w_{ij}^N \mathbb{1}_{I_i^N \times I_j^N}(\xi, \zeta), \quad \xi, \zeta \in [0, 1],$$

where $(w_{ij}^N)_{1 \leq i, j \leq N}$ is the adjacency matrix of G_N , and $D(G_N) = \frac{\#E(G_N)}{\#V(G_N)^2}$ is the edge density of the graph (which tends to zero for sparse graphs), and we have that w^{G_N} has a converging subsequence to a limiting L^p graphon $w \in \mathcal{G}_W^p$ with respect to the cut distance δ_\square . The normalization by the edge density ensures that limiting L^p graphons of sparse graph sequences are not necessarily trivial, contrarily to what happened in the theory of dense graph limits.

◇ **(Graphops)** Graphops were introduced in [5] as an attempt to derive a unifying theory of graphons, graphings and L^p graphons. In this setting, objects are regarded as linear operators

$$\tilde{\mathcal{G}}_{p,q,W} := \left\{ T : L^p(\Omega) \longrightarrow L^q(\Omega) : \begin{array}{l} T \text{ is linear, bounded, positivity-preserving,} \\ \text{self-adjoint, and verifies } \|T\|_{p \rightarrow q} \leq W, \\ \text{and } (\Omega, \Sigma, \mathbb{P}) \text{ is some probability space} \end{array} \right\},$$

where $p \in [1, \infty)$, $q \in [1, \infty)$ and $\|\cdot\|_{p \rightarrow q}$ denotes the operator norm from L^p to L^q . The underlying probability space is not fixed in purpose, as it varies in the limit. In [5, Theorem 2.10] the authors proved that a suitable quotient of $\tilde{\mathcal{G}}_{p,q,W}$ endowed with the action convergence distance d_M (see [5] for further details) is compact.

We remark that graphons $w \in \mathcal{G}_W$ can be realized as graphops through their adjacency operator $T^w : L^p([0, 1]) \longrightarrow L^q([0, 1])$ as above, and also graphings can be realized as graphops. Similarly, finite graphs G_N can also be realized as graphops by defining the (discrete) adjacency operator $T^{G_N} : L^p(\llbracket 1, N \rrbracket) \longrightarrow L^q(\llbracket 1, N \rrbracket)$. Therefore, suitably renormalizing T^{G_N} so that their

operator norms $\|\cdot\|_{p \rightarrow q}$ stay uniformly bounded, we have the existence of a limiting graphop. In [5] it was also observed that action convergence is equivalent to convergence in cut distance when restricted to graphons, and local-global convergence when restricted to graphings.

Graphops are intimately related to s-graphons, which were proposed in [48] as an alternative theory of sparse graphs where the limiting objects consist in symmetric probability measures $\nu \in \mathcal{P}([0, 1] \times [0, 1])$. More precisely, [5, Theorem 6.3] shows that any graphop $T \in \tilde{\mathcal{G}}_{p,q,W}$ on the probability space $(\Omega, \Sigma, \mathbb{P})$ admits a unique representation as a symmetric finite measure $\nu_T \in \mathcal{M}_+(\Omega \times \Omega)$ with marginals absolutely continuous with respect to \mathbb{P} , namely,

$$\nu_T(A \times B) = \int_{\Omega} T[\mathbb{1}_A](\xi) \mathbb{1}_B(\xi) d\mathbb{P}(\xi), \quad A, B \in \Sigma.$$

2.1.3. Mean-field limits on graphs. The above graph limit theories have proven to be extremely useful in the last years to derive rigorous mean-field limits of multi-agent systems with binary interactions such as (1.2) toward a suitable Vlasov-type equation. The fundamental obstruction to apply standard methods of mean-field limits [21, 28, 33] and propagation of chaos [71] is the fact that (1.2) is a non-exchangeable system. Specifically, exchanging two agents modifies the overall dynamics of the groups as interactions are mediated by weights, which introduce a distinguishable feature on agents. Depending on the degree of density or sparseness of the sequence of the maps of connections, graphons, graphops and other extensions have been used.

The first attempts to derive the mean-field limit of interacting particle systems in the case of dense graph sequences were developed in [23] (for Lipschitz graphons) and [43] (for non-Lipschitz graphons), extending Neuzert's method for exchangeable multi-agent systems [61]. The method of proof relies on the stability of the limiting Vlasov equation with respect to the initial datum and the involved graphon in the following metric:

$$d_1(\mu, \bar{\mu}) = \int_0^1 d_{\text{BL}}(\mu^\xi, \bar{\mu}^\xi) d\xi,$$

(see Definition 2.17 for an L^p extension). This estimate involves a continuous dependence on the graphon in the L^1 -norm, which is a stronger requirement than the cut distance. Since the compactness of $(\mathcal{G}_W, \delta_{\square})$ could not be used, the authors prescribed an *ad hoc* discretization of the graphon w by finite graphs G_N to guarantee the stronger convergence $\|w^{G_N} - w\|_{L^1} \rightarrow 0$. In particular, general dense graph sequences are not supported by this argument.

This strategy was later extended in [31] to a sparse setting, using graphops. In this case, the stability estimate of the limiting Vlasov equation was quantified in the analogous metric, but required stronger convergence of graphops than action convergence, namely, the narrow convergence $\nu_{T_n}^\xi \rightarrow \nu_T^\xi$ of the disintegrations (see Theorem 2.15) for \mathbb{P} -a.e. $\xi \in \Omega$. Therefore, an *ad hoc* discretization is needed to guarantee the stronger convergence, which the authors achieved when the probability space $(\Omega, \Sigma, \mathbb{P})$ is a compact Abelian group with its Haar measure.

A further application of this strategy in a sparse setting was recently proposed in [46] to deal with a new class of graph limits called digraph measures, consisting in bounded maps $\xi \in \Omega \mapsto \nu^\xi \in \mathcal{M}_+(\Omega)$. Whilst intimately related to graphops [5] and s-graphons [48], digraph measures do not arise from graph limit theory, and in particular they do not form a compact space. The authors obtained a similar estimate with respect to the distance

$$d_\infty(\mu, \bar{\mu}) = \sup_{\xi \in \Omega} d_{\text{BL}}(\mu^\xi, \bar{\mu}^\xi),$$

which required the convergence of digraph measures $\sup_{\xi \in \Omega} d_{\text{BL}}(\nu_n^\xi, \nu^\xi) \rightarrow 0$. In [46] the authors obtained suitable discretizations of digraph measures by finite graphs under the additional assumption that Ω is a compact subset of an Euclidean space, and the digraph measure $\xi \in \Omega \mapsto \nu^\xi$ is continuous in the bounded-Lipschitz distance.

There are only a handful of results where the graph limit theories in Section 2.1.2 have been used in full, that is, the natural compact topologies on the graph limit spaces have proved enough in the mean-field limit and no preparation on the discretized weights is needed. To the best of our knowledge some examples are [13, 39, 40].

In [13], the authors derived the mean-field limit for a multi-agent system in a 1-dimensional periodic domain over a sequence of dense random graphs. In particular, a stability estimate of the limiting Vlasov equation was derived with respect to the pseudodistance

$$\bar{d}_1(\mu, \bar{\mu}) = d_{\text{BL}} \left(\int_0^1 \mu^\xi d\xi, \int_0^1 \bar{\mu}^\xi d\xi \right),$$

which operates over marginals, and therefore the structure variable $\xi \in [0, 1]$ is disregarded and averaged. Interestingly, the stability estimate depended continuously on the cut distance between the involved graphons, and therefore the compactness of $(\mathcal{G}_W, \delta_\square)$ was enough, and weights did not need to be well prepared.

In [39, 40], the authors defined the space of extended graphons as a new space of graph limits consisting of measure-valued bounded maps

$$\mathcal{EG}_W := \left\{ \nu \in L_\xi^\infty([0, 1], \mathcal{M}_\zeta([0, 1])) \cap L_\zeta^\infty([0, 1], \mathcal{M}_\xi([0, 1])) : \|\nu\|_{L_\xi^\infty \mathcal{M}_\zeta}, \|\nu\|_{L_\zeta^\infty \mathcal{M}_\xi} \leq W \right\}.$$

It was proved that $\mathcal{G}_W \subset \mathcal{EG}_W \subset \tilde{\mathcal{G}}_{p,p,W}$ for any $p \in [1, \infty]$, and the space was endowed with a notion of convergence characterized by the convergence $\tau(T, \nu_n) \rightarrow \tau(T, \nu)$ of the homomorphism density for any tree T . The resulting space was proved compact, which gave access to a notion of convergence of sparse graphs under the simple mild assumptions that

$$\sup_{N \in \mathbb{N}} \max_{1 \leq i \leq N} \sum_{j=1}^N w_{ij}^N + \max_{1 \leq j \leq N} \sum_{i=1}^N w_{ij}^N \leq W.$$

Additionally, the authors obtained some stability estimates for the limiting Vlasov equation depending continuously on the extended graphons with respect to such a topology, thus deriving the mean-field limit for a large class of sparse graphs, and not necessarily well prepared.

We refer to [10, 11, 24, 27, 40, 50, 56, 64] for further recent results on continuum and mean-field limits for non-exchangeable multi-agent systems on graphs, and also to [4] for a comprehensive review of the recent literature.

2.1.4. Hypergraph limit theories. The literature on hypergraph limits is scarce and more recent. To the best of our knowledge, the results available thus far are mostly concentrated in the papers [30, 66, 77, 78], where three types of objects have been studied, namely, uniform hypergraphons, hypergraphons and complexons. We briefly review the main outcomes in this subsection.

◇ **(Uniform hypergraphons)** In [30] (later reformulated also in [77]), a limiting theory of dense uniform hypergraphs (see Figure 2a) was proposed, leading to the so called uniform hypergraphons. Specifically, the sequence of dense hypergraphs were assumed to have a fixed and common cardinality $k \in \mathbb{N}$ on all hyperedges, and the limiting objects were defined by

$$k\text{-}\mathcal{H}_W := \left\{ w \in L_+^\infty([0, 1]^{2^k-2}) : \|w\|_{L^\infty} \leq W, \text{ and } w \text{ is symmetric} \right\}.$$

Elements in $L_+^\infty([0, 1]^{2^k-2})$ have $2^k - 2$ variables indexed by the non-empty and proper subsets of $\llbracket 1, k \rrbracket$, and symmetric means that whenever a permutation $\sigma \in \mathcal{S}_k$ is set, and whenever we rearrange the variables $\xi_1, \xi_2, \dots, \xi_k, \xi_{\{1,2\}}, \xi_{\{1,3\}}$, etc by $\xi_{\sigma(1)}, \xi_{\sigma(2)}, \dots, \xi_{\sigma(k)}, \xi_{\{\sigma(1), \sigma(2)\}}, \xi_{\{\sigma(1), \sigma(3)\}}$, etc, the value of w does not change. The need for additional coordinates is related to the need for suitable regularity partitions for hypergraphs [34].

Similarly to graphons (*i.e.*, $k = 2$), this space was endowed with the topology induced by the convergence of the homomorphism densities $\tau(F, w_n) \rightarrow \tau(F, w)$, that is,

$$\tau(F, w) = \int \prod_{e \in E(F)} w(\xi_e) d\xi,$$

for any k -uniform hypergraph F . Above we denote $\xi = (\xi_s : \emptyset \neq s \subset V(F), \#s \leq k-1)$ the collection of variables indexed by all nonempty subsets of $V(F)$ with size smaller than k , and $\xi_e = (\xi_s : \emptyset \neq s \subsetneq e)$ the subcollection of variables indexed by all nonempty and proper subsets of e . As for the binary case, the resulting space is compact, see [77, Corollary 1.8].

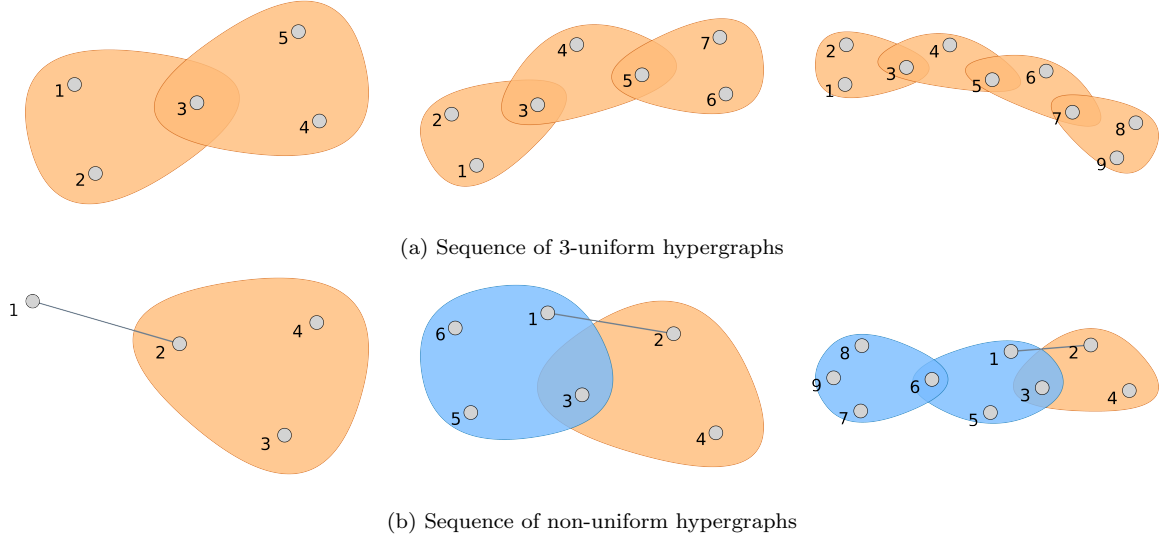


FIGURE 2. Sequences of uniform (top) and non-uniform (bottom) hypergraphs

◇ **(Hypergraphops)** As an extension of graphops [5], hypergraphops were proposed in [78] as a limit theory for non-uniform (see Figure 2b) sparse hypergraphs with bounded rank $r \in \mathbb{N}$:

$$r\text{-}\tilde{\mathcal{H}}_{p,q,W} := \left\{ T : L^p(\Omega)^{r-1} \longrightarrow L^q(\Omega) : \begin{array}{l} T \text{ is multilinear, bounded, positivity-preserving,} \\ \text{symmetric, and verifies } \|T\|_{(L^p)^{r-1} \rightarrow L^q} \leq W, \\ \text{and } (\Omega, \Sigma, \mathbb{P}) \text{ is some probability space} \end{array} \right\},$$

where $p \in [1, \infty)$, $q \in [1, \infty)$ and $\|\cdot\|_{(L^p)^{r-1} \rightarrow L^q}$ denotes the multilinear operator norm from $(L^p)^{r-1}$ to L^q . Again, a suitable quotient of $k\text{-}\tilde{\mathcal{H}}_{p,q,W}$ endowed with the action convergence distance d_M (see [78] for further details) is compact. As for graphs, finite hypergraphs $(H_N)_{N \in \mathbb{N}}$ with bounded rank r can be realized as hypergraphops. In [78] several options were proposed through the definition of the s -action for any $s = 1, \dots, r-1$. Each way of embedding hypergraphs into graphops seems to provide a notion of convergence for hypergraphs that is better adapted to dense, sparse, uniform and non-uniform settings.

◇ **(Complexons)** In [66], a limit theory of dense simplicial complexes of unbounded rank was derived. A simplicial complex is a hypergraph so that any subset of a hyperedge is again a hyperedge. By definition, simplicial complexes are non-uniform and therefore their limit cannot be tackled using k -uniform hypergraphops. Since the rank is also unbounded, hypergraphops are not an option either. The alternative approach in [66] was to define the set of complexons as the set of equivalence classes

$$\mathcal{C}_1 := \left\{ w^\circ = (w_\ell^\circ)_{\ell \in \mathbb{N}} : \begin{array}{l} w = (w_\ell)_{\ell \in \mathbb{N}}, w_\ell \in L_+^\infty([0, 1]^{\ell+1}), \|w_\ell\|_{L^\infty} \leq 1, \\ \text{and } w_\ell \text{ is symmetric for all } \ell \in \mathbb{N} \end{array} \right\}$$

where w° is called the faceted version of w and is defined by

$$w_\ell^\circ(\xi, \xi_1, \dots, \xi_\ell) := \prod_{k=1}^{\ell} \prod_{\{i_0, \dots, i_k\} \subset [0, k]} w_\ell(\xi_{i_0}, \dots, \xi_{i_k}), \quad \xi, \xi_1, \dots, \xi_\ell \in [0, 1],$$

where we denote $\xi_0 := \xi$. The property $w_\ell^\circ \geq w_{\ell+1}^\circ$ for all $\ell \in \mathbb{N}$ is reminiscent of the closure of simplicial complexes under restrictions. \mathcal{C}_1 was endowed with the topology induced by the convergence of the homomorphism densities $\tau(F^\circ, w_n^\circ) \rightarrow \tau(F^\circ, w^\circ)$, that is,

$$\tau(F^\circ, w^\circ) = \int_{[0,1]^{\#V(F)}} \prod_{k \in \mathbb{N}} \prod_{\{i_0, \dots, i_k\} \in E(F^0)} w^\circ(\xi_{i_0}, \dots, \xi_{i_k}) \prod_{i \in V(F^0)} d\xi_0 \dots d\xi_{\#V(F)-1},$$

where for any simplicial complex F , the notation F° stands for the hypergraph with same set of nodes $V(F^\circ) = V(F)$ but only the maximal hyperedges $E(F^\circ) \subset E(F)$ which are not strict subsets of larger hyperedges. So defined, \mathcal{C}_1 is compact as proven in [66, Theorem 13].

The proof is based on a characterization of the topology induced by the homomorphism density, as the topology induced by a suitable higher-order extension of the cut distance of graphons. This cut distance is actually well defined for general non-faceted w and, on this more general space, it induces a compact topology [66, Lemma 33]. Since we do not need to characterize the topology by homomorphism densities, instead of working on the quotient space consisting of complexons, we shall rather work on the primitive space, which we define below.

2.2. Hypergraphons of unbounded rank. In this section we introduce the limit theory for hypergraphs of unbounded rank which we shall use in this paper, and which is inspired by the treatment in [66]. In particular, we introduce their underlying topology, a basic compactness result reminiscent of the classical result for dense graph limits in [54], and some example of convergence of sequences of hypergraphs toward its associated limiting object.

Definition 2.5 (Hypergraphons of unbounded rank). *Given $W > 0$, we define the set of hypergraphons of unbounded rank (or UR-hypergraphons) as follows*

$$\mathcal{H}_W := \left\{ w = (w_\ell)_{\ell \in \mathbb{N}} : \begin{array}{l} w_\ell \in L_+^\infty([0, 1]^{\ell+1}), \|w_\ell\|_{L^\infty} \leq W, \\ \text{and } w_\ell \text{ is symmetric for all } \ell \in \mathbb{N} \end{array} \right\}.$$

For any two UR-hypergraphons $w, \bar{w} \in \mathcal{H}_W$, we define the ℓ -th order labeled cut distance

$$d_{\square, \ell}(w_\ell, \bar{w}_\ell) := \sup_{S, S_1, \dots, S_\ell \subset [0, 1]} \left| \int_{S \times S_1 \times \dots \times S_\ell} (w_\ell - \bar{w}_\ell) d\xi d\xi_1 \dots d\xi_\ell \right|,$$

for every $\ell \in \mathbb{N}$. To comprehend all possible orders, we define the labeled cut distance

$$d_{\square}(w, \bar{w}; (\alpha_\ell)_{\ell \in \mathbb{N}}) := \sum_{\ell=1}^{\infty} \alpha_\ell d_{\square, \ell}(w_\ell, \bar{w}_\ell),$$

where $(\alpha_\ell)_{\ell \in \mathbb{N}}$ is a strictly positive summable sequence, as well as the (unlabeled) cut distance

$$\delta_{\square}(w, \bar{w}; (\alpha_\ell)_{\ell \in \mathbb{N}}) = \inf_{\Phi} d_{\square}(w, \bar{w}^\Phi; (\alpha_\ell)_{\ell \in \mathbb{N}}),$$

where Φ ranges over all bijective measure-preserving maps $\Phi : [0, 1] \rightarrow [0, 1]$, and $\bar{w}^\Phi \in \mathcal{H}_W$ represents the rearranged UR-hypergraphon $\bar{w}_\ell^\Phi(\xi, \xi_1, \dots, \xi_\ell) = \bar{w}_\ell(\Phi(\xi), \Phi(\xi_1), \dots, \Phi(\xi_\ell))$.

As for the binary case, the labeled cut distance on UR-hypergraphs admits an alternative representation as an operator norm. Specifically, the ℓ -th order labeled cut distance can be reformulated in terms of test functions, namely,

$$d_{\square, \ell}(w_\ell, \bar{w}_\ell) = \sup_{\phi, \psi_1, \dots, \psi_\ell : [0, 1] \rightarrow [0, 1]} \left| \iint_{[0, 1]^{\ell+1}} \phi(\xi) \psi_1(\xi_1) \cdots \psi_\ell(\xi_\ell) (w_\ell - \bar{w}_\ell) d\xi d\xi_1 \dots d\xi_\ell \right|,$$

and therefore we obtain the following useful equivalence with multilinear operator norms

$$d_{\square, \ell}(w_\ell, \bar{w}_\ell) \leq \|T^{w_\ell} - T^{\bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1} \leq 2^\ell d_{\square, \ell}(w_\ell, \bar{w}_\ell),$$

where above the multilinear operator $T^{w_\ell} : L^\infty([0, 1])^\ell \rightarrow L^1([0, 1])$ stands for the ℓ -th order adjacency operator of w_ℓ , which consists in the bounded multilinear operator defined by

$$T^{w_\ell}[\psi_1, \dots, \psi_\ell](\xi) := \int_{[0, 1]^\ell} w_\ell(\xi, \xi_1, \dots, \xi_\ell) \psi_1(\xi_1) \cdots \psi_\ell(\xi_\ell) d\xi_1 \dots d\xi_\ell, \quad \xi \in [0, 1], \quad (2.1)$$

for each $\psi_1, \dots, \psi_\ell \in L^\infty([0, 1])$. The operator $T^{\bar{w}_\ell}$ is defined similarly. All the above implies the following equivalence between the labeled cut-distance and the multilinear operator norms.

Proposition 2.6. *For any strictly positive and summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$ we have*

$$d_{\square}(w, \bar{w}; (\alpha_\ell)_{\ell \in \mathbb{N}}) \leq \sum_{\ell=1}^{\infty} \alpha_\ell \|T^{w_\ell} - T^{\bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1} \leq d_{\square}(w, \bar{w}; (2^\ell \alpha_\ell)_{\ell \in \mathbb{N}}),$$

for all $w, \bar{w} \in \mathcal{H}_W$, where the ℓ -adjacency operators T^{w_ℓ} and $T^{\bar{w}_\ell}$ are defined by (2.1).

As mentioned above, the interest in the space of UR-hypergraphons endowed with the cut distance is the following compactness result from [66, Lemma 33].

Proposition 2.7 (Compactness of UR-hypergraphons). *The unlabeled cut distance δ_\square induces topologically equivalent pseudometrics on the space of UR-hypergraphons \mathcal{H}_W which do not depend on the summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$. When induced on the quotient \mathcal{H}_W / \sim that identifies UR-hypergraphons identical modulo rearrangement, the space becomes a compact metric space.*

Similarly to the case of sequences of dense graphs, one can characterize the convergence of a sequence of hypergraphs via the unlabeled cut-distance. For any sequence of hypergraphs $(H_N)_{N \in \mathbb{N}}$ with an increasing number of nodes N , we can associate a UR-hypergraphon $w = (w_\ell^{H_N})_{\ell \in \mathbb{N}}$ via the family of piecewise-constant functions

$$w_\ell^{H_N}(\xi, \xi_1, \dots, \xi_\ell) = N^\ell \sum_{i, j_1, \dots, j_\ell=1}^N w_{i, j_1, \dots, j_\ell}^{\ell, N} \mathbb{1}_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N}(\xi, \xi_1, \dots, \xi_\ell), \quad (2.2)$$

for all $(\xi, \xi_1, \dots, \xi_\ell) \in [0, 1]^{\ell+1}$, where $(w_{i, j_1, \dots, j_\ell}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ is the adjacency tensor of order $\ell+1$ of H_N , and $I_i^N := [\frac{i-1}{N}, \frac{i}{N})$ for all $1 \leq i \leq N$. In view of the scaling condition (1.12) which we assume on our hypergraphs to derive the mean-field limit, we have intentionally rescaled the coupling weights $w_{i, j_1, \dots, j_\ell}^{\ell, N}$ by N^ℓ in the above piecewise definition. Note that it ensures that $w^{H_N} \in \mathcal{H}_W$ for the same W given in Assumption (1.12). This suggests the following notion of convergence of hypergraphs (possibly with unbounded rank) toward a UR-hypergraphon.

Definition 2.8 (Convergence of hypergraphs). *A sequence of hypergraphs $(H_N)_{N \in \mathbb{N}}$ is said to converge to a UR-hypergraphon $w \in \mathcal{H}_W$ when $\lim_{N \rightarrow \infty} \delta_\square(w, w^{H_N}; (\alpha_\ell)_{\ell \in \mathbb{N}}) = 0$ for some (and then all) positive and summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$.*

The above Proposition 2.7 ensures that any such sequence of hypergraphs $(H_N)_{N \in \mathbb{N}}$ satisfying the scaling condition (1.12), must converge toward some limiting UR-hypergraphon up to a subsequence. In fact, as it happens in the binary case, the family of finite hypergraphs is dense in UR-hypergraphons. In Propositions 2.9 and 2.10 below we provide two different constructive methods to show that UR-hypergraphons can be suitably approximated in the cut distance by sequences of finite hypergraphs: L^1 -approximations, and pointwise approximations.

Proposition 2.9 (L^1 -approximations of UR-hypergraphons). *Let $w = (w_\ell)_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ be a UR-hypergraphon. Let $(H_N)_{N \in \mathbb{N}}$ be the sequence of hypergraphs whose adjacency tensors $(w^{\ell, N})_{\ell \in \mathbb{N}}$ are defined by L^1 -approximations of the UR-hypergraphon w , as follows:*

$$w_{i, j_1, \dots, j_\ell}^{\ell, N} = \begin{cases} N \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} w_\ell(\xi, \xi_1, \dots, \xi_\ell) d\xi d\xi_1 \dots d\xi_\ell, & \text{if } \ell \in \llbracket 1, N-1 \rrbracket, \\ 0, & \text{otherwise} \end{cases}$$

for all $(i, j_1, \dots, j_\ell) \in \llbracket 1, N \rrbracket^{\ell+1}$. Then, the sequence of hypergraphs $(H_N)_{N \in \mathbb{N}}$ converges as N tends to infinity to the UR-hypergraphon w in the labeled cut-distance d_\square (and hence also in the unlabeled cut-distance δ_\square).

Proof. Let $w = (w_\ell)_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ and let $(w_\ell^{H_N})_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ denote the piecewise-constant hypergraphon obtained from the hypergraph H_N , as defined in (2.2). We aim to prove that for any positive summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$, $\lim_{N \rightarrow \infty} d_\square(w, w^{H_N}; (\alpha_\ell)_{\ell \in \mathbb{N}}) = 0$. We actually show the convergence of the sequence of hypergraphons in the stronger L^1 -norm, recalling that for any $\ell \in \mathbb{N}$, the labeled and unlabeled cut-distances of order ℓ satisfy:

$$\delta_{\square, \ell}(w_\ell, w_\ell^{H_N}) \leq d_{\square, \ell}(w_\ell, w_\ell^{H_N}) \leq \|w_\ell - w_\ell^{H_N}\|_{L^1(I^{\ell+1})}.$$

Let $\varepsilon > 0$. For each $\ell \in \mathbb{N}$, by density of Lipschitz functions in $L^1([0, 1]^{\ell+1})$, consider $w_\ell^\varepsilon \in \text{Lip}([0, 1]^{\ell+1})$ so that

$$\|w_\ell - w_\ell^\varepsilon\|_{L^1} \leq \varepsilon.$$

Let us denote by $S_\ell^N : L^1([0, 1]^{\ell+1}) \rightarrow L^1([0, 1]^{\ell+1})$ the operator that to any w_ℓ associates its stepfunction, *i.e.*, the hypergraphon $w_\ell^{H_N}$ built piecewise by averaging on hypercubes of the partition on N pieces. One can easily show that S_ℓ^N is linear and bounded with $\|S_\ell^N\|_{L^1 \rightarrow L^1} \leq 1$. Then, by the triangle inequality we obtain

$$\begin{aligned} \|w_\ell - S_\ell^N[w_\ell]\|_{L^1} &\leq \|w_\ell - w_\ell^\varepsilon\|_{L^1} + \|w_\ell^\varepsilon - S_\ell^N[w_\ell^\varepsilon]\|_{L^1} + \|S_\ell^N[w_\ell^\varepsilon] - S_\ell^N[w_\ell]\|_{L^1} \\ &\leq 2\varepsilon + \|w_\ell^\varepsilon - S_\ell^N[w_\ell^\varepsilon]\|_{L^1}. \end{aligned}$$

We can then compute

$$\begin{aligned} \|w_\ell^\varepsilon - S_\ell^N[w_\ell^\varepsilon]\|_{L^1(I^{\ell+1})} &= \int_{I^{\ell+1}} |w_\ell^\varepsilon(\xi, \xi_1, \dots, \xi_\ell) - S_\ell^N[w_\ell^\varepsilon](\xi, \xi_1, \dots, \xi_\ell)| d\xi d\xi_1 \dots d\xi_\ell \\ &= \sum_{i, j_1, \dots, j_\ell=1}^N \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} |w_\ell^\varepsilon(\xi, \xi_1, \dots, \xi_\ell) - S_\ell^N[w_\ell^\varepsilon](\xi, \xi_1, \dots, \xi_\ell)| d\xi d\xi_1 \dots d\xi_\ell \\ &= \sum_{i, j_1, \dots, j_\ell=1}^N \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} \left| w_\ell^\varepsilon(\xi, \xi_\ell) - \left(N^\ell \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} w_\ell^\varepsilon(\xi', \xi'_\ell) d\xi' d\xi'_\ell \right) \right| d\xi d\xi_\ell \\ &\leq \sum_{i, j_1, \dots, j_\ell=1}^N \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} N^{\ell+1} \int_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N} |w_\ell^\varepsilon(\xi, \xi_\ell) - w_\ell^\varepsilon(\xi', \xi'_\ell)| d\xi' d\xi'_\ell d\xi d\xi_\ell \\ &\leq N^{\ell+1} \sum_{i, j_1, \dots, j_\ell=1}^N \int_{(I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N)^2} \frac{\sqrt{\ell+1}}{N} [w_\ell^\varepsilon]_{\text{Lip}} d\xi' d\xi'_\ell d\xi d\xi_\ell \\ &\leq \frac{\sqrt{\ell+1}}{N} [w_\ell^\varepsilon]_{\text{Lip}}, \end{aligned}$$

for each $\ell \in \llbracket 1, N-1 \rrbracket$, where $[w_\ell^\varepsilon]_{\text{Lip}}$ denotes the Lipschitz semi-norm of w_ℓ^ε .

Keeping $\varepsilon > 0$ fixed and passing to the limit as $N \rightarrow \infty$, it holds

$$\limsup_{N \rightarrow \infty} \|w_\ell - w_\ell^{H_N}\|_{L^1} \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain

$$\lim_{N \rightarrow \infty} d_{\square, \ell}(w_\ell, w_\ell^{H_N}) = \lim_{N \rightarrow \infty} \|w_\ell - w_\ell^{H_N}\|_{L^1} = 0.$$

Now, note that for any summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$, it holds

$$\alpha_\ell d_{\square, \ell}(w_\ell, w_\ell^{H_N}) \leq 2W\alpha_\ell,$$

for all $\ell \in \mathbb{N}$ and all $N \in \mathbb{N}$, which gives a uniform-in- N domination by the summable sequence $(2W\alpha_\ell)_{\ell \in \mathbb{N}}$. Additionally, we have proved that we have the pointwise convergence

$$\lim_{N \rightarrow 0} \alpha_\ell d_{\square, \ell}(w_\ell, w_\ell^{H_N}) = 0,$$

for all $\ell \in \mathbb{N}$. Then, by the dominated convergence theorem,

$$\lim_{N \rightarrow \infty} d_{\square}(w, w^{H_N}; (\alpha_\ell)_{\ell \in \mathbb{N}}) = \lim_{N \rightarrow \infty} \sum_{\ell=1}^{\infty} \alpha_\ell d_{\square, \ell}(w_\ell, w_\ell^{H_N}) = \sum_{\ell=1}^{\infty} \alpha_\ell \lim_{N \rightarrow \infty} d_{\square, \ell}(w_\ell, w_\ell^{H_N}) = 0,$$

which concludes the proof. \square

Proposition 2.10 (Pointwise approximation of UR-hypegraphons). *Let $w = (w_\ell)_{\ell \in \mathbb{N}} \in \mathcal{HW}$ be a UR-hypergraphon, and suppose that w_ℓ is continuous for all $\ell \in \mathbb{N}$. For all $N \in \mathbb{N}$, let $(\bar{\xi}_i^N)_{i \in \llbracket 1, N \rrbracket}$ be a sequence of points satisfying $\bar{\xi}_i^N \in I_i^N$ for all $i \in \llbracket 1, N \rrbracket$, where $I_i^N = [\frac{i-1}{N}, \frac{i}{N})$. Let $(\bar{H}_N)_{N \in \mathbb{N}}$ be the sequence of hypergraphs whose adjacency tensors $(\bar{w}^{\ell, N})_{\ell \in \mathbb{N}}$ are defined by evaluating the UR-hypergraphon w on the grid points $(\bar{\xi}_i^N)_{i \in \llbracket 1, N \rrbracket}^{\ell+1}$, as follows:*

$$\bar{w}_{i_1 \dots i_\ell}^{\ell, N} = \begin{cases} \frac{1}{N^\ell} w_\ell(\bar{\xi}_{i_1}^N, \bar{\xi}_{i_2}^N, \dots, \bar{\xi}_{i_\ell}^N), & \text{if } \ell \in \llbracket 1, N-1 \rrbracket, \\ 0, & \text{otherwise} \end{cases}$$

for all $(i, j_1, \dots, j_\ell) \in \llbracket 1, N \rrbracket^{\ell+1}$. Then, the sequence of hypergraphs $(\bar{H}_N)_{N \in \mathbb{N}}$ converges as N tends to infinity to the UR-hypergraphon w in the labeled cut-distance d_\square (and hence also in the unlabeled cut-distance δ_\square).

Proof. As in the proof of Proposition 2.9, we will prove the stronger convergence result in L^1 -norm. Since w_ℓ is continuous and $I^{\ell+1}$ is compact, then w_ℓ is uniformly continuous for all $\ell \in \mathbb{N}$. Denoting by $\varpi_\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the modulus of continuity of w_ℓ , it holds

$$\begin{aligned} \|w_\ell - w_\ell^{\bar{H}_N}\|_{L^1(I^{\ell+1})} &= \sum_{i, j_1, \dots, j_\ell=1}^N \int_{I_i^N \times \dots \times I_{j_\ell}^N} |w_\ell(\xi, \boldsymbol{\xi}_\ell) - N^\ell \bar{w}_{i, \mathbf{j}_\ell}^{\ell, N}| d\xi d\boldsymbol{\xi}_\ell \\ &= \sum_{i, j_1, \dots, j_\ell=1}^N \int_{I_i^N \times \dots \times I_{j_\ell}^N} |w_\ell(\xi, \boldsymbol{\xi}_\ell) - w_\ell(\bar{\xi}_i^N, \bar{\xi}_{j_1}^N, \dots, \bar{\xi}_{j_\ell}^N)| d\xi d\boldsymbol{\xi}_\ell \\ &\leq \varpi_\ell \left(\frac{\sqrt{\ell+1}}{N} \right), \end{aligned}$$

for all $\ell \in \mathbb{N}$. By definition, ϖ is continuous and $\varpi(0) = 0$. Taking limits as $N \rightarrow \infty$ implies

$$\lim_{N \rightarrow \infty} d_{\square, \ell}(w_\ell, w_\ell^{H_N}) = \lim_{N \rightarrow \infty} \|w_\ell - w_\ell^{H_N}\|_{L^1} = 0,$$

for all $\ell \in \mathbb{N}$. The same argument as in the proof of Proposition 2.9, based on the dominated convergence theorem, concludes the proof. \square

In this paper we study the mean-field limit of the multi-agent system (1.3) over dense non-uniform hypergraphs of unbounded rank (note that $r = N$). We rely on the above hypergraph limit theory of UR-hypergraphons, which we show is compatible with the mean-field limit without any *ad hoc* preparations of the weights. Of course our approach also works when weights are prepared according to Propositions 2.9 or 2.10.

To the best of our knowledge, none of the above hypergraph limit theories has been exploited previously in the community of mean-field limits. There is only one result available in the literature where a mean-field limit of a multi-agent system with higher-order interactions has been studied [47]. Inspired by the previous work [46] by the same authors on digraph measures, a new class of limits of hypergraphs with bounded rank $r \in \mathbb{N}$ was proposed in terms of directed hypergraph measures, which consist of bounded maps $\xi \in \Omega \mapsto \nu_\ell^\xi \in \mathcal{M}_+(\Omega^\ell)$ for $\ell \in 1, \dots, r-1$. Directed hypergraph measures are related to hypergraphons [78], but do not arise from hypergraph limit theory. As in the binary case, the authors' strategy was to design a method to well-prepare the hypergraphs as to approximate any given directed hypergraph measure in a strong enough topology, compatible with their stability estimate.

2.3. Some examples of multi-agent systems over hypergraphs. To illustrate the definitions of hypergraph and hypergraphon (*cf.* Definitions 2.1, 2.3 and 2.5), we provide a few examples inspired from the existing literature.

2.3.1. Examples of hypergraphs.

\diamond **(Uniform hypergraphs):** The simplest example of hypergraphs is the uniform hypergraph, in which all hyperedges have the same cardinality. When the cardinality is $r = 2$, all hyperedges are in fact binary edges, so that the hypergraph is a graph.

\diamond **(Simplicial complex from real-world networks):** In [69], the authors propose a model for oscillator dynamics on two different simplicial complexes, which are designed from real datasets. The Macaque brain dataset consists of 242 interconnected regions of the brain, while the UK power grid network consists of 120 nodes and 165 transmission lines. From the simple graphs $(w_{k_1 k_2}^{2, N})_{1 \leq k_1, k_2 \leq N}$ given by these datasets, hyperedges are constructed using the following simple criterion. For any $\ell \in \{1, \dots, N-1\}$,

$$w_{ij_1 \dots j_\ell}^{\ell, N} = \begin{cases} 1 & \text{if } w_{k_1 k_2}^{2, N} = 1 \text{ for all } k_1, k_2 \in \{i, j_1, \dots, j_\ell\}, \\ 0 & \text{otherwise.} \end{cases}$$

that is a hyperedge exists in the hypergraph when all possible pairs of nodes in the hyperedge are edges of the graph. The resulting hypergraph is of rank $r = N$ if and only if the network is fully coupled. Otherwise, the hypergraph is of bounded rank $r < N$.

◊ **(Hypergraph for homogeneous groups)**: For $\theta \in (0, 1]$, the θ -fixed radius near neighbor graph is defined by drawing an edge between each pair of agents whose labels (i, j) satisfy $|i - j| \leq \theta N$. It is an example of an undirected and unweighted graph. We can extend this notion to hypergraphs by drawing a hyperedge between any group of agents whose diameter is bounded by θN , defining

$$w_{i_1 \dots j_\ell}^{\ell, N} = \begin{cases} \frac{1}{N^\ell}, & \text{if } \max_{k_1, k_2 \in \{i, j_1, \dots, j_\ell\}} |k_1 - k_2| \leq \theta N, \\ 0, & \text{otherwise,} \end{cases} \quad (2.3)$$

for each $\ell \in \{1, \dots, N-1\}$. As mentioned above, the scaling by N^ℓ in our weighted hypergraphs is natural in view of the scaling condition (1.12) which ensures that the associated sequence of UR-hypergraphons $(w^{H_N})_{N \in \mathbb{N}}$ defined by (2.2) all lie in \mathcal{H}_W (indeed with $W = 1$). A pixel representation of this hypergraph for the hyperedges of dimension $\ell = 1$ and $\ell = 2$ is given in Figure 3. If the labels represent agents' identities, this hypergraph models the idea that agents interact if and only if they have similar identities, *i.e.* if they form a homogeneous group. Notice that if $\theta = 1$, the hypergraph is fully connected, which represents all-to-all coupling.

Note that although we use the terminology “fixed radius near-neighbor” and “group diameter”, the distance considered to build the hypergraph are in the space of labels. Confusion should not be made with so-called “bounded confidence” models for *exchangeable* particle systems [37], in which the interaction kernel has bounded support, which implies that agents interact if and only if their distance *in the state space* is small enough.

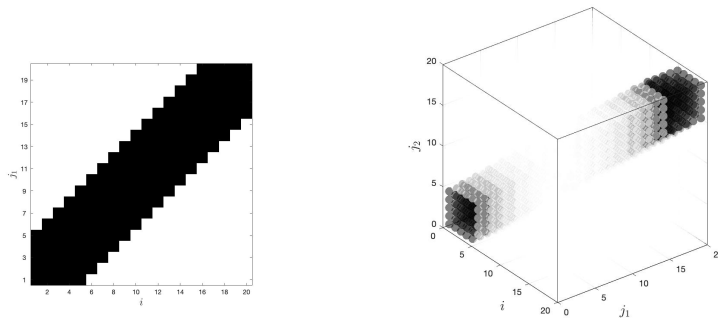


FIGURE 3. Pixel representation of the matrix and 3-tensor respectively corresponding to $w^{1, N}$ and $w^{2, N}$ given by (2.3), with $\theta = 0.3$ and $N = 20$. As this is an unweighted hypergraph, the presence of a hyperedge is represented in black. Transparency was used to indicate depth in the 3-dimensional representation.

Remark 2.11. *The sequence of hypergraphs for homogeneous groups defined in Equation (2.3) can be shown to be obtained by evaluating the UR-hypergraphon $w = (w_\ell)_{\ell \in \mathbb{N}}$, defined by*

$$w_\ell(\xi_0, \dots, \xi_\ell) = \begin{cases} 1 & \text{if } \max_{i, j \in \{0, \dots, \ell\}} |\xi_i - \xi_j| \leq \theta, \\ 0 & \text{otherwise} \end{cases} \quad (2.4)$$

on the grid $(\bar{\xi}_i)_{i \in [1, N]}^{\ell+1}$, where $\bar{\xi}_i := \frac{i-1}{N}$ for all $i \in [1, N]$. Note that Proposition 2.10 cannot be used to prove the convergence of the hypergraphs (2.3) toward the UR-hypergraphon (2.4) since, so defined, w_ℓ is discontinuous for all $\ell \in \mathbb{N}$. Whilst the continuity assumption of Proposition 2.10 could seem restrictive, it is not always necessary and can be relaxed in many cases. Indeed, convergence can still be proven in this case, as shown in Proposition A.1 (see Appendix A).

◇ **(Weighted hypergraph for balanced groups)**: In this example, the hypergraph is also of unbounded rank ($r = N$), and each hyperedge's weight is assumed to be a decreasing function of the distance between the average value of its nodes' labels and the average value of *all* nodes' labels (*i.e.* $\frac{N+1}{2}$). More precisely, given a decreasing function $f : [0, \frac{1}{2}] \rightarrow \mathbb{R}_+$, we define

$$w_{j_0 j_1 \dots j_\ell}^{\ell, N} = \frac{1}{N^\ell} f \left(\frac{1}{N} \left| \frac{1}{\ell+1} \sum_{k=0}^{\ell} j_k - \frac{N+1}{2} \right| \right), \quad (2.5)$$

for all $\ell \in \mathbb{N}$. A pixel representation of such a hypergraph's hyperedges of dimension $\ell = 1$ and $\ell = 2$ is given in Figure 4. This hypergraph models the idea that the more balanced a group's identities are (*i.e.* the average label of the group is close to $\frac{N+1}{2}$), the larger the group interaction.

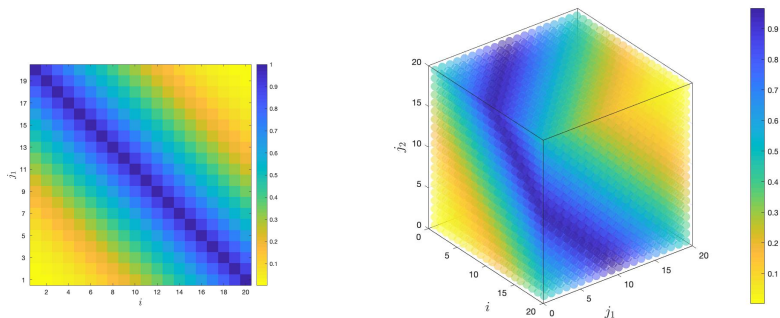


FIGURE 4. Pixel representation of the matrix and 3-tensor respectively corresponding to $w^{1,N}$ and $w^{2,N}$ given by (2.5), with $f : x \mapsto 4(x - \frac{1}{2})^2$ and $N = 20$. For this weighted hypergraph, the value of each hyperedge's weight is represented on a color scale.

Remark 2.12. *The sequence of hypergraphs for balanced groups defined in Equation (2.5) can be shown to be obtained by evaluating the UR-hypergraphon $w = (w_\ell)_{\ell \in \mathbb{N}}$, defined by*

$$w_\ell(\xi_0, \dots, \xi_\ell) = f \left(\left| \frac{1}{\ell+1} \sum_{i=0}^{\ell} \xi_i - \frac{1}{2} \right| \right), \quad (2.6)$$

on the grid $(\bar{\xi}_i)_{i \in [1, N]}^{\ell+1}$, where $\bar{\xi}_i := \frac{i-\frac{1}{2}}{N}$ for all $i \in [1, N]$. Then, whenever the function f is continuous, convergence of convergence of the hypergraphs (2.5) toward the UR-hypergraphon (2.6) is a direct application of Proposition 2.10.

2.3.2. Models of multi-agent dynamics on hypergraphs. Various examples of higher-order multi-agent dynamics have recently been proposed in the literature. Applications range from opinion dynamics to contagion propagation, synchronization of oscillators, animal communication, and evolutionary game dynamics. We present some examples, and refer to [7, 8, 25, 38, 55, 67, 59, 60, 69, 73, 74, 75, 76] for further reading.

◇ **(Higher-order synchronization models)**: Several works studied generalizations of the Kuramoto model of coupled oscillators with higher-order interactions [7, 55, 74, 75, 76]. Denoting by θ_i^N the phase and Ω_i^N the natural frequency of the i -th oscillator, its evolution is prescribed by an equation of the form:

$$\frac{d\theta_i^N(t)}{dt} = \Omega_i^N + \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{i j_1 \dots j_\ell}^{\ell, N} \sin \left(\sum_{k=1}^{\ell} \theta_{j_k}^N(t) - \ell \theta_i^N(t) \right).$$

We remark that the interaction kernels $K_\ell(x, x_1, \dots, x_\ell) = \sin(x_1 + \dots + x_\ell - \ell x)$ satisfy the symmetry assumption (1.15). Notice that the second-order term (for $\ell = 1$) is exactly that

encountered in the classical Kuramoto model [49]:

$$\sum_{j=1}^N w_{ij}^{1,N} \sin(\theta_j^N(t) - \theta_i^N(t)).$$

In [69], the authors studied another extension of the Kuramoto model for coupled oscillators with higher order interactions arising from phase-reductions of limit-cycle oscillators. More precisely, the evolution of the N oscillators is given by

$$\begin{aligned} \frac{d\theta_i^N(t)}{dt} = & \Omega_i^N \sum_{j_1=1}^N w_{ij_1}^{1,N} \sin(\theta_{j_1}^N(t) - \theta_i^N(t)) + \sum_{j_1=1}^N \sum_{j_2=1}^N w_{ij_1 j_2}^{2,N} \sin(2\theta_{j_1}^N(t) - \theta_{j_2}^N(t) - \theta_i^N(t)) \\ & + \sum_{j_1=1}^N \sum_{j_2=1}^N \sum_{j_3=1}^N w_{ij_1 j_2 j_3}^{3,N} \sin(\theta_{j_1}^N(t) + \theta_{j_2}^N(t) - \theta_{j_3}^N(t) - \theta_i^N(t)). \end{aligned}$$

Note that the interaction kernels $K_2(x, x_1, x_2) = \sin(2x_1 - x_2 - x)$ and $K_3(x, x_1, x_2, x_3) = \sin(x_1 + x_2 - x_3 - x)$ do not satisfy the symmetry assumption (1.15). The underlying hypergraph $(w^{\ell,N})_{\ell \in \{1, \dots, 3\}}$ is of rank 3, and is taken to be a simplicial complex constructed from real network datasets, as explained above.

◊ **(Higher-order opinion dynamics models):** In [60], a model for higher-order opinion dynamics is proposed on a uniform hypergraph of rank 2. Denoting $x_i^N(t)$ the opinion of the i -th oscillator, its evolution is given by the following dynamics:

$$\frac{dx_i^N(t)}{dt} = \sum_{j_1=1}^N \sum_{j_2=1}^N w_{ij_1 j_2}^{2,N} e^{\lambda |x_{j_1}^N(t) - x_{j_2}^N(t)|} \left(\frac{x_{j_1}^N(t) + x_{j_2}^N(t)}{2} - x_i^N(t) \right).$$

These dynamics model the idea that each agent i is attracted towards the average opinion of the joint pair (j_1, j_2) , with an intensity that depends on the distance between the opinions of j_1 and j_2 . More precisely, if $\lambda < 0$, then the pair of agents (j_1, j_2) will have a larger influence if their opinions x_{j_1} and x_{j_2} are similar (*i.e.* $|x_{j_1} - x_{j_2}|$ is small). On the opposite, if $\lambda > 0$, then the pair of agents (j_1, j_2) will have a larger influence if their opinions x_{j_1} and x_{j_2} are dissimilar (*i.e.* $|x_{j_1} - x_{j_2}|$ is large). Notice that the interaction kernel $K_2(x, x_1, x_2) = e^{\lambda |x_1 - x_2|} \left(\frac{x_1 + x_2}{2} - x \right)$ does satisfy the symmetry assumption (1.15).

One could further generalize this model to include interactions of groups of all sizes, with the following dynamics:

$$\frac{dx_i^N(t)}{dt} = \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{ij_1 \dots j_\ell}^{\ell,N} e^{\lambda \text{diam}(x_{j_1}^N(t), \dots, x_{j_\ell}^N(t))} \left(\frac{1}{\ell} \sum_{k=1}^{\ell} x_{j_k}^N(t) - x_i^N(t) \right),$$

where $\text{diam}(x_{j_1}, \dots, x_{j_\ell}) := \max_{k_1, k_2 \in \{j_1, \dots, j_\ell\}} |x_{k_1} - x_{k_2}|$. Here, each agent i is attracted towards the average opinion of the group $\{j_1, \dots, j_\ell\}$, with an intensity that depends on the diameter of the group.

2.4. Functional setting. In this section, we set the functional setting which we will use all along the paper. As mentioned above, solutions to the Vlasov equation (1.4)-(1.5) will consist of parametrized families $(\mu^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$ of probability densities. We shall then study existence and uniqueness of distributional solutions living in suitable measure-valued spaces. This suggests the following definition.

Definition 2.13 (Borel family of probability measures). *Consider any $\nu \in \mathcal{P}([0, 1])$, and let $(\mu^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$ be a parametrized family of probability measures defined for ν -a.e. $\xi \in [0, 1]$. We say that $(\mu^\xi)_{\xi \in [0,1]}$ is a Borel family if the map $\xi \in [0, 1] \mapsto \mu^\xi(B)$ is Borel-measurable for every Borel set $B \subset \mathbb{R}^d$.*

As explained in [65, Proposition 2.14], the following three conditions are equivalent for a parametrized family $(\mu^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$ or probability measures:

- (i) (*Borel family I*) $(\mu^\xi)_{\xi \in [0,1]}$ is a Borel family as in Definition 2.13.
- (ii) (*Borel family II*) The following scalar function

$$\xi \in [0, 1] \longmapsto \int_{\mathbb{R}^d} \phi(x) d\mu^\xi(x).$$

is Borel-measurable for every bounded and Borel-measurable $\phi : \mathbb{R}^d \longrightarrow \mathbb{R}$.

- (iii) (*Random probability measure*) The following measured-valued map

$$\xi \in [0, 1] \longmapsto \mu^\xi \in \mathcal{P}(\mathbb{R}^d),$$

is Borel-measurable when $\mathcal{P}(\mathbb{R}^d)$ is endowed with its narrow topology.

Therefore, Borel families of probability measures can be alternatively regarded as various different objects, among them: *Markov transition kernels* or also *random probability measures* (i.e., random variables with values in $\mathcal{P}(\mathbb{R}^d)$), also called *Young measures*. As we shall see below, an alternative reformulation, which will prove useful in our approach is the following.

Definition 2.14 (Fibered probability measures). *Consider any $\nu \in \mathcal{P}([0, 1])$. We define the space of fibered probability measures by*

$$\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]) := \{\mu \in \mathcal{P}(\mathbb{R}^d \times [0, 1]) : \pi_{\xi\#}\mu = \nu\},$$

where $\pi_\xi(x, \xi) = \xi$ is the projection on the second component, and therefore $\pi_{\xi\#}\mu$ stands for the marginal of μ in the second component.

So defined, fibered probability measures seem to be slightly different objects than the above Borel families of probability measures. However, it turns out that Definition 2.14 is also an alternative representation of the Borel families of probability measures in Definition 2.13 as clarified by the following classical result.

Theorem 2.15 (Disintegration theorem). *Consider any $\nu \in \mathcal{P}([0, 1])$ and $\mu \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, then there exists a ν -a.e. uniquely defined Borel family $(\mu^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$ so that*

$$\iint_{\mathbb{R}^d \times [0,1]} \varphi(x, \xi) d\mu(x, \xi) = \int_0^1 \left(\int_{\mathbb{R}^d} \varphi(x, \xi) d\mu^\xi(x) \right) d\nu(\xi),$$

for every bounded Borel-measurable map $\varphi : \mathbb{R}^d \times [0, 1] \longrightarrow \mathbb{R}$. Conversely, given any Borel family $(\mu^\xi)_{\xi \in [0,1]}$, then we can associate a unique fibered probability measure $\mu \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ so that the above formula holds true, and for simplicity, we shall write $\mu(x, \xi) = \mu^\xi(x) \otimes \nu(\xi)$.

For this reason, without loss of generality we will often identify measures $\mu \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ with their associated Borel families of (ν -a.e. defined) measures $(\mu^\xi)_{\xi \in [0,1]} \subset \mathcal{P}(\mathbb{R}^d)$.

We would like to endow the space of fibered probability measures (or Borel families of probability measures, or Markov kernels, or random measures, or Young measures, etc) with a proper metric structure. Note that the set of fibered probability measures $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ in Definition 2.14 is a closed subspace of $\mathcal{P}(\mathbb{R}^d \times [0, 1])$ endowed with the narrow topology (which is metrizable). Hence, we may consider inducing such a (metrizable) topology on $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ so that we have $\mu_n \rightarrow \mu$ narrowly if, and only if,

$$\iint_{\mathbb{R}^d \times [0,1]} \varphi(x, \xi) d\mu_n(x, \xi) \rightarrow \iint_{\mathbb{R}^d \times [0,1]} \varphi(x, \xi) d\mu(x, \xi),$$

for all $\varphi \in C_b(\mathbb{R}^d \times [0, 1])$.

Remark 2.16 (Narrow topology is not appropriate). *The use of narrow topology on $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ was discussed in [65], and it was observed that it coincides with the canonical stable topology [6, 22, 36, 72] when the marginal in the second component is fixed, see [41, Corollary 2.9], [22, Theorem 2.1.1(D)] or [9, Lemma 2.1]. Unfortunately, a strong drawback of choosing the narrow topology is that it is not stable under disintegration, meaning that if $\mu_n \rightarrow \mu$ narrowly, then it is not necessary true that $\mu_n^\xi \rightarrow \mu^\xi$ narrowly for ν -a.e. $\xi \in [0, 1]$, see [65, Remark 2.16].*

Instead of using narrow topology, we shall consider a stronger topology as follows.

Definition 2.17. Consider any $\nu \in \mathcal{P}([0, 1])$ and any $p \in [1, \infty]$, we define

$$\mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]) := \left\{ \mu \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]) : \int_0^1 d_{\text{BL}}^p(\mu^\xi, \delta_0) d\nu(\xi) < \infty \right\}.$$

$$d_{p,\nu}(\mu_1, \mu_2) := \left(\int_0^1 d_{\text{BL}}^p(\mu_1^\xi, \mu_2^\xi) d\nu(\xi) \right)^{1/p}, \quad \mu_1, \mu_2 \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]).$$

A similar fibered distance was proposed in [65] by the second author with the bounded-Lipschitz distance d_{BL} on $\mathcal{P}(\mathbb{R}^d)$ replaced by the Wasserstein distance W_2 on $\mathcal{P}_2(\mathbb{R}^d)$, and with $p = 2$. The resulting space was proven to have a weakly Riemannian structure reminiscent of the classical one on the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ proposed by OTTO [63], and an abstract theory of gradient flows for functionals on such a space was derived. In this paper, we stick to the choice of bounded-Lipschitz distance and general $p \in [1, \infty]$. We emphasize though that the ultimate goal of considering general $p \in [1, \infty]$ is not simple generality of the well-posedness results of (1.4)-(1.5), but actually, the common choice $p = \infty$ in previous literature (see *e.g.*, [46, 47]) will prove insufficient to obtain suitable stability estimates in Section 5 involving a continuous dependence on the underlying UR-hypergraphon with respect to the cut distance. For this reason the case $p = \infty$ is discarded in the statement of our main result Theorem 1.1.

We note that the space in Definition 2.17 encode a metric-valued L^p space, and as such, it is a new metric space. Specifically, we have the following result.

Proposition 2.18. For any $\nu \in \mathcal{P}([0, 1])$ and $p \in [1, \infty]$ the space $(\mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]), d_{p,\nu})$ in Definition 2.17 is a complete metric space. In addition, it is separable when $p \in [1, \infty)$.

We omit its proof since it can be found in [65, Appendix A] for $p = 2$, and its adaptation to general p is straightforward.

Proposition 2.19. For any $\nu \in \mathcal{P}([0, 1])$ and $p \in [0, 1]$ we have

$$d_{\text{BL}}(\pi_{x\#}\mu_1, \pi_{x\#}\mu_2) \leq d_{\text{BL}}(\mu_1, \mu_2) \leq d_{p,\nu}(\mu_1, \mu_2),$$

for every $\mu_1, \mu_2 \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$. Thereby, the $d_{p,\nu}$ topology is finer than the narrow topology.

Proof. \diamond STEP 1: $d_{\text{BL}}(\pi_{x\#}\mu_1, \pi_{x\#}\mu_2) \leq d_{\text{BL}}(\mu_1, \mu_2)$.

Consider any bounded and Lipschitz test function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ with $\|\phi\|_{\text{BL}} \leq 1$. By the definition of the pushforward and bounded-Lipschitz distance on $\mathcal{P}(\mathbb{R}^d \times [0, 1])$, we have

$$\begin{aligned} \int_{\mathbb{R}^d} \phi(x) (d\pi_{x\#}\mu_1(x) - d\pi_{x\#}\mu_2(x)) &= \int_{\mathbb{R}^d \times [0, 1]} \phi(x) (d\mu_1(x, \xi) - d\mu_2(x, \xi)) \\ &\leq \|\phi\|_{\text{BL}} d_{\text{BL}}(\mu_1, \mu_2) \leq d_{\text{BL}}(\mu_1, \mu_2), \end{aligned}$$

and then we end by taking supremum over all ϕ

\diamond STEP 2: $d_{\text{BL}}(\mu_1, \mu_2) \leq d_{p,\nu}(\mu_1, \mu_2)$.

Consider any bounded and Lipschitz test function $\varphi : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$ with $\|\varphi\|_{\text{BL}} \leq 1$. Using the disintegration Theorem 2.15 and Hölder's inequality with $\frac{1}{p} + \frac{1}{q} = 1$ and we have

$$\begin{aligned} \iint_{\mathbb{R}^d \times [0, 1]} \varphi d(\mu_1 - \mu_2) &= \int_0^1 \int_{\mathbb{R}^d} \varphi(x, \xi) (d\mu_1^\xi(x) - d\mu_2^\xi(x)) d\nu(\xi) \\ &\leq \int_0^1 \|\varphi(\cdot, \xi)\|_{\text{BL}} d_{\text{BL}}(\mu_1^\xi, \mu_2^\xi) d\nu(\xi) \leq \left(\int_0^1 \|\varphi(\cdot, \xi)\|_{\text{BL}}^q d\nu(\xi) \right)^{1/q} d_{p,\nu}(\mu_1, \mu_2). \end{aligned}$$

Since $\|\varphi(\cdot, \xi)\|_{\text{BL}} \leq \|\varphi\|_{\text{BL}} \leq 1$, we conclude by arbitrariness of the test function φ . \square

In most cases, we will not have an isolated measure $\mu \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, but a full curve $t \in [0, T] \mapsto \mu_t \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$. Given $t \in [0, T]$ we could disintegrate each μ_t using the above Theorem 2.15, but potentially we may have that the resulting Borel family $(\mu_t^\xi)_{\xi \in [0, 1]}$ is not well defined on a ν -negligible set depending on t . Additionally, this construction does not ensure whether for a.e. $\xi \in [0, 1]$ the family $(\mu_t^\xi)_{t \in [0, T]}$ parametrized by time must be a Borel family, and that will be important later. The following strategy can be used to circumvent both issues.

Theorem 2.20 (Time-dependent disintegrations). *Consider any $\nu \in \mathcal{P}([0, 1])$ and any Borel family of probability measures $(\mu_t)_{t \in [0, T]} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$. Then, there exists a Borel family $(\mu_t^\xi)_{(t, \xi) \in [0, T] \times [0, 1]} \subset \mathcal{P}(\mathbb{R}^d)$ defined for $dt \otimes \nu$ -a.e. $(t, \xi) \in [0, T] \times [0, 1]$ such that*

$$\int_0^T \left(\iint_{\mathbb{R}^d \times [0, 1]} \psi(t, x, \xi) d\mu_t(x, \xi) \right) dt = \iint_{[0, T] \times [0, 1]} \left(\int_{\mathbb{R}^d} \psi(t, x, \xi) d\mu_t^\xi(x) \right) dt d\nu(\xi), \quad (2.7)$$

for every bounded Borel-measurable map $\psi : [0, T] \times \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$. In particular,

- (i) For a.e. $t \in [0, T]$, the slice $(\mu_t^\xi)_{\xi \in [0, 1]}$ is a Borel family defined for ν -a.e. $\xi \in [0, 1]$, and it corresponds to a possible disintegration of μ_t in the sense of Theorem 2.15.
- (ii) For ν -a.e. $\xi \in [0, 1]$, the slice $(\mu_t^\xi)_{t \in [0, T]}$ is a Borel family defined for a.e. $t \in [0, T]$.

Proof. Since $(\mu_t)_{t \in [0, T]} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ is a Borel family, Theorem 2.15 shows allows finding a measure $\hat{\mu} := \frac{1}{T} dt|_{[0, T]} \otimes \mu_t(x, \xi) \in \mathcal{P}([0, T] \times \mathbb{R}^d \times [0, 1])$ such that

$$\iiint_{[0, T] \times \mathbb{R}^d \times [0, 1]} \psi(t, x, \xi) d\hat{\mu}(t, x, \xi) = \int_0^T \left(\iint_{\mathbb{R}^d \times [0, 1]} \psi(t, x, \xi) d\mu_t(x, \xi) \right) \frac{dt}{T}, \quad (2.8)$$

for every bounded Borel-measurable map $\psi : [0, T] \times \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$. Consider the marginal

$$\hat{\nu} := \pi_{(t, \xi)\#} \hat{\mu} = \frac{1}{T} dt|_{[0, T]} \otimes \nu(\xi) \in \mathcal{P}([0, T] \times [0, 1]),$$

and disintegrate $\hat{\mu}$ with respect (t, ξ) as in Theorem 2.15 to obtain a Borel $\hat{\nu}$ -a.e. defined family $(\mu_t^\xi)_{(t, \xi) \in [0, T] \times [0, 1]}$, which is Borel-measurable jointly in the two indices (t, ξ) , and such that

$$\iiint_{[0, T] \times \mathbb{R}^d \times [0, 1]} \psi(t, x, \xi) d\hat{\mu}(t, x, \xi) = \iint_{[0, T] \times [0, 1]} \left(\int_{\mathbb{R}^d} \psi(t, x, \xi) d\mu_t^\xi(x) \right) \frac{dt}{T} \otimes \nu(\xi), \quad (2.9)$$

for every bounded Borel-measurable map $\psi : [0, T] \times \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$. Equating (2.8) and (2.9) and multiplying by T implies (2.7).

Let us consider the $dt \otimes \nu$ -negligible set $\mathcal{N} \subset [0, T] \times [0, 1]$ consisting of the points where the Borel family $(\mu_t^\xi)_{(t, \xi) \in [0, T] \times [0, 1]}$ is not defined and note that \mathcal{N} can be reformulated as

$$\mathcal{N} = \{(t, \xi) \in [0, T] \times [0, 1] : t \in \mathcal{N}_\xi^1\} = \{(t, \xi) \in [0, T] \times [0, 1] : \xi \in \mathcal{N}_t^2\},$$

where \mathcal{N}_ξ^1 and \mathcal{N}_t^2 are the slice in the first and second components:

$$\begin{aligned} \mathcal{N}_\xi^1 &:= \{t \in [0, T] : (t, \xi) \in \mathcal{N}\}, & \xi \in [0, 1], \\ \mathcal{N}_t^2 &:= \{\xi \in [0, 1] : (t, \xi) \in \mathcal{N}\}, & t \in [0, T]. \end{aligned}$$

By Fubini theorem, we have that

$$0 = \hat{\nu}(\mathcal{N}) = \int_{[0, T] \times [0, 1]} \mathbf{1}_{\mathcal{N}}(t, \xi) d\hat{\nu}(t, \xi) = \frac{1}{T} \int_0^T \nu(\mathcal{N}_t^2) dt = \frac{1}{T} \int_0^1 |\mathcal{N}_\xi^1| d\nu(\xi).$$

Hence, $\nu(\mathcal{N}_t^2) = 0$ for a.e. $t \in [0, T]$ and $|\mathcal{N}_\xi^1| = 0$ for ν -a.e. $\xi \in [0, 1]$. Therefore, for a.e. $t \in [0, T]$ the slice $(\mu_t^\xi)_{\xi \in [0, 1]}$ is defined except on a t -dependent ν -negligible set, and for ν -a.e. $\xi \in [0, 1]$ the slice $(\mu_t^\xi)_{t \in [0, T]}$ is defined except on a ξ -dependent Lebesgue negligible set. The Borel-measurability of each slice follows from the joint measurability in (t, ξ) .

Finally, note that $(\mu_t^\xi)_{\xi \in [0, 1]}$ must be a disintegration of μ_t for a.e. $t \in [0, T]$. To this end, we consider any dense sequence $\{\varphi_n\}_{n \in \mathbb{N}} \subset C_c(\mathbb{R}^d \times [0, 1])$, which exists by separability, and consider (2.7) for the special test function $\psi(t, x, \xi) := \eta(t)\varphi_n(x, \xi)$ with $\eta \in C([0, T])$. Then,

$$\int_0^T \eta(t) \left(\iint_{\mathbb{R}^d \times [0, 1]} \varphi_n(x, \xi) d\mu_t(x, \xi) \right) dt = \int_0^T \eta(t) \left(\int_0^1 \int_{\mathbb{R}^d} \varphi_n(x, \xi) d\mu_t^\xi(x) d\nu(\xi) \right) dt,$$

for all $\psi \in C([0, T])$ and all $n \in \mathbb{N}$. Since η is arbitrary and the time integrand is integrable, then the fundamental calculus of variations shows that for every $n \in \mathbb{N}$ there exists a Lebesgue-negligible set $\mathcal{N}_n \subset [0, T]$ so that

$$\int_{\mathbb{R}^d \times [0, 1]} \varphi_n(x, \xi) d\mu_t(x, \xi) = \int_0^1 \int_{\mathbb{R}^d} \varphi_n(x, \xi) d\mu_t^\xi(x) d\nu(\xi),$$

for all $t \in [0, T] \setminus \mathcal{N}_n$ and all $n \in \mathbb{N}$. Defining the new set $\mathcal{N}_\infty := \cup_{n \in \mathbb{N}} \mathcal{N}_n$, which is Lebesgue-negligible again and by density of the sequence $\{\phi_n\}_{n \in \mathbb{N}}$, we have that for every $t \in [0, T] \setminus \mathcal{N}_\infty$ the Borel family $(\mu_t^\xi)_{\xi \in [0, 1]}$ is a disintegration of μ_t in the sense of Theorem 2.15. \square

In the above result we may be tempted to further claim that if $\mu \in C([0, T], \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]))$, then we also have $\mu^\xi \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ for ν -a.e. fiber $\xi \in [0, 1]$. However, this is not true in general as noted in Remark 2.16 because the narrow topology is not disintegrable. Moving from $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ to any $\mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1])$ does not solve the issue neither.

Remark 2.21 (On the reference measure ν). *In this paper we shall restrict to $\nu = d\xi_{[0, 1]}$, that is, the Lebesgue measure on $[0, 1]$. The fundamental reason is that, as for the classical theory of dense graph limits [54], the limit theory of dense simplicial complex in [66] focuses on hypergraphs with weighted hyperedges, but non-weighted vertices. Alternatively, we may think that vertices are weighted uniformly, which explains the uniform density $\nu = d\xi_{[0, 1]}$. However, we could have added heterogeneous weights (mass) to vertices, thus changing the underlying measure ν . This was treated in previous literature [46, 47] for Vlasov-type equations like (1.4)-(1.5) restricted to binary, or higher-order interactions of bounded order.*

In order to make the language of Borel families of probability measures $(\mu_t^\xi)_{\xi \in [0, 1]} \subset \mathcal{P}(\mathbb{R}^d)$ in Definition 2.13 compatible with the language of fibered probability measures $\mu_t \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ in Definition 2.14 at the level of the Vlasov equation, we reformulate (1.4)-(1.5) as follows:

$$\begin{aligned} \partial_t \mu_t + \operatorname{div}_x(F_w[\mu_t] \mu_t) &= 0, \quad t \geq 0, x \in \mathbb{R}^d, \xi \in [0, 1], \\ \mu_{t=0} &= \mu_0, \end{aligned} \tag{2.10}$$

with $\mu_0 \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, where the mean-field force reads

$$\begin{aligned} F_w[\mu_t](x, \xi) &:= \sum_{\ell=1}^{\infty} \int_{[0, 1]^\ell} w_\ell(\xi, \xi_1, \dots, \xi_\ell) \\ &\quad \times \left(\int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\mu_t^{\xi_1}(x_1) \cdots d\mu_t^{\xi_\ell}(x_\ell) \right) d\xi_1, \dots, d\xi_\ell, \end{aligned} \tag{2.11}$$

We note that measure-valued solutions to (2.10)-(2.11) are indeed expected to take values in the subspace of probability measures $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$. Specifically, since the divergence in (2.10)₁ only affects the variable x , then the dynamics takes place in a fiberwise, that is, measure-valued solutions should satisfy $\pi_{\xi \neq \#} \mu_t = \pi_{\xi \neq \#} \mu_0 = \nu$ for all $t \geq 0$. This further justifies our choice of proposing a theory of measure-valued solutions living in fibered spaces $\mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ and $\mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1])$. We refer to Section 4.2 for precise definitions of distributional solutions on the above space, where in particular the equivalence of (1.4)-(1.5) and (2.10)-(2.11) is shown.

3. PROPAGATION OF INDEPENDENCE OVER HYPERGRAPHS

Inspired by the approach developed in [39], we propose the following auxiliary multi-agent system, which we expect to approximate the original multi-agent system (1.3) as $N \rightarrow \infty$ (under suitable scaling assumptions on the coupling weights):

$$\begin{cases} \frac{d\bar{X}_i^N}{dt} = \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{ij_1 \dots j_\ell}^{\ell, N} \mathbb{E}_i^K K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N), \\ \bar{X}_i^N(0) = X_{i,0}^N. \end{cases} \tag{3.1}$$

Above, $\mathbb{E}_i^N = \mathbb{E}[\cdot | \bar{\mathcal{F}}_i^N]$ denotes the expectation conditioned to the natural filtration $\bar{\mathcal{F}}_i^N$ of \bar{X}_i^N , that is, the family of time-dependent sub- σ -algebras defined by

$$\bar{\mathcal{F}}_i^N(t) = \sigma(\{\bar{X}_i^N(s) : 0 \leq s \leq t\}). \quad (3.2)$$

An analogue of the intermediary system (3.1) was first introduced in [39] for multi-agent systems with binary interactions (that is $\ell = 1$). It consists in the non-exchangeable version of the McKean process proposed in [71] to derive propagation of chaos of exchangeable multi-agent systems with bounded-Lipschitz binary interaction kernels via the trajectorial approach. Here, we extend it to operate over non-exchangeable multi-agent systems evolving under higher-order interactions characterized by hypergraphs of unbounded order.

First, we study the well-posedness of the auxiliary multi-agent system (3.1).

Lemma 3.1 (Well-posedness of the intermediary system). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7). Then, for any $(X_{1,0}^N, \dots, X_{N,0}^N)$ such that $\mathbb{E}|X_{i,0}^N| < \infty$, and for any $(w_{ij_1 \dots j_\ell}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ with $\ell = 1, \dots, N-1$, there is a global-in-time solution $(\bar{X}_1^N, \dots, \bar{X}_N^N)$ to (3.1) which is unique pathwise and in law.*

We omit the proof since it is an elementary extension of the classical case with binary interactions and constant weights, see [42, Proposition 1.3.] and [71, Theorem 1.1].

We now find a quantitative error estimate about the deviation between the original system (1.3) and the approximate one (3.1). Contrarily to other approaches (e.g., [39, 71]), where only ℓ_∞ -based estimates were obtained, we find general ℓ_p -based error estimates.

Lemma 3.2 (Error estimate). *Assume that the kernels K_ℓ satisfy the assumptions (1.6) and (1.7), and the weights $(w_{ij_1 \dots j_\ell}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ with $\ell = 1, \dots, N-1$ satisfy (1.11). Suppose additionally that the symmetry hypotheses (1.14) and (1.15) are satisfied. For any $(X_{1,0}^N, \dots, X_{N,0}^N)$ with independent $X_{i,0}^N$ such that $\mathbb{E}|X_{i,0}^N|^p < \infty$ for some $p \in [1, 2]$, consider the unique solutions (X_1^N, \dots, X_N^N) to (1.3) and $(\bar{X}_1^N, \dots, \bar{X}_N^N)$ to (3.1) as in Lemma 3.1. Then,*

$$\left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p} \leq e^{(\tilde{C}_\infty^N + C_p^N)t} \varepsilon_p^N, \quad (3.3)$$

for all $t \geq 0$, where the constants \tilde{C}_∞^N , C_p^N and ε_p^N are defined by

$$\tilde{C}_\infty^N := \max_{1 \leq i \leq N} \sum_{\ell=1}^{N-1} L_\ell \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i\mathbf{j}_\ell}^{\ell, N}, \quad (3.4)$$

$$C_p^N := \left(\sum_{i=1}^N \left(\sum_{\ell=1}^{N-1} L_\ell \sum_{k=1}^{\ell} \sum_{\mathbf{j}_{\ell, k} \in \llbracket 1, N \rrbracket^{\ell-1}} \left(\sum_{j_k=1}^N (w_{i\mathbf{j}_\ell}^{\ell, N})^q \right)^{1/q} \right)^p \right)^{1/p}, \quad (3.5)$$

$$\varepsilon_p^N := 2 \left(\frac{1}{N} \sum_{i=1}^N \left(\sum_{\ell=1}^{N-1} \sqrt{\ell!} B_\ell \left(\sum_{j_1, \dots, j_\ell=1}^N (w_{ij_1 \dots j_\ell}^{\ell, N})^2 \right)^{1/2} \right)^p \right)^{1/p}, \quad (3.6)$$

and $q \in [2, \infty]$ is related to $p \in [1, 2]$ by $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We give a proof only for $1 < p < \infty$, but the same argument goes through for $p = 1$ and $p = \infty$ by suitably replacing sums by max over i when needed, and then we omit the details. We start by comparing the original X_i^N and auxiliary multi-agent system \bar{X}_i^N . Specifically, we integrate (1.3) and (3.1) over the time interval $[0, t]$, take their difference, and add and subtract the intermediary term $K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)$ to obtain

$$\begin{aligned} & |X_i^N(t) - \bar{X}_i^N(t)|^p \\ &= |X_i^N(t) - \bar{X}_i^N(t)|^{p-1} |X_i^N(t) - \bar{X}_i^N(t)| \end{aligned}$$

$$\leq \int_0^t |X_i^N(t) - \bar{X}_i^N(t)|^{p-1} A_i^N(s) ds + \int_0^t |X_i^N(t) - \bar{X}_i^N(t)|^{p-1} B_i^N(s) ds,$$

where each term takes the form

$$A_i^N := \sum_{\ell=1}^{N-1} \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i, \mathbf{j}_\ell}^{\ell, N} |K_\ell(X_i^N, X_{j_1}^N, \dots, X_{j_\ell}^N) - K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)|, \quad (3.7)$$

$$B_i^N := \sum_{\ell=1}^{N-1} \left| \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i, \mathbf{j}_\ell}^{\ell, N} (K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)) \right|. \quad (3.8)$$

Taking expectations, and using Hölder's inequality for the expectation with the exponents p and q on the products in the right hand side, we obtain

$$\mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p \leq (\mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p)^{\frac{1}{q}} \int_0^t \left((\mathbb{E}|A_i^N(s)|^p)^{1/p} + (\mathbb{E}|B_i^N(s)|^p)^{1/p} \right) ds,$$

which readily implies

$$(\mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p)^{1/p} \leq \int_0^t \left((\mathbb{E}|A_i^N(s)|^p)^{1/p} + (\mathbb{E}|B_i^N(s)|^p)^{1/p} \right) ds.$$

Taking ℓ_p norms with respect to the index i , and by Minkowski's inequality, it holds

$$\begin{aligned} & \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p} \\ & \leq \int_0^t \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|A_i^N(s)|^p \right)^{1/p} ds + \int_0^t \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|B_i^N(s)|^p \right)^{1/p} ds. \end{aligned} \quad (3.9)$$

In the sequel we focus on finding suitable bounds for $\mathbb{E}|A_i^N|^p$ and $\mathbb{E}|B_i^N|^p$.

◇ STEP 1: Control of $(\mathbb{E}|A_i^N|^p)^{1/p}$.

On the one hand, using the Lipschitz-continuity assumption (1.7) of K_ℓ on (3.7), taking global expectations, and using Hölder's inequality and Minkowski's inequality, we obtain

$$(\mathbb{E}|A_i^N|^p)^{1/p} \leq A_{i,1}^N + A_{i,2}^N,$$

where

$$A_{i,1}^N := \sum_{\ell=1}^{N-1} L_\ell \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i, \mathbf{j}_\ell}^{\ell, N} (\mathbb{E}|X_i^N - \bar{X}_i^N|^p)^{1/p}$$

and

$$A_{i,2}^N := \sum_{\ell=1}^{N-1} L_\ell \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i, \mathbf{j}_\ell}^{\ell, N} \sum_{k=1}^{\ell} (\mathbb{E}|X_{j_k}^N - \bar{X}_{j_k}^N|^p)^{1/p}.$$

Taking ℓ_p norms with respect to i in the first term $A_{i,1}^N$ yields

$$\left(\frac{1}{N} \sum_{i=1}^N (A_{i,1}^N)^p \right)^{1/p} \leq \tilde{C}_\infty^N \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_i^N - \bar{X}_i^N|^p \right)^{1/p},$$

where the constant \tilde{C}_∞^N is explicitly given by (3.4). The second term $A_{i,2}^N$ can be controlled by

$$A_{i,2}^N = \sum_{\ell=1}^{N-1} L_\ell \sum_{k=1}^{\ell} \sum_{\mathbf{j}_{\ell,k} \in \llbracket 1, N \rrbracket^{\ell-1}} \sum_{j_k=1}^N w_{i, \mathbf{j}_\ell}^{\ell, N} (\mathbb{E}|X_{j_k}^N - \bar{X}_{j_k}^N|^p)^{1/p}$$

$$\leq \sum_{\ell=1}^{N-1} L_\ell \sum_{k=1}^{\ell} \sum_{\mathbf{j}_{\ell,k} \in \llbracket 1, N \rrbracket^{\ell-1}} \left(\sum_{j_k=1}^N (w_{i\mathbf{j}_\ell}^{\ell,N})^q \right)^{1/q} \left(\sum_{j=1}^N \mathbb{E} |X_j^N - \bar{X}_j^N|^p \right)^{1/p},$$

where in the last inequality we have used Hölder's inequality with parameters p and q on the sum in the index j . Then, taking ℓ_p norms with respect to i yields

$$\left(\frac{1}{N} \sum_{i=1}^N (A_{i,2}^N)^p \right)^{1/p} \leq C_p^N \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_i^N - \bar{X}_i^N|^p \right)^{1/p},$$

where the constant C_p^N is also explicitly given in (3.5). Therefore, we readily obtain a bound on the first term of the right hand side of (3.9):

$$\left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} |A_i^N|^p \right)^{1/p} \leq (\tilde{C}_\infty^N + C_p^N) \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_i^N - \bar{X}_i^N|^p \right)^{1/p}. \quad (3.10)$$

◇ STEP 2: Control of $(\mathbb{E} |B_i^N|^p)^{1/p}$.

In this second step, we extend the approach proposed in [71] for exchangeable multi-agent system and also in [39] for non-exchangeable multi-agent systems with binary interactions to our non-exchangeable multi-agent system governed by higher-order interactions. Additionally, we extend the ℓ_∞ -based estimates on the decay of correlations by more general ℓ_p -based estimates.

Taking expectations in (3.8), using Minkowski's inequality for the sum over ℓ , and using Jensen's inequality on each of the terms of the sum over ℓ to control p -th order moments by the second order moments (which can be done because $1 \leq p \leq 2$), we can write

$$(\mathbb{E} |B_i^N|^p)^{1/p} \leq \sum_{\ell=1}^{N-1} R_i^{\ell,N},$$

where the errors $R_i^{\ell,N}$ are defined by

$$R_i^{\ell,N} := \left(\mathbb{E} \left| \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i\mathbf{j}_\ell}^{\ell,N} (K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)) \right|^2 \right)^{1/2}.$$

Squaring $R_i^{\ell,N}$ and expanding the involved square we obtain

$$\begin{aligned} (R_i^{\ell,N})^2 &= \sum_{\mathbf{j}_\ell \in \llbracket 1, N \rrbracket^\ell} \sum_{\mathbf{j}'_\ell \in \llbracket 1, N \rrbracket^\ell} w_{i\mathbf{j}_\ell}^{\ell,N} w_{i\mathbf{j}'_\ell}^{\ell,N} \mathbb{E} \left[(K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)) \right. \\ &\quad \left. \cdot (K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N)) \right]. \end{aligned} \quad (3.11)$$

We now distinguish two different cases for the multi-indices \mathbf{j}_ℓ and \mathbf{j}'_ℓ depending on whether $\{j_1, \dots, j_\ell\}$ can be rearranged into $\{j'_1, \dots, j'_\ell\}$ by a permutation or not.

▷ CASE 1: There exists a permutation $\sigma \in \mathcal{S}_\ell$ such that $\sigma(\mathbf{j}'_\ell) = \mathbf{j}_\ell$.

In this case, by the symmetry assumptions (1.14) on $(w_{ij_1 \dots j_\ell}^{\ell,N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$, it holds

$$w_{i\mathbf{j}'_\ell}^{\ell,N} = w_{i\sigma(\mathbf{j}'_\ell)}^{\ell,N} = w_{i\mathbf{j}_\ell}^{\ell,N}.$$

Similarly, by the symmetry (1.15) for the kernels K_ℓ we also have

$$\begin{aligned} &K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N) \\ &= K_\ell(\bar{X}_i^N, \bar{X}_{\sigma(j'_1)}^N, \dots, \bar{X}_{\sigma(j'_\ell)}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{\sigma(j'_1)}^N, \dots, \bar{X}_{\sigma(j'_\ell)}^N) \\ &= K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N). \end{aligned}$$

Therefore, the corresponding diagonal term in the sum (3.8) simplifies into

$$(w_{i\mathbf{j}_\ell}^{\ell,N})^2 \mathbb{E} |K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)|^2 \leq 4B_\ell^2 (w_{i\mathbf{j}_\ell}^{\ell,N})^2.$$

▷ CASE 2: There is no permutation $\sigma \in \mathcal{S}_\ell$ such that $\sigma(\mathbf{j}'_\ell) = \mathbf{j}_\ell$.

In this case, there exists $0 \leq m < \ell$ and two permutations $\sigma, \sigma' \in \mathcal{S}_\ell$ so that $\sigma(\mathbf{j}_\ell)$ and $\sigma'(\mathbf{j}'_\ell)$ share exactly their first m components, and the last $\ell - m$ components are disjoint, that is,

$$\forall i \in \llbracket 1, m \rrbracket, j_{\sigma(i)} = j'_{\sigma'(i)} \quad \text{and} \quad \{j_{\sigma(m+1)}, \dots, j_{\sigma(\ell)}\} \cap \{j'_{\sigma'(m+1)}, \dots, j'_{\sigma'(\ell)}\} = \emptyset. \quad (3.12)$$

Similarly to \mathbb{E}_i^N , we define the new expectation $\mathbb{E}_{i,m}^N = \mathbb{E}[\cdot | \bar{\mathcal{F}}_{i,m}^N]$ which is conditioned to the natural filtration $\bar{\mathcal{F}}_{i,m}^N$ of the vector $(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N)$, that is

$$\bar{\mathcal{F}}_{i,m}^N(t) = \sigma(\{\bar{X}_i^N(s), \bar{X}_{j_{\sigma(1)}}^N(s), \dots, \bar{X}_{j_{\sigma(m)}}^N(s) : 0 \leq s \leq t\}).$$

The law of iterated expectations $\mathbb{E} = \mathbb{E} \mathbb{E}_{i,m}^N$ along with the above symmetry assumption (1.15) on the kernels imply

$$\begin{aligned} & \mathbb{E} \left[(K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)) \right. \\ & \quad \cdot (K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N) - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j'_1}^N, \dots, \bar{X}_{j'_\ell}^N)) \left. \right] \\ &= \mathbb{E} \left\{ \mathbb{E}_{i,m}^N \left[(K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \right. \right. \\ & \quad \left. \left. - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \right) \right. \\ & \quad \cdot (K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j'_{\sigma'(m+1)}}^N, \dots, \bar{X}_{j'_{\sigma'(\ell)}}^N) \\ & \quad \left. \left. - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j'_{\sigma'(m+1)}}^N, \dots, \bar{X}_{j'_{\sigma'(\ell)}}^N) \right) \right] \left. \right\}. \end{aligned}$$

By (3.12) and the independence of all \bar{X}_i^N we observe that the random vectors $(\bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N)$ and $(\bar{X}_{j'_{\sigma'(m+1)}}^N, \dots, \bar{X}_{j'_{\sigma'(\ell)}}^N)$ are independent and therefore, the random variable

$$\begin{aligned} & K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \\ & \quad - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \end{aligned}$$

and the random variable

$$\begin{aligned} & K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j'_{\sigma'(m+1)}}^N, \dots, \bar{X}_{j'_{\sigma'(\ell)}}^N) \\ & \quad - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j'_{\sigma'(m+1)}}^N, \dots, \bar{X}_{j'_{\sigma'(\ell)}}^N) \end{aligned}$$

are independent conditioned to the sub- σ -algebra $\bar{\mathcal{F}}_{i,m}^N$. Hence, the conditional expectation $\mathbb{E}_{i,m}^N$ of their product is the product of their conditional expectations. The conditional expectations of those random variables separately take the value

$$\begin{aligned} & \mathbb{E}_{i,m}^N \left[K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \right. \\ & \quad \left. - \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \right] \\ & \mathbb{E}_{i,m}^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) \\ & \quad - \mathbb{E}_{i,m}^N \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_{\sigma(1)}}^N, \dots, \bar{X}_{j_{\sigma(m)}}^N, \bar{X}_{j_{\sigma(m+1)}}^N, \dots, \bar{X}_{j_{\sigma(\ell)}}^N) = 0, \end{aligned}$$

where in the last step we have used that since $\bar{\mathcal{F}}_{i,m}^N \subset \bar{\mathcal{F}}_i^N$ then the law of iterated expectations yields $\mathbb{E}_{i,m}^N \mathbb{E}_i^N = \mathbb{E}_i^N$, and similarly for the second random variable.

Altogether this shows that if \mathbf{j}_ℓ and \mathbf{j}'_ℓ lie in CASE 2, the corresponding non-diagonal term in (3.11) necessarily vanishes. Therefore only multi-indices as in CASE 1 leading to diagonal terms can persist. Since given \mathbf{j}_ℓ there are $\ell!$ possibilities for \mathbf{j}'_ℓ as in CASE 1 (that is, all possible

permutations of the ℓ components), then we obtain the following bound for $R_i^{\ell,N}$ in (3.11):

$$R_i^{\ell,N} \leq 2B_\ell \sqrt{\ell!} \left(\sum_{j_1, \dots, j_\ell=1}^N (w_{ij_1 \dots j_\ell}^{\ell,N})^2 \right)^{1/2}.$$

Hence, we have the following control of $(\mathbb{E}|B_i^N|^p)^{1/p}$ in (3.8):

$$(\mathbb{E}|B_i^N|^p)^{1/p} \leq 2 \sum_{\ell=1}^{N-1} B_\ell \sqrt{\ell!} \left(\sum_{j_1, \dots, j_\ell=1}^N (w_{ij_1 \dots j_\ell}^{\ell,N})^2 \right)^{1/2}.$$

Taking ℓ_p norms with respect to i yields

$$\left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|B_i^N|^p \right)^{1/p} \leq \varepsilon_p^N, \quad (3.13)$$

where ε_p^N is given in (3.6). Plugging (3.10) and (3.13) into (3.9) implies the integral inequality:

$$\begin{aligned} & \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p} \\ & \leq (\tilde{C}_\infty^N + C_p^N) \int_0^t \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_i^N(s) - \bar{X}_i^N(s)|^p \right)^{1/p} ds + \varepsilon_p^N t, \end{aligned}$$

for all $t \geq 0$. Thereby, since $X_i^N(0) = \bar{X}_i^N(0)$, Grönwall's lemma ends the proof. \square

As we observe in the proof, the only motivation to restrict to $p \in [1, 2]$ appears in STEP 2, in order to interpolate $(\mathbb{E}|B_i^N|^p)^{1/p}$ by the power $p = 2$. By doing so, the above is controlled by a squared expectation error term as in the seminal work [71], where an analogous cancellation property is obtained thanks to the independence of the random variables \bar{X}_i^N . We note that the power $p = 2$ is not the only special case where this can be done, as notice in [42, Proposition 2.3], where the authors also studied the case $p = 4$. However, for the sake of simplicity we preferred to restrict to the case $p = 2$.

Remark 3.3 (Propagation of independence). *We observe that Lemma 3.2 applies to general hypergraphs satisfying the symmetry condition (1.14). In order for the right hand side of (3.3) to decay as $N \rightarrow \infty$ one needs to assume additionally the following (general) conditions*

$$\sup_{N \in \mathbb{N}} (\tilde{C}_\infty^N + C_p^N) < \infty, \quad \lim_{N \rightarrow \infty} \varepsilon_p^N = 0.$$

A particular case where the above two conditions are met is when the scaling assumption (1.12) of weights and the scaling assumption (1.9) of kernels hold true. Specifically, note that the former guarantees the following control of the above constants

$$\tilde{C}_\infty^N \leq W \sum_{\ell=1}^{\infty} L_\ell, \quad C_p^N \leq W \sum_{\ell=1}^{\infty} \ell L_\ell, \quad \varepsilon_p^N \leq 2W \sum_{\ell=1}^{\infty} \frac{\sqrt{\ell!} B_\ell}{N^{\ell/2}}.$$

For binary interactions, the first term $\ell = 1$ in the above control of ε_p^N coincides with the classical decay $N^{-1/2}$ in propagation of chaos [71]. Interestingly, when multi-body interactions are considered, we find that the higher ℓ , the quicker the decay of the corresponding term $N^{-\ell/2}$. When all possible multi-body interactions are superposed, the scaling assumption (1.9)₂ guarantees the uniform-in- N bound of \tilde{C}_∞^N and C_p^N , and also (1.9)₁ implies

$$\varepsilon_p^N \leq \left(2W \sum_{\ell=1}^{\infty} \frac{\sqrt{\ell!} B_\ell}{\eta^\ell} \right) \frac{\eta^{1/2}}{N^{1/2}},$$

for all $N \geq \eta$. Hence, we are able to recover the global decay $N^{-1/2}$, which coincides with the weakest of the decays among all possible multi-body interactions.

In the following, we show that for independent initial data $X_{i,0}^N$, and therefore independent \bar{X}_i^N solving (3.1), their laws fulfill a closed system of coupled PDEs.

Lemma 3.4 (Coupled PDE system). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), and the weights $(w_{i_{j_1 \dots j_\ell}}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ with $\ell = 1, \dots, N-1$ satisfy (1.11). For any $(X_{1,0}^N, \dots, X_{N,0}^N)$ with independent $X_{i,0}^N$ such that $\mathbb{E}|X_{i,0}^N| < \infty$, consider the unique solution $(\bar{X}_1^N, \dots, \bar{X}_N^N)$ to (3.1) as in Lemma 3.1, and define their associated laws*

$$\bar{\lambda}_t^{N,i} := \text{Law}(\bar{X}_i^N(t)), \quad t \geq 0, 1 \leq i \leq N. \quad (3.14)$$

Then, $(\bar{\lambda}^{N,i})_{1 \leq i \leq N}$ is a solution in distributional sense to the following coupled PDE system

$$\begin{aligned} \partial_t \bar{\lambda}_t^{N,i} + \text{div}_x (F_i^N[\bar{\lambda}_t^{N,1}, \dots, \bar{\lambda}_t^{N,N}] \bar{\lambda}_t^{N,i}) &= 0, \quad t \geq 0, x \in \mathbb{R}^d, 1 \leq i \leq N, \\ \bar{\lambda}_0^{N,i} &= \text{Law}(X_{i,0}^N), \end{aligned} \quad (3.15)$$

where the forces F_i^N take the form

$$F_i^N[\bar{\lambda}_t^{N,1}, \dots, \bar{\lambda}_t^{N,N}](x) = \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{i_{j_1 \dots j_\ell}}^{\ell, N} \int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,1}(x_1) \cdots d\bar{\lambda}_t^{N,N}(x_N). \quad (3.16)$$

Proof. Assume that $i \notin \{j_1, \dots, j_\ell\}$ and note that all $\bar{X}_j^N(t)$ can be regarded as measurable functions $\bar{X}_j^N(t) : \Omega_N \rightarrow \mathbb{R}^d$ on a common probability space $(\Omega_N, \Sigma_N, \mathbb{P}_N)$. Then, by the independency of \bar{X}_i^N and $(\bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)$ and the definition of conditional expectation we have

$$\begin{aligned} &\mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N)(\omega) \\ &= \mathbb{E} [K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) | \bar{X}_i^N = \bar{X}_i^N(\omega)] = \mathbb{E} K_\ell(\bar{X}_i^N(\omega), \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) \\ &= \int_{\mathbb{R}^{d\ell}} K_\ell(\bar{X}_i^N(\omega), x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,j_1}(x_1) \cdots d\bar{\lambda}_t^{N,j_\ell}(x_\ell), \end{aligned}$$

for all $\omega \in \Omega_N$, where in the last step we have used the law of the unconscious statistician. Hence, we obtain the following identity

$$w_{i_{j_1 \dots j_\ell}}^{\ell, N} \mathbb{E}_i^N K_\ell(\bar{X}_i^N, \bar{X}_{j_1}^N, \dots, \bar{X}_{j_\ell}^N) = w_{i_{j_1 \dots j_\ell}}^{\ell, N} \int_{\mathbb{R}^{d\ell}} K_\ell(\bar{X}_i^N, x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,j_1}(x_1) \cdots d\bar{\lambda}_t^{N,j_\ell}(x_\ell),$$

for all $i, j_1, \dots, j_\ell \in \llbracket 1, N \rrbracket$ because we are assuming that $w_{i_{j_1 \dots j_\ell}}^{\ell, N} = 0$ for all $i \in \{j_1, \dots, j_\ell\}$ by (1.11), and therefore the multi-agent system (3.1) can be reformulated in an alternative way closer to the formulation of the McKean process introduced in [71] for exchangeable systems:

$$\begin{aligned} \frac{d\bar{X}_i^N}{dt} &= \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{i_{j_1 \dots j_\ell}}^{\ell, N} \int_{\mathbb{R}^{d\ell}} K_\ell(\bar{X}_i^N, x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,j_1}(x_1) \cdots d\bar{\lambda}_t^{N,j_\ell}(x_\ell), \\ \bar{X}_i^N(0) &= X_{i,0}^N. \end{aligned} \quad (3.17)$$

Then, by a standard argument based on the law of unconscious statistician again we obtain that the laws $\bar{\lambda}_t^{N,i}$ must satisfy the coupled PDE system (3.15). \square

Note that according to Remark 3.3, Lemma 3.2 allows approximating the multi-agent system (1.3), consisting of non-exchangeable and dependent random variables $X_i^N(t)$, by an intermediary multi-agent system (3.1), consisting of still non-exchangeable but independent random variables $\bar{X}_i^N(t)$ under general enough conditions on the hypergraphs and the interaction kernels (e.g., (1.9) and (1.12)). The independency can also be observed in Lemma 3.4, where the individual laws $\bar{\lambda}_t^{N,i}$ of each approximate agent $\bar{X}_i^N(t)$ solve a closed system of N coupled PDEs (3.15)-(3.16). Whilst the degree of complexity has been reduced enormously once correlations have been destroyed, this coupled system of PDEs is still highly complex since the number of PDEs grows with N and one would like to have a simpler representation, in particular as $N \rightarrow \infty$. This was already explored in [39] by using a graphon reformulation of (3.15)-(3.16).

Definition 3.5 (Graphon reformulation). *For every $N \in \mathbb{N}$, and $t \in \mathbb{R}_+$ we define*

$$\mu_t^N \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]), \quad \bar{\mu}_t^N \in \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]), \quad w^N = (w_\ell^N)_{\ell \in \mathbb{N}},$$

where each element is given as follows

$$\mu_t^{N,\xi} := \sum_{i=1}^N \mathbb{1}_{I_i^N}(\xi) \delta_{X_i^N(t)}, \quad \xi \in [0, 1], \quad (3.18)$$

$$\bar{\mu}_t^{N,\xi} := \sum_{i=1}^N \mathbb{1}_{I_i^N}(\xi) \bar{\lambda}_t^{N,i}, \quad \xi \in [0, 1], \quad (3.19)$$

$$w_\ell^N(\xi, \xi_1, \dots, \xi_\ell) := \sum_{i, j_1, \dots, j_\ell=1}^N \mathbb{1}_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N}(\xi, \xi_1, \dots, \xi_\ell) N^\ell w_{ij_1 \dots j_\ell}^{\ell, N}, \quad \xi, \xi_1, \dots, \xi_\ell \in [0, 1], \quad (3.20)$$

for all $1 \leq \ell \leq N-1$, $w_\ell^N \equiv 0$ for all $\ell \geq N$, and $I_i^N = [\frac{i-1}{N}, \frac{i}{N})$ for all $1 \leq i \leq N$.

On the one hand, note that actually $\bar{\mu}_t^N \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$, and also $\mu_t^N \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$ for each realization. On the other hand, $w_\ell^N \in L^\infty([0, 1]^{\ell+1})$ for each $\ell \in \mathbb{N}$ and their L^∞ norms are uniformly bounded with respect to $\ell \in \mathbb{N}$. If we further assume the additional symmetry hypothesis (1.13) on the weights, then $w^N = (w_\ell^N)_{\ell \in \mathbb{N}}$ are UR-hypergraphons (cf. Definition 2.5) for all $N \in \mathbb{N}$, though L^∞ norms may blow up as $N \rightarrow \infty$ unless we additionally assume the fundamental hypothesis (1.12) on the coupling weights, which will become important in Section 6. Then, we have the enough measurability properties to check whether the graphon reformulation $(\bar{\mu}^N, w^N)$ associated to the intermediary multi-agent system (3.1) is a genuine distributional solution to the Vlasov-equation (2.10)-(2.11).

Lemma 3.6. *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), and the weights $(w_{ij_1 \dots j_\ell}^{\ell, N})_{1 \leq i, j_1, \dots, j_\ell \leq N}$ with $\ell = 1, \dots, N-1$ satisfy (1.11). For any $(X_{1,0}^N, \dots, X_{N,0}^N)$ with independent $X_{i,0}^N$ such that $\mathbb{E}|X_{i,0}^N| < \infty$, consider the unique solution $(\bar{X}_1^N, \dots, \bar{X}_N^N)$ to (3.1) as in Lemma 3.1, their associated laws $(\bar{\lambda}^{N,i})_{1 \leq i \leq N}$ as in (3.14) and the graphon reformulation $(\bar{\mu}^N, w^N)$ in Definition 3.5. Then, $\bar{\mu}^N$ is a distributional solution to the Vlasov equation (2.10)-(2.11) with hypergraphon $w^N = (w_\ell^N)_{\ell \in \mathbb{N}}$ and initial datum $\bar{\mu}_{t=0}^{N,\xi} = \sum_{i=1}^N \mathbb{1}_{I_i^N}(\xi) \text{Law}(X_{i,0})$.*

Proof. Consider any $\xi \in [0, 1]$, fix $i \in \llbracket 1, N \rrbracket$ such that $\xi \in I_i^N$ and note that $\bar{\mu}_t^{N,\xi} = \bar{\lambda}_t^{N,i}$. Then, anything we have to prove is that the following identity holds true

$$F_i^N[\bar{\lambda}_t^{N,1}, \dots, \bar{\lambda}_t^{N,N}] = F_{w^N}[\bar{\mu}_t^N](\cdot, \xi).$$

Indeed, bearing in mind the scaling N^ℓ present in the definition of piecewise constant functions w_ℓ^N and the fact that $|I_{j_1} \times \dots \times I_{j_\ell}| = N^\ell$, then one recognises that the discrete sums below correspond to integrals of piecewise constant functions and therefore

$$\begin{aligned} & F_i^N[\bar{\lambda}_t^{N,1}, \dots, \bar{\lambda}_t^{N,N}](x) \\ &= \sum_{\ell=1}^{N-1} \sum_{j_1, \dots, j_\ell=1}^N w_{ij_1 \dots j_\ell}^{\ell, N} \int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,1}(x_1) \dots d\bar{\lambda}_t^{N,N}(x_N) \\ &= \sum_{\ell=1}^{N-1} \frac{1}{N^\ell} \sum_{j_1, \dots, j_\ell=1}^N N^\ell w_{ij_1 \dots j_\ell}^{\ell, N} \int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\bar{\lambda}_t^{N,1}(x_1) \dots d\bar{\lambda}_t^{N,N}(x_N) \\ &= \sum_{\ell=1}^{N-1} \int_{[0,1]^\ell} w_\ell^N(\xi, \xi_1, \dots, \xi_\ell) \int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\bar{\mu}_t^{N,\xi_1}(x_1) \dots d\bar{\mu}_t^{N,\xi_\ell}(x_\ell) d\xi_1 \dots d\xi_\ell, \end{aligned}$$

and then one finds $F_{w^N}[\bar{\mu}_t^N](x, \xi)$ in the right hand side, thus ending the proof. \square

Whilst the precise notion of distributional solution to (2.10)-(2.11) and also (3.15)-(3.16) is standard, in Section 4 we formulate the definition more precisely and we show that the Vlasov equation admits unique solutions living in appropriate measure-valued spaces.

4. WELL-POSEDNESS OF THE VLASOV EQUATION OVER UR-HYPERGRAPHS

In this section we study the well-posedness of the Vlasov equation (2.10)-(2.11). We will not restrict to a well-posedness theory of (2.10)-(2.11) over UR-hypergraphs $w = (w_\ell)_{\ell \in \mathbb{N}}$ only, but actually the theory will be applicable to a broader class of w where L^∞ norms are not available, thus opening the way to a sparse setting.

4.1. Properties of the force and well-posedness of characteristics. In this section we introduce some properties of the force $F_w[\mu_t]$, which will prove useful to solve the kinetic equation (2.10)-(2.11) by the method of characteristics based on the Cauchy-Lipschitz theorem.

Proposition 4.1 (Properties of the force I). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), consider any Borel family of probability measures $(\mu_t)_{t \in \mathbb{R}_+} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, any $w = (w_\ell)_{\ell \in \mathbb{N}}$, and suppose that*

$$B_F := \left\| \sum_{\ell=1}^{\infty} B_\ell \|w_\ell\|_{L^1_{\xi_\ell}} \right\|_{L^\infty} < \infty, \quad L_F := \left\| \sum_{\ell=1}^{\infty} L_\ell \|w_\ell\|_{L^1_{\xi_\ell}} \right\|_{L^\infty} < \infty. \quad (4.1)$$

Then, the force $F_w[\mu_t](x, \xi)$ in (1.5) is well-defined for all $t \in \mathbb{R}_+$, all $x \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$, and it verifies the following three properties:

(i) (Boundedness)

$$|F_w[\mu_t](x, \xi)| \leq B_F,$$

for all $t \in \mathbb{R}_+$, all $x \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$.

(ii) (Lipschitz-continuity with respect to x)

$$|F_w[\mu_t](x, \xi) - F_w[\mu_t](\tilde{x}, \xi)| \leq L_F |x - \tilde{x}|,$$

for all $t \in \mathbb{R}_+$, all $x, \tilde{x} \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$.

(iii) (Measurability)

$$(t, \xi) \in \mathbb{R}_+ \times [0, 1] \mapsto F_w[\mu_t](x, \xi),$$

is Borel-measurable for all $x \in \mathbb{R}^d$.

Proof. We first show boundedness condition, which in particular ensures that $F_w[\mu_t](x, \xi)$ is well defined for all $t \in \mathbb{R}_+$, all $x \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$. Specifically, take any $t \in \mathbb{R}_+$, any $x \in \mathbb{R}^d$ and any for which all $w_\ell(\xi, \cdot)$ are defined (therefore a.e. $\xi \in [0, 1]$), and note that

$$\begin{aligned} |F_w[\mu_t](x, \xi)| &\leq \sum_{\ell=1}^{\infty} \int_{[0,1]^\ell} |w_\ell(\xi, \xi_1, \dots, \xi_\ell)| \left(\int_{\mathbb{R}^d} B_\ell d\mu_t^{\xi_1}(x_1) \cdots d\mu_t^{\xi_\ell}(x_\ell) \right) d\xi_1 \dots d\xi_\ell \\ &= \sum_{\ell=1}^{\infty} B_\ell \int_{[0,1]^\ell} |w_\ell(\xi, \xi_1, \dots, \xi_\ell)| d\xi_1 \dots d\xi_\ell \leq B_F, \end{aligned}$$

where in the first step we have used the boundedness assumption (1.6) of the interaction kernels K_ℓ , in the second step we have use that $\mu_t^{\xi_1}, \dots, \mu_t^{\xi_\ell} \in \mathcal{P}(\mathbb{R}^d)$, and in the last step we have taken essential supremum with respect to $\xi \in [0, 1]$.

Similarly, for all $t \geq 0$, $x, \tilde{x} \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$, the Lipschitz-continuity (1.7) implies

$$\begin{aligned} &|F_w[\mu_t](x, \xi) - F_w[\mu_t](\tilde{x}, \xi)| \\ &\leq \sum_{\ell=1}^{\infty} \int_{[0,1]^\ell} |w_\ell(\xi, \xi_1, \dots, \xi_\ell)| \left(\int_{\mathbb{R}^d} L_\ell |x - \tilde{x}| d\mu_t^{\xi_1}(x_1) \cdots d\mu_t^{\xi_\ell}(x_\ell) \right) d\xi_1 \dots d\xi_\ell \\ &= \left(\sum_{\ell=1}^{\infty} L_\ell \int_{[0,1]^\ell} |w_\ell(\xi, \xi_1, \dots, \xi_\ell)| d\xi_1 \dots d\xi_\ell \right) |x - \tilde{x}| \leq L_F |x - \tilde{x}|, \end{aligned}$$

Finally, since $(\mu_t)_{t \in \mathbb{R}_+}$ is a Borel family, then so is $(\mu_t^{\otimes \ell})_{t \in \mathbb{R}_+}$ (we give a prove below). Note that each $w_\ell(\xi, \cdot) K_\ell(x, \cdot)$ is a bounded and Borel measurable. Then each

$$(t, \xi) \in \mathbb{R}_+ \times [0, 1] \mapsto \int_{(\mathbb{R}^d \times [0, 1])^\ell} w_\ell(\xi, \cdot) K_\ell(x, \cdot) d\mu_t^{\otimes \ell},$$

is Borel-measurable for all $\ell \in \mathbb{N}$, then so is $(t, \xi) \in \mathbb{R}_+ \times [0, 1] \mapsto F_w[\mu_t](x, \xi)$ since it can be obtained as the sum over ℓ of all the above Borel-measurable functions.

To conclude, we recall why the tensor product of Borel families of probability measures is a new Borel family, which has been used above, and it is based on a standard application of Dynskin's $\pi - \lambda$ theorem [16, Theorem 3.2]. For any $\ell \in \mathbb{N}$, we define the sets

$$\mathcal{P}_\ell := \{B_1 \otimes \cdots \otimes B_\ell \subset (\mathbb{R}^d)^\ell : B_\ell \subset \mathbb{R}^d \text{ is a Borel set}\},$$

$$\Lambda_\ell := \{E_\ell \subset (\mathbb{R}^d)^\ell \text{ Borel sets} : t \in \mathbb{R}_+ \rightarrow \mu_t^{\otimes \ell}(E_\ell) \text{ is Borel-measurable}\}.$$

On the one hand, it is clear that $\mathcal{P}_\ell \subset \Lambda_\ell$ because for any collection of Borel sets $B_1, \dots, B_\ell \subset \mathbb{R}^d$ we have that $t \in \mathbb{R}_+ \mapsto \mu_t^{\otimes \ell}(B_1 \otimes \cdots \otimes B_\ell) = \mu_t(B_1) \cdots \mu_t(B_\ell)$ is Borel-measurable because it is the product of ℓ measurable functions by hypothesis. Additionally, it is clear that \mathcal{P}_ℓ is a π -set, *i.e.*, it is closed under finite intersections, and Λ_ℓ is a λ -system. Specifically, we have that $(\mathbb{R}^d)^\ell \in \Lambda_\ell$, if $E_\ell^1 \subset E_\ell^2 \in \Lambda_\ell$ then $E_\ell^2 \setminus E_\ell^1 \in \Lambda_\ell$, and also if $\{E_\ell^n\}_{n \in \mathbb{N}} \subset \Lambda_\ell$ is an increasing sequence, then $\cup_{n \in \mathbb{N}} E_\ell^n \in \Lambda_\ell$. Therefore, the $\pi - \lambda$ theorem ensures that $\sigma(\mathcal{P}_\ell) \subseteq \Lambda_\ell$. Since $\sigma(\mathcal{P}_\ell)$ is the Borel σ -algebra of $(\mathbb{R}^d)^\ell$, we conclude by definition of Λ_ℓ . \square

Proposition 4.2 (Well-posed characteristics). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), consider any Borel family of probability measures $(\mu_t)_{t \in \mathbb{R}_+} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, any $w = (w_\ell)_{\ell \in \mathbb{N}}$, and suppose that (4.1) holds. Then, there is a unique global-in-time Caratheodory solution $X_w[\mu](t, x, \xi)$ to the characteristic system*

$$\begin{aligned} \frac{d}{dt} X_w[\mu](t, x, \xi) &= F_w[\mu_t](X_w[\mu](t, x, \xi), \xi), \quad t \geq 0, \\ X_w[\mu](0, x, \xi) &= x, \end{aligned} \tag{4.2}$$

for all $x \in \mathbb{R}$ and a.e. $\xi \in [0, 1]$. For simplicity of notation, let us denote

$$\mathcal{T}_t^\xi[w, \mu](x) := X_w[\mu](t, x, \xi), \quad t \geq 0, x \in \mathbb{R}^d, \text{ a.e. } \xi \in [0, 1]. \tag{4.3}$$

Then, the following properties hold true:

- (i) For every $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, the map $\xi \in [0, 1] \mapsto \mathcal{T}_t^\xi[w, \mu](x)$ is Borel-measurable.
- (ii) For a.e. $\xi \in [0, 1]$, the map $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathcal{T}_t^\xi[w, \mu](x)$ is continuous.
- (iii) For every $t \in \mathbb{R}_+$ and a.e. $\xi \in [0, 1]$, the map $x \in \mathbb{R}^d \mapsto \mathcal{T}_t^\xi[w, \mu](x)$ is Lipschitz-continuous with Lipschitz constant $e^{L_F t}$.

Proof. By Proposition 4.1 we have that for a.e. $\xi \in [0, 1]$, the map

$$(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto F_w[\mu_t](x, \xi),$$

satisfies the Caratheodory conditions. Specifically:

- (a) For every $t \geq 0$, the map $x \in \mathbb{R}^d \mapsto F_w[\mu_t](x, \xi)$ is Lipschitz-continuous.
- (b) For every $x \in \mathbb{R}^d$, the map $t \in \mathbb{R}_+ \mapsto F_w[\mu_t](x, \xi)$ is Borel-measurable.
- (c) There exists some $m \in L_{loc}^1(\mathbb{R}_+)$ so that $|F_w[\mu_t](x, \xi)| \leq m(t)$ for all $t \geq 0$ and all $x \in \mathbb{R}^d$ (*e.g.*, take $m(t) = B_F$ for all $t \geq 0$ with B_F in (4.1)).

Therefore, by Caratheodory's existence theorem, for every initial datum $x \in \mathbb{R}^d$ there is a unique absolutely continuous trajectory $t \in \mathbb{R}_+ \mapsto X_w[\mu](t, x)$ solving (4.2) in the sense of Caratheodory, that is, the differential equations holds for a.e. $t \geq 0$.

Since F_w is uniformly bounded by Proposition 4.1, and therefore it has sublinear growth, the trajectories are indeed defined globally in time. Additionally, the continuity in $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ in item (ii) is clear by the Caratheodory theory, and the Borel-measurability with respect to

$\xi \in [0, 1]$ in item (i) is also true by the measurability of $F_w[\mu]$ in Proposition 4.1. Regarding the Lipschitz-continuity with respect to $x \in \mathbb{R}^d$ in item (iii) note that

$$\begin{aligned} & |X_w[\mu](t, x, \xi) - X_w[\mu](t, \tilde{x}, \xi)| \\ &= \left| \int_0^t (F_w[\mu_s](X_w[\mu](s, x, \xi), \xi) - F_w[\mu_s](X_w[\mu](s, \tilde{x}, \xi), \xi)) ds + (x - \tilde{x}) \right| \\ &\leq \int_0^t |F_w[\mu_s](X_w[\mu](s, x, \xi), \xi) - F_w[\mu_s](X_w[\mu](s, \tilde{x}, \xi), \xi)| ds + |x - \tilde{x}| \\ &\leq \int_0^t L_F |X_w[\mu](s, x, \xi) - X_w[\mu](s, \tilde{x}, \xi)| ds + |x - \tilde{x}|, \end{aligned}$$

for all $t \in \mathbb{R}_+$, all $x, \tilde{x} \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$, where in the last step we have used the Lipschitz continuity of $F_w[\mu](x, \xi)$ with respect to $x \in \mathbb{R}^d$ in item (ii) of Proposition 4.1. Thus, by Grönwall's lemma, we obtain

$$|\mathcal{T}_t^\xi[w, \mu](x) - \mathcal{T}_t^\xi[w, \mu](\tilde{x})| \leq e^{L_F t} |x - \tilde{x}|,$$

which implies that $\mathcal{T}_t^\xi[w, \mu]$ is Lipschitz-continuous for all $t \in \mathbb{R}_+$ and a.e. $\xi \in [0, 1]$, and we further deduce that following control on the Lipschitz constant

$$[\mathcal{T}_t^\xi[w, \mu]]_{\text{Lip}} \leq e^{L_F t}.$$

□

Remark 4.3. We remark that the scaling conditions on the kernels K_ℓ

$$\sum_{\ell=1}^{\infty} B_\ell < \infty, \quad \sum_{\ell=1}^{\infty} L_\ell < \infty, \quad (4.4)$$

which is weaker than (1.9), together with the scaling condition

$$\sup_{\ell \in \mathbb{N}} \|w_\ell\|_{L^\infty} \leq W,$$

(e.g., $w \in \mathcal{H}_W$ by Definition 2.5) ensure the above hypothesis (4.1) since

$$B_F \leq W \sum_{\ell=1}^{\infty} B_\ell, \quad L_F \leq W \sum_{\ell=1}^{\infty} L_\ell.$$

4.2. Notion of distributional solution.

Definition 4.4 (Distributional solutions of (2.10)-(2.11)). Consider any curve of probability measures $\mu \in C([0, T], \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]))$ -narrow for some $T > 0$, and any $w = (w_\ell)_{\ell \in \mathbb{N}}$. We say that μ is a distributional solution of the system (2.10)-(2.11) if

$$\begin{aligned} & \int_0^T \iint_{\mathbb{R}^d \times [0, 1]} (\partial_t \psi(t, x, \xi) + F_w[\mu_t](t, x, \xi) \cdot \nabla_x \psi(t, x, \xi)) d\mu_t(x, \xi) dt \\ &= - \iint_{\mathbb{R}^d \times [0, 1]} \psi(0, x, \xi) d\mu_0(x, \xi), \end{aligned} \quad (4.5)$$

for all $\psi \in C_c^1([0, T] \times \mathbb{R}^d \times [0, 1])$.

We note that, under the assumptions (4.1), the force $F_w[\mu]$ is bounded and Borel-measurable jointly in (t, x, ξ) by Proposition 4.1 because it is Borel-measurable in (t, ξ) and continuous in the variable x . Therefore, so is the integrand $\partial_t \psi + F[\mu] \cdot \nabla_x \psi$ in the above weak formulation in Definition 4.4, and therefore distributional solutions are well defined. There is an additional subtlety coming from the fact that $F_w[\mu]$ is defined for a.e. $\xi \in [0, 1]$, which could seem problematic at first glance when integrated against μ_t if different representatives take different values over the atoms of ν . Nevertheless, in this paper we restrict to the Lebesgue measure $\nu = d\xi_{[0, 1]}$ as justified in Remark 2.21, which eliminates this issue.

Also note that in Definition 4.4 we may have weakened the time-continuity to simply Borel-measurability of the family $(\mu_t)_{t \in [0, T]} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, and the above notion of distributional solutions would still be well defined. However, it is well known that whenever a Borel family $(\mu_t)_{t \in [0, T]} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$ solves a continuity equation in distributional sense and

$$\int_0^T \iint_{\mathbb{R}^d \times [0, 1]} |F_w[\mu_t](x, \xi)| d\mu_t(x, \xi) dt < \infty, \quad (4.6)$$

which in our case it holds true by Proposition 4.1, then there must exist a time representative which is narrowly continuous, see [1, Theorem 8.2.1]. For this reason we restrict to solutions living in $C([0, T], \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]))$ without loss of generality.

Proposition 4.5 (Distributional solutions of (2.10)-(2.11)). *Consider any curve of probability measures $\mu \in C([0, T], \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]))$ -narrow, and its associated Borel family $(\mu_t^\xi)_{(t, \xi) \in [0, T] \times [0, 1]}$ as in the disintegration Theorem 2.20. Then, the following conditions are equivalent:*

- (i) μ is a distributional solution (2.10)-(2.11).
- (ii) For a.e. $\xi \in [0, 1]$, $(\mu_t^\xi)_{t \in [0, T]}$ is a distributional solution to (1.4)-(1.5), i.e.,

$$\int_0^T \int_{\mathbb{R}^d} (\partial_t \varphi(t, x) + F_w[\mu_t](t, x, \xi) \cdot \nabla_x \varphi(t, x)) d\mu_t^\xi(x) dt = - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0^\xi(x), \quad (4.7)$$

for all $\varphi \in C_c^1([0, T] \times \mathbb{R}^d)$.

- (iii) For a.e. $\xi \in [0, 1]$, $(\mu_t^\xi)_{t \in [0, T]}$ is the push forward along the flow map, i.e.,

$$\mu_t^\xi = \mathcal{T}_t^\xi[w, \mu] \# \mu_0^\xi, \quad \text{a.e. } t \geq 0, \quad (4.8)$$

where \mathcal{T}_t^ξ is given in (4.3).

Proof. The proof of the equivalence between (i) and (ii) follows from a density argument identical to the one used in the proof of the disintegration Theorem 2.20 applied to the integrals in (4.5), and then we omit it. We then focus on the proof of the equivalence between (ii) and (iii).

Assume that (ii) holds and let us fix a.e. $\xi \in [0, 1]$ so that the Borel family $(\mu_t^\xi)_{t \in [0, T]}$, the force $F_w[\mu_t](x, \xi)$ in (2.10), and its flow map $\mathcal{T}_t^\xi[w, \mu](x)$ in (4.3) are all defined. As discussed below Theorem 2.20, we do not expect $\mu^\xi \in C([0, T], \mathcal{P}_\nu(\mathbb{R}^d))$. However, note that

$$\int_0^T \int_{\mathbb{R}^d} |F_w[\mu_t](x, \xi)| d\mu_t^\xi(x) dt < \infty,$$

and therefore [1, Lemma 8.1.2] again ensures the existence of a narrowly continuous representative of the Borel family $(\mu_t^\xi)_{t \in [0, T]}$, namely $\tilde{\mu}^\xi \in C([0, T], \mathcal{P}(\mathbb{R}^d))$ -narrow such that $\mu_t^\xi = \tilde{\mu}_t^\xi$ for a.e. $t \in [0, T]$. Since the characteristic system associated to $F_w[\mu](x, \xi)$ for such a ξ has unique solutions for all $x \in \mathbb{R}^d$, then a standard argument shows that

$$\tilde{\mu}_t = \mathcal{T}_t^\xi[w, \mu] \# \tilde{\mu}_0^\xi, \quad \forall t \geq 0,$$

see [1, Proposition 8.1.8]. Note that $\tilde{\mu}^\xi$ clearly solves (4.7) for the same initial datum μ_0^ξ . Then, $\tilde{\mu}_0^\xi = \mu_0^\xi$ and therefore (4.8) holds true for a.e. $t \geq 0$.

Conversely, assume that (iii) holds true and take any $\varphi \in C_c^1([0, T] \times \mathbb{R}^d)$. Then, using (4.8) on the left hand side of (4.7) yields

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^d} (\partial_t \varphi(t, x) + F_w[\mu_t](t, x, \xi) \cdot \nabla_x \varphi(t, x)) d\mu_t^\xi(x) dt \\ &= \int_0^T \int_{\mathbb{R}^d} \left(\partial_t \varphi(t, \mathcal{T}_t^\xi[w, \mu](x)) + F_w[\mu_t](t, \mathcal{T}_t^\xi[w, \mu](x), \xi) \cdot \nabla_x \varphi(t, \mathcal{T}_t^\xi[w, \mu](x)) \right) d\mu_0^\xi(x) dt \\ &= \int_0^T \int_{\mathbb{R}^d} \frac{d}{dt} \varphi(t, \mathcal{T}_t^\xi[w, \mu](x)) d\mu_0^\xi(x) dt = \int_0^T \frac{d}{dt} \int_{\mathbb{R}^d} \varphi(t, \mathcal{T}_t^\xi[w, \mu](x)) d\mu_0^\xi(x) dt \\ &= - \int_{\mathbb{R}^d} \varphi(0, x) d\mu_0^\xi(x), \end{aligned}$$

where we have used the dominated convergence theorem, the fundamental theorem of calculus, and the fact that φ vanishes at $t = T$. \square

4.3. Fixed point argument. We show the well-posedness of the system (2.10)-(2.11) (or equivalently, (1.4)-(1.5)) through a fixed point. In view of Definition 4.4 and Proposition 4.5, solutions could be built in the more general space $C([0, T], \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1]))$. However, in Section 5 we will restrict to solutions living in either of the subspaces $C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$. For this reason, we decided to restrict to a well-posedness theory in the later subspaces. Our main concern is that a quantitative stability estimate similar to the one derived in Section 5, but operating over solutions lying in the former larger space, seems to require additional Lipschitz-continuity assumptions w_ℓ , which we do not want. We remark that this additional assumption was already present in previous literature for binary interactions [23, 64], as it stems from the classical Dobrushin stability estimate [28].

In the fixed point argument, we will employ the following additional Lipschitz-continuity property of the force $F_w[\mu]$ in the variable μ with respect to the $d_{p,\nu}$ distance.

Proposition 4.6 (Properties of the force II). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), consider any Borel family of probability measures $(\mu_t)_{t \in \mathbb{R}_+} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$, any $w = (w_\ell)_{\ell \in \mathbb{N}}$, and suppose that*

$$C_p := \left\| \sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right\|_{L_\xi^p} < \infty, \quad (4.9)$$

for some $p \in [1, \infty]$ and $q \in [1, \infty]$ so that $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$|F_w[\mu_t](x, \xi) - F_w[\bar{\mu}_t](x, \xi)| \leq \left(\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right) d_{p,\nu}(\mu_t, \bar{\mu}_t),$$

for all $t \in \mathbb{R}_+$, all $x \in \mathbb{R}^d$, a.e. $\xi \in [0, 1]$.

Proof. Note that condition (4.9) in particular implies that the above coefficient

$$\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1},$$

is finite for a.e. $\xi \in [0, 1]$. Then, we have

$$\begin{aligned} & |F_w[\mu_t](x, \xi) - F_w[\bar{\mu}_t](x, \xi)| \\ &= \sum_{\ell=1}^{\infty} \sum_{k=1}^{\ell} \int_{[0,1]^\ell} |w_\ell(\xi, \xi_\ell)| \\ & \quad \times \int_{\mathbb{R}^{d(\ell-1)}} \left| \int_{\mathbb{R}^d} K_\ell(x, \mathbf{x}_\ell) (d\mu_t^{\xi_k}(x_k) - d\bar{\mu}_t^{\xi_k}(x_k)) \right| \prod_{j < k} d\mu_t^{\xi_j}(x_j) \prod_{j > k} d\bar{\mu}_t^{\xi_j}(x_j) d\xi_\ell \\ & \leq \sum_{\ell=1}^{\infty} \sum_{k=1}^{\ell} \int_{[0,1]^\ell} |w_\ell(\xi, \xi_\ell)| \int_{\mathbb{R}^{d(\ell-1)}} BL_\ell d_{\text{BL}}(\mu_t^{\xi_k}, \bar{\mu}_t^{\xi_k}) \prod_{j < k} d\mu_t^{\xi_j}(x_j) \prod_{j > k} d\bar{\mu}_t^{\xi_j}(x_j) d\xi_\ell \\ &= \sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \int_0^1 \|w_\ell(\xi, \cdot)\|_{L_{\xi_{\ell,k}}^1} d_{\text{BL}}(\mu_t^{\xi_k}, \bar{\mu}_t^{\xi_k}) d\xi_k \\ & \leq \left(\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right) d_{p,\nu}(\mu_t, \bar{\mu}_t), \end{aligned}$$

for all $t \in [0, T]$ and all $x \in \mathbb{R}^d$, where we recall that $BL_\ell = \max\{B_\ell, L_\ell\}$ is the bounded-Lipschitz kernel of each kernel K_ℓ as in (1.8), and where in the last step we have used Hölder's inequality in the integral with respect to ξ_k for the exponents $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. \square

Theorem 4.7 (Well-posedness of the Vlasov equation). *Assume that the kernels K_ℓ verify the assumptions (1.6) and (1.7), consider any $w = (w_\ell)_{\ell \in \mathbb{N}}$, and suppose that hypothesis (4.1) and (4.9) hold for some $p \in [1, \infty]$. Then, for every $\mu_0 \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$ there exists a unique $\mu \in C(\mathbb{R}_+, \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ global-in-time distributional solution to (2.10)-(2.11) issued at μ_0 .*

Proof. We shall restrict our proof to the case $1 < p < \infty$, but clear adaptations can be done for the limiting cases $p = 1$ and $p = \infty$, and then we will omit those details.

Fix any $T > 0$ and any initial probability measure $\mu_* \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$. For any curve $\mu \in C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ we define $\mathcal{F}[\mu] \in C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ by

$$\mathcal{F}[\mu]_t^\xi := \mathcal{T}_t^\xi[w, \mu]_{\#} \mu_*^\xi, \quad t \in [0, T], \text{ a.e. } \xi \in [0, 1], \quad (4.10)$$

where \mathcal{T}_t^ξ is the flow-map (4.3). We recall that by Proposition 4.2 the characteristic system (4.2) is globally-in-time well posed for a.e. $\xi \in [0, 1]$ under the assumptions (4.1). Therefore, the family of measures $(\mathcal{F}[\mu]_t^\xi)_{t \in [0, T] \times [0, 1]}$ is defined for all $t \in [0, T]$ and a.e. $\xi \in [0, 1]$.

◇ STEP 1: We show that $\mathcal{F}[\mu] \in C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$.

First, by Proposition 4.2 we have that $\mathcal{T}_t^\xi[\mu]$ is continuous with respect to $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ and Borel-measurable with respect to $\xi \in [0, 1]$. Additionally, $(\mu_*^\xi)_{\xi \in [0, 1]}$ is also a Borel-measurable. Altogether implies that, $(\mathcal{F}[\mu]_t^\xi)_{t \in [0, T] \times [0, 1]} \subset \mathcal{P}(\mathbb{R}^d)$ is a Borel family, and then so is the family $(\mathcal{F}[\mu]_t)_{t \in [0, T]} \subset \mathcal{P}_\nu(\mathbb{R}^d \times [0, 1])$. Second, for a.e. $\xi \in [0, 1]$ and all $0 \leq t_1 \leq t_2 \leq T$ we have

$$\begin{aligned} d_{\text{BL}}(\mathcal{F}[\mu]_{t_1}^\xi, \mathcal{F}[\mu]_{t_2}^\xi) &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \phi(x) (d\mathcal{F}[\mu]_{t_1}^\xi(x) - d\mathcal{F}[\mu]_{t_2}^\xi(x)) \\ &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} (\phi(\mathcal{T}_{t_1}^\xi[w, \mu](x)) - \phi(\mathcal{T}_{t_2}^\xi[w, \mu](x))) d\mu_*^\xi(x) \\ &\leq \|\mathcal{T}_{t_1}^\xi[w, \mu] - \mathcal{T}_{t_2}^\xi[w, \mu]\|_{L^\infty} \leq \int_{t_1}^{t_2} \|F_w[\mu_t](\mathcal{T}_t^\xi[\mu], \xi)\|_{L^\infty} dt \leq B_F |t_1 - t_2|, \end{aligned}$$

where above we have used the definition of the bounded-Lipschitz metric, the fact that $\mathcal{T}_t^\xi[w, \mu]$ is the flow map of $F_w[\mu_t](\cdot, \xi)$ and the triangle inequality, along with the uniform bound of $F_w[\mu]$ by B_F in Proposition 4.1. Taking L^p -norms with respect to $\xi \in [0, 1]$ yields

$$d_{p,\nu}(\mathcal{F}[\mu]_{t_1}, \mathcal{F}[\mu]_{t_2}) \leq B_F |t_1 - t_2|,$$

for all $t_1, t_2 \in [0, T]$. Since $\mathcal{F}_0[\mu] = \mu_* \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$, the above along with the triangle inequality implies that $\mathcal{F}[\mu]_t \in \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$ for all $t \in [0, T]$ and, additionally, $\mathcal{F}[\mu] \in C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ (in fact Lipschitz-continuous in time).

◇ STEP 2: We show that \mathcal{F} is contractive for small T .

Consider any couple $\mu, \bar{\mu} \in C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$. For a.e. $\xi \in [0, 1]$ and all $t \in [0, T]$ a similar argument as above yields

$$\begin{aligned} d_{\text{BL}}(\mathcal{F}[\mu]_t^\xi, \mathcal{F}[\bar{\mu}]_t^\xi) &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \phi(x) (d\mathcal{F}[\mu]_t^\xi(x) - d\mathcal{F}[\bar{\mu}]_t^\xi(x)) \\ &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} (\phi(\mathcal{T}_t^\xi[w, \mu](x)) - \phi(\mathcal{T}_t^\xi[w, \bar{\mu}](x))) d\mu_*^\xi(x) \\ &\leq \int_{\mathbb{R}^d} |X_w[\mu](t, x, \xi) - X_w[\bar{\mu}](t, x, \xi)| d\mu_*^\xi(x). \end{aligned} \quad (4.11)$$

We then need a continuous dependence of the flow map with respect to μ . To this end, note that by definition of the characteristic system (4.2) we have

$$|X_w[\mu](t, x, \xi) - X_w[\bar{\mu}](t, x, \xi)| \leq I_1 + I_2, \quad (4.12)$$

where each factor reads

$$I_1 := \int_0^t |F_w[\mu_s](X_w[\mu](s, x, \xi), \xi) - F_w[\mu_s](X_w[\bar{\mu}](s, x, \xi), \xi)| ds, \quad (4.13)$$

$$I_2 := \int_0^t |F_w[\mu_s](X_w[\bar{\mu}](s, x, \xi), \xi) - F_w[\bar{\mu}_s](X_w[\bar{\mu}](s, x, \xi), \xi)| ds. \quad (4.14)$$

For the first term I_1 we use the Lipschitz-continuity of the force $F_w[\mu]$ with respect to x in Proposition 4.1, which implies

$$I_1 \leq L_F \int_0^t |X_w[\mu](s, x, \xi) - X_w[\bar{\mu}](s, x, \xi)| ds. \quad (4.15)$$

Regarding the second factor, we use Proposition 4.6, which applied to (4.14) yields

$$I_2 \leq \left(\sum_{\ell=1}^{\infty} BL_{\ell} \sum_{k=1}^{\ell} \|w_{\ell}(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right) \int_0^t d_{p,\nu}(\mu_s, \bar{\mu}_s) ds. \quad (4.16)$$

Plugging (4.15)-(4.16) into (4.12) implies

$$\begin{aligned} |X_w[\mu](t, x, \xi) - X_w[\bar{\mu}](t, x, \xi)| &\leq L_F \int_0^t |X_w[\mu](s, x, \xi) - X_w[\bar{\mu}](s, x, \xi)| ds \\ &\quad + \left(\sum_{\ell=1}^{\infty} BL_{\ell} \sum_{k=1}^{\ell} \|w_{\ell}(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right) \int_0^t d_{p,\nu}(\mu_s, \bar{\mu}_s) ds, \end{aligned}$$

for all $t \in [0, T]$. Since $X_w[\mu](0, x, \xi) = x = X_w[\bar{\mu}](0, x, \xi)$, then Grönwall's lemma implies

$$|X_w[\mu](t, x, \xi) - X_w[\bar{\mu}](t, x, \xi)| \leq \left(\sum_{\ell=1}^{\infty} BL_{\ell} \sum_{k=1}^{\ell} \|w_{\ell}(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell,k}}^1} \right) \int_0^t e^{L_F(t-s)} d_{p,\nu}(\mu_s, \bar{\mu}_s) ds, \quad (4.17)$$

for all $t \in [0, T]$ all $x \in \mathbb{R}^d$ and a.e. $\xi \in [0, 1]$. Integrating (4.17) with respect to $\mu_*^{\xi}(x)$ in the variable x , plugging it in (4.11), and taking L^p norms with respect to $\xi \in [0, 1]$ yields

$$d_{p,\nu}(\mathcal{F}[\mu]_t, \mathcal{F}[\bar{\mu}]_t) \leq C_p \int_0^t e^{L_F(t-s)} d_{p,\nu}(\mu_s, \bar{\mu}_s) ds,$$

where the constant C_p is defined in (4.9). Taking the uniform norm with respect to t implies

$$\sup_{t \in [0, T]} d_{p,\nu}(\mathcal{F}[\mu]_t, \mathcal{F}[\bar{\mu}]_t) \leq \frac{C_p}{L_F} (e^{L_F T} - 1) \sup_{t \in [0, T]} d_{p,\nu}(\mu_t, \bar{\mu}_t). \quad (4.18)$$

◇ STEP 3: Global-in-time well posedness.

Note that the above ensures that $\mathcal{F} : C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])) \rightarrow C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ is well defined and contractive as long as

$$T < \frac{1}{L_F} \log \left(1 + \frac{L_F}{C_p} \right).$$

Under this condition, by the Banach contraction principle we have that the operator \mathcal{F} has a unique fixed point in $C([0, T], \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$. Bearing in mind the particular form of \mathcal{F} in (4.10) and the characterization of distributional solutions in Proposition 4.5, we have that the fixed point corresponds to the unique distributional solution to (2.10)-(2.11) issued at μ_* defined in the interval $[0, T]$. Iterating the construction on the intervals $[kT, (k+1)T]$ for all $k \in \mathbb{N}$, which can be done because the lifespan $[0, T]$ of the above local-in-time solution does not depend on the initial datum μ_* , one obtains the unique global-in-time distributional solution. \square

Remark 4.8. *We remark that the scaling conditions on the kernels K_{ℓ}*

$$\sum_{\ell=1}^{\infty} \ell BL_{\ell} < \infty, \quad (4.19)$$

which is weaker than (1.9) but stronger than (4.4), together with the scaling condition

$$\sup_{\ell \in \mathbb{N}} \|w_{\ell}\|_{L^{\infty}} \leq W,$$

(e.g., $w \in \mathcal{H}_W$ by Definition 2.5) ensure the above hypothesis (4.1) and (4.9) since we have

$$B_F \leq W \sum_{\ell=1}^{\infty} B_{\ell}, \quad L_F \leq W \sum_{\ell=1}^{\infty} L_{\ell}, \quad C_p \leq W \sum_{\ell=1}^{\infty} \ell B L_{\ell}.$$

Our remark that Theorem 4.7 does not only operate over absolutely continuous initial measures, but also over initial measures μ_0 having an atomic part leads to the following corollary for initial measures supported on the graph of a function $X_0 : [0, 1] \rightarrow \mathbb{R}^d$.

Corollary 4.9 (Well-posedness of the continuum-limit equation). *Under the hypothesis in Theorem 4.7, assume that $X_0 \in L^p([0, 1], \mathbb{R}^d)$ and set $\mu_0^{\xi} = \delta_{X_0(\xi)}$. Then, the unique solution μ to the Vlasov equation (2.10)-(2.11) must have the form $\mu_t^{\xi} = \delta_{X(t, \xi)}$ where $X \in C(\mathbb{R}_+, L^p([0, 1], \mathbb{R}^d))$ is the unique solution to the continuum-limit equation*

$$\begin{cases} \partial_t X(t, \xi) = \sum_{\ell=1}^{\infty} \int_{[0, 1]^{\ell+1}} w_{\ell}(\xi, \xi_1, \dots, \xi_{\ell}) K_{\ell}(X(t, \xi), X(t, \xi_1), \dots, X(t, \xi_{\ell})) d\xi_1 \dots d\xi_{\ell}, \\ X(0, \cdot) = X_0. \end{cases} \quad (4.20)$$

Equation (4.20) can be regarded as the natural higher-order extension of the version with binary interactions introduced in [58] in the context of non-linear heat equations on dense graphs. We also refer to [2, 3, 32, 64] for further extensions based on binary interactions. A higher-order extension similar to (4.20) was recently introduced in [15] to study bifurcation and stability properties of twisted states for the Kuramoto model restricted to ternary interactions.

5. STABILITY ESTIMATE OF THE VLASOV EQUATION OVER UR-HYPERGRAPHS

In this section we study the stability of the solutions to the Vlasov equation (2.10)-(2.11) with respect to the initial datum $\mu_0 \in \mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1])$ and the involved UR-hypergraphon $w = (w_{\ell})_{\ell \in \mathbb{N}}$. As in Section 4, we shall not restrict to UR-hypergraphons only, but actually the stability estimate will be applicable to a broader class of w .

Theorem 5.1 (Stability estimate for the Vlasov equation). *Assume that the kernels K_{ℓ} verify the assumptions (1.6) and (1.7), and also that $K_{\ell} \in L^1(\mathbb{R}^{d(\ell+1)})$ for all $\ell \in \mathbb{N}$. Consider $w = (w_{\ell})_{\ell \in \mathbb{N}}$ and $\bar{w} = (\bar{w}_{\ell})_{\ell \in \mathbb{N}}$ satisfying the assumptions (4.1) and (4.9) for common constants $B_F, L_F, C_p > 0$, and additionally suppose that*

$$\sum_{\ell=1}^{\infty} 4^{\ell} \|\hat{K}_{\ell}\|_{L^1} < \infty, \quad D_{\infty} := \sum_{\ell=1}^{\infty} 2^{\ell} (\|w_{\ell}\|_{L^{\infty}} + \|\bar{w}_{\ell}\|_{L^{\infty}}) \|\hat{K}_{\ell}\|_{L^1} < \infty, \quad (5.1)$$

For any initial data $\mu_0, \bar{\mu}_0 \in \mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1])$ with $p \in [1, \infty)$, let $\mu, \bar{\mu} \in C(\mathbb{R}_+, \mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1]))$ be the unique global-in-time distributional solutions to (2.10)-(2.11) issued at μ_0 with given w (respectively, $\bar{\mu}_0$ and \bar{w}) as in Theorem 4.7. Then, we have

$$d_{p, \nu}(\mu_t, \bar{\mu}_t) \leq e^{(C_p + L_F)t} \left(d_{p, \nu}(\mu_0, \bar{\mu}_0) + \frac{D_{\infty}^{1/q}}{L_F} d_{\square}(w, \bar{w}; (4^{\ell} \|\hat{K}_{\ell}\|_{L^1})_{\ell \in \mathbb{N}})^{1/p} \right),$$

for every $t \geq 0$, where $\frac{1}{p} + \frac{1}{q} = 1$ and d_{\square} is the labeled cut distance in Definition 2.5.

We note that under the hypothesis (5.1), the sequence $(4^{\ell} \|\hat{K}_{\ell}\|)_{\ell \in \mathbb{N}}$ is summable and therefore the labeled cut distance $d_{\square}(w, \bar{w}; (4^{\ell} \|\hat{K}_{\ell}\|_{L^1})_{\ell \in \mathbb{N}})$ in the stability estimate is well defined. Whilst above $w = (w_{\ell})_{\ell \in \mathbb{N}}$ do not restrict to UR-hypergraphons only, we anticipate that in order for the stability estimate to be useful in the proof of the main Theorem 1.1 in Section 6 we will have to restrict to UR-hypergraphons. More particularly, from the results in the hypergraph limit theories in Section 2.2 we can only guarantee the decay of the cut distance in the class \mathcal{H}_W of UR-hypergraphons as described in Proposition 2.7.

Remark 5.2 (On the Sobolev regularity assumption). *We remark that the regularity-scaling conditions (1.10) on the kernels K_ℓ , and the scaling condition on the weights w_ℓ*

$$\sup_{\ell \in \mathbb{N}} \|w_\ell\|_{L^\infty} \leq W,$$

(i.e., $w \in \mathcal{H}_W$ by Definition 2.5) ensure the above hypothesis (5.1) since we have

$$D_\infty \leq W \sum_{\ell=1}^{\infty} 4^\ell \|\hat{K}_\ell\|_{L^1} \lesssim \frac{4^\ell \pi^{\frac{d\ell}{4}}}{\sqrt{\Gamma(\frac{d\ell}{2})}} \|K_\ell\|_{H^{\frac{d(\ell+1)}{2} + \varepsilon}}.$$

Specifically, note that for any $\ell \in \mathbb{N}$ and setting $k = \frac{d(\ell+1)}{2} + \varepsilon$ for some $\varepsilon > 0$ arbitrarily small, the Cauchy-Schwarz inequality implies $\|\hat{K}_\ell\|_{L^1} \leq \alpha_\ell^{1/2} \beta_\ell^{1/2}$, where

$$\begin{aligned} \alpha_\ell &:= \int_{\mathbb{R}^{d(\ell+1)}} \frac{1}{(1 + |z|^2 + |\mathbf{z}_\ell|^2)^k} dz d\mathbf{z}_\ell, \\ \beta_\ell &:= \int_{\mathbb{R}^{d(\ell+1)}} (1 + |z|^2 + |\mathbf{z}_\ell|^2)^k |\hat{K}_\ell(z, \mathbf{z}_\ell)|^2 dz d\mathbf{z}_\ell. \end{aligned}$$

On the one hand, the first integral is finite because $2k > d(\ell+1)$ by our choice of k . In fact, it can be calculated explicitly in polar coordinated leading to

$$\alpha_\ell = \omega_{d(\ell+1)-1} \int_0^\infty \frac{r^{d(\ell+1)-1}}{(1+r^2)^k} dr = \frac{\omega_{d(\ell+1)-1}}{2} B\left(\frac{d(\ell+1)}{2}, \frac{\varepsilon}{2}\right),$$

where $\omega_{d(\ell+1)-1} = \frac{2\pi^{d(\ell+1)/2}}{\Gamma(d(\ell+1)/2)}$ is the area of the sphere of $\mathbb{R}^{d(\ell+1)}$ and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^\infty \frac{s^{a-1}}{(1+s)^{a+b}} ds$ is the Beta function. Therefore,

$$\alpha_\ell = \frac{\pi^{\frac{d(\ell+1)}{2}} \Gamma(\frac{\varepsilon}{2})}{\Gamma(\frac{d(\ell+1)}{2} + \varepsilon)} \lesssim \frac{\pi^{\frac{d\ell}{2}}}{\Gamma(\frac{d\ell}{2})}.$$

On the other hand, by definition of the fractional Sobolev through Bessel potentials we have

$$\beta_\ell^{1/2} = \|\hat{K}_\ell\|_{H^k} = \|\hat{K}_\ell\|_{H^{\frac{d(\ell+1)}{2} + \varepsilon}}.$$

Proof of Theorem 5.1. We restrict our proof to the case $p > 1$ but clear adaptations can be done to prove the case $p = 1$.

◇ STEP 1: We obtain a first bound for the bounded-Lipschitz distance between μ_t^ξ and $\bar{\mu}_t^\xi$. By item (iii) of Proposition 4.5 we know that for a.e. $\xi \in [0, 1]$ the distributional solutions μ and $\bar{\mu}$ must satisfy

$$\mu_t^\xi = \mathcal{T}_t^\xi[w, \mu] \# \mu_0^\xi, \quad \bar{\mu}_t^\xi = \mathcal{T}_t^\xi[\bar{w}, \bar{\mu}] \# \bar{\mu}_0^\xi,$$

for a.e. $t \in \mathbb{R}_+$, and in fact, the time-continuous representatives obtained in Theorem 4.7 satisfy the above for all $t \in \mathbb{R}_+$. Therefore, we have

$$\begin{aligned} d_{\text{BL}}(\mu_t^\xi, \bar{\mu}_t^\xi) &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \phi(\mathcal{T}_t^\xi[w, \mu](x)) d\mu_0^\xi(x) - \int_{\mathbb{R}^d} \phi(\mathcal{T}_t^\xi[\bar{w}, \bar{\mu}](x)) d\bar{\mu}_0^\xi(x) \\ &\leq \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \left(\phi(\mathcal{T}_t^\xi[w, \mu](x)) - \phi(\mathcal{T}_t^\xi[\bar{w}, \bar{\mu}](x)) \right) d\mu_0^\xi(x) \\ &\quad + \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \phi(\mathcal{T}_t^\xi[\bar{w}, \bar{\mu}](x)) d(\mu_0^\xi - \bar{\mu}_0^\xi)(x) \\ &\leq \int_{\mathbb{R}^d} |X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)| d\mu_0^\xi(x) + e^{L_F t} d_{\text{BL}}(\mu_0^\xi, \bar{\mu}_0^\xi), \end{aligned} \tag{5.2}$$

where in the last line we have observed that since $\phi \in \text{BL}_1(\mathbb{R}^d)$, for all $t \in \mathbb{R}_+$ and a.e. $\xi \in [0, 1]$ the maps $x \in \mathbb{R}^d \mapsto \phi(\mathcal{T}_t^\xi[w, \mu](x))$ are bounded and Lipschitz-continuous with BL norm bounded by $e^{L_F t}$ thanks to item (iii) in Proposition 4.2.

◇ STEP 2: We control the difference $|X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)|$.

This step is similar to STEP 2 in the proof of Theorem 4.7, where a similar estimate was obtained in (4.17) in the particular case that $w = \bar{w}$ and $\mu_0 = \bar{\mu}_0$. Since this is the key step of the proof, which involves a certain continuity of the flow map with respect to the underlying w and \bar{w} , we provide a proof. For all $t \in [0, T]$ and a.e. $\xi \in [0, 1]$, we get

$$\begin{aligned} & |X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)| \\ &= \left| \int_0^t (F_w[\mu_s](X_w[\mu](s, x, \xi), \xi) - F_{\bar{w}}[\bar{\mu}_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi)) ds \right| \\ &\leq \int_0^t |F_w[\mu_s](X_w[\mu](s, x, \xi), \xi) - F_w[\mu_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi)| ds \\ &\quad + \int_0^t |F_w[\mu_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi) - F_{\bar{w}}[\bar{\mu}_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi)| ds \\ &=: I_3 + I_4. \end{aligned} \tag{5.3}$$

Regarding the first term I_3 , the Lipschitz continuity property of the force $F_w[\mu_s](x, \xi)$ with respect to x in item (ii) of Proposition 4.1 implies

$$I_3 \leq L_F \int_0^t |X_w[\mu](s, x, \xi) - X_{\bar{w}}[\bar{\mu}](s, x, \xi)| ds. \tag{5.4}$$

Regarding the second term I_4 , we can rewrite it as $I_4 = I_{41} + I_{42}$, where

$$\begin{aligned} I_{41} &:= \int_0^t |F_w[\mu_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi) - F_w[\bar{\mu}_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi)| ds, \\ I_{42} &:= \int_0^t |F_w[\bar{\mu}_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi) - F_{\bar{w}}[\bar{\mu}_s](X_{\bar{w}}[\bar{\mu}](s, x, \xi), \xi)| ds. \end{aligned}$$

From Proposition 4.6, we know that

$$I_{41} \leq \int_0^t \left(\sum_{\ell=1}^{\infty} BL_{\ell} \sum_{k=1}^{\ell} \|w_{\ell}(\xi, \cdot)\|_{L_{\xi_k}^q L_{\xi_{\ell, k}}^1} \right) d_{p, \nu}(\mu_s, \bar{\mu}_s) ds. \tag{5.5}$$

Besides, using the expression (1.5) of the forces $F_w[\bar{\mu}_s]$ and $F_{\bar{w}}[\bar{\mu}_s]$ we get

$$\begin{aligned} I_{42} &\leq \int_0^t \sum_{\ell=1}^{\infty} \left| \int_{[0, 1]^{\ell}} (w_{\ell}(\xi, \xi_1, \dots, \xi_{\ell}) - \bar{w}_{\ell}(\xi, \xi_1, \dots, \xi_{\ell})) \right. \\ &\quad \left. \times \left(\int_{\mathbb{R}^{d\ell}} K_{\ell}(X_{\bar{w}}[\bar{\mu}](s, x, \xi), x_1, \dots, x_{\ell}) d\bar{\mu}_s^{\xi_1}(x_1) \cdots d\bar{\mu}_s^{\xi_{\ell}}(x_{\ell}) \right) d\xi_1 \dots d\xi_{\ell} \right| ds. \end{aligned} \tag{5.6}$$

Each of the integrals on $[0, 1]^{\ell}$ is reminiscent of the action of the ℓ -th order adjacency operator $T^{w_{\ell} - \bar{w}_{\ell}}$ over the function

$$(\xi_1, \dots, \xi_{\ell}) \in [0, 1]^{\ell} \mapsto \int_{\mathbb{R}^{d\ell}} K_{\ell}(X_{\bar{w}}[\bar{\mu}](s, x, \xi), x_1, \dots, x_{\ell}) d\bar{\mu}_s^{\xi_1}(x_1) \cdots d\bar{\mu}_s^{\xi_{\ell}}(x_{\ell}).$$

However, the later does not have separate variables and therefore we cannot readily invoke the ℓ -th order cut distance. To this end, inspired by [13] (where the binary case $\ell = 1$ on a one-dimensional periodic domain was considered) or also [29, 62], we write K_{ℓ} as a superposition of functions with separate variables via the Fourier inversion formula

$$K_{\ell}(x, x_1, \dots, x_{\ell}) = \int_{\mathbb{R}^{d(\ell+1)}} e^{2\pi i(x \cdot z + x_1 \cdot z_1 + \dots + x_{\ell} \cdot z_{\ell})} \hat{K}_{\ell}(z, z_1, \dots, z_{\ell}) dz dz_1 \dots dz_{\ell},$$

with \hat{K}_{ℓ} the Fourier transform. We note that \hat{K}_{ℓ} are all well defined because $K_{\ell} \in L^1(\mathbb{R}^{d(\ell+1)})$ by our assumptions, and also the inversion formula holds with an absolutely convergent integral because we are further assuming that $\hat{K}_{\ell} \in L^1(\mathbb{R}^{d(\ell+1)})$. Defining the function

$$g(s, z, \xi) := \int_{\mathbb{R}^d} e^{2\pi i x \cdot z} d\bar{\mu}_s^{\xi}(x), \quad s \in \mathbb{R}_+, z \in \mathbb{R}^d, \xi \in [0, 1],$$

and since $|e^{2\pi i X_{\bar{\mu}}[s, x, \xi]}| = 1$, we can reformulate the term I_{42} as follows

$$\begin{aligned} I_{42} &\leq \int_0^t \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \int_{[0,1]^\ell} (w_\ell(\xi, \boldsymbol{\xi}_\ell) - \bar{w}_\ell(\xi, \boldsymbol{\xi}_\ell)) \prod_{k=1}^{\ell} g(s, z_k, \xi_k) d\boldsymbol{\xi}_\ell \right| dz d\mathbf{z}_\ell ds \\ &= \int_0^t \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)](\xi) \right| dz d\mathbf{z}_\ell ds. \end{aligned} \quad (5.7)$$

Above, and since g is complex-valued, we have extended the definition of the adjacency operator (2.1) to complex valued-functions as follows, $\tilde{T}^{w_\ell - \bar{w}_\ell} : L^\infty([0, 1], \mathbb{C})^\ell \rightarrow L^1([0, 1], \mathbb{C})$ such that

$$\tilde{T}^{w_\ell - \bar{w}_\ell} [g_1, \dots, g_\ell] = \int_{[0,1]^\ell} (w_\ell(\xi, \boldsymbol{\xi}_\ell) - \bar{w}_\ell(\xi, \boldsymbol{\xi}_\ell)) g_1(\xi_1) \cdots g_\ell(\xi_\ell) d\xi d\boldsymbol{\xi}_\ell,$$

for all $g_1, \dots, g_\ell \in L^\infty([0, 1], \mathbb{C})$. Thus, putting (5.4), (5.5) and (5.7) into (5.3) we get

$$\begin{aligned} &|X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)| \\ &\leq L_F \int_0^t |X_w[\mu](s, x, \xi) - X_{\bar{w}}[\bar{\mu}](s, x, \xi)| ds \\ &\quad + \int_0^t \left(\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\boldsymbol{\xi}_{\ell,k}}^1} \right) d_{p,\nu}(\mu_s, \bar{\mu}_s) ds \\ &\quad + \int_0^t \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)](\xi) \right| dz d\mathbf{z}_\ell ds. \end{aligned} \quad (5.8)$$

◇ STEP 3: We deduce a bound on the integrated difference.

We integrate the above inequality (5.8) with respect to $\mu_0^\xi(x)$. Thus, we get

$$\begin{aligned} &\int_{\mathbb{R}^d} |X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)| d\mu_0^\xi(x) \\ &\leq L_F \int_0^t \int_{\mathbb{R}^d} |X_w[\mu](s, x, \xi) - X_{\bar{w}}[\bar{\mu}](s, x, \xi)| d\mu_0^\xi(x) ds \\ &\quad + \int_0^t \left(\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\boldsymbol{\xi}_{\ell,k}}^1} \right) d_{p,\nu}(\mu_s, \bar{\mu}_s) ds \\ &\quad + \int_0^t \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)](\xi) \right| dz d\mathbf{z}_\ell ds. \end{aligned}$$

By applying Grönwall's lemma, we obtain

$$\begin{aligned} &\int_{\mathbb{R}^d} |X_w[\mu](t, x, \xi) - X_{\bar{w}}[\bar{\mu}](t, x, \xi)| d\mu_0^\xi(x) \\ &\leq \int_0^t e^{L_F(t-s)} \left\{ \left(\sum_{\ell=1}^{\infty} BL_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L_{\xi_k}^q L_{\boldsymbol{\xi}_{\ell,k}}^1} \right) d_{p,\nu}(\mu_s, \bar{\mu}_s) \right. \\ &\quad \left. + \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)](\xi) \right| dz d\mathbf{z}_\ell \right\} ds. \end{aligned} \quad (5.9)$$

◇ STEP 4: We obtain a bound for the operator norm of $\tilde{T}^{w_\ell - \bar{w}_\ell}$.

We start by obtaining a bound for the operator norm of the real-valued operators $T^{w_\ell - \bar{w}_\ell}$. By the assumption (5.1) on w and \bar{w} we have that $T^{w_\ell - \bar{w}_\ell}$ is a bounded multilinear operator from $L^\infty([0, 1])^\ell$ to $L^1([0, 1])$, but also from $L^\infty([0, 1])^\ell$ to $L^\infty([0, 1])$. By interpolation, $T^{w_\ell - \bar{w}_\ell}$ is a bounded multilinear operator from $L^\infty([0, 1])^\ell$ to any $L^p([0, 1])$. More precisely, we have

$$\|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell]\|_{L_\xi^p} \leq \|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell]\|_{L_\xi^\infty}^{1/q} \|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell]\|_{L_\xi^1}^{1/p},$$

for any $\psi_1, \dots, \psi_\ell \in L^\infty([0, 1])$. Besides,

$$\begin{aligned} & \|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell] \|_{L^\infty_\xi} \\ &= \operatorname{ess\,sup}_{\xi \in [0, 1]} \left| \int_{[0, 1]^\ell} (w_\ell(\xi, \xi_1, \dots, \xi_\ell) - \bar{w}_\ell(\xi, \xi_1, \dots, \xi_\ell)) \psi_1(\xi_1) \cdots \psi_\ell(\xi_\ell) d\xi_1 \dots d\xi_\ell \right| \\ &\leq (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty}) \|\psi_1\|_{L^\infty} \cdots \|\psi_\ell\|_{L^\infty}, \end{aligned}$$

and also

$$\begin{aligned} & \|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell] \|_{L^1_\xi} \\ &= \int_{[0, 1]} \left| \int_{[0, 1]^\ell} (w_\ell(\xi, \xi_1, \dots, \xi_\ell) - \bar{w}_\ell(\xi, \xi_1, \dots, \xi_\ell)) \psi_1(\xi_1) \cdots \psi_\ell(\xi_\ell) d\xi_1 \dots d\xi_\ell \right| d\xi \\ &\leq \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1} \|\psi_1\|_{L^\infty} \cdots \|\psi_\ell\|_{L^\infty}, \end{aligned}$$

by definition of the operator norm. Thus, we deduce that

$$\|T^{w_\ell - \bar{w}_\ell} [\psi_1, \dots, \psi_\ell] \|_{L^p_\xi} \leq (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} \|\psi_1\|_{L^\infty} \cdots \|\psi_\ell\|_{L^\infty}, \quad (5.10)$$

for every $\psi_1, \dots, \psi_\ell \in L^\infty([0, 1])$.

Now, we move to the bound of $\tilde{T}^{w_\ell - \bar{w}_\ell}$. Given $g_1, \dots, g_\ell \in L^\infty([0, 1], \mathbb{C})$, we write $g_k = \varphi_k + i\psi_k$ for real and imaginary parts $\varphi_k, \psi_k \in L^\infty([0, 1])$. Expanding the products $g_1(\xi_1) \cdots g_\ell(\xi_\ell)$ in $\tilde{T}^{w_\ell - \bar{w}_\ell}[g_1, \dots, g_\ell]$ into real and imaginary parts through some combinatorics, taking L^p norm, using the triangly inequality and applying (5.10) on each term we have

$$\begin{aligned} & \|\tilde{T}^{w_\ell - \bar{w}_\ell}[g_1, \dots, g_\ell] \|_{L^p_\xi} \leq (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} \\ & \times \left(\sum_{k=0}^{\lfloor \ell/2 \rfloor} \sum_{\substack{I \subset [1, \ell] \\ \#I=2k}} \prod_{j \notin I} \|\varphi_j\|_{L^\infty} \prod_{j \in I} \|\psi_j\|_{L^\infty} + \sum_{k=0}^{\lfloor (\ell-1)/2 \rfloor} \sum_{\substack{I \subset [1, \ell] \\ \#I=2k+1}} \prod_{j \notin I} \|\varphi_j\|_{L^\infty} \prod_{j \in I} \|\psi_j\|_{L^\infty} \right). \end{aligned}$$

In particular, if $\|g_k\|_{L^\infty} \leq 1$ then also $\|\varphi_k\|_{L^\infty} \leq 1$ and $\|\psi_k\|_{L^\infty} \leq 1$, which implies

$$\begin{aligned} & \|\tilde{T}^{w_\ell - \bar{w}_\ell}[g_1, \dots, g_\ell] \|_{L^p_\xi} \leq (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} \\ & \times \left(\sum_{k=0}^{\lfloor \ell/2 \rfloor} \binom{\ell}{2k} + \sum_{k=0}^{\lfloor (\ell-1)/2 \rfloor} \binom{\ell}{2k+1} \right). \end{aligned}$$

Since the above sums of binomial coefficients equals 2^ℓ then we have

$$\|\tilde{T}^{w_\ell - \bar{w}_\ell}[g_1, \dots, g_\ell] \|_{L^p_\xi} \leq 2^\ell (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p}, \quad (5.11)$$

for all $g_1, \dots, g_\ell \in L^\infty([0, 1], \mathbb{C})$ with $\|g_k\|_{L^\infty} \leq 1$.

◇ STEP 5: Stability estimate.

Plugging (5.9) into (5.2) implies

$$\begin{aligned} d_{\text{BL}}(\mu_t^\xi, \bar{\mu}_t^\xi) &\leq e^{L_F t} d_{\text{BL}}(\mu_0^\xi, \bar{\mu}_0^\xi) \\ &+ \int_0^t e^{L_F(t-s)} \left\{ \left(\sum_{\ell=1}^{\infty} B L_\ell \sum_{k=1}^{\ell} \|w_\ell(\xi, \cdot)\|_{L^q_{\xi_k} L^1_{\xi_{\ell, k}}} \right) d_{p, \nu}(\mu_s, \bar{\mu}_s) \right. \\ &\quad \left. + \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left| \tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)] (\xi) \right| dz d\mathbf{z}_\ell \right\} ds. \end{aligned}$$

Taking L^p norms with respect to ξ and by Minkowski's integral inequality, we deduce

$$d_{p, \nu}(\mu_t, \bar{\mu}_t) \leq e^{L_F t} d_{p, \nu}(\mu_0, \bar{\mu}_0)$$

$$\begin{aligned}
& + \int_0^t e^{L_F(t-s)} \left\{ C_p d_{p,\nu}(\mu_s, \bar{\mu}_s) \right. \\
& \quad \left. + \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d(\ell+1)}} |\hat{K}_\ell(z, \mathbf{z}_\ell)| \left\| |\tilde{T}^{w_\ell - \bar{w}_\ell} [g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)] \right\|_{L_\xi^p} dz d\mathbf{z}_\ell \right\} ds \\
& \leq e^{L_F t} d_{p,\nu}(\mu_0, \bar{\mu}_0) \\
& \quad + \int_0^t e^{L_F(t-s)} \left\{ C_p d_{p,\nu}(\mu_s, \bar{\mu}_s) + \sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \right\} ds.
\end{aligned}$$

where in the last inequality we have used the interpolation inequality (5.11) with test functions $g_k(\xi) = g(s, z_k, \xi)$ for $k = 1, \dots, \ell$, which belong to $L^\infty([0, 1], \mathbb{C})$ and have norm bounded by 1.

We multiply both sides by $e^{-L_F t}$ and get

$$\begin{aligned}
e^{-L_F t} d_{p,\nu}(\mu_t, \bar{\mu}_t) & \leq d_{p,\nu}(\mu_0, \bar{\mu}_0) \\
& \quad + \int_0^t C_p e^{-L_F s} d_{p,\nu}(\mu_s, \bar{\mu}_s) ds \\
& \quad + \left(\int_0^t e^{-L_F s} ds \right) \sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \\
& \leq d_{p,\nu}(\mu_0, \bar{\mu}_0) + \frac{1}{L_F} \sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \\
& \quad + \int_0^t C_p e^{-L_F s} d_{p,\nu}(\mu_s, \bar{\mu}_s) ds.
\end{aligned}$$

We finally apply Grönwall's lemma and obtain

$$\begin{aligned}
e^{-L_F t} d_{p,\nu}(\mu_t, \bar{\mu}_t) & \leq \left(d_{p,\nu}(\mu_0, \bar{\mu}_0) + \frac{1}{L_F} \sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1}^{1/p} (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty})^{1/q} \right) e^{C_p t}.
\end{aligned}$$

Multiplying the above by $e^{L_F t}$ and using Hölder's inequality on the sum over $\ell \in \mathbb{N}$ we deduce

$$\begin{aligned}
d_{p,\nu}(\mu_t, \bar{\mu}_t) & \leq e^{(C_p + L_F)t} \left\{ d_{p,\nu}(\mu_0, \bar{\mu}_0) \right. \\
& \quad \left. + \frac{1}{L_F} \left(\sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty}) \right)^{1/q} \left(\sum_{\ell=1}^{\infty} 2^\ell \|\hat{K}_\ell\|_{L^1} \|T^{w_\ell - \bar{w}_\ell}\|_{(L^\infty)^\ell \rightarrow L^1} \right)^{1/p} \right\},
\end{aligned}$$

for all $t \geq 0$, which ends the proof by the definition of labeled cut distance $d_\square(w, \bar{w}; (4^\ell \|\hat{K}_\ell\|_{L^1})_{\ell \in \mathbb{N}})$ in Definition 2.5, its relation to the multilinear operator norms in Proposition 2.6, and by definition of the constant D_∞ . \square

Remark 5.3 (On the order-dependent regularity). *We note that assumption (5.1) is imposing as it is order-dependent, and therefore the larger $\ell \in \mathbb{N}$, the stronger the decay we need on \hat{K}_ℓ , or alternatively the smoother K_ℓ (cf. Remark 5.2). This is due to the high generality of the kernels $K_\ell(x, x_1, \dots, x_\ell)$, which have no special structure. However, for special kernels like*

$$K_\ell(x, x_1, \dots, x_\ell) = G_\ell \left(x - \frac{x_1 + \dots + x_\ell}{\ell} \right), \quad (x, x_1, \dots, x_\ell) \in \mathbb{R}^{d(\ell+1)},$$

we could have followed an alternative approach for the separation of variables argument in STEP 2 of the proof of Theorem 5.1. Specifically, as long as $G_\ell \in L^1$ and also $\hat{G}_\ell \in L^1$ (e.g. by

assuming that $G_\ell \in H^{\frac{d}{2}+\varepsilon}$) the Fourier inversion formula of G_ℓ allows writing

$$\begin{aligned} & \int_{\mathbb{R}^{d\ell}} K_\ell(X_{\bar{w}}[\bar{\mu}](s, x, \xi), x_1, \dots, x_\ell) d\bar{\mu}_s^{\xi_1}(x_1) \cdots d\bar{\mu}_s^{\xi_\ell}(x_\ell) \\ &= \int_{\mathbb{R}^d} e^{2\pi i X_{\bar{w}}[\bar{\mu}](s, x, \xi) \cdot z} g(s, z, \xi_1) \cdots g(s, z, \xi_\ell) \hat{G}_\ell(z) dz, \end{aligned}$$

and the above follows with $\tilde{T}_{w_\ell - \bar{w}_\ell}[g(s, z_1, \cdot), \dots, g(s, z_\ell, \cdot)]$ replaced by $\tilde{T}_{w_\ell - \bar{w}_\ell}[g(s, z, \cdot), \dots, g(s, z, \cdot)]$.

Remark 5.4 (On kernels in the Wiener algebra). *The proof of Theorem 5.1 actually does not require that $\hat{K}_\ell \in L^1(\mathbb{R}^{d(\ell+1)})$, but rather that K_ℓ is the Fourier transform of a finite Borel measure. This condition still guarantees the good separation of variables which we exploited above. Specifically, it is enough if $K_\ell \in A(\mathbb{R}^{d(\ell+1)})$, where $A(\mathbb{R}^d)$ represents the Wiener algebra of absolutely convergent Fourier integrals of complex finite Borel measures, that is,*

$$A(\mathbb{R}^d) := \left\{ f : f(x) = \int_{\mathbb{R}^{d(\ell+1)}} e^{-2\pi i x \cdot y} d\mu(y) \text{ for all } x \in \mathbb{R}^d, \mu \in \mathcal{M}(\mathbb{R}^d) \right\},$$

with norm $\|f\|_A := \|\mu\|_{\text{TV}}$, see [51] for further details. In this case, (5.1) can be replaced by

$$\sum_{\ell=1}^{\infty} 4^\ell \|K_\ell\|_A < \infty, \quad \sum_{\ell=1}^{\infty} 2^\ell (\|w_\ell\|_{L^\infty} + \|\bar{w}_\ell\|_{L^\infty}) \|K_\ell\|_A < \infty,$$

This condition relaxes the integrability assumptions of K_ℓ and some of its derivatives as we observe for example in the following radial function

$$f(x) = 1 - \frac{1}{(1 + 4\pi^2|x|^2)^{\alpha/2}}, \quad x \in \mathbb{R}^d,$$

for any $0 < \alpha < \frac{d}{2}$. Specifically, $f \in A(\mathbb{R}^d)$ with associated finite Borel measure

$$\mu(x) = \delta_0(x) + G_\alpha(x) dx,$$

where G_α is the Bessel kernel

$$G_\alpha(x) = \frac{1}{(4\pi)^{\alpha/2} \Gamma(\frac{\alpha}{2})} \int_0^\infty \frac{e^{-\frac{\pi|x|^2}{r} - \frac{r}{4\pi}}}{r^{\frac{d-\alpha}{2}+1}} dr, \quad x \in \mathbb{R}^d,$$

see [70, Chapter 5, Section 2, Proposition 2]. However, $f \notin H^{\frac{d}{2}+\varepsilon}(\mathbb{R}^d)$ since it does not even decay at infinity. Indeed $f \notin L^2(\mathbb{R}^d)$ and more generally $f \notin \dot{H}^s(\mathbb{R}^d)$ unless $s > \frac{d}{2} - \alpha$.

Remark 5.5 (On the necessity of $p < \infty$). *We also note that the case $p = \infty$ has been excluded in Theorem 5.1. Whilst the same argument works, the dependency on the labeled cut distance $d_\square(w_\ell, \bar{w}_\ell; (4^\ell \|\hat{K}_\ell\|_{L^1})_{\ell \in \mathbb{N}})$ is lost (as it appears raised to the power $1/p$), meaning that Theorem 5.1 is no longer able to quantify stability with respect to the involved w in the cut distance.*

As discussed in Section 2.1.3, other analogous stability estimates studied thus far both for the case of binary [46], and higher-order interactions [47] rely on L^∞ estimates, which explains why stability of the Vlasov equation with respect to the graphon in the cut distance (a natural topology in graph limits theory) has not been much explored except for probably in the recent paper [13] and also [39, 40], but rather with respect to stronger distances.

Slightly changing how we control the factor (5.6) in STEP 2, we may still produce an L^∞ version of the stability estimate involving an $L_\xi^\infty L_\xi^1$ norm on $w_\ell - \bar{w}_\ell$

$$d_{\infty, \nu}(\mu_t, \bar{\mu}_t) \leq e^{(C_\infty + L_F)t} \left(d_{\infty, \nu}(\mu_0, \bar{\mu}_0) + \frac{1}{L_F} \left\| \sum_{\ell=1} B_\ell \|w_\ell - \bar{w}_\ell\|_{L_\xi^1} \right\|_{L_\xi^\infty} \right),$$

for all $t \geq 0$, or a weaker uniform bounded-Lipschitz distance as in [46, 47], when restricted to solutions continuous in $\xi \in [0, 1]$. Nevertheless, in doing so we would no longer be in position to exploit the natural compactness property of the cut distance studied in [66].

Following the same train of thoughts as in Corollary 4.9, we note that Theorem 5.1 can also be applied to special solutions $\mu_t^\xi = \delta_{X(t,\xi)}$ and $\bar{\mu}_t^\xi = \delta_{\bar{X}(t,\xi)}$ to (1.4)-(1.5), where X and \bar{X} solve the corresponding continuum-limit equation (4.20).

Corollary 5.6 (Stability of the continuum-limit equation). *Under the hypothesis in Theorem 5.1, consider $X, \bar{X} \in C(\mathbb{R}_+, L^p([0, 1], \mathbb{R}^d))$ solutions to the continuum-limit equation (4.20) with given w and \bar{w} . Then, we have*

$$\|X(t, \cdot) - \bar{X}(t, \cdot)\|_{L^p} \leq e^{(C_p + L_F)t} \left(\|X_0 - \bar{X}_0\|_{L^p} + \frac{D_\infty^{1/q}}{L_F} d_\square(w, \bar{w}; (4^\ell \|\hat{K}_\ell\|_{L^1})_{\ell \in \mathbb{N}})^{1/p} \right),$$

for all $t \in \mathbb{R}_+$.

We refer to [58] and also [2, 3, 32, 64] for stability estimates in analogous continuum-limit equations restricted to binary interactions only. However, we emphasize that, to the best of our knowledge, Corollary 5.6 is the first result that allows for a graph-limit based distance in place of the stronger L^p based distances in previous literature.

6. PROOF OF MAIN THEOREM 1.1

In this section we focus on the proof of the main result of the paper, namely, Theorem 1.1, which we state below in a more rigorous formulation for clarity of the presentation.

Theorem 6.1 (Mean-field limit). *Assume that the kernels K_ℓ and the weights $w_{i_{j_1 \dots j_\ell}}^{\ell, N}$ verify Assumptions 1, 2 and 3. For any $(X_{1,0}^N, \dots, X_{N,0}^N)$ with i.i.d. $X_{i,0}^N$ (but N dependent law) verifying Assumption 4 for some $p \in [1, 2]$, consider the unique solutions (X_1^N, \dots, X_N^N) to (1.3), and define the pair (μ^N, w^N) as in Definition 3.5, i.e.,*

$$\begin{aligned} \mu_t^{N,\xi} &:= \sum_{i=1}^N \mathbb{1}_{I_i^N}(\xi) \delta_{X_i^N(t)}, \quad \xi \in [0, 1], \\ w_\ell^N(\xi, \xi_1, \dots, \xi_\ell) &:= \sum_{i, j_1, \dots, j_\ell=1}^N \mathbb{1}_{I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N}(\xi, \xi_1, \dots, \xi_\ell) N^\ell w_{i_{j_1 \dots j_\ell}}^{\ell, N}, \quad \xi, \xi_1, \dots, \xi_\ell \in [0, 1]. \end{aligned}$$

Then, there are a subsequence $N_k \rightarrow \infty$, some measure-preserving maps $\Phi_k : [0, 1] \rightarrow [0, 1]$ and a distributional solution $\mu \in C(\mathbb{R}_+, \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1]))$ to (2.10)-(2.11) for some UR-hypergraphon $w = (w_\ell)_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ such that

$$\sup_{t \in [0, T]} \mathbb{E} d_{p,\nu}(\mu_t^{N_k, \Phi_k}, \mu_t) \rightarrow 0, \quad d_\square(w^{N_k, \Phi_k}, w) \rightarrow 0 \quad (6.1)$$

for all $T \in \mathbb{R}_+$, where $\mu_t^{N_k, \Phi_k}$ and $w^{N_k, \Phi_k} = (w_\ell^{N_k, \Phi_k})_{\ell \in \mathbb{N}}$ represent the rearrangements

$$\mu_t^{N_k, \Phi_k, \xi} = \mu_t^{N_k, \Phi_k}(\xi), \quad w_\ell^{N_k, \Phi_k}(\xi, \xi_1, \dots, \xi_\ell) = w^{N_k}(\Phi_k(\xi), \Phi_k(\xi_1), \dots, \Phi_k(\xi_\ell)).$$

Remark 6.2 (On the i.i.d. assumption of initial data). *We observe that whilst throughout the paper initial data $(X_{1,0}^N, \dots, X_{N,0}^N)$ were only assumed independent, in the statement of our main Theorem 6.1 we further impose that they are identically distributed according to the same law (possibly depending on N but not i). Similar assumptions are considered in previous literature, see [13]. As we shall show, this hypothesis is only a sufficient condition which helps proving compactness of initial data in STEP 1 of the proof, as in that case initial data consist of measures with constant disintegrations, and then compactness in $\mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$ amounts to compactness in $\mathcal{P}(\mathbb{R}^d)$. However, it could be possible to remove the i.i.d. property and also prove compactness of non-constant initial data, in a compatible way with the compactness of the UR-hypergraphons w^N associated to the finite weights $w_{i_{j_1 \dots j_\ell}}^{\ell, N}$. We leave this question for future work.*

Proof of Theorem 6.1. Let $(\bar{X}_1^N, \dots, \bar{X}_N^N)$ be the unique solution to (3.1) in Lemma 3.1, $(\bar{\lambda}^{N,i})_{1 \leq i \leq N}$ their associated laws (3.14), and $(\bar{\mu}^{N,\xi})_{\xi \in [0,1]}$ the graphon reformulation in Definition 3.5, *i.e.*,

$$\bar{\mu}_t^{N,\xi} := \sum_{i=1}^N \mathbb{1}_{I_i^N}(\xi) \bar{\lambda}_t^{N,i}, \quad \xi \in [0,1].$$

The proof is divided into four steps. First, we define the limiting μ and w solving the Vlasov equation (2.10)-(2.11). Second, we study the fluctuation of fibered measures μ^N associated with the original multi-agent system (1.3) with respect to the fibered measures $\bar{\mu}^N$ associated with the intermediate multi-agent system (3.1). Third, we study the fluctuations of $\bar{\mu}^N$ with respect to the obtained limiting solution μ . Finally, we compare μ^N and μ .

◇ STEP 1: Definition of the limiting w and μ .

Regarding the definition of the limiting w , we note that by assumptions (1.12) and (1.13) we have that $w^N = (w_\ell^N)_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ with uniform bound $W > 0$. Therefore, by [66, Lemma 33] there exists $w = (w_\ell)_{\ell \in \mathbb{N}} \in \mathcal{H}_W$ with the same W such that $\delta_\square(w^{N_k}, w) \rightarrow 0$ for a suitable subsequence N_k . Since δ_\square represents the unlabeled cut distance, the above means that there exists $\Phi_k : [0,1] \rightarrow [0,1]$ measure-preserving maps such that

$$d_\square(w^{N_k, \Phi_k}, w) \rightarrow 0, \quad (6.2)$$

where d_\square is the labeled cut distance and $w^{N_k, \Phi_k} = (w_\ell^{N_k, \Phi_k})_{\ell \in \mathbb{N}}$ is the rearranged UR-hypergraphon

$$w_\ell^{N_k, \Phi_k}(\xi, \xi_1, \dots, \xi_\ell) = w^{N_k}(\Phi_k(\xi), \Phi_k(\xi_1), \dots, \Phi_k(\xi_\ell)).$$

For practical reasons we similarly define the rearranged μ^{N_k, Φ_k} and $\bar{\mu}^{N_k, \Phi_k}$ as follows

$$\mu_t^{N_k, \Phi_k, \xi} = \mu_t^{N_k, \Phi_k}(\xi), \quad \bar{\mu}_t^{N_k, \Phi_k, \xi} = \bar{\mu}_t^{N_k, \Phi_k}(\xi).$$

Regarding the definition of the limiting μ , we first prescribe its initial datum μ_0 . Since $(X_{i,0}^N)_{1 \leq i \leq N}$ are distributed identically according to some $\lambda^N \in \mathcal{P}(\mathbb{R}^d)$, then $\bar{\mu}_0^N$ has constant disintegrations, and for this reason so does the rearranged $\bar{\mu}_0^{N_k, \Phi_k}$ namely,

$$\bar{\mu}_0^{N_k, \Phi_k, \xi} = \lambda^{N_k}, \quad \xi \in [0,1],$$

for all $k \in \mathbb{N}$. We also note that by assumption (1.16) we have $\sup_{k \in \mathbb{N}} \int_{\mathbb{R}^d} |x| d\lambda^{N_k}(x) < \infty$, and therefore $(\lambda^{N_k})_{k \in \mathbb{N}}$ is uniformly tight. Hence, by Prokhorov's theorem there is some subsequence, which we still denote by the same letter for simplicity, and there is $\lambda \in \mathcal{P}(\mathbb{R}^d)$ such that $d_{\text{BL}}(\lambda^{N_k}, \lambda) \rightarrow 0$. We define the initial datum

$$\mu_0^\xi := \lambda, \quad \xi \in [0,1],$$

which is also constant in $\xi \in [0,1]$. Note that this implies

$$d_{p,\nu}(\bar{\mu}^{N_k, \Phi_k}, \mu_0) \rightarrow 0. \quad (6.3)$$

We finally define $\mu \in C(\mathbb{R}_+, \mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0,1]))$ to be the distributional solution to (2.10)-(2.11) with given w and initial datum μ_0 as in Theorem 4.7.

◇ STEP 2: Comparing μ^N and $\bar{\mu}^N$.

Given $\xi \in [0,1]$, let us fix $i \in \llbracket 1, N \rrbracket$ such that $\xi \in I_i^N$. Then, we have that $\mu_t^{N,\xi} = \delta_{X_i^N(t)}$ and $\bar{\mu}_t^{N,\xi} = \bar{\lambda}_t^{N,i}$. Then, we have that

$$\begin{aligned} d_{\text{BL}}(\mu_t^{N,\xi}, \bar{\mu}_t^{N,\xi}) &= d_{\text{BL}}(\delta_{X_i^N(t)}, \bar{\lambda}_t^{N,i}) \\ &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} \phi(x) d(\delta_{X_i^N(t)}(x) - \bar{\lambda}_t^{N,i}(x)) \\ &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \left(\phi(X_i^N(t)) - \int_{\mathbb{R}^d} \phi(x) d\bar{\lambda}_t^{N,i}(x) \right) \\ &= \sup_{\|\phi\|_{\text{BL}} \leq 1} \int_{\mathbb{R}^d} (\phi(X_i^N(t)) - \phi(x)) d\bar{\lambda}_t^{N,i}(x) \end{aligned}$$

$$\begin{aligned} &\leq \int_{\mathbb{R}^d} |X_i^N(t) - x| d\bar{\lambda}_t^{N,i}(x) \\ &\leq (\mathbb{E}_i^N |X_i^N(t) - \bar{X}_i^N(t)|^p)^{1/p}, \end{aligned}$$

for all $1 \leq i \leq N$, where $\mathbb{E}_i^N = \mathbb{E}[\cdot | \bar{\mathcal{F}}_i^N]$ is the expectation conditioned to the natural filtration (3.2) of \bar{X}_i^N , and in the last step we have used Jensen's inequality. Taking L^p norms we have

$$d_{p,\nu}(\mu_t^N, \bar{\mu}_t^N) \leq \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}_i^N |X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p}.$$

Finally, taking global expectation, and by Jensen's inequality we have

$$\mathbb{E} d_{p,\nu}(\mu_t^N, \bar{\mu}_t^N) \leq \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} \mathbb{E}_i^N |X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p} = \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p}, \quad (6.4)$$

where in last step we have used the law of iterated expectation $\mathbb{E} = \mathbb{E} \mathbb{E}_i^N$.

◇ STEP 3: Comparing $\bar{\mu}^{N_k, \Phi_k}$ and μ .

First, since $\bar{\mu}^{N_k}$ is a distributional solution to (2.10)-(2.11) given w^{N_k} by virtue of Lemma 3.6, and since Φ_k are measure-preserving maps, then $\bar{\mu}^{N_k, \Phi_k}$ is also a distributional solution to (2.10)-(2.11) given w^{N_k, Φ_k} . Therefore, we can use the stability estimate in Theorem 5.1 to compare it to the distributional solution μ with given w defined in STEP 1. Specifically, we obtain

$$\begin{aligned} &\sup_{t \in [0, T]} d_{p,\nu}(\bar{\mu}_t^{N_k, \Phi_k}, \mu_t) \\ &\leq e^{(C_p + L_F)T} \left(d_{p,\nu}(\bar{\mu}_0^{N_k, \Phi_k}, \mu_0) + \frac{D_\infty^{1/q}}{L_F} d_\square(w^{N_k, \Phi_k}, w; (2^\ell \|\hat{K}_\ell\|_{L^1})_{\ell \in \mathbb{N}})^{1/p} \right), \end{aligned} \quad (6.5)$$

for all $T \in \mathbb{R}_+$.

◇ STEP 4: Comparing μ^{N_k, Φ_k} and μ .

Using the triangle inequality and taking expectations and supremum over $t \in [0, T]$ we have

$$\begin{aligned} &\sup_{t \in [0, T]} \mathbb{E} d_{p,\nu}(\mu^{N_k, \Phi_k}, \mu_t) \\ &\leq \sup_{t \in [0, T]} \mathbb{E} d_{p,\nu}(\mu_t^{N_k, \Phi_k}, \bar{\mu}_t^{N_k, \Phi_k}) + \sup_{t \in [0, T]} d_{p,\nu}(\bar{\mu}_t^{N_k, \Phi_k}, \mu_t) \\ &= \sup_{t \in [0, T]} \mathbb{E} d_{p,\nu}(\mu_t^{N_k}, \bar{\mu}_t^{N_k}) + \sup_{t \in [0, T]} d_{p,\nu}(\bar{\mu}_t^{N_k, \Phi_k}, \mu_t) \\ &\leq \sup_{t \in [0, T]} \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E} |X_i^N(t) - \bar{X}_i^N(t)|^p \right)^{1/p} \\ &\quad + e^{(C_p + L_F)T} \left(d_{p,\nu}(\bar{\mu}_0^{N_k, \Phi_k}, \mu_0) + \frac{D_\infty^{1/q}}{L_F} d_\square(w^{N_k, \Phi_k}, w; (2^\ell \|\hat{K}_\ell\|_{L^1})_{\ell \in \mathbb{N}})^{1/p} \right), \end{aligned}$$

where in the second step we have used that L^p norms are invariant under measure-preserving rearrangements, and in the last step we have used the control (6.4) from STEP 2 and (6.5) from STEP 3. Using the convergence of initial data in (6.3) and the convergence of UR-hypergraphons in (6.2) which we have by construction of μ_0 and w , along with Lemma 3.2 and more particularly the decay in Remark 3.3, we end the proof. \square

The above mean-field limit in Theorem 6.1 is precise enough to retain the full structure in the limit as $N \rightarrow \infty$, and not only averaged observables like in [39] for extended graphons, and [13] for dense random graphs. However, by virtue of Proposition 2.19 it is shown that

$$\mathbb{E} d_{\text{BL}}(\pi_{x\#} \mu_t^{N_k, \Phi_k}, \pi_{x\#} \mu_t) \leq \mathbb{E} d_{p,\nu}(\mu_t^{N_k, \Phi_k}, \mu_t).$$

Since $\pi_{x\#}\mu_t^{N_k, \Phi_k} = \frac{1}{N} \sum_{i=1}^N \delta_{X_i^{N_k}(t)}$ reduce to the standard empirical measures, then Theorem 6.1 has the following straightforward consequence.

Corollary 6.3 (Convergence of the empirical measures). *Under the assumptions in Theorem 6.1, the standard empirical measures $\hat{\mu}_t^{N_k} := \frac{1}{N} \sum_{i=1}^N \delta_{X_i^{N_k}(t)}$ verify*

$$\sup_{t \in [0, T]} \mathbb{E} d_{\text{BL}}(\hat{\mu}_t^{N_k}, \hat{\mu}_t) \rightarrow 0,$$

for all $T \in \mathbb{R}_+$, and $\hat{\mu}_t = \pi_{x\#}\mu_t \in C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d)$ -narrow) solves

$$\partial_t \hat{\mu}_t + \text{div}_x \left(\int_0^1 F_w[\mu_t](\cdot, \xi) \mu_t^\xi d\xi \right) = 0, \quad t \geq 0, x \in \mathbb{R}^d. \quad (6.6)$$

Remark 6.4 (The exchangeable case). *The equation (6.6) for the limit $\hat{\mu}$ of empirical measures is generally not closed as it depends on the full structured limit μ . However, for the all-to-all UR-hypergraphon $w_\ell \equiv 1$, the system becomes exchangeable and $\hat{\mu}$ solves a closed system*

$$\begin{cases} \partial_t \hat{\mu}_t + \text{div}_x(\hat{F}[\hat{\mu}_t]\hat{\mu}_t) = 0, & t \geq 0, x \in \mathbb{R}^d, \\ \hat{F}[\hat{\mu}_t](x) := \sum_{\ell=1}^{\infty} \int_{\mathbb{R}^{d\ell}} K_\ell(x, x_1, \dots, x_\ell) d\hat{\mu}_t(x_1) \dots d\hat{\mu}_t(x_\ell), \end{cases}$$

which corresponds to the standard Vlasov equation for exchangeable systems.

For the general non-exchangeable case, the above observable $\hat{\mu}_t = \pi_{x\#}\mu_t$ is only the first level of a larger hierarchy of observables indexed by directed hypertrees. Namely, the unsigned finite measure in the divergence term in (6.6) actually has a nice structure, and its evolution involves higher-order observables that are easy to characterize. Collecting all the family of involved observables we have the following result, which can be seen as a higher-order extension of the hierarchy of observables indexed by trees formulated in the binary case in [39, 40].

Remark 6.5 (A hierarchy indexed by hypertrees). *Under the assumptions in Theorem 6.1, let us define the family of observables associated with the limit (w, μ)*

$$\begin{aligned} & \tau(H, w, \mu_t)(x_1, \dots, x_{\#V(H)}) \\ & := \int_{[0,1]^{\#V(H)}} \prod_{\ell \in \mathbb{N}} \prod_{(i; j_1, \dots, j_\ell) \in E(H)} w_\ell(\xi_i, \xi_{j_1}, \dots, \xi_{j_\ell}) \prod_{i \in V(H)} \mu_t^{\xi_i}(x_i) \prod_{i \in V(H)} d\xi_i, \end{aligned} \quad (6.7)$$

for H any directed hypertree. Then, $\tau(H, w, \mu)$ solves the hierarchy of coupled equations

$$\begin{aligned} & \partial_t \tau(H, w, \mu_t)(x_1, \dots, x_{\#V(H)}) \\ & + \sum_{i=1}^{\#V(H)} \sum_{\ell=1}^{\infty} \text{div}_{x_i} \left(\int_{\mathbb{R}^{d\ell}} K_\ell(x_i, y_1, \dots, y_\ell) \tau(H_i^\ell, w, \mu_t)(x_1, \dots, x_{\#V(H)}, dy_1, \dots, dy_\ell) \right) = 0, \end{aligned} \quad (6.8)$$

in distributional sense, where H_i^ℓ consist in the new directed hypertree constructed from H by adding a directed hyperedge of cardinality $\ell + 1$ to the node $i \in V(H)$ (see Figure 5).

In [39], the space of pairs (w, μ) was endowed with the topology of the convergence of the observables (6.7). In this paper, pairs (w, μ) are endowed with the product topology of the cut-distance on the set of UR-hypergraphons \mathcal{H}_W and the $d_{p, \nu}$ distance on the space of probability measures $\mathcal{P}_{p, \nu}(\mathbb{R}^d \times [0, 1])$. At this stage, it is not clear which of them is finer. In particular, an interesting open question is to characterize whether $\tau(H, w^{N_k}, \mu^{N_k}) \rightarrow \tau(H, w, \mu)$ is implied by the convergence (6.1), or conversely. We leave this question to future work.

7. NUMERICAL SIMULATIONS

To illustrate the convergence result of Theorem 1.1, we provide some numerical simulations based on two examples of hypergraphs presented in Section 2.3. For computational reasons, we set dimension $d = 1$, and we only consider hypergraphs of bounded rank $r = 3$.

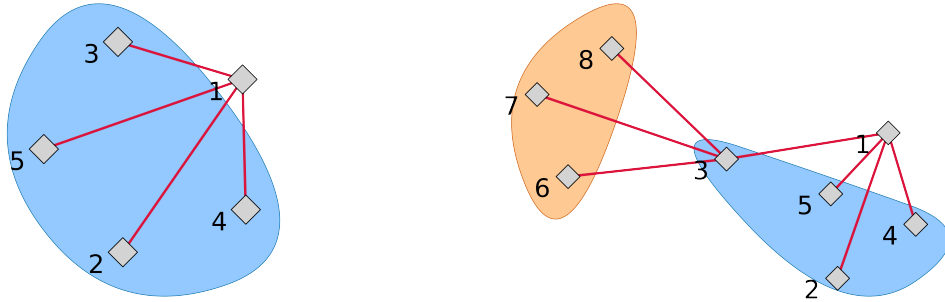


FIGURE 5. Directed hypertree H with nodes $V(H) = \{1, 2, 3, 4, 5\}$ and a single directed hyperedge $e = \{1; 2, 3, 4, 5\}$, and associated directed hypertree H_3^3 .

7.1. Hypergraph for a homogeneous group of rank 3. Consider the particle system (1.3) posed on the hypergraph for a homogeneous group of rank 3, that is,

$$w_{ij_1j_2}^{2,N} = \begin{cases} \frac{1}{N^2}, & \text{if } \max_{k_1, k_2 \in \{i, j_1, j_2\}} |k_1 - k_2| \leq \theta N, \\ 0, & \text{otherwise,} \end{cases} \quad (7.1)$$

for some $\theta \in (0, 1)$, and $w_{ij_1 \dots j_\ell}^{\ell, N} = 0$ for all $\ell \neq 2$ and $i, j_1, \dots, j_\ell \in \llbracket 1, N \rrbracket$. This hypergraph was introduced in Equation (2.3) in Section 2.3, and is represented in Figure 3 (right).

We also consider the linear interaction kernel given by

$$K_2(x, x_1, x_2) = \frac{x_1 + x_2}{2} - x. \quad (7.2)$$

A slight simplification of the proof of Proposition A.1 allows to prove that the above sequence of hypergraphs (7.1) converges as N tends to infinity to the limit UR-hypergraphon (with actually bounded rank) given by $w_\ell \equiv 0$ for all $\ell \neq 2$, and

$$w_2(\xi_0, \xi_1, \xi_2) = \begin{cases} 1, & \text{if } \max_{i, j \in \{0, 1, 2\}} |\xi_i - \xi_j| \leq \theta, \\ 0, & \text{otherwise.} \end{cases} \quad (7.3)$$

In all that follows, the parameter θ is fixed to the numerical value $\theta = 0.1$. Arguing as our main Theorem 6.1, we then have that the solution $(X_i^N)_{i \in \{1, \dots, N\}}$ to (1.3)-(7.1)-(7.2) converges as N tends to infinity to the solution μ to the Vlasov equation (1.4)-(7.3)-(7.2).

In Figure 6, the solution μ to (1.4)-(7.3)-(7.2) with uniform initial condition $\mu_0 = d\xi_{[[0,1]} dx_{[[0,1]}$ is represented for different time steps. The solution was computed using a finite difference scheme and numerical integration. Notice that due to the attractive nature of the interaction kernel K_2 in (7.2), we observe a concentration phenomenon. Moreover, the convergence speed seems faster for the central labels $\xi \in [\theta, 1 - \theta]$ and slower for edge labels ξ close to 0 or 1. This differentiated convergence speed can be explained by the hypergraphon w_2 in (7.3), which promotes interaction for central labels, and demotes interactions for edge labels. Indeed, an agent with label $\xi \in [\theta, 1 - \theta]$ belongs to all hyperedges (ξ, ξ_1, ξ_2) such that $\xi_1, \xi_2 \in [\xi - \theta, \xi + \theta]$. On the other hand, an agent with label $\xi = 0$ belongs to hyperedges (ξ, ξ_1, ξ_2) such that $\xi_1, \xi_2 \in [0, \theta]$, which reduces its chances of interaction.

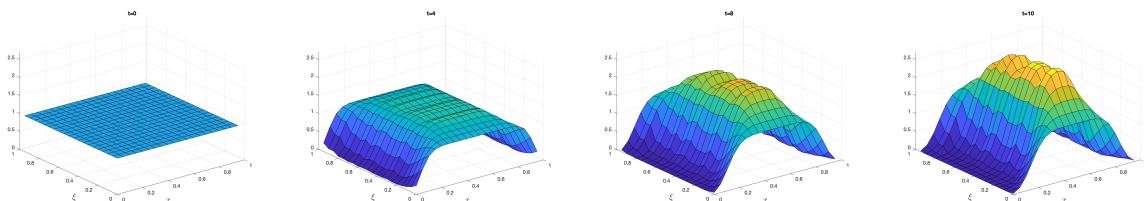


FIGURE 6. Evolution of μ_t , solution to the limit equation (1.4)-(7.3)-(7.2), at times $t = 0$, $t = 4$, $t = 8$ and $t = 10$.

To compare the solutions of the discrete and continuum systems, in Figure 7 we superimpose such a solution μ_t to (1.4)-(7.3)-(7.2) with uniform initial condition on $[0, 1] \times [0, 1]$ (represented by a color gradient) with the solution $(X_i^N(t))_{i \in \{1, \dots, N\}}$ to (1.3)-(7.1)-(7.2) with initial condition drawn randomly from the uniform distribution on $[0, 1] \times [0, 1]$. For comparison, in Figure 8 we also display the binned density computed by summing the number of agents in each square of dimension 0.1×0.1 .

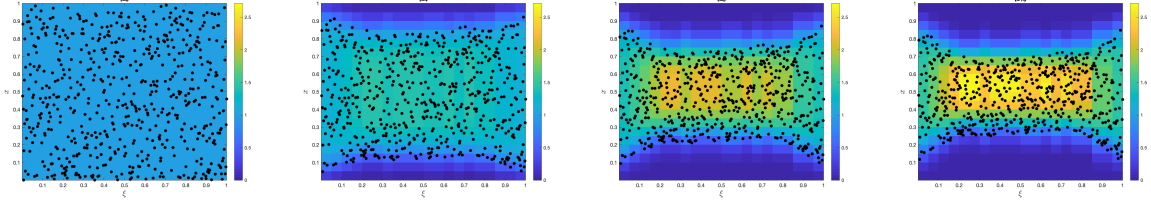


FIGURE 7. Evolution of μ_t , solution to the limit equation (1.4)-(7.3)-(7.2) (color gradient) at times $t = 0, t = 4, t = 8$ and $t = 10$, superimposed with the solution $X_i^N(t)$ of the microscopic system (1.3)-(7.1)-(7.2) for $N = 600$ (black dots).

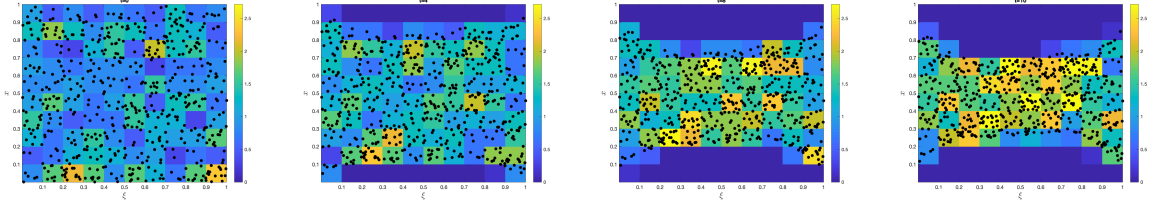


FIGURE 8. Evolution of the solution $X_i^N(t)$ of the microscopic system (1.3)-(7.1)-(7.2) for $N = 600$ (black dots) at times $t = 0, t = 4, t = 8$ and $t = 10$, together with the binned density obtained by counting the number of agents in each square of dimension 0.1×0.1 (color gradient).

In Figure 9, we illustrate the convergence of the empirical measure to the solution of the limit equation by providing a comparison of the binned density of X_i^N at the final time $t = 10$ for different numbers of agents: $N = 100, N = 200, N = 400$ and $N = 600$.

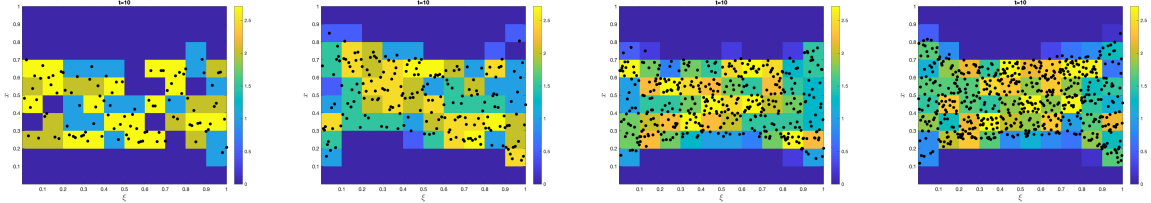


FIGURE 9. Solution $X_i^N(T)$ of the microscopic system (1.3)-(7.1)-(7.2) (black dots) at final time $T = 10$, together with the binned density obtained by counting the number of agents in each square of dimension 0.1×0.1 (color gradient) for $N = 100, N = 200, N = 400$ and $N = 600$ (left to right).

7.2. Hypergraph for a balanced group of rank 3. We now consider the particle system (1.3) posed on the hypergraph for a balanced group of rank 3, that is,

$$w_{j_0 j_1 j_2}^{2,N} = f \left(\frac{1}{N} \left| \frac{1}{3} \sum_{k=0}^2 j_k - \frac{N+1}{2} \right| \right). \quad (7.4)$$

and $w_{i_j_1 \dots i_{j_\ell}}^{\ell, N} = 0$ for all $\ell \neq 2$ and $i, j_1, \dots, j_\ell \in \llbracket 1, N \rrbracket$, where $f : [0, \frac{1}{2}] \rightarrow \mathbb{R}_+$ is a continuous decreasing function. This hypergraph was introduced in the example in Equation (2.5) in Section 2.3, and is represented in Figure 4 (right).

Since f is continuous, it can be shown by Proposition 2.10 (see also Remark 2.12) that the above sequence of hypergraphs (7.4) converges as N tends to infinity to the limit UR-hypergraphon (with actually bounded rank) given by $w_\ell \equiv 0$ for all $\ell \neq 2$, and

$$w_2(\xi_0, \xi_1, \xi_2) = f\left(\left|\frac{\xi_0 + \xi_1 + \xi_2}{3} - \frac{1}{2}\right|\right). \quad (7.5)$$

As in the previous section, we consider the linear interaction kernel K_2 given by (7.2). The function f is taken to be the continuous function $f : x \mapsto 4(x - \frac{1}{2})^2$. Then, arguing as in Theorem 6.1 we can show that the particle system (1.3)-(7.4)-(7.2) converges to the solution to the Vlasov equation (1.4)-(7.5)-(7.2), as a direct application of Proposition 2.10.

In Figure 10, the solution μ to (1.4)-(7.5)-(7.2) with uniform initial condition $\mu_0 = d\xi_{|[0,1]} dx_{|[0,1]}$ is represented for different time steps.

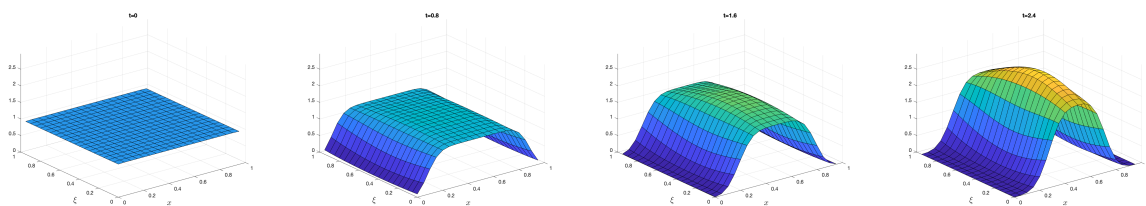


FIGURE 10. Evolution of μ_t , solution to the limit equation (1.4)-(7.5)-(7.2), at times $t = 0$, $t = 0.8$, $t = 1.6$ and $t = 2.4$.

To compare with the solutions to the discrete and continuum systems, in Figure 11 we superimpose the solution μ_t to (1.4)-(7.5)-(7.2) with uniform initial condition on $[0, 1] \times [0, 1]$ (represented by a color gradient) with the solution $(X_i^N(t))_{i \in \{1, \dots, N\}}$ solution to (1.3)-(7.4)-(7.2) with an initial condition drawn randomly from the uniform distribution on $[0, 1] \times [0, 1]$. For comparison, in Figure 12 we also display the binned density computed by counting the number of agents in each square of dimension 0.1×0.1 .

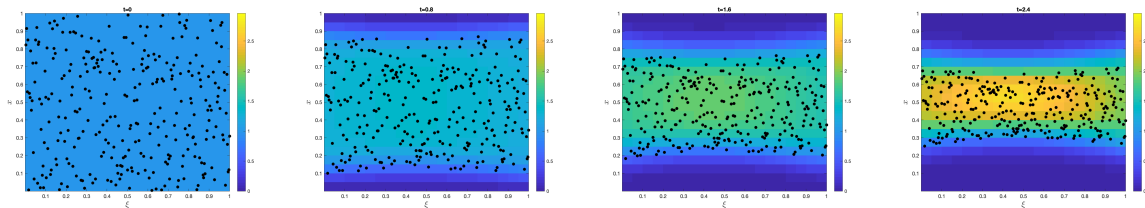


FIGURE 11. Evolution of μ_t , solution to the limit equation (1.4)-(7.5)-(7.2) (color gradient) at times $t = 0$, $t = 0.8$, $t = 1.6$ and $t = 2.4$, superimposed with the solution $X_i^N(t)$ of the microscopic system (1.3)-(7.4)-(7.2) for $N = 300$ (black dots).

8. CONCLUSIONS AND PERSPECTIVES

In this paper, we derived rigorously the convergence of an interacting particle system with higher-order interactions toward a Vlasov-type equation on an unbounded-rank hypergraphon. The convergence was established taking advantage of a natural topology arising from the graph theory, *i.e.* the cut-distance, and its corresponding compactness properties. The innovation of our work lies both in the derivation of a model which allows for unbounded rank interactions, and in the exploitation of a natural topology coming from graph theory. There are however a

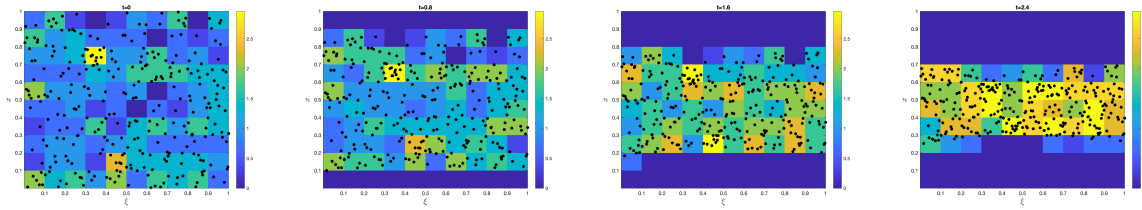


FIGURE 12. Evolution of the solution $X_i^N(t)$ of the microscopic system (1.3)-(7.4)-(7.2) for $N = 300$ (black dots) at times $t = 0$, $t = 0.8$, $t = 1.6$ and $t = 2.4$, together with the binned density obtained by counting the number of agents in each square of dimension 0.1×0.1 (color gradient).

certain number of directions that we would like to pursue in order to enrich this first result. We leave these questions to future work.

Firstly, we observe that whilst throughout the paper initial data were only assumed independent, in the statement of our main Theorem 6.1 we further impose that they are identically distributed according to the same law (possibly depending on N but not i). This hypothesis is only a sufficient condition which helps to prove compactness of initial data. However, it could be possible to remove the *i.i.d.* property and also prove compactness of non-constant $\bar{\mu}_0^N$ associated to independent (but not *i.i.d.*) initial data $X_{1,0}^N, \dots, X_{N,0}^N$, in a compatible way with the compactness of the UR-hypergraphons w^N associated to the finite weights $w_{ij_1 \dots j_\ell}^{\ell, N}$.

Secondly, but closely related to the above question, in this paper pairs (w, μ) are endowed with the product topology of the cut-distance on the set of UR-hypergraphons \mathcal{H}_W and the $d_{p,\nu}$ distance on the space of probability measures $\mathcal{P}_{p,\nu}(\mathbb{R}^d \times [0, 1])$, while in [39] the space of pairs (w, μ) was endowed with the topology of the convergence of observables analogous to $\tau(H, w, \mu)$ in (6.7). An interesting open question is to characterize whether the convergence $\tau(H, w^N, \mu^N) \rightarrow \tau(H, w, \mu)$ for all directed hypertree H is implied by the convergence $\delta_\square(w^N, w) \rightarrow 0$ and $d_{p,\nu}(\mu^N, \mu) \rightarrow 0$ in the above-mentioned product topology, or conversely.

Lastly, models for multi-agent dynamics on hypergraphs have recently attracted the attention of the Physics' community, with a focus on exploring how equilibrium states and speed of convergence are influenced by the presence of higher-order interactions (whether it be consensus in opinion dynamics, synchronization in Kuramoto-type oscillator dynamics, or spread of epidemics in SIR-type models). In that aim, the mean-field formulation provides a main advantage, as its formalism usually simplifies the study of the multi-agent system's equilibrium states. Hence, provided that the long-time behavior of the macroscopic model approximates well that of the microscopic system, one could infer from it important information concerning the collective dynamics and the phase transitions of the model.

APPENDIX A. CONVERGENCE OF HYPERGRAPHS FOR HOMOGENEOUS GROUPS

Proposition A.1. *Consider the sequence of hypergraphs $(H_N)_{N \in \mathbb{N}}$ for homogeneous groups defined for some radius $\theta \in (0, 1)$ by the hyperedge weights $(w_{ij_1 \dots j_\ell}^{\ell, N})_{i, j_1, \dots, j_\ell \in [1, N]}$ given by*

$$w_{ij_1 \dots j_\ell}^{\ell, N} = \begin{cases} \frac{1}{N^\ell}, & \text{if } \max_{k_1, k_2 \in \{i, j_1, \dots, j_\ell\}} |k_1 - k_2| \leq \theta N, \\ 0, & \text{otherwise,} \end{cases}$$

for each $N \in \mathbb{N}$ and $\ell \in [1, N - 1]$. Then, $(H_N)_{N \in \mathbb{N}}$ converges in the labeled cut-distance d_\square (and hence also in the unlabeled cut-distance δ_\square) to the UR-hypergraph $w = (w_\ell)_{\ell \in \mathbb{N}}$ given by

$$w_\ell(\xi_0, \dots, \xi_\ell) = \begin{cases} 1, & \text{if } \max_{k_1, k_2 \in \{1, \dots, \ell\}} |\xi_{k_1} - \xi_{k_2}| \leq \theta, \\ 0, & \text{otherwise,} \end{cases}$$

for all $\ell \in \mathbb{N}$.

Proof. We begin by noticing that the sequence of hypergraphs H_N can be obtained from the UR-hypergraphon $w = (w_\ell)_{\ell \in \mathbb{N}}$ by evaluating the hypergraphon w_ℓ on the grid points as in Proposition 2.10. However, the hypergraphon w does not satisfy the continuity assumption in Proposition 2.10. We will see that, in this case, convergence of $(H_N)_{N \in \mathbb{N}}$ still holds.

Consider the piecewise-constant UR-hypergraphon w^{H_N} defined from the hypergraph weights as in Equation (2.2), *i.e.*

$$\begin{aligned} w_\ell^{H_N}(\xi_0, \boldsymbol{\xi}_\ell) &= N^\ell w_{i_{j_1 \dots j_\ell}}^{\ell, N} = N^\ell \frac{1}{N^\ell} w_\ell \left(\frac{i-1}{N}, \dots, \frac{j_\ell-1}{N} \right) = N^\ell \frac{1}{N^\ell} w_\ell \left(\frac{\lfloor N\xi_0 \rfloor}{N}, \dots, \frac{\lfloor N\xi_\ell \rfloor}{N} \right) \\ &= \begin{cases} 1, & \text{if } \max_{k_1, k_2 \in \llbracket 0, \ell \rrbracket} \left| \frac{\lfloor N\xi_{k_1} \rfloor}{N} - \frac{\lfloor N\xi_{k_2} \rfloor}{N} \right| \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

for all $(i, j_1, \dots, j_\ell) \in \llbracket 1, N \rrbracket^{\ell+1}$, for all $(\xi_0, \xi_1, \dots, \xi_\ell) \in I_i^N \times I_{j_1}^N \times \dots \times I_{j_\ell}^N$.

We aim to show that $\lim_{N \rightarrow \infty} \|w_\ell - w_\ell^{H_N}\|_{L^1(I^{\ell+1})} = 0$ for all $\ell \in \mathbb{N}$. To this end we shall argue differently for $I^{\ell+1} \setminus U$ and on U , where U is the subset defined by

$$U := \left\{ (\xi_0, \boldsymbol{\xi}_\ell) \in I^{\ell+1} : \theta - \frac{1}{N} \leq \max_{(i,j) \in \llbracket 0, N \rrbracket^2} |\xi_i - \xi_j| \leq \theta + \frac{1}{N} \right\}.$$

We begin to show that for all $(\xi_0, \boldsymbol{\xi}_\ell) \in I^{\ell+1} \setminus U$, it holds $w_\ell(\xi_0, \boldsymbol{\xi}_\ell) = w_\ell^{H_N}(\xi_0, \boldsymbol{\xi}_\ell)$.

Firstly, consider $(\xi_0, \boldsymbol{\xi}_\ell) \in I^{\ell+1} \setminus U$, and suppose that $\max_{(i,j) \in \llbracket 0, N \rrbracket^2} |\xi_i - \xi_j| < \theta - \frac{1}{N}$. Then, $w_\ell(\xi_0, \boldsymbol{\xi}_\ell) = 1$. Moreover, for all $(i, j) \in \llbracket 0, N \rrbracket^2$, it holds

$$\left| \frac{\lfloor N\xi_i \rfloor}{N} - \frac{\lfloor N\xi_j \rfloor}{N} \right| = \left| \xi_i + \frac{\{\!-\!N\xi_i\}}{N} - \xi_j - \frac{\{\!-\!N\xi_j\}}{N} \right| \leq |\xi_i - \xi_j| + \frac{1}{N} |\{\!-\!N\xi_i\} - \{\!-\!N\xi_j\}| < \theta - \frac{1}{N} + \frac{1}{N} = \theta,$$

where for $z \in \mathbb{R}_+$, we denoted $\{z\} := z - \lfloor z \rfloor \in [0, 1)$. Thus from the definition of $w_\ell^{H_N}$ above, it holds $w_\ell^{H_N}(\xi_0, \boldsymbol{\xi}_\ell) = 1 = w_\ell(\xi_0, \boldsymbol{\xi}_\ell)$.

Secondly, if $(\xi_0, \boldsymbol{\xi}_\ell) \in I^{\ell+1} \setminus U$, and suppose that there exists $(i, j) \in \llbracket 0, N \rrbracket^2$ such that $|\xi_i - \xi_j| > \theta + \frac{1}{N}$, then $w_\ell(\xi_0, \boldsymbol{\xi}_\ell) = 0$. Moreover,

$$\left| \frac{\lfloor N\xi_i \rfloor}{N} - \frac{\lfloor N\xi_j \rfloor}{N} \right| \geq \left| |\xi_i - \xi_j| - \frac{1}{N} |\{\!-\!N\xi_i\} - \{\!-\!N\xi_j\}| \right| > \theta + \frac{1}{N} - \frac{1}{N} = \theta,$$

so that $w_\ell^{H_N}(\xi_0, \boldsymbol{\xi}_\ell) = 0 = w_\ell(\xi_0, \boldsymbol{\xi}_\ell)$.

Then, altogether implies

$$\begin{aligned} \|w_\ell - w_\ell^{H_N}\|_{L^1(I^{\ell+1})} &= \int_U |w_\ell(\xi_0, \boldsymbol{\xi}_\ell) - w_\ell^{H_N}(\xi_0, \boldsymbol{\xi}_\ell)| d\xi_0 d\boldsymbol{\xi}_\ell \leq \int_U d\xi_0 d\boldsymbol{\xi}_\ell \\ &\leq \binom{\ell+1}{2} \int_{I^{\ell-1}} \int_{\theta - \frac{1}{N} \leq \xi_0 - \xi_1 \leq \theta + \frac{1}{N}} d\xi_0 d\xi_1 \cdots d\xi_\ell \leq \frac{4}{N} \binom{\ell+1}{2} = 2 \frac{\ell(\ell+1)}{N}. \end{aligned}$$

Hence, $\lim_{N \rightarrow \infty} \|w_\ell - w_\ell^{H_N}\|_{L^1(I^{\ell+1})} = 0$ for each $\ell \in \mathbb{N}$, and concluding with the dominated convergence theorem as in Proposition 2.9, we deduce that for any summable sequence $(\alpha_\ell)_{\ell \in \mathbb{N}}$, it holds $\lim_{N \rightarrow \infty} d_\square(w_\ell, w_\ell^{H_N}; (\alpha_\ell)_{\ell \in \mathbb{N}}) = 0$. \square

REFERENCES

1. L. Ambrosio, N. Gigli, and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser, Basel, 2008.
2. N. Ayi and N. Pouradier Duteil, Mean-field and graph limits for collective dynamics models with time-varying weights, *J. Differ. Equ.* **299** (2021), 65–110.
3. ———, Graph limit for interacting particle systems on weighted random graphs, 2023, arXiv:2307.12801.
4. ———, Large-population limits of non-exchangeable particle systems, 2024, arXiv:2401.07748.
5. A. Backhausz and B. Szegedy, Action convergence of operators and graphs, *Canad. J. Math.* **74** (2020), no. 1, 1–50.
6. E. J. Balder, New fundamentals of Young measure convergence, *Calculus of Variations and Optimal Control* (A. Ioffe, S. Reich, and I. Shafir, eds.), Research Notes in Mathematics, vol. 411, Chapman and Hall/CRC, New York, 1999, pp. 24–48.

7. F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, *Phys. Rep.* **874** (2020), 1–92.
8. F. Battiston and G. Petri, Higher-order systems, *Understanding Complex Systems*, Springer International Publishing, 2022.
9. M. Beiglböck and D. Lacker, Denseness of adapted processes among causal couplings, 2018, arXiv:1805.03185.
10. I. Ben Porat, J. A. Carrillo, and S. T. Galtung, Mean field limit for one dimensional opinion dynamics with Coulomb interaction and time dependent weights, *Nonlinear Anal. Theory Methods Appl.* **240** (2024), 113462.
11. I. Ben Porat, J. A. Carrillo, and P.-E. Jabin, The graph limit for a pairwise competition model, *J. Differ. Equ.* **413** (2024), 329–369.
12. I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs, *Electron. J. Probab.* **6** (2001), no. 23, 1–13.
13. G. Bet, F. Coppini, and F. R. Nardi, Weakly interacting oscillators on dense random graphs, *Canad. J. Math.* **61** (2024), no. 1, 255–278.
14. B. Bianconi, Higher-order networks: an introduction to simplicial complexes, *Elements in the structure and dynamics of complex networks*, Cambridge University Press, Cambridge, 2021.
15. C. Bick and C. Böhle, Kuehn, Phase oscillator networks with nonlocal higher-order interactions: twisted states, stability, and bifurcations, *SIAM J. Appl. Dyn. Syst.* **22** (2023), no. 3, 1590–1638.
16. P. Billingsley, Probability and measure, John Wiley & Sons, New York, 1995.
17. S. Boccaletti, P. De Lellis, C.I. del Genio, K. Alfaro-Bittner, R. Criado, S. Jalan, and M. Romance, The structure and dynamics of networks with higher order interactions, *Phys. Rep.* **1018** (2023), 1–64, *The structure and dynamics of networks with higher order interactions*.
18. B. Bollobás and O. Riordan, Sparse graphs: Metrics and random models, *Random Struct. Algorithms* **39** (2011), 1–38.
19. C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence, *Ann. Probab.* **46** (2018), no. 1, 337–396.
20. ———, An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions, *Trans. Am. Math. Soc.* **372** (2019), no. 5, 3019–3062.
21. W. Braun and K. Hepp, The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles, *Comm. Math. Phys.* **56** (1977), no. 2, 101–113.
22. C. Castaing, P. Raynaud de Fitte, and M. Valadier, Young Measures on Topological Spaces With Applications in Control Theory and Probability Theory, *Mathematics and Its Applications*, vol. 571, Springer, Netherlands, 2004.
23. H. Chiba and G. S. Medvedev, The mean field analysis for the kuramoto model on graphs I. The mean field equation and transition point formulas, *Discrete Contin. Dyn. Syst. Ser. A* **39** (2019), no. 1, 131–155.
24. F. Coppini, H. Dietert, and G. Giacomin, A law of large numbers and large deviations for interacting diffusions on Erdős-rényi graphs, *Stoch. Dyn.* **20** (2020), 2050010.
25. E. Cruciani, E. L. Giacomelli, and J. Lee, On the convergence of nonlinear averaging dynamics with three-body interactions on hypergraphs, *SIAM J. Appl. Dyn. Syst.* **23** (2021), no. 3, 2364–2406.
26. E. N. Davison, B. O. Turner, K. J. Schlesinger, S. T. Miller, M. B. Abd Grafton, D. S. Bassett, and J. M. Carlson, Individual differences in dynamic functional brain connectivity across the human lifespan, *PLoS Comput. Biol.* **12** (2016), no. 11, 1–29.
27. S. Delattre, G. Giacomin, and E. Luçon, A note on dynamical models on random graphs and Fokker-Planck equations, *J. Stat. Phys.* **165** (2016), no. 4, 785–798.
28. R. Dobrushin, Vlasov equations, *Funct. Anal. Appl.* **13** (1979), no. 2, 115–123.
29. P. Dupuis and G. S. Medvedev, The large deviation principle for interacting dynamical systems on random graphs, *Commun. Math. Phys.* **390** (2022), 545–575.
30. G. Elek and B. Szegedy, A measure-theoretic approach to the theory of dense hypergraphs, *Adv. Math.* **231** (2012), 1731–1772.
31. M. A. Gkogkas and C. Kuehn, Graphop mean-field limits for Kuramoto-type models, *SIAM J. Appl. Dyn. Syst.* **21** (2022), no. 1, 248–283.
32. M. A. Gkogkas, C. Kuehn, and C. Xu, Continuum limits for adaptive network dynamics, *Commun. Math. Sci.* **21** (2023), 83–106.
33. F. Golse, On the dynamics of large particle systems in the mean field limit, *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity* (A. Muntean, J. Rademacher, and A. Zagaris, eds.), *Lecture Notes in Applied Mathematics and Mechanics*, vol. 3, Springer, Cham, 2016, pp. 1–144.
34. W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. Math.* **166** (2007), 897–946.
35. H. Hatami, L. Lovász, and B. Szegedy, Limits of locally-globally convergent graph sequences, *Geom. Funct. Anal.* **24** (2014), no. 1, 269–296.
36. E. Häusler and H. Luschgy, Stable Convergence and Stable Limit Theorems, *Probability Theory and Stochastic Modelling*, vol. 74, Springer, Switzerland, 2015.

37. R. Hegselmann and U. Krause, Opinion Dynamics and Bounded Confidence Models, *Analysis and Simulation, J. Artif. Soc. Soc. Simul.* **5** (2002).
38. I. Iacopini, J. R. Foote, N. H. Fefferman, E. P. Derryberry, and M. J. Silk, Not your private tete-à-tete: leveraging the power of higher-order networks to study animal communication, *Phil. Trans. R. Soc. B* **379** (2024), no. 1905, 20230190.
39. P.-E. Jabin, D. Poyato, and J. Soler, Mean-field limit of non-exchangeable systems, 2021, arXiv:2112.15406.
40. P.-E. Jabin and D. Zhou, The mean-field limit of sparse networks of integrate and fire neurons, 2023, arXiv:2309.04046.
41. J. Jacod and J. M. Mémin, Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité, *Séminaire de probabilités de Strasbourg* **15** (1981), 529–546.
42. B. Joudain and S. Méléard, Propagation of chaos and fluctuations for a moderate model with smooth initial data, *Ann. Inst. Henri Poincaré Probab.* **34** (1998), no. 6, 727–766.
43. D. Kaliuzhnyi-Verbovetskyi and G. S. Medvedev, The mean field equation for the kuramoto model on graph sequences with non-Lipschitz limit, *SIAM J. Math. Anal.* **50** (2018), 2441–2465.
44. J. Ko, Y. Kook, and K. Shin, Growth patterns and models of real-world hypergraphs, *Knowl. Inf. Syst.* **64** (2022), no. 11, 2883–2920.
45. Y. Kook, J. Ko, and K. Shin, Evolution of real-world hypergraphs: Patterns and models without oracles, *Proc. - IEEE Int. Conf. Data Min. ICDM (2020)*, 272–281.
46. C. Kuehn and C. Xu, Vlasov equations on digraph measures, *J. Differ. Equ.* **339** (2022), 261–349.
47. ———, Vlasov equations on directed hypergraph measures, 2022, arXiv:2207.03806.
48. D. Kunszenti-Kovács, L. Lovász, and B. Szegedy, Measures on the square as sparse graph limits, *J. Comb. Theory. Ser. B* **138** (2019), 1–40.
49. Y. Kuramoto, Self-entrainment of a population of coupled non-linear oscillators, *International Symposium on Mathematical Problems in Theoretical Physics (H. Araki, ed.)*, *Lecture Notes in Physics*, vol. 30, Springer-Verlag, Kyoto, Japan, 1975, pp. 420–422.
50. P. Le Bris and C. Poquet, A note on uniform in time mean-field limit in graphs, *ALEA, Lat. Am. J. Probab. Math. Stat.* **21** (2024), 775–790.
51. E. Liflyand, S. Samko, and R. Trigub, The Wiener algebra of absolutely convergent Fourier integrals: an overview, *Anal. Math. Phys.* **2** (2012), 1–68.
52. Q. F. Lotito, M. Contisciani, C. De Bacco, L. Di Gaetano, L. Gallo, A. Montresor, F. Musciotto, N. Ruggeri, and F. Battiston, Hypergraphx: a library for higher-order network analysis, *J. Complex Netw.* **11** (2023), no. 3, 2051–1329.
53. L. Lovász, Large Networks and Graph Limits, *Colloquium Publications*, vol. 60, American Mathematical Society, Providence, Rhode Island, 2012.
54. L. Lovász and B. Szegedy, Limits of dense graph sequence, *J. Comb. Theory. Ser. B* **96** (2006), 933–957.
55. M. Lucas, G. Cencetti, and F. Battiston, Multiorder Laplacian for synchronization in higher-order networks, *Phys. Rev. Res.* **2** (2020), 033410.
56. E. Luçon, Quenched asymptotics for interacting diffusions on inhomogeneous random graphs, *Stoch. Process. Their Appl.* **130** (2020), no. 11, 6783–6842.
57. S. Majhi, M. Perc, and D. Ghosh, Dynamics on higher-order networks: a review, *J. R. Soc. Interface* **19** (2022), no. 188, 20220043.
58. G. S. Medvedev, The nonlinear heat equation on dense graphs and graph limits, *SIAM J. Math. Analysis* **46** (2014), 2743–2766.
59. A. P. Millán, J. J. Torres, and G. Bianconi, Explosive higher-order kuramoto dynamics on simplicial complexes, *Phys. Rev. Lett.* **124** (2020), 218301.
60. L. Neuhäuser, R. Lambiotte, and M. T. Schaub, Consensus dynamics and opinion formation on hypergraphs, *Higher-Order Systems (F. Battiston and G. Petri, eds.)*, *Understanding Complex Systems*, Springer, Cham, 2022, pp. 347–376.
61. H. Neunzert and J. Wick, Die approximation der lösung von integro-differentialgleichungen durch endliche punktmengen, *Numerische Behandlung nichtlinearer Integrodifferential-und Differentialgleichungen (Berlin, Heidelberg) (R. Ansorge and W. Törnig, eds.)*, Springer Berlin Heidelberg, 1974, pp. 275–290.
62. R.I. Oliveira and G.H. Reis, Interacting diffusions on random graphs with diverging average degrees: hydrodynamics and large deviations, *J. Stat. Phys.* **176** (2019), 1057–1087.
63. F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Comm. Part. Differ. Equat.* **26** (2001), no. 1–2, 101–174.
64. T. Paul and E. Trélat, From microscopic to macroscopic scale equations: mean field, hydrodynamic and graph limits, 2022, arXiv:2209.08832.
65. J. Peszek and D. Poyato, Heterogeneous gradient flows in the topology of fibered optimal transport, *Calc. Var. Partial Differ. Equ.* **62** (2023), 258.
66. T. M. Roddenberry and S. Segarra, Limits of dense simplicial complexes, *J. Mach. Learn. Res.* **24** (2023), 1–42.
67. R. Sahasrabudde, L. Neuhäuser, and R. Lambiotte, Modelling non-linear consensus dynamics on hypergraphs, *J. Phys.: Complex.* **2** (2021), no. 2, 025006.

68. H. Schawe and L. Hernández, Higher order interactions destroy phase transitions in deffuant opinion dynamics model, *Commun. Phys.* **5** (2022), no. 1, 32.
69. P. S. Skardal and A. Arenas, Higher order interactions in complex networks of phase oscillators promote abrupt synchronization switching, *Commun. Phys.* **3** (2020), no. 1, 218.
70. E. M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, 1970.
71. A.-S. Sznitman, Topics in propagation of chaos, *Ecole d'Été de Probabilités de Saint-Flour XIX –1989* (P. L. Hennequin, ed.), *Lecture Notes in Mathematics*, vol. 1464, Springer, Berlin, Heidelberg, 1991, pp. 165–251.
72. M. Valadier, Young Measures, *Methods of Nonconvex Analysis* (A. Cellina, ed.), *Lecture Notes in Mathematics*, vol. 1446, Springer, Berlin, Heidelberg, 1989, pp. 152–188.
73. X. Wang, C. Xu, and Z. Zheng, Phase transition and scaling in kuramoto model with high-order coupling, *Nonlinear Dyn.* **103** (2021), 2721–2732.
74. C. Xu and P. S. Skardal, Spectrum of extensive multiclusters in the kuramoto model with higher-order interactions, *Phys. Rev. Res.* **3** (2021), 013013.
75. C. Xu, X Wang, and P. S. Skardal, Bifurcation analysis and structural stability of simplicial oscillator populations, *Phys. Rev. Res.* **2** (2020), 023281.
76. Y. Zhang, M. Lucas, and F. Battiston, Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes, *Nat. Commun.* **14** (2023), 1605.
77. Y. F. Zhao, Hypergraph limits: A regularity approach, *Random Struct. and Alg.* **47** (2015), 205–226.
78. G. Zucal, Action convergence of general hypergraphs and tensors, 2023, arXiv:2308.00226.

SORBONNE UNIVERSITÉ, UNIVERSITÉ PARIS-DIDEROT SPC, CNRS, INRIA, INSTITUT UNIVERSITAIRE DE FRANCE, LABORATOIRE JACQUES-LOUIS LIONS, PARIS, FRANCE

Email address: `nathalie.ayi@sorbonne-universite.fr`

SORBONNE UNIVERSITÉ, INRIA, UNIVERSITÉ PARIS-DIDEROT SPC, CNRS, LABORATOIRE JACQUES-LOUIS LIONS, PARIS, FRANCE

Email address: `nastassia.pouradier.duteil@sorbonne-universite.fr`

DEPARTAMENTO DE MATEMÁTICA APLICADA AND RESEARCH UNIT “MODELING NATURE” (MNAT), FACULTAD DE CIENCIAS, UNIVERSIDAD DE GRANADA, 18071 GRANADA, SPAIN

Email address: `dauidpoyato@ugr.es`