Innov. Graph Theory **2**, 2025, pp. 25–117
https://doi.org/10.5802/igt.7

PROBABILITY-GRAPHONS: LIMITS OF LARGE DENSE WEIGHTED GRAPHS

by Romain ABRAHAM, Jean-François DELMAS & Julien WEIBEL

Abstract. — We introduce probability-graphons which are probability kernels that generalize graphons to the case of weighted graphs. Probability-graphons appear as the limit objects to study sequences of large weighted graphs whose distribution of subgraph sampling converge. The edge-weights are taken from a general Polish space, which also covers the case of decorated graphs. Here, graphs can be either directed or undirected. Starting from a distance d_{m} inducing the weak topology on measures, we define a cut distance on probability-graphons, making it a Polish space, and study the properties of this cut distance. In particular, we exhibit a tightness criterion for probability-graphons related to relative compactness in the cut distance. We also prove that under some conditions on the distance d_{m}, which are satisfied for some well-know distances like the Lévy–Prokhorov distance, and the Fortet–Mourier and Kantorovitch–Rubinshtein norms, the topology induced by the cut distance on the space of probability-graphons is independent from the choice of d_{m}. Eventually, we prove that this topology coincides with the topology induced by the convergence in distribution of the sampled subgraphs.

1. Introduction

1.1. Motivation and literature review

Networks appear naturally in a wide variety of contexts, including for example: biological networks [19, 45], epidemics processes [21, 40], electrical power grids [1] and social networks [2, 50]. Most of those problems involve large dense graphs, that is graphs that have a large number of vertices and a number of edges that scales as the square of the number of vertices. Those graphs are too large to be represented entirely in the targeted applications.

Keywords: random graphs, stochastic networks, graphons, probability-graphons, dense graph limits, weighted graphs, decorated graphs.
2020 *Mathematics Subject Classification:* 05C80, 60B10.

The idea is then to go from a combinatorial representation given by the graph to an infinite continuum representation.

In the case of unweighted graphs (i.e. graphs without edge-weights), a theory was developed to study the asymptotic behaviour of large dense graphs, with the limit objects being the so-called *graphons*. The properties of graphons were studied in a series of articles started by [47, 26, 15, 16, 18]. We shall refer to the monograph [46] which exposes in details the theory of graphons developed in this series of articles. Graphons can be used to define models of random graphs with latent vertex-type variables (called W-random graphs) generalizing the Erdös-Rényi graph and the stochastic block model (SBM). The space of graphons can be equipped with the so-called *cut distance*, making it a compact space, and whose topology is that of the convergence in distribution for all sampled subgraphs, or equivalently of the convergence for subgraph homomorphism densities.

In recent years, graphons have been used in several application context: non-parametric estimation methods and algorithms for massive networks [17], SIS epidemic models [20], the study of transferability properties for Graph Neural Networks [39]. Furthermore, there has been recent developments in the study of mean-field systems using graphons: stochastic games and their Nash equilibria [43], opinion dynamic on a graphon [3], cooperative multi-agent reinforcement learning [32], to cite a few.

However, most real-world phenomena on the above networks involve weighted networks, where each edge in the graph carries additional information such as intensity or frequency of interaction, or transfer capacity.

There exists many models of random weighted graphs. For example configuration models with edges having independent exponential weights have been considered in [10, 5, 6], see also [29, 34] where the distribution of the weight of an edge depends on the types of its end-points. Random geometric graphs with vertices and edges having independent Gaussian weights have been considered in [4].

Weighted SBMs (sometimes also called labeled SBMs), in which each edge independently receives a random weight whose distribution depends on the community labels of its end-points, have been studied to solve community detection in [44] (see also [53] for more general models where vertex-labels come from a compact space), and exact community recovery in [36], and to get bounds on the number of misclassified vertices in [55, 54]. Note that weighted SBMs correspond to a special case of the probability-graphons we study in this article where the space of vertex-labels is finite (they correspond to the stepfunction probability-graphons we define in Section 3).

Concomitantly to our work, in [9], the authors studied mean-field equations on large real-weighted graphs modeling interactions with a probability kernel from $[0,1]^2$ to $\mathcal{M}_1(\mathbb{R})$ the set of probability measures on $\mathbb{R}$, but they did not study the topological properties of the set of those probability kernels. Recently, in [31], the authors studied the limit of the total weight of the minimum spanning tree (MST) for a sequence of random weighted graphs. Following what has been done for the uniform spanning tree in [30, 7], one expects the local and scaling limits of the MST to be directly constructed from the limit of the random weighted graphs.

Motivated by those examples, we shall consider *probability-graphons* as possible limits of large weighted graphs; they are defined as maps from $[0,1]^2$ to the space of probability measures $\mathcal{M}_1(\mathbf{Z})$ on a Polish space $\mathbf{Z}$. When $\mathbf{Z}$ is compact, this question has been considered in [49] and in [46, Section 17.1] using convergence of homomorphism densities of subgraphs decorated with real functions defined on $\mathbf{Z}$, see also [41] on multigraphs where $\mathbf{Z} = \mathbb{N}$, but the metric properties of the set of probability-graphons $\mathcal{W}_1$ have only been established when $\mathbf{Z}$ is finite, see [25]. Concomitantly to our work, in [8], the authors established the metric properties for the set of probability-graphons $\mathcal{W}_1$ when $\mathbf{Z} = [-1,1]$, where they consider the Kantorovitch–Rubinshtein cut distance (denoted here by $\delta_{\square,\mathrm{KR}}$). However their proof can not be extended directly to a general Polish space. The work [42] is an extension of [49] where $\mathcal{M}_1(\mathbf{Z})$ is replaced by the dual space $\mathcal{Z}$ of a separable Banach space $\mathcal{B}$. As $\mathcal{M}_1(\mathbf{Z})$ is a subset of the dual of the space $C_b(\mathbf{Z})$ of real-valued continuous bounded functions on $\mathbf{Z}$, this approach covers our setting when $C_b(\mathbf{Z})$ is separable, that is, $\mathbf{Z}$ compact (see Section 2 below). The norm introduced on the space of $\mathcal{Z}$-valued graphons therein implies the convergence of homomorphisms densities of $\mathcal{B}$-decorated sub-graphs, however there is no equivalence *a priori*. In [28], probability-graphons with $\mathbf{Z} = [0,1]$ are used as a part of latinons, which are limit objects for latin squares (square matrix of size $n \times n$ where each row and column contains each integer from 1 to n exactly once). The authors propose an interesting method to prove compactness of the space of latinons for the cut distance; but this does not extend directly to the case of a general Polish space $\mathbf{Z}$.

In this paper we study the topological properties of the space of probability-graphons $\widetilde{\mathcal{W}}_1$ when $\mathbf{Z}$ is a general Polish space: the space $\widetilde{\mathcal{W}}_1$ is a Polish topological space and we give "natural" cut distances on $\widetilde{\mathcal{W}}_1$ which are complete. One of the main difficulties is that the space of probability measures $\mathcal{M}_1(\mathbf{Z})$ can be endowed with many distances which induce the topology of weak convergence, each of them giving rise to a different cut distance on $\widetilde{\mathcal{W}}_1$. Our main result is:

a) The topology induced on $\widetilde{\mathcal{W}}_1$ does not depend on the initial choice of the distance on $\mathcal{M}_1(\mathbf{Z})$, provided this distance satisfies some simple general conditions. Those conditions are satisfied if the distance is *quasi-convex* (a property that generalizes the convexity of a norm).
b) This topology coincides with the topology induced by the convergence in distribution of the sampled subgraphs with random weights on the edges (or equivalently the convergence of the homomorphism densities of $C_b(\mathbf{Z})$-decorated subgraphs).
c) Similarly to the graphon setting, the sequence of large sampled weighted subgraphs from a probability-graphon W convergence in distribution to W.
d) We also provide a tightness criterion for studying the convergence of weighted graphs towards probability-graphons.

In conclusion, we believe that the unified framework developed here is easy-to-work-with and will allow to use probability-graphons to study large (random) weighted graphs.

1.2. New contribution

Through the article, *measure* will always be used to denote a positive measure.

1.2.1. Definition of probability-graphons

In this article, we define an analogue of graphons for weighted graphs, which we call *probability-graphons*, and study their properties. To avoid any confusion, in the rest of the article we say *real-valued graphons* instead of graphons. We consider the general case where weighted graphs take their edge-weights in a Polish space $\mathbf{Z}$ (e.g. $\mathbb{Z}$, $\mathbb{R}$ or $\mathbb{R}^d$), which thus also covers the case of decorated graphs, multi-graphs (graphs with possibly multiple edges between two vertices) and dynamical graphs (where edge-weights evolve over time).

We define a *probability-graphon* as a probability kernel $W : [0,1]^2 \to \mathcal{M}_1(\mathbf{Z})$, where $\mathcal{M}_1(\mathbf{Z})$ is the space of probability measures on $\mathbf{Z}$. A probability-graphon can be interpreted as follows: for two "vertex type" x and y in $[0,1]$, the weight z of an edge between two vertices of type x and y is distributed as the probability measure $W(x,y;\mathrm{d}z)$. In particular, the special case $\mathbf{Z} = \{0,1\}$ allows to recover real-valued graphons: as any real-valued graphon $w : [0,1]^2 \to [0,1]$ can be represented as a probability-graphon $W(x,y;\cdot) = w(x,y)\delta_1 + (1-w(x,y))\delta_0$, where δ_z denotes the Dirac mass located at z. Let us mention that it is possible to define the probability-graphons on a more general probability space $(\Omega, \mathcal{A}, \mu)$ than $[0,1]$ for the vertex-types, see Remark 3.4 for details. In this article, we also define and study the properties of *signed measure-valued kernels* which are bounded (in total mass/total variation norm) measurable functions $W : [0,1]^2 \mapsto \mathcal{M}_{\pm}(\mathbf{Z})$ whose values are signed measures, but for brevity we mainly focus on probability-graphons in this introduction.

As probability-graphons are measurable functions, we identify probability-graphons that are equal for almost every $(x,y) \in [0,1]^2$, and we denote by $\mathcal{W}_1$ the space of probability-graphons. Moreover, as we consider weighted graphs that are unlabeled (that is vertices are unordered), we need to consider probability-graphons up to "relabeling": for a measure-preserving map $\varphi : [0,1] \to [0,1]$ (relabeling map for probability-graphons), we define $W^{\varphi}(x,y;\cdot) = W(\varphi(x), \varphi(y);\cdot)$; we say that two probability-graphons are weakly isomorphic if there exists measure-preserving maps $\varphi, \psi : [0,1] \to [0,1]$ such that $U^{\varphi} = W^{\psi}$ for a.e. $(x,y) \in [0,1]^2$. We denote by $\widetilde{\mathcal{W}}_1$ the space of probability-graphons where we identity probability-graphons that are weakly isomorphic.

We can always assume that weighted graphs are complete graphs by adding all missing edges and giving them a weight/decoration ∂ which is a cemetery point added to $\mathbf{Z}$. Any weighted graph G can be represented as a probability-graphon W_G in the following way: denote by n the number of vertices of G and divide the unit interval $[0,1]$ into n intervals $I_1, \cdots, I_n$ of equal lengths, then W_G is defined for $(x,y) \in I_i \times I_j$ as $W_G(x,y;\cdot) = \delta_{M(i,j)}$, where $M(i,j)$ is the weight on the edge (i,j) in G. Thus a weighted graph G, with its edges decorated by elements of $\mathbf{Z}$, is associated with a $\mathcal{M}_1(\mathbf{Z})$-valued graphon W_G where the decorations are replaced by the corresponding Dirac masses. Note that weighted graphs can be either directed or undirected, in the case of undirected weighted graphs their limit objects are symmetric probability-graphons, that is probability-graphons W such that $W(x,y;\cdot) = W(y,x;\cdot)$.

1.2.2. The cut distance for probability-graphons and its properties

While there is a usual distance on the field of reals $\mathbb{R}$, this is not the case for probability measures, measures or signed measures endowed with the weak topology. Some commonly used distances include the Lévy–Prokhorov distance d_{LP} which can be defined on measures, and the Kantorovitch–Rubinshtein norm $\|\cdot\|_{\mathrm{KR}}$ (sometimes also called the bounded Lipschitz norm) and the Fortet–Mourier norm $\|\cdot\|_{\mathrm{FM}}$ defined on signed measures but metrizing the weak topology on measures. (Note that in general the weak topology is not metrizable on signed measures, see Section 2 below.) We also use a norm $\|\cdot\|_{\mathcal{F}}$ based on a convergence determining sequence $\mathcal{F} \subset C_b(\mathbf{Z})$ (that is, a sequence containing enough functions to characterize the convergence of measures for the weak topology, see page 37 for a more formal definition). See Section 3.8 for definitions of those distances. To define an analogue of the cut norm for probability-graphons, we first need to choose a distance d_{m} that metrizes the weak topology on the space of sub-probability measures $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (i.e. measures with total mass at most 1); we then define the *cut distance* $d_{\square,\mathrm{m}}$ for probability-graphons as:

$$d_{\square,\mathrm{m}}(U,W) = \sup_{S,T\subset[0,1]} d_{\mathrm{m}}\Big(U(S\times T;\cdot), W(S\times T;\cdot)\Big),$$

where the supremum is taken over all measurable subsets S and T of $[0,1]$, and where $W(S\times T;\cdot) = \int_{S\times T} W(x,y;\cdot)\,\mathrm{d}x\mathrm{d}y$ is a sub-probability measure, and similarly for U. Moreover, if the distance d_{m} is derived from a norm N_{m} defined on the space of signed measures $\mathcal{M}_{\pm}(\mathbf{Z})$, then the cut distance $d_{\square,\mathrm{m}}$ derives from the *cut norm* $N_{\square,\mathrm{m}}$ defined on signed measure-valued kernels:

$$N_{\square,\mathrm{m}}(W) = \sup_{S,T\subset[0,1]} N_{\mathrm{m}}\Big(W(S\times T;\cdot)\Big).$$

We then define the unlabeled *cut distance* $\delta_{\square,\mathrm{m}}$ on the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$ as:

$$\delta_{\square,\mathrm{m}}(U,W) = \inf_{\varphi} d_{\square,\mathrm{m}}(U,W^{\varphi}) = \min_{\varphi,\psi} d_{\square,\mathrm{m}}(U^{\varphi},W^{\psi}),$$

where the infimum is taken over all measure-preserving maps φ and ψ, see Proposition 3.18 for alternative expressions of $\delta_{\square,\mathrm{m}}$ (including proof that the minimum exists for the second expression) and see Theorem 3.17 that states that $\delta_{\square,\mathrm{m}}$ is indeed a distance on $\widetilde{\mathcal{W}}_1$. In Proposition 4.13, we prove an equivalent of the weak regularity lemma for probability-graphons.

We define the notion of a quasi-convex distance, which generalizes the convexity of a norm.

Definition 1.1 (Quasi-convex distance). — *Let* (X, d) *be a metric space which is a convex subset of a vector space. The distance* d *is* quasi-convex *if for all* $x_1, x_2, y_1, y_2 \in X$ *and all* $\alpha \in [0, 1]$, *we have:*

$$d(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \leqslant \max(d(x_1, y_1), d(x_2, y_2)).$$

In particular, any distance (on a convex subset of a vector space) which derive from a norm is quasi-convex. Moreover, the Lévy–Prokhorov distance d_{LP} is quasi-convex (see Lemma 3.21).

An interesting fact is that under some conditions on d_{m} (including the case when d_{m} is quasi-convex), the topology induced by the associated cut distance $\delta_{\square,\mathrm{m}}$ does not depend on the particular choice of d_{m}. The following proposition is a particular case of Theorem 5.5 together with Corollary 4.14.

Proposition 1.2. — *Let* d_{m} *be a quasi-convex distance on* $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ *that induces the weak topology. The cut distances* $\delta_{\square,\mathrm{m}}$, $\delta_{\square,\mathrm{LP}}$, $\delta_{\square,\mathrm{KR}}$, $\delta_{\square,\mathrm{FM}}$ *and* $\delta_{\square,\mathcal{F}}$ *induce the same topology on the space of probability-graphons* $\widetilde{\mathcal{W}}_1$.

Recall that $\mathbf{Z}$ is a Polish space. We now state that $\widetilde{\mathcal{W}}_1$ is also Polish for the distance $\delta_{\square,\mathrm{LP}}$ (but not for $\delta_{\square,\mathcal{F}}$!), and we refer to Theorem 5.10 for other distances.

Theorem 1.3. — *The space of probability-graphons* $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{LP}})$ *is a Polish metric space.*

We prove an analogue of Prokhorov's theorem with a tightness criterion for probability-graphons. We say that a subset of probability-graphons $\mathcal{K} \subset \widetilde{\mathcal{W}}_1$ is *tight* if the set of probability measures $\{M_W : W \in \mathcal{K}\}$ is tight (in the sense of probability measures), where $M_W(\cdot) = W([0,1]^2; \cdot)$. This tightness criterion appears in [41] for multigraphs, corresponding to the particular case $\mathbf{Z} = \mathbb{N}$. The next result is consequence of Theorem 5.1 and Proposition 5.2 as well as Corollary 4.14.

Theorem 1.4 (Compactness property). — *Consider the topology on* $\widetilde{\mathcal{W}}_1$ *from Proposition 1.2.*

(i) *If a sequence of elements of* $\widetilde{\mathcal{W}}_1$ *is tight, then it has a converging subsequence.*
(ii) *A subset* $\mathcal{K} \subset \widetilde{\mathcal{W}}_1$ *is relatively compact is and only if it is tight.*
(iii) *If* $\mathbf{Z}$ *is compact, then the space* $\widetilde{\mathcal{W}}_1$ *is compact.*

1.2.3. Sampling from probability-graphons and its link with the cut distance

Finally, we link the topology of the cut distance $\delta_{\square,\mathrm{m}}$ with subgraph sampling. The probability-graphons allow to define models of random weighted graphs (the W-random graph model) which generalize weighted SBM random graphs, and which plays the role of sampled subgraphs for probability-graphons. The W-random graph (or sampled subgraph of size k) $\mathbb{G}(k, W)$ has two parameters, a number of vertices k and a probability-graphon W for edge-weights, and is defined as follows: first let $X_1, \cdots, X_k$ be k independent random "vertex-types" uniformly distributed over $[0, 1]$; then given $X_1, \cdots, X_k$, each edge receives a weight independently, where the weight of the edge (i, j) is distributed as $W(X_i, X_j; \cdot)$. By identifying $z \in \mathbf{Z}$ with the Dirac mass at z, we can identify the random weighted graph $\mathbb{G}(k, W)$ with its corresponding probability-graphon (see Section 1.2.1), notice that this random probability-graphon is unique up to a weak isomorphism, and with a slight abuse we shall denote it by $\mathbb{G}(k, W)$.

We also provide the a.s. convergence of sampled subgraphs for the topology from Proposition 1.2, see Theorem 6.13 together with Corollary 5.6.

Theorem 1.5. — *Let W be a probability-graphon. Then, a.s. the sequence of sampled subgraphs $(\mathbb{G}(k, W))_{k\in\mathbb{N}^*}$ converges to W for the topology from Proposition 1.2.*

To prove this theorem, we adapt the proof scheme of [46, Sections 10.5 and 10.6] relying on the first and second sampling lemmas for real-valued graphons. The proof is done using the cut distance $\delta_{\square,\mathcal{F}}$ because of the good approximation properties of $\|\cdot\|_{\mathcal{F}}$.

In the case of unweighted graphs, the homomorphism numbers $\hom(F, G)$ count the number of occurences of a graph F (often called a *motif* or a *graphlet*) as an induced subgraph of G, and their normalized counterparts, the homomorphism densities $t(F, G)$ allow to characterize a graph (up to relabeling and twin-vertices expansion), and also characterize the topology on real-valued graphons. In the case of weighted graphs and probability-graphons, we need to replace absence/presence of edges (which is 0-1 valued) by test functions from $C_b(\mathbf{Z})$ decorating the edges. Hence, we define the *homomorphism density* of a $\mathcal{G}$-graph F^g which is a finite graph $F = (V, E)$ whose edges are decorated with a family of functions $g = (g_e)_{e\in E}$ from a subset $\mathcal{G} \subset C_b(\mathbf{Z})$ (in practice, we only consider the cases $\mathcal{G} = C_b(\mathbf{Z})$ or $\mathcal{G} = \mathcal{F} \subset C_b(\mathbf{Z})$ a convergence determining sequence), in a probability-graphon W as:

$$t(F^g, W) = M_W^F(g) := \int_{[0,1]^V} \prod_{(i,j)\in E} W(x_i, x_j; g_{i,j}) \prod_{i\in V} \mathrm{d}x_i,$$

where $W(x, y; f) = \int_{\mathbf{Z}} f(z)\, W(x, y; \mathrm{d}z)$. Moreover, M_W^F defines a measure on $\mathbf{Z}^E$ (which we still denote by M_W^F) which is defined by $M_W^F(\otimes_{e\in E} g_e) = M_W^F(g)$ for $g = (g_e)_{e\in E}$. Note that when F is the complete graph with k vertices, M_W^F is the joint measure of all the edge-weights of the random graph $\mathbb{G}(k, W)$, and thus characterizes the random graph $\mathbb{G}(k, W)$.

In the counting Lemma 7.5 and the weak counting Lemma 7.7, we prove that the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ allows to control the homomorphism densities. Conversely, in the inverse counting Lemma 7.8, we prove that the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ can be controlled by the homomorphism densities. In particular, the topology of the cut distance turns out to be exactly the topology of convergence in distribution for sampled subgraphs of any given size; the next result is a direct consequence of Theorem 7.11.

Theorem 1.6 (Characterization of the topology). — *Let $(W_n)_{n\in\mathbb{N}}$ and W be unlabeled probability-graphons from $\widetilde{\mathcal{W}}_1$. The following properties are equivalent:*

(i) *$(W_n)_{n\in\mathbb{N}}$ converges to W for the topology from Proposition 1.2.*
(ii) *$\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$ for all $C_b(\mathbf{Z})$-graphs F^g.*
(iii) *$\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$ for all $\mathcal{F}$-graphs F^g, for some convergence determining sequence $\mathcal{F}$.*
(iv) *For all $k \geqslant 2$, the sequence of sampled subgraphs $(\mathbb{G}(k, W_n))_{n\in\mathbb{N}}$ converges in distribution to $\mathbb{G}(k, W)$.*

Now, we can turn back to the initial problem of finding a limit object for a convergent sequence of weighted graphs $(G_n)_{n\in\mathbb{N}}$; here convergent means that for all $k \geqslant 2$, the sequence $(\mathbb{G}(k, G_n) = \mathbb{G}(k, W_{G_n}))_{n\in\mathbb{N}}$ of sampled subgraphs of size k (defined above) converges in distribution (to some limit random graph). Note that the tightness criterion for a sequence of probability-graphons $(W_n)_{n\in\mathbb{N}}$ can be equivalently rephrased as tightness of the sequence $(\mathbb{G}(2, W_n))_{n\in\mathbb{N}}$ of sampled subgraphs of size 2. Hence, the convergence in distribution of the sequence $(\mathbb{G}(2, G_n))_{n\in\mathbb{N}}$ implies its tightness, and thus the tightness of the sequence of probability-graphons $(W_{G_n})_{n\in\mathbb{N}}$. Then, Theorem 1.4 guarantees the existence of a probability-graphon W which is a subsequential limit of the sequence $(W_{G_n})_{n\in\mathbb{N}}$ in the cut distance $\delta_{\square,\mathcal{F}}$, and then Theorem 1.6 guarantees that for all $k \geqslant 2$, the sequence $(\mathbb{G}(k, G_n))_{n\in\mathbb{N}}$ converges in distribution to $\mathbb{G}(k, W)$.

As a consequence, probability-graphons are precisely the limit objects for sequences of weighted graphs $(G_n)_{n\in\mathbb{N}}$ (and also for random weighted graphs) whose number of vertices goes to infinity (otherwise the limit would simply be a weighted graph) and such that for each size $k \geqslant 2$, the sequence of sampled subgraphs $(\mathbb{G}(k, G_n))_{n\in\mathbb{N}}$ converges in distribution.

Remark 1.7 (Extension to vertex-weights). — The framework we have developed for probability-graphons could easily be extended to add weights on the vertices, or equivalently to allow for self-loops (i.e. edges linking a vertex to itself). In this case, weighted graphs and probability-graphons have a two-variable kernel (probability-graphon) W^{e} for edge-weights as before, and a one-variable kernel $W^{\mathrm{v}} : [0,1] \to \mathcal{M}_1(\mathbf{Z})$ for vertex-weights. Note that this implies, as expected, that the same measure-preserving map $\varphi : [0,1] \to [0,1]$ must be used for both kernels W^{v} and W^{e} when relabeling.

1.3. Organization of the paper

The rest of the paper is organized as follows. In Section 2, we define some notations used throughout the paper, and recall some properties of the weak topology on the space of signed measures. In Section 3, we define probability-graphons and signed-measure valued kernels, we then define the cut distance and the cut norm and study their properties, and we also give some exemple of distances with the Lévy–Prokhorov distance d_{LP}, the Kanrorovitch-Rubinstein and Fortet–Mourier norms $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$, and the norm $\|\cdot\|_{\mathcal{F}}$ based on a convergence determining sequence. In Section 4, we define the steppings of a probability-graphon (which are stepfunction approximations corresponding to conditional expectations on $[0,1]^2$), we define the tightness criterion for probability-graphons, and we prove the weak regularity property of the cut distance. In Section 5, we prove the theorem linking the tightness criterion with relative compactness for the cut distance, we prove that under some conditions the topology of the cut distance does not depend on the choice of the initial distance d_{m}, and we prove that the space of probability-graphons with the cut distance is a Polish space. In Section 6, we define the subgraph $\mathbb{G}(k, W)$ sampled from a probability-graphon W, we then prove approximation bound in the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ between probability-graphons and their sampled subgraphs. In Section 7, we prove the counting lemmas linking the cut distance with the homomorphism densities, and prove that the topology induced by the cut distance coincides with the topology of convergence in distribution for all the sampled subgraphs.

An index of notations used in the paper is provided in Section 9.

Contents

2. Notations and topology on the space of signed measures

Throughout the article, *measure* will always be used to denote a positive measure.

Let $\mathbb{N}$ be the set of non-negative integers, $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$ the set of positive integers, and, for $n \in \mathbb{N}^*$, we define the integer set $[n] = \{1, \dots, n\}$. For $k \in \mathbb{N}^*$, the set $[0,1]^k$ is endowed with the Borel σ-field and the Lebesgue measure λ_k; and we write λ for λ_k when the context is clear. The supremum of a real-valued function f defined on $[0,1]^k$ is denoted by $\|f\|_\infty = \sup_{x\in[0,1]^k} f(x)$.

Let d be a distance on a topological space $(X, \mathcal{O})$.

(i) The distance d is *continuous* w.r.t. the topology $\mathcal{O}$ if the identity map from $(X, \mathcal{O})$ to (X, d) is continuous.

(ii) The distance d is *sequentially continuous* w.r.t. the topology $\mathcal{O}$ if for any sequence $(x_n)_{n\in\mathbb{N}}$ in X which converges to some limit x for the topology $\mathcal{O}$, we also have that $\lim_{n\to\infty} d(x_n, x) = 0$.

Let d and d' be two distances on a space X. We say that d' is continuous (resp. uniformly continuous) w.r.t. d if the identity map from (X, d) to (X, d') is continuous (resp. uniformly continuous).

Remark 2.1. — If the topology $\mathcal{O}$ is metrizable (i.e. can be generated by a distance on the space X), then the topology on X induced by the distance d is equivalent to $\mathcal{O}$ if and only if for every sequence with values in X, convergence for d is equivalent to convergence for $\mathcal{O}$ (see [23, Theorem 4.1.2]). Moreover, when the topology is metrizable, then topological notions and their sequential counterparts coincides (e.g. compact and sequentially compact sets, closed and sequentially closed sets, see [23, Proposition 4.1.1 and Theorem 4.1.17]).

Remark 2.2. — For a function, continuity always implies sequential continuity; and the converse is also true when the topology is metrizable.

A map $\varphi : \Omega_1 \to \Omega_2$ between two probability spaces $(\Omega_i, \mathcal{A}_i, \pi_i)$, $i = 1, 2$, is measure-preserving if it is measurable and if for every $A \in \mathcal{A}_2$, $\pi_2(A) = \pi_1(\varphi^{-1}(A))$. In this case, for every measurable non-negative function $f : \Omega_2 \to \mathbb{R}$, we have:

$$\int_{\Omega_1} f(\varphi(x))\ \pi_1(\mathrm{d}x) = \int_{\Omega_2} f(x)\ \pi_2(\mathrm{d}x). \tag{2.1}$$

We denote by $S_{[0,1]}$ the set of bijective measure-preserving maps from $[0,1]$ with the Lebesgue measure to itself, and by $\bar{S}_{[0,1]}$ the set of measure-preserving maps from $[0,1]$ with the Lebesgue measure to itself.

Let $(\mathbf{Z}, \mathcal{O}_\mathbf{Z})$ be some (non-empty) Polish space, and let $\mathcal{B}(\mathbf{Z})$ be the Borel σ-field on $\mathbf{Z}$ generated by the topology $\mathcal{O}_\mathbf{Z}$. We denote by $C_b(\mathbf{Z})$ the space of real-valued continuous bounded functions on $(\mathbf{Z}, \mathcal{O}_\mathbf{Z})$. We denote by $\mathcal{M}_\pm(\mathbf{Z})$ the space of finite signed measures on $(\mathbf{Z}, \mathcal{B}(\mathbf{Z}))$; $\mathcal{M}_+(\mathbf{Z})$ the subspace of measures; $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ the subspace of measures with total mass at most 1; and $\mathcal{M}_1(\mathbf{Z})$ the subspace of probability measures. We have:

$$\mathcal{M}_1(\mathbf{Z}) \subset \mathcal{M}_{\leqslant 1}(\mathbf{Z}) \subset \mathcal{M}_+(\mathbf{Z}) \subset \mathcal{M}_\pm(\mathbf{Z}).$$

For a signed measure $\mu \in \mathcal{M}_\pm(\mathbf{Z})$, we recall the definition of the Hahn–Jordan decomposition $\mu = \mu^+ - \mu^-$ where $\mu^+, \mu^- \in \mathcal{M}_+(\mathbf{Z})$ are mutually singular measures (that is $\mu^+(A) = 0$ and $\mu^-(A^c) = 0$ for some measurable set A), as well as the total variation measure of μ which is defined as $|\mu| = \mu^+ + \mu^- \in \mathcal{M}_+(\mathbf{Z})$. Note that for a measure $\mu \in \mathcal{M}_+(\mathbf{Z})$, we simply have $|\mu| = \mu$. For a signed-measure $\mu \in \mathcal{M}_\pm(\mathbf{Z})$ and a real-valued measurable function f defined on $\mathbf{Z}$, we write $\mu(f) = \langle \mu, f\rangle = \int f \, \mathrm{d}\mu = \int_\mathbf{Z} f(x)\, \mu(\mathrm{d}x)$ the integral of f w.r.t. μ whenever it is well defined. For a signed measure $\mu \in \mathcal{M}_\pm(\mathbf{Z})$, we denote by $\|\mu\|_\infty = \mu^+(\mathbf{Z}) + \mu^-(\mathbf{Z})$ its total mass, which is also equal to the supremum of $\mu(f)$ over all measurable functions f with values in $[-1, 1]$.

We endow $\mathcal{M}_\pm(\mathbf{Z})$ with the topology of weak convergence, that is the smallest topology for which the maps $\mu \mapsto \mu(f)$ are continuous for all $f \in C_b(\mathbf{Z})$. In particular, a sequence of signed measures $(\mu_n)_{n\in\mathbb{N}}$ weakly converges to some $\mu \in \mathcal{M}_\pm(\mathbf{Z})$ if and only if, for every function $f \in C_b(\mathbf{Z})$, we have $\lim_{n\to+\infty} \mu_n(f) = \mu(f)$. Let us recall that $\mathcal{M}_+(\mathbf{Z})$ and $\mathcal{M}_1(\mathbf{Z})$ endowed with the topology of weak convergence are Polish spaces.

Remark 2.3 (The weak topology on $\mathcal{M}_\pm(\mathbf{Z})$). — The topology of weak convergence on the set of signed measures $\mathcal{M}_\pm(\mathbf{Z})$ is equivalent to the weak-$*$ topology on $\mathcal{M}_\pm(\mathbf{Z})$ seen as a subspace of the topological dual of $C_b(\mathbf{Z})$ (see the paragraph after Definition 3.1.1 in [14]). As usual in probability theory, this topology will be simply called the weak topology (this is also consistent with [14]).

We recall that a sequence of $[0, 1]$-valued functions $\mathcal{F} = (f_k)_{k\in\mathbb{N}}$ in $C_b(\mathbf{Z})$, with $f_0 = \mathbb{1}$ the constant function equal to one, is:

(i) *Separating* if for every measures μ, ν from $\mathcal{M}_\pm(\mathbf{Z})$ (or equivalently just from $\mathcal{M}_+(\mathbf{Z})$) such that for every $k \in \mathbb{N}$, $\mu(f_k) = \nu(f_k)$, then $\mu = \nu$.

(ii) *Convergence determining* if for every $(\mu_n)_{n\in\mathbb{N}}$ and μ measures from $\mathcal{M}_+(\mathbf{Z})$ such that we have $\lim_{n\to+\infty} \mu_n(f_k) = \mu(f_k)$ for all $k \in \mathbb{N}$, then $(\mu_n)_{n\in\mathbb{N}}$ weakly converges to μ.

Notice that a convergence determining sequence is also separating. A sequence of functions is separating if and only if it separates the points of $\mathbf{Z}$ (see [24, Theorem 3.4.5]). There always exists a convergence determining sequence on Polish spaces, see [14, Corollary 2.2.6] or the proof of Proposition 3.4.4 in [24] (which are stated for probability measures but can be extended to finite positive measures as we required that $\mathbb{1}$ belongs to $\mathcal{F}$). Note that there does not exist a convergence determining sequence for $\mathcal{M}_\pm(\mathbf{Z})$ as the weak topology is not metrizable on $\mathcal{M}_\pm(\mathbf{Z})$ (see Remark 2.6 below).

Remark 2.4 (The Borel σ-field on $\mathcal{M}_\pm(\mathbf{Z})$). — By [14, Corollary 5.1.9], the Borel σ-field on $\mathcal{M}_\pm(\mathbf{Z})$, associated with the weak topology, is countably generated and can be generated by either:

- the family of maps $\mu \mapsto \mu(f_n)$ where the sequence $(f_n)_{n\in\mathbb{N}}$ of functions from $C_b(\mathbf{Z})$ is separating;
- the family of maps $\mu \mapsto \mu(B)$ where $B \in \mathcal{A}$ and the subset $\mathcal{A} \subset \mathcal{B}(\mathbf{Z})$ is countable and generates the whole σ-field $\mathcal{B}(\mathbf{Z})$ (such subset $\mathcal{A}$ always exists, see [13, Corollary 6.7.5]).

Note that the Borel σ-field of a Polish space is generated by any family of Borel functions that separates points (see [13, Theorem 6.8.9]).

Furthermore, the maps $\mu \mapsto \mu^+$ and $\mu \mapsto \mu^-$ (and thus also $\mu \mapsto |\mu|$) are measurable (see [22, Theorem 2.8] and Remark 2.4). As a consequence, the map $\mu \mapsto \|\mu\|_\infty$ is also measurable (in fact it is even lower semicontinuous by [14, Theorem 2.7.4]). Note that $\mathcal{M}_1(\mathbf{Z})$ and $\mathcal{M}_+(\mathbf{Z})$ are closed, and thus measurable, subsets of $\mathcal{M}_\pm(\mathbf{Z})$.

We define the following two important properties for subsets of signed measures, which are related to relative compactness (see Lemma 2.8 below).

Definition 2.5. — *Let $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$ be a subset of signed measures.*

(i) The set $\mathcal{M}$ is bounded *(in total variation) if:*

$$\sup_{\mu\in\mathcal{M}} \|\mu\|_\infty < +\infty.$$

(ii) The set $\mathcal{M}$ is tight *if for all $\varepsilon > 0$, there exists a compact set $K \subset \mathbf{Z}$ such that:*

$$\sup_{\mu\in\mathcal{M}} |\mu|(K^c) \leqslant \varepsilon.$$

Remark 2.6 (On the compact sets and metrizability of the weak topology). — Recall that $\mathbf{Z}$ is a Polish space. We stress that the weak topology on signed measures is not metrizable unless it coincides with the strong topology (see [52, Theorem 4.1]), which happens only when the initial space $\mathbf{Z}$ is finite (see [14, Proposition 3.1.8]).

Moreover, the closed norm ball $\{\mu \in \mathcal{M}_{\pm}(\mathbf{Z}) : \|\mu\|_\infty \leqslant 1\}$ of $\mathcal{M}_{\pm}(\mathbf{Z})$ is metrizable if and only if $\mathbf{Z}$ is compact (see [14, Proposition 3.1.8 and Theorem 3.1.9]).

Let $\mathcal{M} \subset \mathcal{M}_{\pm}(\mathbf{Z})$. The following properties are equivalent (see [14, Theorems 2.3.4 and 3.1.9]):

(i) $\mathcal{M}$ is weakly compact (i.e. $\mathcal{M}$ is compact for the weak topology);
(ii) $\mathcal{M}$ is sequentially weakly compact (that is every sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$ has a subsequence that converges to some limit $\mu \in \mathcal{M}$);
(iii) $\mathcal{M}$ is compact for the sequential weak topology (for which sets are closed if and only if they are closed under weak convergence).

Moreover, when any of those is true, $\mathcal{M}$ is tight, bounded, and metrizable in the weak topology. Furthermore, the Kantorovitch–Rubinshtein and Fortet-Mouriet norms $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$ (defined in Section 3.8.2) can be used to generate the weak topology on a weakly compact set (see [14, Remark 3.2.5]).

Nevertheless, the weak topology on the unit sphere $\{\mu \in \mathcal{M}_{\pm}(\mathbf{Z}) : \|\mu\|_\infty = 1\}$ of $\mathcal{M}_{\pm}(\mathbf{Z})$ is always metrizable with a complete metric, making the unit sphere a Polish space, however, the Kantorovitch–Rubinshtein and Fortet-Mouriet norms $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$ do not provide a complete metrization in this case (see [14, Theorem 3.2.8]).

Remark 2.7 (On the compactness of $\mathcal{M}_1(\mathbf{Z})$). — Let $\mathcal{M}$ be either $\mathcal{M}_1(\mathbf{Z})$, $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ or the closed norm ball $\{\mu \in \mathcal{M}_{\pm}(\mathbf{Z}) : \|\mu\|_\infty \leqslant 1\}$ of $\mathcal{M}_{\pm}(\mathbf{Z})$. Then, $\mathcal{M}$ is weakly compact if and only if $\mathbf{Z}$ is compact.

We give a short proof of this statement. As $\mathcal{M}_1(\mathbf{Z})$ is closed in $\mathcal{M}_{\pm}(\mathbf{Z})$ for the weak topology, if $\mathcal{M}$ is weakly compact, then $\mathcal{M}_1(\mathbf{Z})$ is also weakly compact, and thus $\mathbf{Z}$ is compact by [52, Theorem 3.4]. Conversely, if $\mathbf{Z}$ is compact, then by [14, Theorem 1.3.3], we know that $\mathcal{M}_{\pm}(\mathbf{Z})$ (endowed with the weak topology) is the topological dual space of $C_b(\mathbf{Z})$ (endowed with the uniform convergence topology), thus using Banach–Alaoglu theorem (see [14, Theorem 1.3.6]), we get that the closed unit norm-ball of $\mathcal{M}_{\pm}(\mathbf{Z})$, and thus $\mathcal{M}$, are compact for the weak topology.

We recall the following result, which is an equivalent of Prokhorov's theorem for signed measures.

LEMMA 2.8 (Prokhorov's theorem for signed measures, [14, Theorems 2.3.4 and 3.1.9]). — *Let $\mathbf{Z}$ be a Polish space, and let $\mathcal{M} \subset \mathcal{M}_{\pm}(\mathbf{Z})$ be a* *subset of signed measures on* $\mathbf{Z}$*. Then the following conditions are equivalent:*

(i) $\mathcal{M}$ *is relatively sequentially compact, that is every sequence* $(\mu_n)_{n\in\mathbb{N}}$ *in* $\mathcal{M}$ *contains a subsequence which weakly converges in* $\mathcal{M}_\pm(\mathbf{Z})$*.*

(ii) $\mathcal{M}$ *is relatively compact for the weak topology, that is the closure of* $\mathcal{M}$ *is compact for the weak topology.*

(iii) The family $\mathcal{M}$ *is tight and bounded.*

Remark 2.9 (On the weak sequential topology). — When the space $\mathbf{Z}$ is infinite, the weak topology does not coincide with the weak sequential topology on $\mathcal{M}_\pm(\mathbf{Z})$ (but recall from Remark 2.6 that their compact sets are the same). Recall that if the space $\mathbf{Z}$ is compact, then the unit norm ball of $\mathcal{M}_\pm(\mathbf{Z})$ is metrizable, and thus the weak topology and the weak sequential topology coincide on it. However, if the space $\mathbf{Z}$ is non-compact, then the weak topology and the weak sequential topology do not coincide on the unit norm ball of $\mathcal{M}_\pm(\mathbf{Z})$.

We give a short proof of those statements according to $\mathbf{Z}$ being compact or not.

(i) Recall that when $\mathbf{Z}$ is an infinite compact space (for instance $\mathbf{Z} = [0,1]$), the Banach space $C_b(\mathbf{Z})$ is infinite-dimensional and separable (using Stone–Weierstrass theorem), and its topological dual is $(C_b(\mathbf{Z}))^* = \mathcal{M}_\pm(\mathbf{Z})$ (see [14, Theorem 1.3.3]). Thus, using [33, Theorem 2.5], we get the existence of a countable subset which is weak sequentially closed yet weak dense in $\mathcal{M}_\pm(\mathbf{Z})$. In particular, the weak sequential topology and the weak topology do not coincide on $\mathcal{M}_\pm(\mathbf{Z})$.

(ii) Assume that the space $\mathbf{Z}$ is non-compact. Thus, $\mathbf{Z}$ contains a countable closed subset F whose points are at mutual distances uniformly bounded away from zero. By [14, Remark 3.1.7], the weak topology on $\mathcal{M}_\pm(F)$ for a closed subset F coincides with the trace of the weak topology on the whole space. By [14, Section 3.1, p. 102], $\mathcal{M}_\pm(F)$ is homeomorphic to ℓ^1 both endowed with their weak topology, weak convergence on ℓ^1 is equivalent to norm convergence, and the weak topology on ℓ^1 is not sequential, even on the unit norm ball. Hence, the weak topology on $\mathcal{M}_\pm(\mathbf{Z})$ is not sequential, even on the unit norm ball.

We define the notion of a quasi-convex distance, which generalizes the convexity of a norm.

Definition 2.10 (Quasi-convex distance). — *Let* (X, d) *be a metric space which is a convex subset of a vector space. The distance* d *is* quasi-convex *if for all* $x_1, x_2, y_1, y_2 \in X$ *and all* $\alpha \in [0, 1]$, *we have:*

$$d(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \leqslant \max(d(x_1, y_1), d(x_2, y_2)).$$

In particular, any distance (on a convex subset of a vector space) which derives from a norm is quasi-convex.

Lemma 2.11. — *Let* d_{m} *be distance on* $\mathcal{M}_\epsilon(\mathbf{Z})$ *with* $\epsilon \in \{+, \pm\}$ *which is quasi-convex and sequentially continuous with respect to the weak topology. Then,* d_{m} *is uniformly continuous with respect to* $\|\cdot\|_\infty$ *on* $\mathcal{M}_\epsilon(\mathbf{Z})$.

Proof. — We shall simply consider the case $\mathcal{M} = \mathcal{M}_+(\mathbf{Z})$, the other case being simpler. We first check that for all $\mu \in \mathcal{M}$ and $\varepsilon > 0$, there exists $\eta > 0$ such that for all $\nu \in \mathcal{M}$, we have that $\|\mu - \nu\|_\infty < \eta$ implies $d_{\mathrm{m}}(\mu, \nu) < \varepsilon$. As d_{m} is sequentially continuous w.r.t. the weak topology, it is also (sequentially) continuous w.r.t. the strong topology. Let $\mu \in \mathcal{M}$ and $\varepsilon > 0$. Then, the set $\{\nu \in \mathcal{M} \ : \ d_{\mathrm{m}}(\mu, \nu) < \varepsilon\}$ is an open set of $\mathcal{M}$ containing μ both for d_{m} and for the strong topology. Thus, it contains a neighborhood of μ for the strong topology $\{\nu \in \mathcal{M} \ : \ \|\mu - \nu\|_\infty < \eta\}$ for $\eta > 0$ small enough. This proves the claim.

As d_{m} is quasi-convex and $\mathcal{M}$ is a cone, for $\mu, \nu \in \mathcal{M}$ we have:

$$\begin{aligned} d_{\mathrm{m}}(\mu, \mu + \nu) &= d_{\mathrm{m}}\left(\frac{1}{2} \cdot (2\mu + 0), \frac{1}{2} \cdot (2\mu + 2\nu)\right) \\ &\leqslant \max(d_{\mathrm{m}}(2\mu, 2\mu), d_{\mathrm{m}}(0, 2\nu)) \\ &= d_{\mathrm{m}}(0, 2\nu). \end{aligned}$$

Let $\varepsilon > 0$ be fixed. We choose $\eta \in (0, 1)$ such that $\|\nu\|_\infty < \eta$, with $\nu \in \mathcal{M}$, implies $d_{\mathrm{m}}(0, \nu) < \varepsilon$. Let $\mu, \nu \in \mathcal{M}$ be such that $\|\mu - \nu\|_\infty < \eta/2$. Let $\lambda' = \mu + \nu$ and f (resp. g) the density of μ (resp. ν) with respect to λ'. We set $\pi = \min(f, g)\, \lambda'$, $\mu' = (f - g)_+\, \lambda'$ and $\nu' = (f - g)_-\, \lambda'$ so that $\pi, \mu', \nu' \in \mathcal{M}$, $\mu = \pi + \mu'$ and $\nu = \pi + \nu'$. Since $\mu' - \nu' = \mu - \nu$ and μ' and ν' are mutually singular, we deduce that $\|\mu'\|_\infty + \|\nu'\|_\infty < \eta/2$. We get:

$$\begin{aligned} d_{\mathrm{m}}(\mu, \nu) = d_{\mathrm{m}}(\pi + \mu', \pi + \nu') &\leqslant d_{\mathrm{m}}(\pi, \pi + \mu') + d_{\mathrm{m}}(\pi, \pi + \nu') \\ &\leqslant d_{\mathrm{m}}(0, 2\mu') + d_{\mathrm{m}}(0, 2\nu') \\ &\leqslant 2\varepsilon. \end{aligned}$$

Hence, the distance d_{m} is uniformly continuous with respect to $\|\cdot\|_\infty$ on $\mathcal{M}$. □

3. Measure-valued graphons and the cut distance

In Section 3.1, we introduce the measure-valued graphons, which are a generalization of real-valued graphons (i.e. $[0,1]$-valued measurable functions defined on $[0,1]^2$). We refer to the monography [46] on real-valued graphons for more details. Contrary to the case of $[0,1]$ with its usual topology, there is no canonical distance inducing the weak topology on $\mathcal{M}_1(\mathbf{Z})$. In Sections 3.2, 3.3 and 3.4, starting from a distance d_{m} inducing the weak topology on measures, we introduce the cut distance, and its unlabeled variant, on the space of measure-valued graphons which are analogous to the ones for real-valued graphons (see [46, Chapter 8]). In Section 3.5, we define a weak isomorphism relation for measure-valued graphons based on this distance, which is analogous to the weak isomorphism relation for real-valued graphons (see [46, Sections 7.3 and 10.7]). Then, in Section 3.6, we give an alternative combinatorial formulation of the cut distance for stepfunctions.

3.1. Definition of measure-valued graphons

We start by defining measure-valued kernels and graphons which are a generalization of real-valued kernels and graphons. Recall that $\mathbf{Z}$ is a Polish space and $\mathcal{M}_{\pm}(\mathbf{Z})$ is the space of *finite* signed measures.

Definition 3.1 (Signed measure-valued kernels). — *A* signed measure-valued kernel *or* $\mathcal{M}_{\pm}(\mathbf{Z})$-valued kernel *is a map W from $[0,1]^2$ to $\mathcal{M}_{\pm}(\mathbf{Z})$, such that:*

(i) W is measurable *in (x,y): for every measurable set $A \subset \mathbf{Z}$, the function $(x,y) \mapsto W(x,y;A)$ defined on $[0,1]^2$ is measurable.*

(ii) W is bounded*:*

$$\|W\|_\infty := \sup_{x,y\in[0,1]} \|W(x,y;\cdot)\|_\infty < +\infty. \tag{3.1}$$

We denote by $\mathcal{W}_1$ (resp. $\mathcal{W}_{\leqslant 1}$, resp. $\mathcal{W}_+$, resp. $\mathcal{W}_{\pm}$) the space of probability measure-valued kernels or simply probability-graphons (resp. subprobability measure-valued kernels, resp. measure-valued kernels, resp. signed measure-valued kernels), where we identify kernels that are equal a.e. on $[0,1]^2$, with respect to the Lebesgue measure. Then, (3.1) should be read with an essential supremum instead of a supremum. In what follows, we always assume for simplicity that we choose representatives of measure-valued kernels such that $\|W\|_\infty$ is also the essential supremum of $(x, y) \mapsto \|W(x, y; \cdot)\|_\infty$.

For $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$, we denote by $\mathcal{W}_\mathcal{M}$ the subset of signed measure-valued kernels $W \in \mathcal{W}_\pm$ which are $\mathcal{M}$-valued: $W(x, y; \cdot) \in \mathcal{M}$ for every $(x, y) \in [0, 1]^2$.

Remark 3.2 (On real-valued kernels). — Let $\mathbf{Z} = \{0, 1\}$ be equipped with the discrete topology. Every real-valued graphon w can be represented using a probability-graphon W defined for every $x, y \in [0, 1]$ by $W(x, y; \mathrm{d}z) = w(x, y)\delta_1(\mathrm{d}z) + (1 - w(x, y))\delta_0(\mathrm{d}z)$, where δ_z is the Dirac mass located at z. In particular we have that $w(x, y) = W(x, y; \{1\})$ for $x, y \in [0, 1]$.

Let $W \in \mathcal{W}_\pm$ be a signed measure-valued kernel. Define the map $W^+ : [0, 1]^2 \to \mathcal{M}_+(\mathbf{Z})$ to be the positive part of W, i.e. for every $(x, y) \in [0, 1]^2$, $W^+(x, y; \cdot)$ is the positive part of the measure $W(x, y; \cdot)$. Similarly define $W^- : [0, 1]^2 \to \mathcal{M}_+(\mathbf{Z})$ the negative part of W; and then define $|W| = W^+ + W^-$ the total variation of W and $\|W\| = |W|(\mathbf{Z})$ the total mass of W.

Lemma 3.3 (The positive part W^+ of a kernel). — *The maps W^+, W^- and $|W|$ are all measure-valued kernels, and the map $\|W\| : (x, y) \mapsto \|W(x, y; \cdot)\|_\infty$ is measurable.*

Proof. — The statements for $|W|$ and $\|W\|$ are immediate consequences of the statements for W^+ and W^-; and as the proof for W^+ and W^- are similar, we only need to prove that W^+ is a measure-valued kernel. It is immediate that W^+ is bounded and that for every $(x, y) \in [0, 1]^2$, $W^+(x, y; \cdot)$ is a measure in $\mathcal{M}_+(\mathbf{Z})$. Thus, we are left to prove the measurability of W^+ in (x, y). By [22, Proposition 2.1] and Remark 2.4, a signed measure-valued kernel U is measurable in (x, y) (i.e. for every $A \in \mathcal{B}(\mathbf{Z})$, the map $(x, y) \mapsto U(x, y; A)$ is measurable) if and only if the map $(x, y) \mapsto U(x, y; \cdot)$ is measurable from $[0, 1]^2$ (with its Borel σ-field) to $\mathcal{M}_\pm(\mathbf{Z})$ equipped with the Borel σ-field generated by the weak topology. By [22, Theorem 2.8], the map $\mu \mapsto \mu^+$, that associates to a signed measure the positive part of its Hahn–Jordan decomposition, is measurable from $\mathcal{M}_\pm(\mathbf{Z})$ to $\mathcal{M}_+(\mathbf{Z})$ both endowed with the Borel σ-field generated by the weak topology. Considering the composition of W and $\mu \mapsto \mu^+$, we get that W^+ is measurable in (x, y) and is thus a measure-valued kernel. □

Remark 3.4 (Probability-graphons $W : \Omega \times \Omega \to \mathcal{M}_1(\mathbf{Z})$). — Similarly to the case of real-valued graphons, it is possible to replace the vertex-type space $[0, 1]$ by any standard probability space $(\Omega, \mathcal{A}, \pi)$ that might be more appropriate to represent vertex-types for some applications, and to consider probability-graphons of the form $W : \Omega \times \Omega \to \mathcal{M}_1(\mathbf{Z})$. We recall that a standard probability space $(\Omega, \mathcal{A}, \pi)$ is a probability space such that there exists a measure-preserving map $\varphi : [0,1] \to \Omega$, where $[0,1]$ is endowed with the Borel σ-field and the Lebesgue measure. In particular, every Polish space endowed with its Borel σ-field is a standard probability space. As an example, the space $[0,1]^2$ equipped with the Borel σ-field and the Lebesgue measure λ_2 is a standard probability space; we will reuse this fact later.

Using the measure preserving map φ, it is then possible to consider an unlabeled version W^φ of W constructed on $\Omega' = [0,1]$, and to modify the definition of the cut distance $\delta_{\square,\mathrm{m}}$ similarly as in [35, Theorem 6.9] to allow each probability-graphon to be constructed on different standard probability spaces. For simplicity, in this article we only consider the equivalent case where all probability-graphons are constructed on $\Omega = [0,1]$.

Remark 3.5 (Symmetric kernels). — We shall consider non-symmetric measure-valued kernels and probability-graphons in order to handle directed graphs whose adjacency matrices are thus *a priori* non-symmetric. We say that a measure-valued kernel or graphon W is symmetric if for a.e. $x, y \in [0,1]$, $W(x, y; \cdot) = W(y, x; \cdot)$.

We define stepfunction measure-valued kernels which are often used for approximation.

Definition 3.6 (Signed measure-valued stepfunctions). — *A signed measure-valued kernel $W \in \mathcal{W}_\pm$ is a* stepfunction *if there exists a finite partition of $[0,1]$ into measurable (possibly empty) sets, say $\mathcal{P} = \{S_1, \cdots, S_k\}$, such that W is constant on the sets $S_i \times S_j$, for $1 \leqslant i, j \leqslant k$. We say that W and the partition $\mathcal{P}$ are* adapted *to each other. We write $|\mathcal{P}| = k$ the number of elements of the partition $\mathcal{P}$.*

Throughout the article, we shall only consider partitions composed of measurable subsets.

3.2. The cut distance

We define a distance and a norm on signed measure-valued graphons and kernels, called the *cut distance* and the *cut norm* respectively which are analogous to the cut norm for real-valued graphons and kernels, see [46, Chapter 8]. For a signed measure-valued kernel $W \in \mathcal{W}_\pm$ and a measurable subsets $A \subset [0,1]^2$, we denote by $W(A;\cdot)$ the signed measure on $\mathbf{Z}$ defined by:

$$W(A;\cdot) = \int_A W(x,y;\cdot)\,\mathrm{d}x\mathrm{d}y.$$

Definition 3.7 (The cut distance $d_{\square,\mathrm{m}}$). — *Let d_m be a quasi-convex distance on $\mathcal{M}$ a convex subset of $\mathcal{M}_\pm(\mathbf{Z})$ containing the zero measure. The associated* cut distance $d_{\square,\mathrm{m}}$ *is the function defined on $\mathcal{W}_\mathcal{M}^2$ by:*

$$d_{\square,\mathrm{m}}(U,W) = \sup_{S,T\subset[0,1]} d_\mathrm{m}\Big(U(S\times T;\cdot), W(S\times T;\cdot)\Big), \tag{3.2}$$

where the supremum is taken over all measurable subsets S and T of $[0,1]$.

Notice that the right-hand side of (3.2) is well defined as $\mathcal{M}$ contains the zero measure (and thus if U belongs to $\mathcal{W}_\mathcal{M}$ then $U(A;\cdot)$ belongs to $\mathcal{M}$).

Definition 3.8 (The cut norm $N_{\square,\mathrm{m}}$). — *The* cut norm $N_{\square,\mathrm{m}}$ *associated with a norm N_m on $\mathcal{M}_\pm(\mathbf{Z})$ is the function defined on $\mathcal{W}_\pm$ by:*

$$N_{\square,\mathrm{m}}(W) = \sup_{S,T\subset[0,1]} N_\mathrm{m}\Big(W(S\times T;\cdot)\Big),$$

where the supremum is taken over all measurable subsets S and T of $[0,1]$.

The next proposition states that the cut distance (resp. norm) is indeed a distance (resp. norm); its extension to distances on $\mathcal{M}_+(\mathbf{Z})$ and $\mathcal{M}_\pm(\mathbf{Z})$ is immediate.

Proposition 3.9 ($d_{\square,\mathrm{m}}$ is a distance, $N_{\square,\mathrm{m}}$ is a norm). — *The cut distance $d_{\square,\mathrm{m}}$ associated with a distance d_m on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$) is a distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$). The cut norm $N_{\square,\mathrm{m}}$ associated with a norm N_m on $\mathcal{M}_\pm(\mathbf{Z})$ is a norm on $\mathcal{W}_\pm$.*

Moreover, when the distance d_m on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$) derives from a norm N_m on $\mathcal{M}_\pm(\mathbf{Z})$, then the distance $d_{\square,\mathrm{m}}$ derives also from the norm $N_{\square,\mathrm{m}}$.

Proof. — Let d_m be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (the proof for the case $\mathcal{M}_+(\mathbf{Z})$ is similar). It is clear that $d_{\square,\mathrm{m}}$ is symmetric and satisfies the triangular inequality. Thus, we only need to prove that $d_{\square,\mathrm{m}}$ is separating. Let U and W be two probability-graphons such that $d_{\square,\mathrm{m}}(U,W) = 0$. Then, for every measurable subsets $S,T\subset[0,1]$, we have $U(S\times T;\cdot) = W(S\times T;\cdot)$. Let $\mathcal{F} = (f_k)_{k\in\mathbb{N}}$ be a separating sequence. For every $k\in\mathbb{N}$, and for every measurable subsets $S,T\subset[0,1]$, we have that $U(S\times T; f_k) = W(S\times T; f_k)$. This implies that $U(x,y;f_k)\,\mathrm{d}x\mathrm{d}y = W(x,y;f_k)\,\mathrm{d}x\mathrm{d}y$ for all $k\in\mathbb{N}$. Hence, we deduce that for all $k\in\mathbb{N}$, $U(x,y;f_k) = W(x,y;f_k)$ for almost every

$(x,y) \in [0,1]^2$. Thus, $U(x,y;\cdot) = W(x,y;\cdot)$ for almost every $(x,y) \in [0,1]^2$. This implies that $d_{\Box,\mathrm{m}}$ is separating on $\mathcal{W}_1$, and thus a distance on $\mathcal{W}_1$.

The proof for the cut norm is similar. The proof of the last part of the proposition is clear. $\Box$

3.3. Graphon relabeling, invariance and smoothness properties

The analogue of graph relabelings for graphons are measure-preserving maps. Recall the definition of a measure-preserving map from Section 2, and in particular (2.1). Recall $\bar{S}_{[0,1]}$ denotes the set of measure-preserving (measurable) maps from $[0,1]$ to $[0,1]$ endowed with the Lebesgue measure, and $S_{[0,1]}$ denotes its subset of bijective maps.

The relabeling of a signed measure-valued kernel W by a measure-preserving map φ, is the signed measure-valued kernel W^φ defined for every $x,y \in [0,1]$ and every measurable set $A \subset \mathbf{Z}$ by:

$$W^\varphi(x,y;A) = W(\varphi(x),\varphi(y);A) \quad \text{for } x,y \in [0,1] \text{ and } A \subset \mathbf{Z} \text{ measurable.}$$

We say that a subset $\mathcal{K} \subset \mathcal{W}_\pm$ is *uniformly bounded* if:

$$\sup_{W \in \mathcal{K}} \|W\|_\infty < +\infty. \tag{3.3}$$

Definition 3.10 (Invariance and smoothness of a distance on kernels). *Let d be a distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$). We say that the distance d is:*

(i) ***Invariant****: if $d(U,W) = d(U^\varphi, W^\varphi)$ for every bijective measure-preserving map $\varphi \in S_{[0,1]}$ and $U,V \in \mathcal{W}_1$ (resp. U,V belongs to $\mathcal{W}_+$ or $\mathcal{W}_\pm$).*

(ii) ***Smooth****: if a.e. weak convergence implies convergence for d, that is, if $(W_n)_{n\in\mathbb{N}}$ and W are kernels from $\mathcal{W}_1$ (resp. kernels from $\mathcal{W}_+$ or $\mathcal{W}_\pm$ that are uniformly bounded and) such that for a.e. $(x,y) \in [0,1]^2$, $W_n(x,y;\cdot)$ weakly converges to $W(x,y;\cdot)$ as $n \to \infty$, then $\lim_{n\to\infty} d(W_n,W) = 0$.*

We say that a norm N on $\mathcal{W}_\pm$ is invariant (resp. smooth) if its associated distance d on $\mathcal{W}_\pm$ is invariant (resp. smooth).

We shall see in Section 3.8 some examples of distances d_{m} for which the associated cut distance $d_{\Box,\mathrm{m}}$ is invariant and smooth. The invariance property from Definition 3.10 is always satisfied by the cut distance, and thus also by the cut norm.

Lemma 3.11 ($d_{\square,\mathrm{m}}$ is invariant). — *Let* d_{m} *be a distance on* $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ *(resp.* $\mathcal{M}_{+}(\mathbf{Z})$*, resp.* $\mathcal{M}_{\pm}(\mathbf{Z})$*). Then the cut distance* $d_{\square,\mathrm{m}}$ *on* $\mathcal{W}_1$ *(resp.* $\mathcal{W}_+$*, resp.* $\mathcal{W}_\pm$*) is invariant.*

Proof. — For a signed measure-valued kernel W, a bijective measure-preserving map $\varphi \in S_{[0,1]}$, and measurable sets $S, T \subset [0,1]$, we have thanks to (2.1):

$$\int_{S\times T} W^{\varphi}(x,y;\cdot)\,\mathrm{d}x\mathrm{d}y = \int_{S\times T} W(\varphi(x),\varphi(y);\cdot)\,\mathrm{d}x\mathrm{d}y = \int_{\varphi(S)\times\varphi(T)} W(x,y;\cdot)\,\mathrm{d}x\mathrm{d}y.$$

Hence, taking the supremum over every measurable sets $S, T \subset [0,1]$, we get that the cut distance $d_{\square,\mathrm{m}}$ is invariant. □

When a smooth distance on $\mathcal{W}_1$ or $\mathcal{W}_+$ derives from a distance on $\mathcal{M}_1(\mathbf{Z})$ or $\mathcal{M}_+(\mathbf{Z})$, we have the following result.

Lemma 3.12 (Smoothness and the weak topology). — *Let* d_{m} *be a distance on* $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ *(resp.* $\mathcal{M}_{+}(\mathbf{Z})$ *or* $\mathcal{M}_{\pm}(\mathbf{Z})$*) such that the distance* $d_{\square,\mathrm{m}}$ *on* $\mathcal{W}_1$ *(resp.* $\mathcal{W}_+$ *or* $\mathcal{W}_\pm$*) is smooth. Then, the distance* d_{m} *is continuous w.r.t. the weak topology on* $\mathcal{M}_1(\mathbf{Z})$ *(resp.* $\mathcal{M}_+(\mathbf{Z})$*).*

Proof. — Let $(\mu_n)_{n\in\mathbb{N}}$, and μ be measures from $\mathcal{M}_1(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$) such that $(\mu_n)_{n\in\mathbb{N}}$ weakly converges to μ. Consider the constant measure-valued graphons (resp. kernels) $W_n \equiv \mu_n$, $n \in \mathbb{N}$, and $W \equiv \mu$. Then, for every $x, y \in [0,1]$, $W_n(x,y;\cdot)$ weakly converges to $W(x,y;\cdot)$ as $n \to \infty$. As the distance $d_{\square,\mathrm{m}}$ is smooth, we get that $\lim_{n\to\infty} d_{\square,\mathrm{m}}(W_n, W) = 0$. Considering $S = T = [0,1]$ in the cut distance, we deduce that $\lim_{n\to\infty} d_{\mathrm{m}}(\mu_n,\mu) = 0$. □

The next lemma is a partial converse of Lemma 3.12, it gives sufficient conditions for $d_{\square,\mathrm{m}}$ to be smooth. Recall the definition of a quasi-convex distance in Definition 2.10.

Proposition 3.13 ($d_{\square,\mathrm{m}}$ is smooth). — *Let* d_{m} *be distance on* $\mathcal{M}_\epsilon(\mathbf{Z})$ *with* $\epsilon \in \{+,\pm\}$ *which is quasi-convex and sequentially continuous w.r.t. the weak topology (on* $\mathcal{M}_\epsilon(\mathbf{Z})$*). Then, the cut distance* $d_{\square,\mathrm{m}}$ *is smooth.*

Moreover, for all $U, W \in \mathcal{W}_\epsilon$*, and for all measurable* $A \subset [0,1]^2$*, we have:*

$$d_{\mathrm{m}}(U(A;\cdot), W(A;\cdot)) \leqslant \operatorname*{essup}_{(x,y)\in A} d_{\mathrm{m}}(U(x,y;\cdot), W(x,y;\cdot)). \tag{3.4}$$

To prove Proposition 3.13, we first need to prove the following lemma for approximation by $\mathcal{M}$-valued kernels taking finitely many values.

Lemma 3.14. — *Let $W \in \mathcal{W}_\pm$ and consider a subset $A \subset [0,1]^2$. There exists a sequence $(W_n)_{n\in\mathbb{N}}$ in $\mathcal{W}_\pm$ such that $(W_n(A;\cdot))_{n\in\mathbb{N}}$ weakly converges to $W(A;\cdot)$ and for all $n \in \mathbb{N}$, W_n is finitely valued and takes its values in $\{W(x,y;\cdot) : (x,y) \in A\}$.*

Proof. — By scaling, we may assume that $\|W\|_\infty \leqslant 1$. Let $(f_k)_{k\in\mathbb{N}}$ be a convergence determining sequence with $f_0 = \mathbb{1}$ and f_k taking values in $[0,1]$. Thus, for all $(x,y) \in [0,1]^2$, $\epsilon \in \{\pm 1\}$ and $k \in \mathbb{N}$, we have $W_\epsilon(x,y;f_k) \in [0,1]$. For all $n \in \mathbb{N}$, let $(C_{n,j})_{1\leqslant j\leqslant d_n}$ be a partition of $[0,1]^{2(n+1)}$ into $d_n = n^{2(n+1)}$ hypercubes of edge-length $r_n = 1/n$. Then, for all $n \in \mathbb{N}$ and $j \in [d_n]$, define $B_{n,j} = A \cap (W_+(\cdot;(f_j)_{0\leqslant j\leqslant n}), W_-(\cdot;(f_j)_{0\leqslant j\leqslant n}))^{-1}(C_{n,j})$; thus we get a partition $(B_{n,j})_{1\leqslant j\leqslant d_n}$ of A. If $B_{n,j} \neq \emptyset$, fix some $\mu_{n,j} \in \{W(x,y;\cdot) : (x,y) \in B_{n,j}\}$. If $A \neq [0,1]^2$, fix some $\mu_\partial \in \{W(x,y;\cdot) : (x,y) \in [0,1]^2 \setminus A\}$. For $n \in \mathbb{N}$, we define $W_n = \mathbb{1}_{A^c}\, \mu_\partial + \sum_{j=1}^{d_n} \mathbb{1}_{B_{n,j}}\, \mu_{n,j}$, which is finitely valued and takes its values in $\{W(x,y;\cdot) : (x,y) \in A\}$.

Let $k \in \mathbb{N}$ and $\epsilon \in \{\pm\}$. For all $n \geqslant k$, we have:

$$|W_\epsilon(A;f_k) - (W_n)_\epsilon(A;f_k)| \leqslant \sum_{j=1}^{d_n} \int_{B_{n,j}} |W_\epsilon(x,y;f_k) - (\mu_{n,j})_\epsilon| \,\mathrm{d}x\mathrm{d}y \leqslant \frac{1}{n}\cdot$$

As $(f_k)_{k\in\mathbb{N}}$ is convergence determining, this implies that $((W_n)_\epsilon(A;\cdot))_{n\in\mathbb{N}}$ weakly converges to $W_\epsilon(A;\cdot)$ for $\epsilon \in \{\pm\}$. Hence, $(W_n(A;\cdot))_{n\in\mathbb{N}}$ weakly converges to $W(A;\cdot)$. □

Proof of Proposition 3.13. — As d_{m} is quasi-convex, (3.4) is immediate when U and W take only finitely many values. Now, assume that U and W are arbitrary $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels. Let $\varepsilon > 0$. As d_{m} is sequentially continuous w.r.t. the weak topology, using Lemma 3.14, there exist two $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels U' and W' such that $d_{\mathrm{m}}(U'(A;\cdot), U(A;\cdot)) < \varepsilon$ and U' is finitely valued and takes its values in $\{U(x,y;\cdot) : (x,y) \in A\}$, and similarly for W' and W. Thus, we have:

$$d_{\mathrm{m}}(U(A;\cdot), W(A;\cdot)) \leqslant 2\varepsilon + \operatorname*{essup}_{(x,y)\in A} d_{\mathrm{m}}(U(x,y;\cdot), W(x,y;\cdot)),$$

and this being true for all $\varepsilon > 0$, we get (3.4).

Let $(W_n)_{n\in\mathbb{N}}$ and W be $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels which are uniformly bounded by some constant $C < \infty$ and such that for a.e. $(x,y) \in [0,1]^2$, the sequence $((W_n(x,y;\cdot))_{n\in\mathbb{N}}$ converges to $W(x,y;\cdot)$ for the weak topology, and thus also for d_{m}. Let $\varepsilon > 0$ and $S, T \subset [0,1]$. As d_{m} is quasi-convex and sequentially continuous w.r.t. the weak topology, using Lemma 2.11, there exists $\eta > 0$ such that for all $\mu, \nu \in \mathcal{M}_\epsilon(\mathbf{Z})$, we have that $\|\mu - \nu\|_\infty < \eta$

implies $d_{\mathrm{m}}(\mu,\nu) < \varepsilon$. For all $n \in \mathbb{N}$, define the measurable set:

$$A_n = \{(x,y) \in S \times T \ : \ d_{\mathrm{m}}(W_n(x,y;\cdot), W(x,y;\cdot)) < \varepsilon\}.$$

By assumption, we have that $\lim_{n\to\infty} \lambda(A_n) = \lambda(S \times T)$. Let $N \in \mathbb{N}$ be such that for $n \geqslant N$, we have $\lambda((S \times T) \setminus A_n) < \eta/C$. Let $n \geqslant N$. Remark that $W_n((S \times T) \setminus A_n;\cdot)$ and $W((S \times T) \setminus A_n;\cdot)$ have total mass at most $C\lambda(A_n^c) < \eta$. Thus, we have that $d_{\mathrm{m}}(W_n(A_n;\cdot), W_n(S \times T;\cdot)) < \varepsilon$ and $d_{\mathrm{m}}(W(A_n;\cdot), W(S \times T;\cdot)) < \varepsilon$. Hence, using (3.4) we get that:

$$\begin{aligned} d_{\mathrm{m}}(W_n(S \times T;\cdot), W(S \times T;\cdot)) &\leqslant 2\varepsilon + d_{\mathrm{m}}(W_n(A_n;\cdot), W(A_n;\cdot)) \\ &\leqslant 2\varepsilon + \operatorname*{essup}_{(x,y)\in A_n} d_{\mathrm{m}}(W_n(x,y;\cdot), W(x,y;\cdot)) \\ &\leqslant 3\varepsilon. \end{aligned}$$

Taking the supremum over $S, T \subset [0,1]$, we get $d_{\square,\mathrm{m}}(W_n, W) \leqslant 3\varepsilon$. This being true for all $\varepsilon > 0$, we conclude that $(W_n)_{n\in\mathbb{N}}$ converges to W for $d_{\square,\mathrm{m}}$, and thus $d_{\square,\mathrm{m}}$ is smooth. □

3.4. The unlabeled cut distance

We can now define the cut distance for unlabeled graphons.

Definition 3.15 (The unlabeled cut distance $\delta_{\square,\mathrm{m}}$). — *Set $\mathcal{K} \in \{\mathcal{W}_1, \mathcal{W}_+, \mathcal{W}_\pm\}$. Let d be an invariant distance on the kernel space $\mathcal{K}$. The pseudometric $\delta_\square$ on $\mathcal{K}$, also called the* cut distance*, is defined by:*

$$\delta_\square(U,W) = \inf_{\varphi\in S_{[0,1]}} d(U, W^\varphi) = \inf_{\varphi\in S_{[0,1]}} d\left(U^\varphi, W\right). \tag{3.5}$$

Notice that $\delta_\square$ satisfies the symmetry property (as d is invariant) and the triangular inequality. Hence, $\delta_\square$ induces a distance (that we still denote by $\delta_\square$) on the quotient space $\widetilde{\mathcal{K}}_d = \mathcal{K}/\sim_d$ of kernels in $\mathcal{K}$ associated with the equivalence relation $\sim_d$ defined by $U \sim_d W$ if and only if $\delta_\square(U,W) = 0$.

When the metric $d = d_{\square,\mathrm{m}}$ on $\mathcal{K} = \mathcal{W}_1$ (resp. $\mathcal{W}_+$, resp. $\mathcal{W}_\pm$) derives from a metric d_{m} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$, resp. $\mathcal{M}_\pm(\mathbf{Z})$), and is thus invariant thanks to Lemma 3.11, we write $\delta_{\square,\mathrm{m}}$ for $\delta_\square$ and $\widetilde{\mathcal{K}}_{\mathrm{m}}$ for $\widetilde{\mathcal{K}}_{d_{\square,\mathrm{m}}}$. We shall see in Theorem 5.5 and Corollary 5.6 that under some conditions, different choices of distance d_{m}, which induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, lead to the same quotient space, then simply denoted by $\widetilde{\mathcal{W}}_1$, with the same topology.

3.5. Weak isomorphism

Similarly to Theorem 8.13 in [46], when the distance d_m is such that $d_{\square,\mathrm{m}}$ is invariant and smooth, we can rewrite the cut distance $\delta_{\square,\mathrm{m}}$ as a minimum instead of an infimum using measure-preserving maps, see the last equality in (3.6).

We introduce a weak isomorphism relation that allows to "un-label" probability-graphons.

Definition 3.16 (Weak isomorphism). — *We say that two signed measure-valued kernels U and W are* weakly isomorphic (*and we note $U \sim W$) if there exists two measure-preserving maps $\varphi, \psi \in \bar{S}_{[0,1]}$ such that $U^\varphi(x,y;\cdot) = W^\psi(x,y;\cdot)$ for a.e. $x,y \in [0,1]$.*

We denote by $\widetilde{\mathcal{W}}_\mathcal{M} = \mathcal{W}_\mathcal{M}/\sim$ the space of unlabeled $\mathcal{M}$-valued kernels i.e. the space of $\mathcal{M}$-valued kernels where we identify $\mathcal{M}$-valued kernels that are weakly isomorphic. For $\mathcal{M} = \mathcal{M}_\epsilon(\mathbf{Z})$ with $\epsilon \in \{1, \leqslant 1, +, \pm\}$, we simply write $\widetilde{\mathcal{W}}_\epsilon = \widetilde{\mathcal{W}}_\mathcal{M} = \mathcal{W}_\epsilon/\sim$.

Notice that $U \sim W$ implies that $\|U\|_\infty = \|W\|_\infty$ (we recall that signed measure-valued kernels are only defined for a.e. $x, y \in [0,1]$ and that $\|W\|_\infty$ in (3.1) is an essup in general). In particular, the notion of uniformly bounded subset defined in (3.3) naturally extends to $\widetilde{\mathcal{W}}_\pm$. The last part of this section is devoted to the proof of the following key result.

Theorem 3.17 (Weak isomorphism and $\delta_\square$). — *Let d be a distance defined on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) which is invariant and smooth. Then, two kernels are weakly isomorphic, i.e. $U \sim W$, if and only if $U \sim_d W$,* i.e. *$\delta_\square(U,W) = 0$.*

Furthermore, the map $\delta_\square$ is a distance on $\widetilde{\mathcal{W}}_1 = \widetilde{\mathcal{W}}_{1,d}$ (resp. $\widetilde{\mathcal{W}}_+ = \widetilde{\mathcal{W}}_{+,d}$ or $\widetilde{\mathcal{W}}_\pm = \widetilde{\mathcal{W}}_{\pm,d}$).

As a first step in the proof of Theorem 3.17, following [46], we give a nice description of $\delta_\square$ using couplings. We say that a measure μ on $[0,1]^2$ is a coupling measure on $[0,1]^2$ (between two copies of $[0,1]$ each equipped with the Lebesgue measure) if the projection maps on each component $\tau, \rho : [0,1]^2 \to [0,1]$ (where $[0,1]^2$ is equipped with the measure μ and $[0,1]$ with the Lebesgue measure λ) are measure-preserving. In particular, for every kernel W on $([0,1], \mathcal{B}([0,1]), \lambda)$, the function W^τ is a kernel on the probability space $([0,1]^2, \mathcal{B}([0,1]^2), \mu)$, and similarly for the projection ρ.

Let φ be a given measure-preserving map from $[0,1]$ with the Lebesgue measure to $[0,1]^2$ with a coupling measure μ. For an invariant distance d on

$\mathcal{W}_1$ (resp. $\mathcal{W}_\pm$), we define a distance, say d^μ, on kernels on $([0,1]^2, \mathcal{B}([0,1]^2), \mu)$ by:

$$d^\mu(U', W') = d(U'^\varphi, W'^\varphi).$$

It is easy to see that, for U and W kernels on $[0,1]$, we have $d^\mu(U^\tau, W^\tau) = d(U,W)$ as d is invariant and $\tau \circ \varphi$ is a measure-preserving map from $[0,1]$ to itself; and similarly $d^\mu(U^\rho, W^\rho) = d(U,W)$.

A straightforward adaptation of the proof of [46, Theorem 8.13] gives the next result.

Proposition 3.18 (Minima in the cut distance $\delta_\square$). — *Let d be a distance defined on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) which is invariant and smooth. Then, we have the following alternative formulations for the cut distance $\delta_\square$ on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$):*

$$\begin{aligned} \delta_\square(U,W) &= \inf_{\varphi \in S_{[0,1]}} d(U, W^\varphi) = \inf_{\varphi \in \bar{S}_{[0,1]}} d(U, W^\varphi) \\ &= \inf_{\psi \in S_{[0,1]}} d(U^\psi, W) = \inf_{\psi \in \bar{S}_{[0,1]}} d(U^\psi, W) \\ &= \inf_{\varphi,\psi \in S_{[0,1]}} d(U^\psi, W^\varphi) = \min_{\varphi,\psi \in \bar{S}_{[0,1]}} d(U^\psi, W^\varphi), \end{aligned} \tag{3.6}$$

and

$$\delta_\square(U,W) = \min_\mu d^\mu\left(U^\tau, W^\rho\right)$$

where μ range over all coupling measures on $[0,1]^2$.

Proof of Theorem 3.17. — We deduce from the last equality in (3.6) that $\delta_\square(U,W) = 0$ if and only if there exist measure-preserving maps $\varphi, \psi \in \bar{S}_{[0,1]}$ such that $U^\psi(x,y;\cdot) = W^\varphi(x,y;\cdot)$ for a.e. $x,y \in [0,1]$. This gives that the equivalence relations $\sim_d$ and $\sim$ are the same. $\square$

3.6. The cut norm for stepfunctions

For a quasi-convex distance d_{m}, the cut distance d_{m} for stepfunctions can be reformulated using a finite combinatorial optimization. For a collection of subsets $\mathcal{P}$, denote by $\sigma(\mathcal{P})$ the σ-field generated by $\mathcal{P}$.

Lemma 3.19 (Combinatorial optimization of quasi-convex d_{m} for stepfunctions). — *Let d_{m} be a quasi-convex distance on $\mathcal{M}$ a convex subset of $\mathcal{M}_\pm(\mathbf{Z})$ containing the zero measure. Let $U, W \in \mathcal{W}_\mathcal{M}$ be $\mathcal{M}$-valued stepfunctions adapted to the same finite partition $\mathcal{P}$. Then, there exists $S, T \in \sigma(\mathcal{P})$ such that:*

$$d_{\square,\mathrm{m}}(U,W) = d_{\mathrm{m}}(U(S\times T;\cdot), W(S\times T;\cdot)).$$

Proof. — Let $\mathcal{P} = \{S_1, \dots, S_k\}$ with $k = |\mathcal{P}|$ the size of the partition $\mathcal{P}$. First, remark that the quantity $d_{\square,\mathrm{m}}(U, W) = d_{\mathrm{m}}(U(S' \times T'; \cdot), W(S' \times T'; \cdot))$ depends on S' and T' only through the values of $\lambda(S' \cap S_i)$ and $\lambda(T' \cap S_i)$ for $1 \leqslant i \leqslant k$. Thus, the cut distance between U and W can be reformulated as:

$$d_{\square,\mathrm{m}}(U, W) = \sup_{0 \leqslant \alpha_i, \beta_i \leqslant \lambda(S_i);\, 1 \leqslant i \leqslant k} d_{\mathrm{m}} \left(\sum_{1 \leqslant i,j \leqslant k} \alpha_i \beta_j\, \mu_{i,j}(\cdot), \sum_{1 \leqslant i,j \leqslant k} \alpha_i \beta_j\, \nu_{i,j}(\cdot) \right),$$

where $\mu_{i,j}$ (resp. $\nu_{i,j}$) is the constant value of $U(x, y; \cdot)$ (resp. $W(x, y; \cdot)$) when $x \in S_i$ and $y \in S_j$. Moreover, when we fix the value of $\beta = (\beta_i)_{1 \leqslant i \leqslant k}$, the quantity

$$d_{\mathrm{m}} \left(\sum_{1 \leqslant i,j \leqslant k} \alpha_i \beta_j\, \mu_{i,j}(\cdot), \sum_{1 \leqslant i,j \leqslant k} \alpha_i \beta_j\, \nu_{i,j}(\cdot) \right)$$

is a quasi-convex function of $\alpha = (\alpha_i)_{1 \leqslant i \leqslant k}$, and thus realizes its maximum on the extremal points of the hypercube $\prod_{i=1}^{k} [0, \lambda(S_i)]$, i.e. when α_i equals 0 or $\lambda(S_i)$ for every $1 \leqslant i \leqslant k$. By symmetry, a similar argument holds for β. The cut distance can thus be reformulated as the combinatorial optimization:

$$d_{\square,\mathrm{m}}(U, W) = \max_{\mathrm{I},\mathrm{J} \subset [k]} d_{\mathrm{m}} \left(\sum_{i \in \mathrm{I}, j \in \mathrm{J}} \mu_{i,j}(\cdot), \sum_{i \in \mathrm{I}, j \in \mathrm{J}} \nu_{i,j}(\cdot) \right).$$

Let $I, J \subset [k]$ be index sets that maximize this combinatorial optimization, and take $S = \cup_{i \in I} S_i$ and $T = \cup_{j \in J} S_j$ to conclude. □

3.7. The supremum in S and T in the cut distance $d_{\square,\mathrm{m}}$

In this section, we prove that the supremum in the cut distance $d_{\square,\mathrm{m}}$ is achieved by some subsets $S, T \subset [0, 1]$.

For $W \in \mathcal{M}_{\pm}(\mathbf{Z})$ and $f, g : [0, 1] \to [0, 1]$ measurable, we define the signed measure:

$$W(f \otimes g; \cdot) = \int_{[0,1]^2} W(x, y; \cdot) f(x) g(y)\ \mathrm{d}x \mathrm{d}y.$$

Remark that if we have $W \in \mathcal{W}_\epsilon$ with $\epsilon \in \{1, \leqslant 1, +, \pm\}$, then we have $W(f \otimes g; \cdot) \in \mathcal{M}_\epsilon(\mathbf{Z})$.

Note that a similar result was already known for real-valued kernels, see [46, Lemma 8.10].

LEMMA 3.20 (The supremum in the cut distance $d_{\square,\mathrm{m}}$ for quasi-convex distance d_{m}). — *Let d_{m} be a quasi-convex distance on $\mathcal{M}_\epsilon(\mathbf{Z})$ with $\epsilon \in \{+,\pm\}$ that is sequentially continuous w.r.t. the weak topology. Let $U, W \in \mathcal{W}_\epsilon$. Then, there exist measurable subsets $S, T \subset [0,1]$ such that $f = \mathbb{1}_S$ and $g = \mathbb{1}_T$ achieve the supremum in:*

$$\sup_{f,g} d_{\mathrm{m}}\Big(U(f \otimes g; \cdot), W(f \otimes g; \cdot)\Big)$$

where the supremum is taken over measurable functions f, g from $[0,1]$ to itself.

In the proof and later on we shall use the following notation. For $W \in \mathcal{W}_\pm$ and $f \in C_b(\mathbf{Z})$, we denote by $W[f]$ the real-valued kernel defined by:

$$W[f](x,y) = W(x,y;f) = \int_{\mathbf{Z}} f(z)\ W(x,y;\mathrm{d}z). \tag{3.7}$$

Proof. — Define the map $\Psi : (f,g) \mapsto d_{\mathrm{m}}(U(f \otimes g;\cdot), W(f \otimes g;\cdot))$, and denote $C = \sup_{f,g} \Psi(f,g)$, where the supremum is taken over measurable functions f, g from $[0,1]$ to itself. Let $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ be sequences of measurable functions from $[0,1]$ to itself such that $\lim_{n\to\infty} \Psi(f_n, g_n) = C$. As the unit ball of $L^\infty([0,1],\lambda)$ is compact for the weak-$*$ topology (with primal space $L^1([0,1],\lambda)$), upon taking subsequences, we may assume that $(f_n)_{n\in\mathbb{N}}$ (resp. $(g_n)_{n\in\mathbb{N}}$) weak-$*$ converges to some f (resp. g) which take values in $[0,1]$. Thus, $(f_n \otimes g_n)_{n\in\mathbb{N}}$ weak-$*$ converges to $f \otimes g$ in $L^\infty([0,1]^2, \lambda_2)$. In particular, for every $h \in C_b(\mathbf{Z})$, as $W[h]$ is a real-valued kernel, this implies that $\lim_{n\to\infty} W(f_n \otimes g_n; h) = W(f \otimes g; h)$. This being true for every $h \in C_b(\mathbf{Z})$, we get that the sequence $(W(f_n \otimes g_n;\cdot))_{n\in\mathbb{N}}$ in $\mathcal{M}_\epsilon(\mathbf{Z})$ weakly converges to $W(f \otimes g;\cdot) \in \mathcal{M}_\epsilon(\mathbf{Z})$; and similarly for U. As d_{m} is sequentially continuous w.r.t. the weak topology on $\mathcal{M}_\epsilon(\mathbf{Z})$, we get that $C = \lim_{n\to\infty} \Psi(f_n, g_n) = \Psi(f,g)$.

Now, we show that we can replace the functions f and g by functions that only take the values 0 and 1 (i.e. indicator functions). We first fix g and do this for f. Let X be a random variable uniformly distributed over $[0,1]$, and consider the random function $\mathbb{1}_{X \leqslant f}$. Remark that we have $\mathbb{E}[W(\mathbb{1}_{X\leqslant f} \otimes g;\cdot)] = W(f \otimes g;\cdot)$, and similarly for U. As d_{m} is quasi-convex and sequentially continuous w.r.t. the weak topology, we have:

$$\begin{aligned} C &\geqslant \sup_{x\in[0,1]} d_{\mathrm{m}}(U(\mathbb{1}_{x\leqslant f}\otimes g;\cdot), W(\mathbb{1}_{x\leqslant f}\otimes g;\cdot)) \\ &\geqslant d_{\mathrm{m}}\Big(\mathbb{E}[U(\mathbb{1}_{X\leqslant f}\otimes g;\cdot)], \mathbb{E}[W(\mathbb{1}_{X\leqslant f}\otimes g;\cdot)]\Big) \\ &= d_{\mathrm{m}}(U(f\otimes g;\cdot), W(f\otimes g;\cdot)) \\ &= C, \end{aligned}$$

where in the second equality we used the quasi-convex supremum inequality from (3.4) with the $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels $U'(x,y;\cdot) = U(\mathbb{1}_{x\leqslant f}\otimes g;\cdot)$ and $W'(x,y;\cdot) = W(\mathbb{1}_{x\leqslant f}\otimes g;\cdot)$, and $A = [0,1]^2$. All inequalities being equalities, this imposes:

$$\begin{aligned} C &= \sup_{x\in[0,1]} d_{\mathrm{m}}\big(U(\mathbb{1}_{x\leqslant f}\otimes g;\cdot), W(\mathbb{1}_{x\leqslant f}\otimes g;\cdot)\big) \\ &= \lim_{n\to\infty} d_{\mathrm{m}}\big(U(\mathbb{1}_{x_n\leqslant f}\otimes g;\cdot), W(\mathbb{1}_{x_n\leqslant f}\otimes g;\cdot)\big), \end{aligned}$$

for some sequence $(x_n)_{n\in\mathbb{N}}$ in $[0,1]$. Upon taking a subsequence, we may assume that the sequence $(x_n)_{n\in\mathbb{N}}$ monotonically converges to some $x\in[0,1]$. In particular, the sequence of functions $(\mathbb{1}_{x_n\leqslant f})_{n\in\mathbb{N}}$ (monotonically) converges to the function $f' = \mathbb{1}_{x\leqslant f}$ (resp. $f' = \mathbb{1}_{x<f}$) if $(x_n)_{n\in\mathbb{N}}$ is non-decreasing (resp. decreasing), and thus also weak-$*$ converges in $L^\infty([0,1],\lambda)$. Using, as in the first part of the proof, the sequential continuity of the function Ψ w.r.t. the weak-$*$ topology on $L^\infty([0,1],\lambda)$, we get that $\Psi(f',g) = d_{\mathrm{m}}(U(f'\otimes g;\cdot), W(f'\otimes g;\cdot)) = C$, that is we can replace f by the indicator function f'. The same argument allows to replace g by an indicator function. □

3.8. Examples of distance d_{m}

We consider usual distances and norms on $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$ that induce the weak topology on $\mathcal{M}_+(\mathbf{Z})$. All the distances we consider are quasi-convex, and all the norms we consider are sequentially continuous w.r.t. the weak topology on $\mathcal{M}_\pm(\mathbf{Z})$. Thus their associated cut distances are invariant and smooth by Lemma 3.11 and Proposition 3.13. Properties for the cut distances associated with those distances and norms are summarized in Corollaries 4.14 and 5.6.

In this section, we assume that $(\mathbf{Z}, d_0)$ is a Polish metric space, and recall that $\mathcal{B}(\mathbf{Z})$ denotes its Borel σ-field.

3.8.1. The Lévy–Prokhorov distance d_{LP}

The Lévy–Prokhorov distance d_{LP} is a complete distance defined on the set of finite measures $\mathcal{M}_+(\mathbf{Z})$ that induces the weak topology (see [11, Theorem 6.8]). It is defined for $\mu, \nu \in \mathcal{M}_+(\mathbf{Z})$ as:

$$(3.8)\quad d_{\mathrm{LP}}(\mu,\nu) = \inf\{\varepsilon > 0 : \forall A \in \mathcal{B}(\mathbf{Z}),\ \mu(A) \leqslant \nu(A^\varepsilon) + \varepsilon \quad \text{and} \quad \nu(A) \leqslant \mu(A^\varepsilon) + \varepsilon\},$$

where $A^\varepsilon = \{x \in \mathbf{Z} : \exists y \in A, d_0(x,y) < \varepsilon\}$. For probability measures, we only need one inequality in (3.8) to define the Lévy–Prokhorov distance; however for positive measures we need both inequalities as two arbitrary positive measures might not have the same total mass. For $d_{\mathrm{m}} = d_{\mathrm{LP}}$, we use the subscript $\mathrm{m} = \mathcal{P}$. We now prove that the Lévy–Prokhorov distance is quasi-convex.

Lemma 3.21. — *The Lévy–Prokhorov distance d_{LP} is quasi-convex on $\mathcal{M}_+(\mathbf{Z})$.*

Proof. — Let $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}_+(\mathbf{Z})$ and let $\alpha \in [0,1]$. Let $\varepsilon > \max(d_{\mathrm{LP}}(\mu_1,\nu_1), d_{\mathrm{LP}}(\mu_2,\nu_2))$, then for all $i \in \{1,2\}$ and $B \in \mathcal{B}(\mathbf{Z})$, we have that $\mu_i(B) \leqslant \nu_i(B^\varepsilon) + \varepsilon$ and $\nu_i(B) \leqslant \mu_i(B^\varepsilon) + \varepsilon$. Taking a linear combination of those inequalities, we get that for all $B \in \mathcal{B}(\mathbf{Z})$, we have that $\alpha\mu_1(B) + (1-\alpha)\mu_2(B) \leqslant \alpha\nu_1(B^\varepsilon) + (1-\alpha)\nu_2(B^\varepsilon) + \varepsilon$, and similarly when swapping the role (μ_1,μ_2) and (ν_1,ν_2). Hence, we get that $d_{\mathrm{LP}}(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) \leqslant \varepsilon$, and taking the infimum over ε, we get that $d_{\mathrm{m}}(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) \leqslant \max(d_{\mathrm{m}}(\mu_1,\nu_1), d_{\mathrm{m}}(\mu_2,\nu_2))$. □

3.8.2. The Kantorovitch–Rubinshtein and Fortet–Mourier norms

The Kantorovitch–Rubinshtein norm $\|\cdot\|_{\mathrm{KR}}$ (sometimes also called the bounded Lipschitz distance) and the Fortet–Mourier norm $\|\cdot\|_{\mathrm{FM}}$ are two norms defined on $\mathcal{M}_\pm(\mathbf{Z})$ that induce the weak topology on $\mathcal{M}_+(\mathbf{Z})$ (see Section 3.2 in [14] for definition and properties of those norms). They are defined for $\mu \in \mathcal{M}_\pm(\mathbf{Z})$ by:

$$\|\mu\|_{\mathrm{KR}} = \sup\left\{\int_{\mathbf{Z}} f \,\mathrm{d}\mu : f \text{ is 1-Lipschitz and } \|f\|_\infty \leqslant 1\right\},$$

$$\|\mu\|_{\mathrm{FM}} = \sup\left\{\int_{\mathbf{Z}} f \,\mathrm{d}\mu : f \text{ is Lipschitz and } \|f\|_\infty + \mathrm{Lip}(f) \leqslant 1\right\},$$

where $\|f\|_\infty = \sup_{x\in\mathbf{Z}} |f(x)|$ is the infinity norm and $\mathrm{Lip}(f)$ is the smallest constant $L > 0$ such that f is L-Lipschitz. Those two norms are metrically equivalent, see beginning of Section 3.2 in [14]:

$$\|\mu\|_{\mathrm{FM}} \leqslant \|\mu\|_{\mathrm{KR}} \leqslant 2\|\mu\|_{\mathrm{FM}}. \tag{3.9}$$

Note that we have $\|\mu\|_{\mathrm{KR}} \leqslant \|\mu\|_\infty$, and thus those two norms are sequentially continuous w.r.t. the weak topology on $\mathcal{M}_\pm(\mathbf{Z})$.

An easy adaptation of the proof for Theorem 3.2.2 in [14] gives the following comparison between d_{LP}, $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$.

Lemma 3.22 (Comparison of d_{LP}, $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$). — *Let $\mu, \nu \in \mathcal{M}_+(\mathbf{Z})$. Then, we have:*

$$\frac{d_{\mathrm{LP}}(\mu,\nu)^2}{1 + d_{\mathrm{LP}}(\mu,\nu)} \leqslant \|\mu-\nu\|_{\mathrm{FM}} \leqslant \|\mu-\nu\|_{\mathrm{KR}} \leqslant \big(2+\min(\mu(\mathbf{Z}),\nu(\mathbf{Z}))\big)\, d_{\mathrm{LP}}(\mu,\nu).$$

In particular, the Lévy–Prokhorov distance d_{LP} is uniformly continuous w.r.t. $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$ on $\mathcal{M}_+(\mathbf{Z})$; and $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$ are uniformly continuous w.r.t. d_{LP} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$.

For the special choice $N_{\mathrm{m}} = \|\cdot\|_{\mathrm{KR}}$ (resp. $N_{\mathrm{m}} = \|\cdot\|_{\mathrm{FM}}$), we use the subscript $\mathrm{m} = \mathrm{KR}$ (resp. $\mathrm{m} = \mathrm{FM}$).

3.8.3. A norm based on a convergence determining sequence

From a convergence determining sequence $\mathcal{F} = (f_k)_{k\in\mathbb{N}}$, where $f_0 = \mathbb{1}$ and $f_k \in C_b(\mathbf{Z})$ takes values in $[0,1]$, we define a norm on $\mathcal{M}_\pm(\mathbf{Z})$ metrizing the weak topology on $\mathcal{M}_+(\mathbf{Z})$, for $\mu \in \mathcal{M}_\pm(\mathbf{Z})$, by:

$$\|\mu\|_{\mathcal{F}} = \sum_{k\in\mathbb{N}} 2^{-k} |\mu(f_k)|. \tag{3.10}$$

Note that we have $\|\mu\|_{\mathcal{F}} \leqslant 2\|\mu\|_\infty$, and thus $\|\cdot\|_{\mathcal{F}}$ is sequentially continuous w.r.t. the weak topology on $\mathcal{M}_\pm(\mathbf{Z})$. For the special choice $N_{\mathrm{m}} = \|\cdot\|_{\mathcal{F}}$, we use the subscript $\mathrm{m} = \mathcal{F}$.

Even though the norm $\|\cdot\|_{\mathcal{F}}$ is not complete when $\mathbf{Z}$ is not compact (see Lemma 3.23 below), the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ and the cut distance $\delta_{\square,\mathcal{F}}$ will turn out to be very useful in Sections 6 and 7 to link the topology of the cut distance to the homomorphism densities. Recall $d_{\mathcal{F}}$ is the distance derived from the norm $\|\cdot\|_{\mathcal{F}}$.

Lemma 3.23 ($d_{\mathcal{F}}$ is not complete in general). — *Let $\mathcal{F}$ be a convergence determining sequence. Then, the distance $d_{\mathcal{F}}$ is complete over $\mathcal{M}_1(\mathbf{Z})$ if and only if $\mathcal{M}_1(\mathbf{Z})$ is a compact space,* i.e. , *if and only if $\mathbf{Z}$ is compact.*

Proof. — Theorem 3.4 in [52] states that $\mathbf{Z}$ is compact if and only if $\mathcal{M}_1(\mathbf{Z})$ is compact. When this is the case, any distance metrizing the weak topology on $\mathcal{M}_1(\mathbf{Z})$ is complete.

Reciprocally, assume that $d_{\mathcal{F}}$ is a complete metric over $\mathcal{M}_1(\mathbf{Z})$ and write $\mathcal{F} = (f_m)_{m\in\mathbb{N}}$. Let $(\mu_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of probability measures from $\mathcal{M}_1(\mathbf{Z})$. For every $m \in \mathbb{N}$, as f_m takes values in $[0, 1]$, we have for every $n \in \mathbb{N}$ that $\mu_n(f_m) \in [0, 1]$. Hence, using a diagonal extraction, there exists a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ of the sequence $(\mu_n)_{n\in\mathbb{N}}$ such that for every $m \in \mathbb{N}$, the sequence $(\mu_{n_k}(f_m))_{k\in\mathbb{N}}$ converges, that is, $(\mu_{n_k})_{k\in\mathbb{N}}$ is a Cauchy sequence for the distance $d_{\mathcal{F}}$. As we assumed the distance $d_{\mathcal{F}}$ to be complete, this implies that the sequence $(\mu_n)_{n\in\mathbb{N}}$ has a convergent subsequence. The sequence $(\mu_n)_{n\in\mathbb{N}}$ being arbitrary, we conclude that the space $\mathcal{M}_1(\mathbf{Z})$ is sequentially compact, and thus compact by Remark 2.6. □

We denote by $\|\cdot\|_{\square,\mathbb{R}}$ (resp. $\|\cdot\|^+_{\square,\mathbb{R}}$) the cut norm (resp. one-sided version of the cut norm) for real-valued kernels defined as:

$$\text{(3.11)} \qquad \|w\|_{\square,\mathbb{R}} = \sup_{S,T\subset[0,1]} \left| \int_{S\times T} w(x, y)\ \mathrm{d}x\mathrm{d}y \right|$$

$$\text{and} \quad \|w\|^+_{\square,\mathbb{R}} = \sup_{S,T\subset[0,1]} \int_{S\times T} w(x, y)\ \mathrm{d}x\mathrm{d}y,$$

where w is a real-valued kernel w (see [46, Section 8.2], resp. [46, Section 10.3], for definition and properties of those objects).

Recall from (3.7) on page 53 the definition of the real-valued kernel $W[f]$ for $W \in \mathcal{W}_\pm$ and $f \in C_b(\mathbf{Z})$. The following two remarks link the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ of a signed measure-valued kernel W with the cut norm $\|\cdot\|_{\square,\mathbb{R}}$ of the real-valued kernels $W[f]$ for some particular choices of functions $f \in C_b(\mathbf{Z})$. We will reuse those facts in Section 6.

Remark 3.24 (Link between $\|\cdot\|_{\square,\mathcal{F}}$ and $\|\cdot\|^+_{\square,\mathbb{R}}$). — For $\mu \in \mathcal{M}_\pm(\mathbf{Z})$ we have:

$$\text{(3.12)} \qquad \|\mu\|_{\mathcal{F}} = \sup_{\varepsilon\in\{\pm1\}^{\mathbb{N}}} \sum_{n\in\mathbb{N}} 2^{-n}\varepsilon_n \mu(f_n) = \sup_{\varepsilon\in\{\pm1\}^{\mathbb{N}}} \mu\left(\sum_{n\in\mathbb{N}} 2^{-n}\varepsilon_n f_n\right),$$

with $\varepsilon = (\varepsilon_n)_{n\in\mathbb{N}}$. Hence, for a signed measure-valued kernel $W \in \mathcal{W}_\pm$, we have:

$$\|W\|_{\square,\mathcal{F}} = \sup_{\varepsilon\in\{\pm1\}^{\mathbb{N}}} \sup_{S,T\subset[0,1]} W\left(S\times T; \sum_{n\in\mathbb{N}} 2^{-n}\varepsilon_n f_n\right) \tag{3.13}$$
$$= \sup_{\varepsilon\in\{\pm1\}^{\mathbb{N}}} \left\| W\left[\sum_{n\in\mathbb{N}} 2^{-n}\varepsilon_n f_n\right]\right\|^+_{\square,\mathbb{R}}.$$

Remark 3.25 (Inequality with $\|\cdot\|_{\square,\mathcal{F}}$ and $\|\cdot\|_{\square,\mathbb{R}}$). — For a signed measure-valued kernel W, we have:

$$\|W\|_{\square,\mathcal{F}} = \sup_{S,T\subset[0,1]} \sum_{n=0}^{\infty} 2^{-n}\left|\int_{S\times T} W(x,y;f_n)\,\mathrm{d}x\mathrm{d}y\right|$$
$$\leqslant \sum_{n=0}^{\infty} 2^{-n} \sup_{S,T\subset[0,1]} \left|\int_{S\times T} W(x,y;f_n)\,\mathrm{d}x\mathrm{d}y\right|$$
$$= \sum_{n=0}^{\infty} 2^{-n}\|W[f_n]\|_{\square,\mathbb{R}}. \tag{3.14}$$

4. Tightness and weak regularity

In this section, using a conditional expectation approach as for real-valued kernels in [46, Chapter 9], we provide approximations of signed measure-valued kernels and probability-graphons by stepfunctions. In doing so, we prove that the space of probability-graphons (resp. the space of signed measure-valued kernels) is separable. We introduce a tightness criterion for signed measure-valued kernels and probability-graphons similar to the one for random measures; it will be used in the next section to characterise relatively compact subset of probability-graphons. We then prove a weak regularity property for cut distances which is a uniform approximation bound for tight subsets of signed measure-valued kernels and probability-graphons. This weak regularity property is an analogue of the weak regularity lemma for real-valued kernels from [27, Theorem 12] and [46, Lemma 9.15], but without an explicit bound on the quality of the approximation due to the lack of an Euclidean structure for signed measure-valued kernels. Lastly, we introduce some Euclidean structure on $\mathcal{W}_\pm$ linked to the cut distance $d_{\square,\mathcal{F}}$, and we prove a weak regularity lemma with an explicit bound for signed measure-valued kernel.

4.1. Approximation by stepfunctions

We start by introducing the partitioning of a signed measure-valued kernel.

DEFINITION 4.1 (The stepping operator). — *Let $W \in \mathcal{W}_\pm$ be a signed measure-valued kernel and $\mathcal{P} = \{S_1, \cdots, S_k\}$ be a finite partition of $[0,1]$. We define the kernel stepfunction $W_\mathcal{P}$ adapted to the partition $\mathcal{P}$ by averaging W over the partition subsets:*

$$W_\mathcal{P}(x, y; \cdot) = \frac{1}{\lambda(S_i)\lambda(S_j)} W(S_i \times S_j; \cdot) \qquad \text{for } x \in S_i, y \in S_j,$$

when $\lambda(S_i) \neq 0$ and $\lambda(S_j) \neq 0$, and $W_\mathcal{P}(x, y; \cdot) = 0$ the null measure otherwise. We call the map $W \mapsto W_\mathcal{P}$ defined on $\mathcal{W}_\pm$ the stepping operator (associated with the finite partition $\mathcal{P}$).

Since the signed measure-valued kernel are defined up to an a.e. equivalence, the value of $W_\mathcal{P}(x, y; \cdot)$ for $x \in S_i, y \in S_j$ when $\lambda(S_i)\lambda(S_j)$ is unimportant.

Remark 4.2 (Link with conditional expectation). — The stepfunction $W_\mathcal{P}$ can be viewed as the conditional expectation of W w.r.t. the (finite) sigma-field $\sigma(\mathcal{P} \times \mathcal{P})$ on $[0,1]^2$, where $W : [0,1]^2 \to \mathcal{M}_\pm(\mathbf{Z})$ is seen as a random signed measure in $\mathcal{M}_\pm(\mathbf{Z})$ and the probability measure on $[0,1]^2$ is the Lebesgue measure.

Remark 4.3 (Steppings are convex stable). — Let $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$ be a convex subset of measures, for instance $\mathcal{M}$ is $\mathcal{M}_1(\mathbf{Z})$, $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$. Whenever $W \in \mathcal{W}_\pm$ is a $\mathcal{M}$-valued kernel, then by simple computation its stepping $W_\mathcal{P}$ is also a $\mathcal{M}$-valued kernel.

In the following remark, we give a characterization of refining partitions that generate the Borel σ-field of $[0,1]$.

Remark 4.4 (On refining partitions that generates the Borel σ-field). — Let $(\mathcal{P}_k)_{k\in\mathbb{N}}$ be a sequence of refining partitions of $[0,1]$. It generates the Borel σ-field of $[0,1]$ (that is, $\{S : S \in \mathcal{P}_k, k \in \mathbb{N}\}$ generates the Borel σ-field of $[0,1]$) if and only if $(\mathcal{P}_k)_{k\in\mathbb{N}}$ separates points (that is, for every distinct $x, y \in [0,1]$, there exists $k \in \mathbb{N}$ such that x and y belong to different classes of $\mathcal{P}_k$).

Indeed, assume that $(\mathcal{P}_k)_{k\in\mathbb{N}}$ separates points, and consider the countable family of Borel-measurable functions $\mathcal{F} = \{\mathbb{1}_S : S \in \mathcal{P}_k, k \in \mathbb{N}\}$ which separates points. Thus, by [13, Theorem 6.8.9] (remark that a Polish space is a Souslin space, see [13, Definition 6.6.1]), the family $\mathcal{F}$ generates the Borel σ-field of $[0,1]$. This implies that the family of Borel sets $\{S : S \in \mathcal{P}_k,\ k \in \mathbb{N}\}$ generates the Borel σ-field of $[0,1]$.

Conversely, assume there exist $x, y \in [0,1]$ which are not separated by $(\mathcal{P}_k)_{k\in\mathbb{N}}$, i.e. for all $k \in \mathbb{N}$, x and y belong to the same class of $\mathcal{P}_k$. This implies that the set $\{x\}$ does not belong to the σ-field generated by $(\mathcal{P}_k)_{k\in\mathbb{N}}$, and thus $(\mathcal{P}_k)_{k\in\mathbb{N}}$ does not generate the Borel σ-field of $[0,1]$.

Recall the definition of the norm $\|\cdot\|_\infty$ on $\mathcal{W}_\pm$ defined in (3.1). The following lemma allows to approximate any signed measure-valued kernel by its steppings.

Lemma 4.5 (Approximation using the stepping operator). — *Let $W \in \mathcal{W}_\pm$ be a signed measure-valued kernel (which is bounded by definition). Let $(\mathcal{P}_n)_{n\in\mathbb{N}}$ be a refining sequence of finite partitions of $[0,1]$ that generates the Borel σ-field on $[0,1]$. Then, the sequence $(W_{\mathcal{P}_n})_{n\in\mathbb{N}}$ is uniformly bounded by $\|W\|_\infty$, and weakly converges to W almost everywhere (on $[0,1]^2$).*

Proof. — Set $W_n = W_{\mathcal{P}_n}$ for $n \in \mathbb{N}$. By definition of the stepping operator, we have for every $n \in \mathbb{N}$ and every $(x,y) \in [0,1]^2$ that the total mass of $W_n(x,y;\cdot)$ is upper bounded by $\|W\|_\infty$.

Recall that for $W \in \mathcal{W}_\pm$ and $f \in C_b(\mathbf{Z})$, the real-valued kernel $W[f]$ is defined by (3.7). First assume that $W \in \mathcal{W}_+$. Let $\mathcal{F} = (f_k)_{k\in\mathbb{N}}$ be a convergence determining sequence, with by convention $f_0 = \mathbb{1}$. For every $k \in \mathbb{N}$ and $n \in \mathbb{N}$, an immediate computation gives $W_n[f_k] = (W[f_k])_{\mathcal{P}_n}$. For every $k \in \mathbb{N}$, as $W[f_k]$ is a real-valued kernel, we can apply the closed martingale theorem (as $(W[f_k])_{\mathcal{P}_n}$ can be viewed as a conditional expectation, see Remark 4.2), and we get that $\lim_{n\to\infty} W_n[f_k] = W[f_k]$ almost everywhere, since $(\mathcal{P}_n)_{n\in\mathbb{N}}$ generates the Borel σ-field. Hence, as the sequence $(f_k)_{k\in\mathbb{N}}$ is convergence determining, the sequence $(W_n)_{n\in\mathbb{N}}$ weakly converges to W almost everywhere.

Now, for $W \in \mathcal{W}_\pm$, write $W = W^+ - W^-$ where $W^+, W^- \in \mathcal{W}_+$ (see Lemma 3.3). By linearity of the stepping operator, remark that we have $W_n = (W^+)_{\mathcal{P}_n} - (W^-)_{\mathcal{P}_n}$ for all $n \in \mathbb{N}$. By the first case, we have that the sequence $((W^+)_{\mathcal{P}_n})_{n\in\mathbb{N}}$ weakly converges a.e. to W^+, and similarly for $((W^-)_{\mathcal{P}_n})_{n\in\mathbb{N}}$ and W^-. Hence, the sequence $(W_n)_{n\in\mathbb{N}}$ weakly converges to W almost everywhere. ☐

We first provide a separability result on the space of probability -graphons.

Proposition 4.6 (Separability of $\mathcal{W}_1$ and $\widetilde{\mathcal{W}}_1$). — *Let d be a smooth distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$). Then, the space $(\mathcal{W}_1, d)$ (resp. $(\mathcal{W}_+, d)$ or $(\mathcal{W}_\pm, d)$) is separable.*

If furthermore d is invariant (which implies that $\delta_\square$ is a distance), then the space $(\widetilde{\mathcal{W}}_1, \delta_\square)$ (resp. $(\widetilde{\mathcal{W}}_+, \delta_\square)$ or $(\widetilde{\mathcal{W}}_\pm, \delta_\square)$) is separable.

In particular, this proposition can be applied when $d = \delta_{\square,\mathrm{m}}$ and d_m is a quasi-convex distance continuous w.r.t. the weak topology, as then $d_{\square,\mathrm{m}}$ is invariant and smooth (recall Lemma 3.11 and Proposition 3.13).

Proof. — We shall consider the space of probability-graphons $\mathcal{W}_1$, as the proofs for $\mathcal{W}_+$ and $\mathcal{W}_\pm$ are similar. Applying Lemma 4.5 with the sequence of dyadic partitions, for every probability-graphon W, we can find a sequence of probability-graphon stepfunctions adapted to finite dyadic partitions and converging to W almost everywhere on $[0,1]^2$.

As the space $\mathbf{Z}$ is separable, the space of probability measures $\mathcal{M}_1(\mathbf{Z})$ is also separable for the weak topology (see [11, Theorem 6.8]). Let $\mathcal{A} \subset \mathcal{M}_1(\mathbf{Z})$ be an at most countable dense (for the weak topology) subset. Then, for any stepfunction $W \in \mathcal{W}_1$ adapted to a finite dyadic partition, we can approach it everywhere on $[0,1]^2$ by a sequence of $\mathcal{A}$-valued stepfunctions adapted to the same finite dyadic partition.

Hence, for every $W \in \mathcal{W}_1$, there exists a sequence $(W_n)_{n\in\mathbb{N}}$ in the countable set of $\mathcal{A}$-valued stepfunctions adapted to a finite dyadic partition that converges to W almost everywhere on $[0,1]^2$. As d is smooth, we get that this convergence also holds for d. Thus, the space $(\mathcal{W}_1, d)$ is complete.

Recall that by Theorem 3.17, when the distance d is invariant and smooth, then the pseudometric $\delta_\square$ is a distance on $\widetilde{\mathcal{W}}_1$. In that case, convergence for d implies convergence for $\delta_\square$, and thus the space $(\widetilde{\mathcal{W}}_1, \delta_\square)$ is also separable. □

4.2. Tightness

Similarly to the case of signed measures (recall Lemma 2.8), we introduce a tightness criterion for signed measure-valued kernels that characterizes relative compactness, see Proposition 4.8 below. For a signed measure-valued kernel $W \in \mathcal{W}_\pm$, we define the measure $M_W \in \mathcal{M}_+(\mathbf{Z})$ by:

$$M_W(\mathrm{d}z) = |W|([0,1]^2; \mathrm{d}z) = \int_{[0,1]^2} |W|(x, y; \mathrm{d}z)\ \mathrm{d}x\mathrm{d}y, \tag{4.1}$$

where for every $x, y \in [0,1]$, $|W|(x, y; \cdot)$ is the total variation of $W(x, y; \cdot)$ (see Lemma 3.3). In particular, if W is a probability-graphon then M_W is a probability measure from $\mathcal{M}_1(\mathbf{Z})$. Notice also that if W and U are weakly isomorphic, then $M_W = M_U$, so that the application $W \mapsto M_W$ can be seen as a map from $\widetilde{\mathcal{W}}_1$ (resp. $\widetilde{\mathcal{W}}_\pm$) to $\mathcal{M}_1(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$).

Definition 4.7 (Tightness criterion). — *A subset $\mathcal{K} \subset \mathcal{W}_\pm$ (resp. $\mathcal{K} \subset \widetilde{\mathcal{W}}_\pm$) is said to be tight if the subset of measures $\{M_W : W \in \mathcal{K}\} \subset \mathcal{M}_+(\mathbf{Z})$ is tight.*

The following proposition shows the equivalence between a global tightness criterion and a local tightness criterion. Recall that uniformly bounded subsets of $\widetilde{\mathcal{W}}_\pm$ are discussed after Definition 3.16. Recall also λ_2 is the Lebesgue measure on $[0,1]^2$.

Proposition 4.8 (Alternative tightness criterion). — *Let $\mathcal{K} \subset \mathcal{W}_\pm$ (or $\mathcal{K} \subset \widetilde{\mathcal{W}}_\pm$) be a uniformly bounded subset of signed measure-valued kernels. The set $\mathcal{K}$ is tight if and only if for every $\varepsilon > 0$, there exists a compact set $K \subset \mathbf{Z}$, such that for every $W \in \mathcal{K}$ we have:*

$$\lambda_2\Big(\{(x,y) \in [0,1]^2 : |W|(x,y;K^c) \leqslant \varepsilon\}\Big) > 1-\varepsilon. \tag{4.2}$$

Proof. — As the left hand side of (4.2) is invariant by relabeling, it is enough to do the proof for $\mathcal{W}_\pm$. Let $\mathcal{K} \subset \mathcal{W}_\pm$ be uniformly bounded and set $C = \sup_{W\in\mathcal{K}} \|W\|_\infty < \infty$. Assume that for every $\varepsilon > 0$, there exists a compact set $K \subset \mathbf{Z}$, such that (4.2) holds for every $W \in \mathcal{K}$. Let $1 > \varepsilon > 0$. Thus, there exists a compact subset $K \subset \mathbf{Z}$ such that for every $W \in \mathcal{K}$ there exists a subset $A_W \subset [0,1]^2$ with (Lebesgue) measure at least $1-\varepsilon$, such that for every $(x,y) \in A_W$, we have $|W|(x,y;K^c) \leqslant \varepsilon$. We have that for all $W \in \mathcal{K}$:

$$\begin{aligned} M_W(K^c) = \int_{[0,1]^2} |W|(x,y;K^c)\ \mathrm{d}x\mathrm{d}y &\leqslant \|W\|_\infty \lambda_2(A_W^c) + \varepsilon\lambda_2(A_W) \\ &\leqslant (C+1)\varepsilon. \end{aligned}$$

Hence, the subset of measures $\{M_W : W \in \mathcal{K}\} \subset \mathcal{M}_+(\mathbf{Z})$ is tight, that is $\mathcal{K}$ is tight.

Conversely, suppose that $\mathcal{K}$ is tight. Let $\varepsilon > 0$. There exists a compact set $K \subset \mathbf{Z}$ such that for every $W \in \mathcal{K}$, we have $M_W(K^c) < \varepsilon^2$. For $W \in \mathcal{K}$, define $A_W = \{(x,y) \in [0,1]^2 : |W|(x,y;K^c) \leqslant \varepsilon\}$. We have:

$$\varepsilon^2 > M_W(K^c) = \int_{[0,1]^2} |W|(x,y;K^c)\ \mathrm{d}x\mathrm{d}y \geqslant \varepsilon\lambda_2(A_W^c).$$

Hence, $\lambda_2(A_W) > 1-\varepsilon$, and consequently Equation (4.2) holds. □

We end this section on a continuity result of the map $W \mapsto M_W$.

LEMMA 4.9 (Regularity of the map $W \mapsto M_W$). — *Let d_{m} be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$). Then the map $W \mapsto M_W$ is 1-Lipschitz, and thus continuous, from $(\widetilde{\mathcal{W}}_{1,\mathrm{m}}, \delta_{\square,\mathrm{m}})$ (resp. $(\widetilde{\mathcal{W}}_{+,\mathrm{m}}, \delta_{\square,\mathrm{m}})$) to $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ (resp. $(\mathcal{M}_+(\mathbf{Z}), d_{\mathrm{m}})$).*

Proof. — Taking $S = T = [0,1]$ in Definition (3.2) of $d_{\square,\mathrm{m}}$, we get that $d_{\mathrm{m}}(M_U, M_W) \leqslant d_{\square,\mathrm{m}}(W,U)$. As $M_{U^\varphi} = M_U$ for any measure-preserving map φ thanks to (2.1), we deduce from Definition (3.5) of $\delta_{\square,\mathrm{m}}$ that $d_{\mathrm{m}}(M_U, M_W) \leqslant \delta_{\square,\mathrm{m}}(U,W)$. □

4.3. Weak regularity

We shall consider the following extra regularities of distances on the set of signed measure-valued kernels w.r.t. the stepping operator. For a finite partition $\mathcal{P}$, denote by $|\mathcal{P}|$ the size of the partition $\mathcal{P}$, i.e. the number of sets composing $\mathcal{P}$.

DEFINITION 4.10 (Regularities of distances). — *Let d be a distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$).*

(i) ***Weak regularity.*** *The distance d is weakly regular if whenever the subset $\mathcal{K}$ of $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) is tight (resp. tight and uniformly bounded), then for every $\varepsilon > 0$, there exists $m \in \mathbb{N}^*$, such that for every kernel $W \in \mathcal{K}$, and for every finite partition $\mathcal{Q}$ of $[0,1]$, there exists a finite partition $\mathcal{P}$ of $[0,1]$ that refines $\mathcal{Q}$ such that:*

$$|\mathcal{P}| \leqslant m|\mathcal{Q}| \quad \text{and} \quad d(W, W_{\mathcal{P}}) < \varepsilon.$$

(ii) ***Regularity w.r.t. the stepping operator.*** *The distance d is regular w.r.t. the stepping operator if (resp. for any finite constant $C \geqslant 0$) there exists a finite constant $C_0 > 0$ such that for every W, U in $\mathcal{W}_1$ (resp. in $\mathcal{W}_+$ or $\mathcal{W}_\pm$, with $\|W\|_\infty \leqslant C$ and $\|U\|_\infty \leqslant C$) and every finite partition $\mathcal{P}$ of $[0,1]$, then we have:*

$$d(W, W_{\mathcal{P}}) \leqslant C_0\, d(W, U_{\mathcal{P}}). \tag{4.3}$$

We say that a norm N on $\mathcal{W}_\pm$ is weakly regular (resp. regular w.r.t. the stepping operator) if its associated distance d on $\mathcal{W}_\pm$ is weakly regular (resp. regular w.r.t. the stepping operator).

The weak regularity property above is an analogue to the weak regularity lemma for real-valued graphons, see [27, Theorem 12] or [46, Lemma 9.15]. In those two references there is an explicit relation between m and ε (namely that $m = 2^{\lfloor 1/\varepsilon^2 \rfloor}$). This is not the case here because we shall consider it for cut distances $d_{\square,\mathrm{m}}$ associated to a general distance d_{m}, and they are not associated to an Euclidean structure (see also Section 4.4 for a stronger version of the weak regularity property with the cut distance $d_{\square,\mathcal{F}}$ which is indeed associated to an Euclidean structure).

If a distance d is weakly regular, then for a subset $\mathcal{K} \subset \mathcal{M}_{\pm}(\mathbf{Z})$ which is tight and uniformly bounded, every $\mathcal{K}$-valued kernel can be approximated by a stepfunction with a uniform bound. The regularity w.r.t. the stepping operator states that the stepping operator gives an almost optimal way to approximate a signed measure-valued kernel using stepfunctions adapted to a given partition.

4.3.1. An example of cut distance regular w.r.t. the stepping operator

Recall the definition of a quasi-convex distance in Definition 2.10. We first show that the stepping operator is 1-Lipschitz for the cut distance $d_{\square,\mathrm{m}}$ when the distance d_{m} is quasi-convex.

Lemma 4.11 (The stepping operator is 1-Lipschitz). — *Let d_{m} be a quasi-convex distance on $\mathcal{M}$ a convex subset of $\mathcal{M}_{\pm}(\mathbf{Z})$ containing the zero measure. Then, the stepping operator associated with a given finite partition of $[0,1]$ is 1-Lipschitz on $\mathcal{W}_{\mathcal{M}}$ for the cut distance $d_{\square,\mathrm{m}}$.*

Proof. — Let $U, W \in \mathcal{W}_{\mathcal{M}}$ be $\mathcal{M}$-valued kernels, and let $\mathcal{P}$ be a finite measurable partition of $[0,1]$. As $U_{\mathcal{P}}$ and $W_{\mathcal{P}}$ are stepfunctions adapted to the same partition, and as d_{m} is quasi-convex, we can use Lemma 3.19 to get for some $S, T \in \sigma(\mathcal{P})$ that:

$$\begin{aligned} d_{\square,\mathrm{m}}(U_{\mathcal{P}}, W_{\mathcal{P}}) &= d_{\mathrm{m}}(U_{\mathcal{P}}(S \times T;\cdot), W_{\mathcal{P}}(S \times T;\cdot)) \\ &= d_{\mathrm{m}}(U(S \times T;\cdot), W(S \times T;\cdot)) \\ &\leqslant d_{\square,\mathrm{m}}(U, W), \end{aligned}$$

where the second equality comes from the fact that the integrals are equal as $S, T \in \sigma(\mathcal{P})$ and thus the integration is over full steps of the partition. Hence, the stepping operator is 1-Lipschitz on $\mathcal{W}_{\mathcal{M}}$ for the cut distance d_{m}. □

For a quasi-convex distance d_{m}, the cut distance $d_{\square,\mathrm{m}}$ is regular w.r.t. the stepping operator with $C_0 = 2$ in (4.3) (and one can take $C = +\infty$ in Definition 4.10 (ii)).

Lemma 4.12 ($d_{\square,\mathrm{m}}$ is regular w.r.t. the stepping operator). — *Let d_{m} be a quasi-convex distance on $\mathcal{M}$ a convex subset of $\mathcal{M}_{\pm}(\mathbf{Z})$ containing* *the zero measure. Let $W, U \in \mathcal{W}_\epsilon$ be $\mathcal{W}_\mathcal{M}$-valued kernels, and let $\mathcal{P}$ be a finite partition of* $[0,1]$. *Then, we have:*

$$d_{\square,\mathrm{m}}(W, W_\mathcal{P}) \leqslant 2d_{\square,\mathrm{m}}(W, U_\mathcal{P}).$$

Proof. — The proof is similar to the proof of [46, Lemma 9.12]. As d_m is quasi-convex, using Lemma 4.11, we get:

$$d_{\square,\mathrm{m}}(W, W_\mathcal{P}) \leqslant d_{\square,\mathrm{m}}(W, U_\mathcal{P}) + d_{\square,\mathrm{m}}(U_\mathcal{P}, W_\mathcal{P}) \leqslant 2d_{\square,\mathrm{m}}(W, U_\mathcal{P}).$$

□

4.3.2. An example of weakly regular cut distance

We have the following general result. Recall Definitions 3.10 and 4.10 on distances and norms on $\mathcal{W}_\epsilon$, with $\varepsilon \in \{+, \pm\}$, being invariant, smooth, weakly regular and regular w.r.t. the stepping operator.

Proposition 4.13 (Weak regularity of $d_{\square,\mathrm{m}}$). — *Let d_m be a quasi-convex distance on $\mathcal{M}_\epsilon(\mathbf{Z})$, with $\epsilon \in \{+, \pm\}$, which is sequentially continuous w.r.t. the weak topology. Then, the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_\epsilon$ is invariant, smooth, weakly regular and regular w.r.t. the stepping operator.*

Using results from Section 3.8, we directly get the following weak regularity of the cut distance $d_{\square,\mathrm{LP}}$ and the cut norms $\|\cdot\|_{\square,\mathcal{F}}$, $\|\cdot\|_{\square,\mathrm{KR}}$ and $\|\cdot\|_{\square,\mathrm{FM}}$.

Corollary 4.14 (Weak regularity of usual distances and norms). — *The cut norms $\|\cdot\|_{\square,\mathcal{F}}$, $\|\cdot\|_{\square,\mathrm{KR}}$ and $\|\cdot\|_{\square,\mathrm{FM}}$ (resp. the cut distance $d_{\square,\mathrm{LP}}$) on $\mathcal{W}_\pm$ (resp. $\mathcal{W}_+$) are invariant, smooth, weakly regular and regular w.r.t. the stepping operator.*

Proof of Proposition 4.13. — We deduce from Lemmas 3.11 and 4.12, Proposition 3.13 and that the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_\epsilon$ is invariant, smooth and regular w.r.t. the stepping operator. We are left to prove that $d_{\square,\mathrm{m}}$ is weakly regular on $\mathcal{W}_\epsilon$. We prove it by considering in the first step the case $\mathbf{Z}$ compact and in a second step the general case $\mathbf{Z}$ Polish.

Step 1. We assume $\mathbf{Z}$ compact. As in the definition of weak regularity, let $\mathcal{K} \subset \mathcal{W}_\epsilon$ be a subset of $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels that is tight and uniformly bounded by some finite constant C. Let $\mathcal{M} \subset \mathcal{M}_\epsilon(\mathbf{Z})$ be the subset of elements of $\mathcal{M}_\epsilon(\mathbf{Z})$ with total mass at most C; in particular $\mathcal{M}$ is a convex set containing 0 and $\mathcal{K} \subset \mathcal{W}_\mathcal{M}$. As $\mathbf{Z}$ is compact, from Remarks 2.6 and 2.7, we know that the weak topology is metrizable on $\mathcal{M}$ and that $\mathcal{M}$ is compact, and thus sequentially weakly compact. Hence, as d_m is sequentially continuous w.r.t. the weak topology on $\mathcal{M}_\epsilon(\mathbf{Z})$, we have that $(\mathcal{M}, d_{\mathrm{m}})$ is sequentially compact, and thus compact.

Denote by $B(\mu, r) = \{\nu \in \mathcal{M} : d_{\mathrm{m}}(\mu, \nu) < r\}$ the open ball centered at $\mu \in \mathcal{M}$ with radius $r > 0$. Let $\varepsilon > 0$. As $\mathcal{M}$ is compact, there exist $\mu_1, \ldots, \mu_n \in \mathcal{M}$, $n \in \mathbb{N}^*$, such that $\mathcal{M} = \cup_{i=1}^n B(\mu_i, \varepsilon)$. For $1 \leqslant i \leqslant n$, define $A_i = B(\mu_i, \varepsilon) \backslash \cup_{j<i} B(\mu_j, \varepsilon)$, so that $\{A_1, \ldots, A_n\}$ is a finite partition (with possibly some empty sets) of $\mathcal{M}$.

Every $\mathcal{M}$-valued kernel W can be approximated by a $\{\mu_1, \ldots, \mu_n\}$-valued kernel U defined for every $(x, y) \in [0,1]^2$ by $U(x, y; \cdot) = \mu_i$ for i such that $W(x, y; \cdot) \in A_i$. Thus, by construction, we have that for every $(x, y) \in [0,1]^2$, $d_{\mathrm{m}}(W(x, y; \cdot), U(x, y; \cdot)) < \varepsilon$. Applying the quasi-convex supremum inequality from (3.4) to W and U, we get that:

$$d_{\square,\mathrm{m}}(W, U) \leqslant \operatorname*{essup}_{(x,y)\in[0,1]^2} d_{\mathrm{m}}(W(x, y; \cdot), U(x, y; \cdot)) \leqslant \varepsilon.$$

Then, as the stepping operator is 1-Lipschitz for the cut norm, see Lemma 4.11, we have for any finite partition $\mathcal{P}$ of $[0,1]$ that:

$$\begin{aligned} d_{\square,\mathrm{m}}(W, W_{\mathcal{P}}) &\leqslant d_{\square,\mathrm{m}}(W, U) + d_{\square,\mathrm{m}}(U, U_{\mathcal{P}}) + d_{\square,\mathrm{m}}(U_{\mathcal{P}}, W_{\mathcal{P}}) \\ &\leqslant 2\varepsilon + d_{\square,\mathrm{m}}(U, U_{\mathcal{P}}). \end{aligned} \tag{4.4}$$

Hence, to get the weak regularity property for $\mathcal{M}$-valued kernels, we are left to prove it for the much smaller set of $\mathcal{V}$-valued kernels, where $\mathcal{V}$ is the convex hull of $\{\mu_1, \ldots, \mu_n\}$.

As d_{m} is quasi-convex and sequentially continuous w.r.t. the weak topology, using Lemma 2.11, there exists $\eta > 0$ such that for all $\mu, \nu \in \mathcal{M}_\epsilon(\mathbf{Z})$, we have that $\|\mu - \nu\|_\infty < \eta$ implies that $d_{\mathrm{m}}(\mu, \nu) \leqslant \varepsilon$.

As $\mathcal{V}$ is a subset of a vector space with finite dimension n, the norm $\|\cdot\|_\infty$ seen over $\mathcal{V}$ is equivalent to the L_1-norm $\mu = \sum_{i=1}^n \alpha_i \mu_i \mapsto \|\alpha\|_1 = \sum_{i=1}^n |\alpha_i|$. We can now see $\mathcal{V}$-valued kernel as $\mathbb{R}^n$-valued graphon with a cut norm derived from the L_1-norm $\|\cdot\|_1$. In this case the proof for the weak regularity Lemma 9.9 in [46] in $\mathbb{R}$ can easily be adapted to $\mathbb{R}^n$ by first decomposing the $\mathbb{R}^n$-valued kernels into their n dimensional marginals $(w_1, \cdots, w_n)$ which are all real-valued kernels, then applying the weak regularity Lemma 9.9 in [46] to each of the w_i's independently with precision $\varepsilon' = \varepsilon/n$ giving partitions $\mathcal{P}_i$ for $i \in \{1, \ldots, n\}$, and lastly defining $\mathcal{P}$ as the common refinement of the $\mathcal{P}_i$'s and using Lemma 9.12 in [46] for regularity w.r.t. the stepping operator to conclude. Hence, we have the weak regularity property for $\mathcal{V}$-valued kernels: there exists $m \in \mathbb{N}^*$, such that for every $\mathcal{V}$-valued kernel U', and for every finite partition $\mathcal{Q}$ of $[0,1]$ there exists a finite partition $\mathcal{P}$ of $[0,1]$ that refines $\mathcal{Q}$, and such that $|\mathcal{P}| \leqslant m|\mathcal{Q}|$ and $\sup_{S,T\subset[0,1]} \|(U' - U'_{\mathcal{P}})(S\times T;\cdot)\|_\infty < \eta$, and thus $d_{\square,\mathrm{m}}(U', U'_{\mathcal{P}}) \leqslant \varepsilon$.

Taking $U' = U$ in (4.4), we get that $d_{\square,\mathrm{m}}(W, W_{\mathcal{P}}) \leqslant 3\varepsilon$ and $|\mathcal{P}| \leqslant m|\mathcal{Q}|$. This concludes the proof of the lemma when $\mathbf{Z}$ is compact.

Step 2. We consider the general case $\mathbf{Z}$ Polish. We now prove that $d_{\square,\mathrm{m}}$ is weakly regular on $\mathcal{W}_\epsilon$. Let $\mathcal{K} \subset \mathcal{W}_\epsilon$ be a subset of $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels that is tight and uniformly bounded, and denote by $C = \sup_{W\in\mathcal{K}} \|W\|_\infty < \infty$.

Let $\varepsilon > 0$. As d_{m} is quasi-convex and sequentially continuous w.r.t. the weak topology, using Lemma 2.11, there exists $\eta > 0$ such that for all $\mu, \nu \in \mathcal{M}_\epsilon(\mathbf{Z})$, we have that $\|\mu - \nu\|_\infty < \eta$ implies that $d_{\mathrm{m}}(\mu,\nu) < \varepsilon$. Without loss of generality, we assume that $\eta \leqslant \varepsilon$. Let $\eta_C = \min(\eta, \eta/C)$.

As $\mathcal{K}$ is tight, using Proposition 4.8, there exists a compact set $K \subset \mathbf{Z}$, such that for every $W \in \mathcal{K}$ the subset $A_W = \{(x,y) \in [0,1]^2 : |W|(x,y;K^c) \leqslant \eta_C/2\}$ has Lebesgue measure at least $1 - \eta_C/2$. Let $W \in \mathcal{K}$, and define the signed measure-valued kernel U by: $U(x,y;\cdot) = W(x,y;\cdot\cap K)$ for every $(x,y) \in A_W$, and $U(x,y;\cdot) = 0$ otherwise. Let $S,T \subset [0,1]$. We have:

$$\begin{aligned}
\|(W-U)(S\times T;\cdot)\|_\infty &\leqslant \int_{S\times T} \|W(x,y;\cdot) - U(x,y;\cdot)\|_\infty \,\mathrm{d}x\mathrm{d}y \\
&\leqslant \int_{A_W\cap(S\times T)} |W|(x,y;K^c)\,\mathrm{d}x\mathrm{d}y \\
&\qquad + \int_{A_W^c\cap(S\times T)} \|W(x,y;\cdot)\|_\infty \,\mathrm{d}x\mathrm{d}y \\
&\leqslant \eta_C/2 + C\cdot\eta_C/2 \\
&\leqslant \eta.
\end{aligned}$$

Thus, we have that $d_{\mathrm{m}}(W(S\times T;\cdot), U(S\times T;\cdot)) < \varepsilon$. Since this holds for all $S,T \subset [0,1]$, we get that $d_{\square,\mathrm{m}}(W,U) \leqslant \varepsilon$.

Notice that the $\mathcal{M}_\pm(\mathbf{Z})$-valued kernel U is also a $\mathcal{M}_\pm(K)$-valued kernel, where $K \subset \mathbf{Z}$ is a compact set, and that $\|U\|_\infty \leqslant \|W\|_\infty \leqslant C$. Further remark that, using Lemma 4.11, for every $W \in \mathcal{K}$ and every finite partition $\mathcal{P}$ of $[0,1]$, we have that:

$$\begin{aligned}
d_{\square,\mathrm{m}}(W, W_{\mathcal{P}}) &\leqslant d_{\square,\mathrm{m}}(W,U) + d_{\square,\mathrm{m}}(U,U_{\mathcal{P}}) + d_{\square,\mathrm{m}}(U_{\mathcal{P}}, W_{\mathcal{P}}) \\
&\leqslant 2\varepsilon + d_{\square,\mathrm{m}}(U, U_{\mathcal{P}}).
\end{aligned}$$

Hence, to get the weak regularity property for $d_{\square,\mathrm{m}}$ on $\mathcal{K}$ (see Definition 4.10 (i)), it is enough to prove that $d_{\square,\mathrm{m}}$ restricted to $\mathcal{M}_\epsilon(K)$-valued kernels is weakly regular, which is true by Step 1. As a consequence, we get that $d_{\square,\mathrm{m}}$ on $\mathcal{W}_\epsilon$ is weakly regular. □

4.4. A stronger weak regularity lemma for $d_{\square,\mathcal{F}}$

In this subsection, we prove a stronger version of the weak regularity lemma for the special case of the cut distance $d_{\square,\mathcal{F}}$. We shall use this result for the proof of the second sampling Lemma 6.12.

Let $\mathcal{F} = (f_n)_{n\in\mathbb{N}}$, with $f_0 = \mathbb{1}$ and f_n takes values in $[0,1]$, be a convergence determining sequence, which is assumed fixed in this section.

4.4.1. Comparison between $\|\cdot\|_{\square,\mathcal{F}}$ and an Euclidian norm

To better understand the stepping operator, we introduce a scalar product over signed measure-valued kernels. The link between this scalar product and the norm $\|\cdot\|_{\square,\mathcal{F}}$ is given by Lemma 4.15. We define the scalar product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ on signed measure-valued kernels for $U, W \in \mathcal{W}_{\pm}$ by:

$$\langle U, W\rangle_{\mathcal{F}} = \sum_{n\geqslant 0} 2^{-n} \langle U[f_n], W[f_n]\rangle,$$

where for all n the scalar product taken for $U[f_n]$ and $W[f_n]$ is the usual scalar product in $L^2([0,1]^2, \lambda_2)$ for real-valued kernels:

$$\langle U[f_n], W[f_n]\rangle = \int_{[0,1]^2} U[f_n](x,y) W[f_n](x,y)\, \mathrm{d}x\mathrm{d}y.$$

The scalar product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ induces a norm on $\mathcal{W}_{\pm}$ which we denote by $\|\cdot\|_{2,\mathcal{F}}$.

Let $\mathcal{P}$ be a finite partition of $[0,1]$. As the stepping operator for measurable real-valued L^2 functions on $[0,1]^2$ is a linear projection, and is idempotent and symmetric, and by definition of the scalar product $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ for signed measure-valued kernels, we have that the stepping operator for signed measure-valued kernels is linear, idempotent and symmetric for $\langle\cdot,\cdot\rangle_{\mathcal{F}}$. Moreover, the stepping operator is the orthogonal projection for $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ onto the space of stepfunctions with steps in $\mathcal{P}$.

Note that for a probability-graphon $W \in \mathcal{W}_1$, we have $\|W\|_{2,\mathcal{F}} \leqslant \sqrt{2}$ as each f_n takes values in $[0,1]$. The following technical lemma gives a comparison between $\|\cdot\|_{\square,\mathcal{F}}$ and $\|\cdot\|_{2,\mathcal{F}}$.

Lemma 4.15 (Comparison between $\|\cdot\|_{\square,\mathcal{F}}$ and $\|\cdot\|_{2,\mathcal{F}}$). — *For a signed measure-valued kernel $W \in \mathcal{W}_{\pm}$, we have $\|W\|_{\square,\mathcal{F}} \leqslant \sqrt{2}\|W\|_{2,\mathcal{F}}$.*

Proof. — Let $S, T \subset [0,1]$ be measurable subsets. By the Cauchy–Schwarz inequality, we have $|\langle W[f_n], \mathbb{1}_{S\times T}\rangle|^2 \leqslant \|W[f_n]\|_2^2 =$

$\langle W[f_n], W[f_n]\rangle$ for every $n \geqslant 0$. Using this inequality along with Jensen's inequality, we get for every $S, T \subset [0,1]$ that:

$$\begin{aligned}\left(\sum_{n\geqslant 0} 2^{-n}|W(S\times T, f_n)|\right)^2 &= \left(\sum_{n\geqslant 0} 2^{-n}|\langle W[f_n], \mathbb{1}_{S\times T}\rangle|\right)^2 \\ &\leqslant \sum_{n\geqslant 0} 2^{-n+1}|\langle W[f_n], \mathbb{1}_{S\times T}\rangle|^2 \\ &\leqslant \sum_{n\geqslant 0} 2^{-n+1}\langle W[f_n], W[f_n]\rangle \\ &= 2(\|W\|_{2,\mathcal{F}})^2.\end{aligned}$$

Taking the supremum over every measurable subsets $S, T \subset [0,1]$ gives the desired inequality. □

4.4.2. The weak regularity lemma for $\|\cdot\|_{\square,\mathcal{F}}$

The following lemma gives an explicit bound on the approximation of a signed measure-valued kernel, say W, by its steppings $W_{\mathcal{P}}$, with $\mathcal{P}$ a finite partition on $[0,1]$. Its proof is a straightforward adaptation of the proof of the weak regularity lemma for real-valued graphons in [46, Lemma 9.9].

LEMMA 4.16 (Weak regularity lemma for $\|\cdot\|_{\square,\mathcal{F}}$, simple formulation). *For every signed measure-valued kernel $W \in \mathcal{W}_{\pm}$ and $k \geqslant 1$, there exists a finite partition $\mathcal{P}$ of $[0,1]$ such that $|\mathcal{P}| = k$ and:*

$$\|W - W_{\mathcal{P}}\|_{\square,\mathcal{F}} \leqslant \frac{\sqrt{8}}{\sqrt{\log(k)}}\|W\|_{2,\mathcal{F}}.$$

In particular, if $W \in \mathcal{W}_1$ is a probability-graphon, (as $\|W\|_{2,\mathcal{F}} \leqslant \sqrt{2}$) we have:

$$\|W - W_{\mathcal{P}}\|_{\square,\mathcal{F}} \leqslant \frac{4}{\sqrt{\log(k)}}.$$

It is possible in the weak regularity lemma to ask for extra requirements, for instance to start from an already existing partition, or to ask the partition to be balanced, as stated in the following lemma. The proof is a straightforward adaptation of the proof of [46, Lemma 9.15].

LEMMA 4.17 (Weak regularity lemma for $\|\cdot\|_{\square,\mathcal{F}}$, with extra requirements). — *Let $W \in \mathcal{W}_1$ be a probability-graphon, and let $1 \leqslant m < k$.*

(i) For every partition $\mathcal{Q}$ of $[0,1]$ into m classes, there is a partition $\mathcal{P}$ with k classes refining $\mathcal{Q}$ and such that:

$$\|W - W_{\mathcal{P}}\|_{\Box,\mathcal{F}} \leqslant \frac{4}{\sqrt{\log(k/m)}}\cdot$$

(ii) For every partition $\mathcal{Q}$ of $[0,1]$ into m classes, there is an equipartition (i.e. a finite partition into classes with the same measure) $\mathcal{P}$ of $[0,1]$ into k classes and such that:

$$\|W - W_{\mathcal{P}}\|_{\Box,\mathcal{F}} \leqslant 2\|W - W_{\mathcal{Q}}\|_{\Box,\mathcal{F}} + \frac{2m}{k}\cdot$$

5. Compactness and completeness of $\widetilde{\mathcal{W}}_1$

In this section, we study compactness and completeness properties of the space of probability-graphon $\widetilde{\mathcal{W}}_1$, as well as the relation between the topologies induced by the different cut distances $\delta_{\Box,\mathrm{m}}$ (associated to the distance d_{m} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ which induces the weak topology). While the space of real-valued graphons equipped with the cut distance is compact (see [48] or [46, Section 9.3]), this is not true in general for probability-graphons as the space of probability measures on a Polish space is not compact in general. Then, using the tightness criterion introduced in Section 4.2, we characterize subsets of probability-graphons and measure-valued kernels that are relatively compact w.r.t. the cut distance $\delta_{\Box,\mathrm{m}}$, see in Section 5.1. Recall that unlike for $\mathbb{R}$ with its usual topology, there is no canonical distance d_{m} on the space of signed measures $\mathcal{M}_{\pm}(\mathbf{Z})$ inducing the weak topology. For this reason, it is necessary to consider cut distances $\delta_{\Box,\mathrm{m}}$ for signed measure-valued kernels and probability-graphons indexed by several distances d_{m}. Lastly, in Section 5.2, we compare the topologies induced by the cut distance $\delta_{\Box,\mathrm{m}}$ for different choice of d_{m}, and state that under some conditions on d_{m}, those topologies coincide. In Section 5.3, we investigate the completeness of $\widetilde{\mathcal{W}}_1$ endowed with the cut distance $\delta_{\Box,\mathrm{m}}$ and prove that the space of probability-graphons $\widetilde{\mathcal{W}}_1$ is a Polish space (Theorem 5.10), and that it is compact if and only if $\mathbf{Z}$ is compact (Corollary 5.13). The technical proofs are postponed to Section 8.

5.1. Tightness criterion and compactness

Let $\mathcal{M} \subset \mathcal{M}_{\pm}(\mathbf{Z})$ be a subset of signed measures on $\mathbf{Z}$. Recall that $\mathcal{W}_{\mathcal{M}} \subset \mathcal{W}_{\pm}$ denotes the subset of signed measure-valued kernels which are

$\mathcal{M}$-valued. In this section, we shall denote by $\widetilde{\mathcal{W}}_{\mathcal{M}}$ the quotient of $\mathcal{W}_{\mathcal{M}}$ identifying signed measure-valued kernels that are weakly isomorphic.

Recall from Definition 3.15 and Theorem 3.17 that for an invariant, smooth and weakly regular distance d on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$, $\mathcal{W}_\pm$), $\delta_\square$ is defined as $\delta_\square(U,W) = \inf_{\varphi\in S_{[0,1]}} d(U, W^\varphi)$, and is a distance on $\widetilde{\mathcal{W}}_1$ (resp. $\widetilde{\mathcal{W}}_+$, $\widetilde{\mathcal{W}}_\pm$).

We are now ready to formulate the following important theorem, which relates tightness with compactness and convergence for signed measure-valued kernels. We prove this theorem in Section 8.

Theorem 5.1 (Compactness theorem for $\widetilde{\mathcal{W}}_1$). — *Let d be an invariant, smooth and weakly regular distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_\pm$).*

- *(i) If a sequence of elements of $\mathcal{W}_1$ or $\widetilde{\mathcal{W}}_1$ (resp. $\mathcal{W}_\pm$ or $\widetilde{\mathcal{W}}_\pm$) is tight (resp. tight and uniformly bounded), then it has a subsequence converging for $\delta_\square$.*
- *(ii) If $\mathcal{M} \subset \mathcal{M}_1(\mathbf{Z})$ (resp. $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$) is convex and compact (resp. sequentially compact) for the weak topology, then the space $(\widetilde{\mathcal{W}}_{\mathcal{M}}, \delta_\square)$ is convex and compact.*
- *(iii) If $\mathbf{Z}$ is compact, then the space $(\widetilde{\mathcal{W}}_1, \delta_\square)$ is compact.*

In particular, this theorem can be applied when $d = d_{\square,\mathrm{m}}$ and d_m is a quasi-convex distance continuous w.r.t. the weak topology, as then $d_{\square,\mathrm{m}}$ is invariant, smooth and weakly regular (recall Proposition 4.13).

We deduce from this theorem a characterization of relative compactness for subsets of probability-graphons.

Proposition 5.2 (Characterization of relative compactness). — *Let d_m be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$) that induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$). Assume that the distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) is (invariant) smooth and weakly regular.*

- *(i) If a sequence of elements of $\mathcal{W}_1$ or $\widetilde{\mathcal{W}}_1$ (resp. $\mathcal{W}_+$ or $\widetilde{\mathcal{W}}_+$) is converging for $\delta_{\square,\mathrm{m}}$, then it is tight.*
- *(ii) Let $\mathcal{K}$ be a subset of $\widetilde{\mathcal{W}}_1$ (resp. a uniformly bounded subset of $\widetilde{\mathcal{W}}_+$). Then, the set $\mathcal{K}$ is relatively compact for $\delta_{\square,\mathrm{m}}$ if and only if it is tight.*
- *(iii) Let $\mathcal{M}$ be a subset of $\mathcal{M}_+(\mathbf{Z})$ which is bounded, convex and closed for the weak topology. Then the set $\widetilde{\mathcal{W}}_{\mathcal{M}}$ is convex and closed in $\widetilde{\mathcal{W}}_+$.*

Remark that convergence for $\delta_{\square,\mathrm{m}}$ does not necessarily imply tightness on $\mathcal{W}_\pm$ or on $\widetilde{\mathcal{W}}_\pm$.

Proof. — We consider the case where d_{m} is a distance on $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$, the case with $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ is similar.

We prove Point (i). Let $(W_n)_{n\in\mathbb{N}}$ be a convergent sequence of $\mathcal{W}_+$ (and thus of $\widetilde{\mathcal{W}}_+$) for $\delta_{\Box,\mathrm{m}}$. We deduce from the continuity of the map $W \mapsto M_W$, see Lemma 4.9, that the sequence $(M_{W_n})_{n\in\mathbb{N}}$ is converging for d_{m}, and thus is tight as d_{m} induces the weak topology on $\mathcal{M}_+(\mathbf{Z})$. Then, by definition the sequence $(W_n)_{n\in\mathbb{N}}$ is tight.

We prove Point (ii). If $\mathcal{K} \subset \widetilde{\mathcal{W}}_+$ is tight and uniformly bounded, then by Theorem 5.1 (i) every sequence in $\mathcal{K}$ has a subsequence converging for $\delta_{\Box,\mathrm{m}}$, which implies that $\mathcal{K}$ is relatively compact in the metric space $(\widetilde{\mathcal{W}}_+, \delta_{\Box,\mathrm{m}})$ (see Remark 2.1).

Conversely, assume that $\mathcal{K} \subset \widetilde{\mathcal{W}}_+$ is uniformly bounded and relatively compact for $\delta_{\Box,\mathrm{m}}$. Define $\mathcal{M} = \{M_W : W \in \mathcal{K}\} \subset \mathcal{M}_+(\mathbf{Z})$. By Lemma 4.9, the mapping $W \mapsto M_W$ is continuous from $(\widetilde{\mathcal{W}}_1, \delta_{\Box,\mathrm{m}})$ to $(\mathcal{M}_+(\mathbf{Z}), d_{\mathrm{m}})$. Hence, as d_{m} induces the weak topology on $\mathcal{M}_+(\mathbf{Z})$, the set $\mathcal{M}$ is also relatively compact in $\mathcal{M}_+(\mathbf{Z})$ for the weak topology. As the space $\mathbf{Z}$ is Polish, applying Lemma 2.8, we get that $\mathcal{M} \subset \mathcal{M}_+(\mathbf{Z})$ is tight, and by Definition 4.7, the set $\mathcal{K} \subset \widetilde{\mathcal{W}}_+$ is tight.

We postpone the proof of Point (iii) to Section 8 on page 109. □

5.2. Equivalence of topologies induced by $\delta_{\Box,\mathrm{m}}$

The following lemma allows to show a first result on equivalence of the topologies induced by the cut distance $\delta_{\Box,\mathrm{m}}$ for different distances d_{m}, where the sub-script m is used to distinguish different distances. Its proof is given below. Recall from Theorem 3.17 that $d_{\Box,\mathrm{m}}$ must be smooth for $\delta_{\Box,\mathrm{m}}$ to be a distance.

LEMMA 5.3 (Comparison of topologies induced by $d_{\Box,\mathrm{m}}$ and $\delta_{\Box,\mathrm{m}}$). — *Let d_{m} and $d_{\mathrm{m}'}$ be two distances on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that $d_{\mathrm{m}'}$ is uniformly continuous w.r.t. d_{m} (in particular, d_{m} induces a finer topology than $d_{\mathrm{m}'}$ on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$). Then, we have the following properties.*

- *(i) The distance $d_{\Box,\mathrm{m}'}$ is uniformly continuous w.r.t. $d_{\Box,\mathrm{m}}$ on $\mathcal{W}_1$. In particular $d_{\Box,\mathrm{m}}$ induces a finer topology than $d_{\Box,\mathrm{m}'}$ on $\mathcal{W}_1$.*
- *(ii) If the distance $d_{\Box,\mathrm{m}}$ on $\mathcal{W}_1$ is smooth, then the distance $d_{\Box,\mathrm{m}'}$ is also smooth and $\delta_{\Box,\mathrm{m}'}$ is uniformly continuous w.r.t. $\delta_{\Box,\mathrm{m}}$. In particular, $\delta_{\Box,\mathrm{m}}$ induces a finer topology than $\delta_{\Box,\mathrm{m}'}$ on $\widetilde{\mathcal{W}}_1$.*

(iii) *If the distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is weakly regular, then the distance $d_{\square,\mathrm{m}'}$ is also weakly regular.*

(iv) *Assume that the distance $d_{\mathrm{m}'}$ induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, and that the distance $d_{\square,\mathrm{m}}$ is smooth and weakly regular. In particular, the distance d_{m} also induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$. Then, the distances $\delta_{\square,\mathrm{m}}$ and $\delta_{\square,\mathrm{m}'}$ induce the same topology on $\widetilde{\mathcal{W}}_1$.*

We will see some application of Lemma 5.3 in Corollary 5.6 below.

Remark 5.4 (Extension to $\mathcal{W}_\pm$ for topology comparisons). — In Lemma 5.3 (i)-(iii), one can replace $\mathcal{W}_1$ and $\widetilde{\mathcal{W}}_1$ by $\mathcal{W}_+$ and $\widetilde{\mathcal{W}}_+$ or by $\mathcal{W}_\pm$ and $\widetilde{\mathcal{W}}_\pm$ as soon as the distances d_{m} and $d_{\mathrm{m}'}$ are defined on $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$; in this case comparisons of topologies only apply on uniformly bounded subsets. In Lemma 5.3 (iv), one can replace $\widetilde{\mathcal{W}}_1$ by $\widetilde{\mathcal{W}}_{\mathcal{M}}$ with a bounded subset $\mathcal{M} \subset \mathcal{M}_+(\mathbf{Z})$ as soon as the distances d_{m} and $d_{\mathrm{m}'}$ are defined on $\mathcal{M}_+(\mathbf{Z})$.

Proof of Lemma 5.3. — We prove Point (i). Let $\varepsilon > 0$. As $d_{\mathrm{m}'}$ is uniformly continuous w.r.t. d_{m}, there exists $\eta > 0$ such that for every $\mu, \nu \in \mathcal{M}_{\leqslant 1}(\mathbf{Z})$, if $d_{\mathrm{m}}(\mu,\nu) < \eta$, then $d_{\mathrm{m}'}(\mu,\nu) < \varepsilon$. Let $U, W \in \mathcal{W}_1$ such that $d_{\square,\mathrm{m}}(U,W) < \eta$. Then, for every subsets $S, T \subset [0,1]$, we have:

$$d_{\mathrm{m}'}\left(U(S\times T;\cdot), W(S\times T;\cdot)\right) < \varepsilon.$$

Thus, $d_{\square,\mathrm{m}'}(U,W) \leqslant \varepsilon$. Hence, $d_{\square,\mathrm{m}'}$ is uniformly continuous w.r.t. $d_{\square,\mathrm{m}}$.

We prove Point (ii). Assume that $d_{\square,\mathrm{m}}$ is smooth. Let $(W_n)_{n\in\mathbb{N}}$ and W be probability-graphons such that $W_n(x,y;\cdot)$ weakly converges to $W(x,y;\cdot)$ for almost every $x, y \in [0,1]$. Since the cut distance $d_{\square,\mathrm{m}}$ is smooth, we get that $d_{\square,\mathrm{m}}(W_n,W) \to 0$. As $d_{\square,\mathrm{m}'}$ is uniformly continuous (and thus also continuous) w.r.t. $d_{\square,\mathrm{m}}$, we have that $d_{\square,\mathrm{m}'}(W_n,W) \to 0$. Hence, $d_{\square,\mathrm{m}'}$ is smooth.

Furthermore, let $\varepsilon > 0$. Let $\eta > 0$ be such that for every $\mu,\nu \in \mathcal{M}_{\leqslant 1}(\mathbf{Z})$, $d_{\mathrm{m}}(\mu,\nu) < \eta$ implies $d_{\mathrm{m}'}(\mu,\nu) < \varepsilon$. For every $U, W \in \mathcal{W}_1$ such that $\delta_{\square,\mathrm{m}}(U,W) < \eta$, there exists $\varphi \in S_{[0,1]}$ such that $d_{\square,\mathrm{m}}(U,W^\varphi) < \eta$, which implies that $d_{\square,\mathrm{m}'}(U,W^\varphi) < \varepsilon$, which then implies that $\delta_{\square,\mathrm{m}'}(U,W) < \varepsilon$. That is, $\delta_{\square,\mathrm{m}'}$ is uniformly continuous w.r.t. $\delta_{\square,\mathrm{m}}$.

We prove Point (iii). Assume that $d_{\square,\mathrm{m}}$ is weakly regular. Let $\mathcal{K} \subset \mathcal{W}_1$ be tight. Let $\varepsilon > 0$. As $d_{\square,\mathrm{m}'}$ is uniformly continuous w.r.t. $d_{\square,\mathrm{m}}$, there exists $\eta > 0$ such that for every $U, W \in \mathcal{W}_1$, if $d_{\square,\mathrm{m}}(U,W) < \eta$, then $d_{\square,\mathrm{m}'}(U,W) < \varepsilon$. Since $d_{\square,\mathrm{m}}$ is weakly regular, there exists $m \in \mathbb{N}^*$, such that for every probability-graphon $W \in \mathcal{K}$, and for every finite partition $\mathcal{Q}$

of $[0,1]$, there exists a finite partition $\mathcal{P}$ of $[0,1]$ that refines $\mathcal{Q}$ and such that $|\mathcal{P}| \leqslant m|\mathcal{Q}|$ and $d_{\square,\mathrm{m}}(W, W_{\mathcal{P}}) < \eta$; and thus we also have $d_{\square,\mathrm{m}'}(W, W_{\mathcal{P}}) < \varepsilon$. Hence, $d_{\square,\mathrm{m}'}$ is weakly regular.

We prove Point (iv). Assume that $d_{\mathrm{m}'}$ induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ and that $d_{\square,\mathrm{m}}$ is smooth and weakly regular. In particular, the topology induced by d_{m} if finer than the topology induced by $d_{\mathrm{m}'}$, i.e. finer than the weak topology. As $d_{\square,\mathrm{m}}$ is smooth, by Lemma 3.12, d_{m} is continuous w.r.t. the weak topology (i.e. the weak topology if finer than the topology induced by d_{m}), and thus d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$. By Points (ii) and (iii), we get that $d_{\square,\mathrm{m}'}$ is also smooth and weakly regular. By Point (ii), the distance $\delta_{\square,\mathrm{m}}$ induces a finer topology than $\delta_{\square,\mathrm{m}'}$ on $\widetilde{\mathcal{W}}_1$.

We now prove that the topology of $\delta_{\square,\mathrm{m}'}$ is finer than the topology of $\delta_{\square,\mathrm{m}}$. Let $(W_n)_{n\in\mathbb{N}}$ and W be probability-graphons in $\widetilde{\mathcal{W}}_1$, such that W_n converges to W for $\delta_{\square,\mathrm{m}'}$. By Proposition 5.2 (i), we deduce that the sequence $(W_n)_{n\in\mathbb{N}}$ is tight. As $d_{\square,\mathrm{m}}$ is smooth and weakly regular, Theorem 5.1 gives that every subsequence $(W_{n_k})_{k\in\mathbb{N}}$ of the sequence $(W_n)_{n\in\mathbb{N}}$ has a further subsequence $(W_{n'_k})_{k\in\mathbb{N}}$ that converges for $\delta_{\square,\mathrm{m}}$ to a limit, say $U \in \widetilde{\mathcal{W}}_1$. Since $\delta_{\square,\mathrm{m}}$ is finer than $\delta_{\square,\mathrm{m}'}$, we deduce that $(W_{n'_k})_{k\in\mathbb{N}}$ converges also to U for $\delta_{\square,\mathrm{m}'}$; but, as a subsequence, it also converges to W for $\delta_{\square,\mathrm{m}'}$. As $\delta_{\square,\mathrm{m}'}$ is a distance on $\mathcal{W}_1$ thanks to Theorem 3.17, we get $U = W$. Hence, every subsequence of $(W_n)_{n\in\mathbb{N}}$ has a further subsequence that converges to W for $\delta_{\square,\mathrm{m}}$, therefore the whole sequence itself converges to W for $\delta_{\square,\mathrm{m}}$. Consequently, $\delta_{\square,\mathrm{m}'}$ is finer than $\delta_{\square,\mathrm{m}}$, and thus those two distances induce the same topology on $\widetilde{\mathcal{W}}_1$. □

The following theorem states that under appropriate assumptions, the topology induced by $\delta_{\square,\mathrm{m}}$ does not depend on d_{m}. We prove this theorem in Section 8. Recall that under suitable conditions satisfied in the next theorem, the quotient space $\widetilde{\mathcal{W}}_1$ does not depend on the choice of the distance d_{m}, see Theorem 3.17.

Theorem 5.5 (Equivalence of topologies induced by $\delta_{\square,\mathrm{m}}$ on $\widetilde{\mathcal{W}}_1$). — *The topology on the space of probability-graphons $\widetilde{\mathcal{W}}_1$ induced by the distance $\delta_{\square,\mathrm{m}}$ does not depend on the choice of the distance d_{m} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, as long as d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ and the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is (invariant) smooth, weakly regular and regular w.r.t. the stepping operator.*

Recall from Proposition 4.13 that when the distance d_{m} is quasi-convex and continuous w.r.t. the weak topology on $\mathcal{M}_{+}(\mathbf{Z})$ or $\mathcal{M}_{\pm}(\mathbf{Z})$, then the cut distance $d_{\square,\mathrm{m}}$ is invariant, smooth, weakly regular and regular w.r.t. the stepping operator. This is in particular the case of d_{LP}, $\|\cdot\|_{\mathcal{F}}$, $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathrm{FM}}$.

The next corollary is an immediate consequence of Lemma 3.22, Corollary 4.14, Lemma 5.3 and Theorem 5.5. This corollary gathers results comparing the topology induced by the cut distances associated with the distances introduced in Section 3.8. It is yet unclear if the distances $d_{\square,\mathcal{F}}$ induce the same topology on the space of labeled probability-graphons $\mathcal{W}_1$ as the one induced by $d_{\square,\mathrm{LP}}$, $d_{\square,\mathrm{FM}}$ or $d_{\square,\mathrm{KR}}$.

Corollary 5.6 (Topological equivalence of the cut distances associated to d_{LP}, $\|\cdot\|_{\mathrm{FM}}$, $\|\cdot\|_{\mathrm{KR}}$ and $\|\cdot\|_{\mathcal{F}}$). — *The cut distances $d_{\square,\mathrm{LP}}$ on $\mathcal{W}_+$ and $d_{\square,\mathrm{KR}}, d_{\square,\mathrm{FM}}$ and $d_{\square,\mathcal{F}}$ on $\mathcal{W}_\pm$ are invariant, smooth, weakly regular and regular w.r.t. the stepping operator. Moreover, we have the following comparison between the distances introduced in Section 3.8.*

(i) The cut norms $\|\cdot\|_{\square,\mathrm{FM}}$ and $\|\cdot\|_{\square,\mathrm{KR}}$ (resp. the cut distances $\delta_{\square,\mathrm{FM}}$ and $\delta_{\square,\mathrm{KR}}$) are metrically equivalent on $\mathcal{W}_\pm$ (resp. $\widetilde{\mathcal{W}}_\pm$).

(ii) The cut distances $\delta_{\square,\mathrm{FM}}$, $\delta_{\square,\mathrm{KR}}$ and $\delta_{\square,\mathrm{LP}}$ (resp. $d_{\square,\mathrm{FM}}$, $d_{\square,\mathrm{KR}}$ and $d_{\square,\mathrm{LP}}$) are uniformly continuous w.r.t. one another, and thus induce the same topology on $\widetilde{\mathcal{W}}_1$ (resp. $\mathcal{W}_1$) and on every uniformly bounded subset of $\widetilde{\mathcal{W}}_+$ (resp. $\mathcal{W}_+$).

(iii) The cut distances $\delta_{\square,\mathrm{FM}}$, $\delta_{\square,\mathrm{KR}}$, $\delta_{\square,\mathrm{LP}}$ and $\delta_{\square,\mathcal{F}}$, for every choice of the convergence determining sequence $\mathcal{F}$, induce the same topology on $\widetilde{\mathcal{W}}_1$.

Proof. — The first part of the corollary is a re-statement of Corollary 4.14. Point (i) is an immediate consequence of (3.9).

We now prove Point (ii). Thanks to (3.9) and Point (i), it is enough to consider only the Lévy–Prokhorov and the Kantorovitch–Rubinshtein distances. As d_{LP} is uniformly continuous w.r.t. d_{KR} (see Lemma 3.22), applying Lemma 5.3 (recall Corollary 4.14) with Remark 5.4 in mind, we get that $\delta_{\square,\mathrm{LP}}$ (resp. $d_{\square,\mathrm{LP}}$) is uniformly continuous w.r.t. $\delta_{\square,\mathrm{KR}}$ (resp. $d_{\square,\mathrm{KR}}$) on every uniformly bounded subset of $\widetilde{\mathcal{W}}_+$ (resp. $\mathcal{W}_+$) As d_{KR} is also uniformly continuous w.r.t. d_{LP} (see Lemma 3.22), applying again Lemma 5.3, we have that $\delta_{\square,\mathrm{KR}}$ (resp. $d_{\square,\mathrm{KR}}$) is uniformly continuous w.r.t. $\delta_{\square,\mathrm{LP}}$ (resp. $d_{\square,\mathrm{LP}}$) on every uniformly bounded subset of $\widetilde{\mathcal{W}}_+$ (resp. $\mathcal{W}_+$).

Point (iii) is an immediate consequence of Corollary 4.14 and Theorem 5.5, together with Point (ii). □

Remark 5.7 (Extension to uniformly bounded subsets of $\widetilde{\mathcal{W}}_+$). — In Theorem 5.5 and also in Corollary 5.6 (iii), one can replace $\widetilde{\mathcal{W}}_1$ by $\widetilde{\mathcal{W}}_{\mathcal{M}}$ with a bounded subset $\mathcal{M} \subset \mathcal{M}_+(\mathbf{Z})$ as soon as the distance d_{m} is defined on $\mathcal{M}_+(\mathbf{Z})$. (One has in mind the case $\mathcal{M} = \mathcal{M}_{\leqslant 1}(\mathbf{Z})$.) This can be seen by an easy modification in the proof of Theorem 5.5. Alternatively, this can be seen using scaling to reduce the case of general $\mathcal{M}$ to the case of $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, and then adding a cemetery point (for missing mass of measures) to $\mathbf{Z}$ to further reduce to the case of $\mathcal{M}_1(\mathbf{Z})$.

5.3. Completeness

Let d_{m} be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ or $\mathcal{M}_+(\mathbf{Z})$. We shall consider a slight modification of the cut distances $d_{\square,\mathrm{m}}$ and $\delta_{\square,\mathrm{m}}$ to achieve completeness. Recall the measure $M_W \in \mathcal{M}_+(\mathbf{Z})$ defined by (4.1) associated to $W \in \mathcal{W}_+$.

Definition 5.8 (The cut distances $d^{\mathrm{c}}_{\square,m}$ and $\delta^{\mathrm{c}}_{\square,m}$). — *Let d_{m} and d^{c} be two distances on $\mathcal{M}_\epsilon(\mathbf{Z})$ with $\epsilon \in \{\leqslant 1, +\}$. We define the cut distance $d^{\mathrm{c}}_{\square,m}$ on the space of $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels $\mathcal{W}_\epsilon$ as:*

$$d^{\mathrm{c}}_{\square,m}(U,W) = d_{\square,\mathrm{m}}(U,W) + d^{\mathrm{c}}(M_U, M_W),$$

and the cut (pseudo-)distance $\delta^{\mathrm{c}}_{\square,m}$ on the space of unlabeled $\mathcal{M}_\epsilon(\mathbf{Z})$-valued kernels $\widetilde{\mathcal{W}}_\epsilon$ as:

$$\delta^{\mathrm{c}}_{\square,m}(U,W) = \inf_{\varphi\in S_{[0,1]}} d^{\mathrm{c}}_{\square,m}(U,W^\varphi) = \delta_{\square,\mathrm{m}}(U,W) + d^{\mathrm{c}}(M_U, M_W).$$

Notice that by Lemma 3.11 and the definition of M_W, the distance $d^{\mathrm{c}}_{\square,m}$ is invariant.

Lemma 5.9 (Topological equivalence of $\delta_{\square,\mathrm{m}}$ and $\delta^{\mathrm{c}}_{\square,m}$). — *Let d_{m} and d^{c} be two distances on $\mathcal{M}_\epsilon(\mathbf{Z})$, with $\epsilon \in \{\leqslant 1, +\}$, such that d^{c} is continuous w.r.t. d_{m} and that $d_{\square,\mathrm{m}}$ is (invariant and) smooth on $\mathcal{W}_\epsilon$. Then, the cut distance $d^{\mathrm{c}}_{\square,m}$ is invariant and smooth and $\delta^{\mathrm{c}}_{\square,m}$ is a distance on $\widetilde{\mathcal{W}}_\epsilon$. Moreover, the distances $d_{\square,\mathrm{m}}$ and $d^{\mathrm{c}}_{\square,m}$ (resp. $\delta_{\square,\mathrm{m}}$ and $\delta^{\mathrm{c}}_{\square,m}$) induce the same topology on the space $\mathcal{W}_\epsilon$ (resp. $\widetilde{\mathcal{W}}_\epsilon$).*

Proof. — Let $(W_n)_{n\in\mathbb{N}}$ and W be elements of $\mathcal{W}_{\leqslant 1}$ such that $(W_n(x,y;\cdot))_{n\in\mathbb{N}}$ weakly converges to $W(x,y;\cdot)$ for almost every $x,y \in [0,1]$. Since the distance $d_{\square,\mathrm{m}}$ is smooth, we have that $\lim_{n\to\infty} d_{\square,\mathrm{m}}(W_n,W) = 0$. Using Lemma 4.9 on the continuity of the map $W \mapsto M_W$ and that d^{c} is continuous w.r.t. d_{m}, we obtain that $\lim_{n\to\infty} d^{\mathrm{c}}_{\square,m}(W_n,W) = 0$. This gives that the distance $d^{\mathrm{c}}_{\square,m}$ is smooth. Since we have already seen that $d^{\mathrm{c}}_{\square,m}$ is invariant, we deduce from Theorem 3.17 that $\delta^{\mathrm{c}}_{\square,m}$ is a distance on $\widetilde{\mathcal{W}}_1$.

We now prove that the two distances $d_{\square,\mathrm{m}}$ and $\delta^{\mathrm{c}}_{\square,m}$ induce the same topology (which implies that this is also true for $\delta_{\square,\mathrm{m}}$ and $\delta^{\mathrm{c}}_{\square,m}$). As $d_{\square,\mathrm{m}} \leqslant d^{\mathrm{c}}_{\square,m}$, convergence for $d^{\mathrm{c}}_{\square,m}$ implies convergence for $d_{\square,\mathrm{m}}$. Conversely, let $(W_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{W}_\epsilon$ that converges for $d_{\square,\mathrm{m}}$ to a limit, say $W \in \mathcal{W}_\epsilon$. Using again Lemma 4.9 and the continuity of d^{c} w.r.t. d_{m}, we obtain that $\lim_{n\to\infty} d^{\mathrm{c}}(M_{W_n}, M_W) = 0$. This clearly implies that the sequence $(W_n)_{n\in\mathbb{N}}$ converges to W for $d^{\mathrm{c}}_{\square,m}$. Then, the two distances have the same convergent sequences and thus induce the same topology (see Remark 2.1). □

Recall $\mathbf{Z}$ is a Polish space. We already proved in Proposition 4.6 that the space $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is separable; and we now investigate completeness of this space.

Theorem 5.10 ($\widetilde{\mathcal{W}}_1$ is a Polish space). — *Let d_{m} and d^{c} be two distances on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that d^{c} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, d^{c} is complete and continuous w.r.t. d_{m}, and $d_{\square,\mathrm{m}}$ is (invariant) smooth and weakly regular on $\mathcal{W}_1$. Then, the space $(\widetilde{\mathcal{W}}_1, \delta^{\mathrm{c}}_{\square,m})$ is a Polish metric space.*

Note that the assumptions in Theorem 5.10 imply that d_{m} also induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$. Indeed, as d^{c} is continuous w.r.t. d_{m}, the topology induced by d_{m} if finer than the topology induced by d^{c}, i.e. finer than the weak topology. As $d_{\square,\mathrm{m}}$ is smooth, by Lemma 3.12, d_{m} is continuous w.r.t. the weak topology (i.e. the weak topology if finer than the topology induced by d_{m}), and thus d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$.

Also note that Theorem 5.10 can easily be extended to $\mathcal{W}_{\leqslant 1}$ or the space of unlabeled $\mathcal{M}$-valued kernels $\widetilde{\mathcal{W}}_{\mathcal{M}}$ when $\mathcal{M}$ is a bounded convex closed subset of $\mathcal{M}_+(\mathbf{Z})$.

Proof. — From Lemma 5.9, we have that $\delta^{\mathrm{c}}_{\square,m}$ is a distance on $\widetilde{\mathcal{W}}_1$ which induces the same topology as $\delta_{\square,\mathrm{m}}$, and from Proposition 4.6, we have that $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$, and thus $(\widetilde{\mathcal{W}}_1, \delta^{\mathrm{c}}_{\square,m})$, is separable. To get that this latter space is Polish, we are left to prove that the distance $\delta^{\mathrm{c}}_{\square,m}$ is complete.

Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of probability-graphons that is Cauchy for $\delta^{\mathrm{c}}_{\square,m}$. By definition of the cut distance $\delta^{\mathrm{c}}_{\square,m}$, the sequence of probability measures $(M_{W_n})_{n\in\mathbb{N}}$ is Cauchy in $\mathcal{M}_1(\mathbf{Z})$ for the complete distance d^{c}. Thus, the sequence $(M_{W_n})_{n\in\mathbb{N}}$ is weakly convergent as d^{c} induces the weak topology, which implies that it is tight (see Lemma 2.8). Hence, by definition, the sequence of probability-graphons $(W_n)_{n\in\mathbb{N}}$ is tight. By Theorem 5.1 (i), there exists a subsequence $(W_{n_k})_{k\in\mathbb{N}}$ that converges for $\delta_{\square,\mathrm{m}}$ to a limit, say $W \in \widetilde{\mathcal{W}}_1$. This subsequence also converges for $\delta^{\mathrm{c}}_{\square,m}$ to W as $\delta_{\square,\mathrm{m}}$ and $\delta^{\mathrm{c}}_{\square,m}$ induce the same topology. Finally, because the sequence

$(W_n)_{n\in\mathbb{N}}$ is Cauchy for $\delta^{\mathrm{c}}_{\square,m}$ and has a subsequence converging to W for $\delta^{\mathrm{c}}_{\square,m}$, the whole sequence must also converge to W for $\delta^{\mathrm{c}}_{\square,m}$. Consequently, the distance $\delta^{\mathrm{c}}_{\square,m}$ is complete. □

The following lemma shows that every probability measure can be represented as a constant probability-graphon.

Lemma 5.11 ($\mathcal{M}_1(\mathbf{Z})$ seen as a closed subset of $\mathcal{W}_1$). — *Let d_{m} be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that $d_{\square,\mathrm{m}}$ is (invariant and) smooth on $\mathcal{W}_1$. Then, the map $\mu \mapsto W_\mu \equiv \mu$ is an injection from $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ to $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ with a closed range and continuous inverse.*

Proof. — For any $\mu \in \mathcal{M}_1(\mathbf{Z})$ consider the constant probability-graphon $W_\mu \equiv \mu$, and notice that $M_{W_\mu} = \mu$, that $W_\mu(S\times T;\cdot) = \lambda(S)\lambda(T)\,\mu$ for all measurable $S,T \subset [0,1]$, and that $W_\mu^\varphi = W_\mu$ for any measure-preserving map φ. This readily implies that for $\mu \in \mathcal{M}_1(\mathbf{Z})$ and $W \in \mathcal{W}_1$:

$$\begin{aligned}\delta_{\square,\mathrm{m}}(W_\mu, W) &= d_{\square,\mathrm{m}}(W_\mu, W)\\ &= \sup_{S,T\subset[0,1]} d_{\mathrm{m}}(\lambda(S)\lambda(T)\mu, W(S\times T;\cdot))\\ &\geqslant d_{\mathrm{m}}(\mu, M_W).\end{aligned} \tag{5.1}$$

In particular, taking $W = W_\nu$ for $\nu \in \mathcal{M}_1(\mathbf{Z})$ we get that $\delta_{\square,\mathrm{m}}(W_\mu, W_\nu) \geqslant d_{\mathrm{m}}(\mu,\nu)$. This implies that the map $\mathcal{I} : \mu \mapsto W_\mu \equiv \mu$ is an injection, and its inverse, given by the map $W_\mu \mapsto \mu$, is 1-Lipschitz.

Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_1(\mathbf{Z})$ such that the sequence $(W_{\mu_n})_{n\in\mathbb{N}}$ converges for $\delta_{\square,\mathrm{m}}$ to a limit, say W. We deduce from (5.1) that $(\mu_n)_{n\in\mathbb{N}}$ converges for d_{m} to $\mu = M_W$ and that for all measurable $S,T \subset [0,1]$, $(\lambda(S)\lambda(T)\mu_n)_{n\in\mathbb{N}}$ converges for d_{m} to $W(S\times T;\cdot)$. This implies that $W(S\times T;\cdot) = \lambda(S)\lambda(T)\mu(\cdot)$ for all measurable $S,T \subset [0,1]$, that is, $W = W_\mu$. This implies that the image by $\mathcal{I}$ of any closed subset of $\mathcal{M}_1(\mathbf{Z})$ is a closed subset of $\mathcal{W}_1$, and thus the range of $\mathcal{I}$ is closed. □

Remark 5.12 (Extension to isometric representation of $\mathcal{M}_1(\mathbf{Z})$). — If the distance d_{m}, in addition to the hypothesis of Lemma 5.11, is sub-homogeneous, that is, for all $\mu,\nu \in \mathcal{M}_1(\mathbf{Z})$ we have $d_{\mathrm{m}}(\mu,\nu) = \sup_{r\in[0,1]} d_{\mathrm{m}}(r\mu, r\nu)$ (which is the case if d_{m} is quasi-convex), then we deduce from (5.1) that the map $\mu \mapsto W_\mu \equiv \mu$ is isometric from $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ to $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$.

We now state a characterization of compactness and completeness for the space of probability-graphons. Recall $\mathbf{Z}$ is a Polish space.

Corollary 5.13 (Characterization of compactness and completeness for $\widetilde{\mathcal{W}}_1$). — *Let d_{m} be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, which induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, and such that $d_{\square,\mathrm{m}}$ is (invariant) smooth and weakly regular on $\mathcal{W}_1$. We have the following properties.*

(i) $\mathbf{Z}$ *is compact* $\iff$ $(\mathcal{M}_{\leqslant 1}(\mathbf{Z}), d_{\mathrm{m}})$ *is compact* $\iff$ $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ *is compact.*

(ii) If $(\mathcal{M}_{\leqslant 1}(\mathbf{Z}), d_{\mathrm{m}})$ is complete then $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is complete.

(iii) Assume furthermore that d_{m} is sub-homogeneous (see Remark 5.12). If $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is complete, then $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ is complete.

Proof. — We prove Point (i). From Remark 2.7, we already know that $\mathbf{Z}$ is compact if and only if $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ is weakly compact, i.e. compact for d_{m} as d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$.

Now, assume that $(\mathcal{M}_{\leqslant 1}(\mathbf{Z}), d_{\mathrm{m}})$ is compact. Applying Theorem 5.1 (iii), we get that the space $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is also compact.

Conversely, assume that $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is compact. By Lemma 4.9, the mapping $W \mapsto M_W$ is continuous from $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ to $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$, and as $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is compact its image through this mapping is also compact. To conclude, it is enough to check that this mapping is surjective. But this is clear as the image of the constant probability-graphon $W_\mu \equiv \mu$ is $M_{W_\mu} = \mu$. Hence, $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ (and thus $(\mathcal{M}_{\leqslant 1}(\mathbf{Z}), d_{\mathrm{m}})$) is compact.

We prove Point (ii). Assume that $(\mathcal{M}_{\leqslant 1}(\mathbf{Z}), d_{\mathrm{m}})$ is complete. Thus, we can choose $d^{\mathrm{c}} = d_{\mathrm{m}}$ in Definition 5.8, and apply Theorem 5.10 to get that $(\widetilde{\mathcal{W}}_1, \delta^{\mathrm{c}}_{\square,m})$ is complete. As $d^{\mathrm{c}} = d_{\mathrm{m}}$, we have $\delta_{\square,\mathrm{m}} \leqslant \delta^{\mathrm{c}}_{\square,m} \leqslant 2\delta_{\square,\mathrm{m}}$. Hence, $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is also complete.

We prove Point (iii). Assume that $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is complete. Let $(\mu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence of probability measures in $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$. By Remark 5.12, the sequence of constant probability-graphons $(W_{\mu_n})_{n\in\mathbb{N}}$ is also Cauchy for $\delta_{\square,\mathrm{m}}$. As $(\widetilde{\mathcal{W}}_1, \delta_{\square,\mathrm{m}})$ is complete, there exists a probability-graphon $W \in \widetilde{\mathcal{W}}_1$ such that $(W_{\mu_n})_{n\in\mathbb{N}}$ converges to W for the cut distance $\delta_{\square,\mathrm{m}}$. Thanks to Lemma 5.11, W is constant equal to some $\mu \in \mathcal{M}_1(\mathbf{Z})$, and $(\mu_n)_{n\in\mathbb{N}}$ converges to μ for d_{m}. Hence, $(\mathcal{M}_1(\mathbf{Z}), d_{\mathrm{m}})$ is complete. □

6. Sampling from probability-graphons

Measure-valued graphons allow to define models for generating random weighted graphs that are more general than the models based on real-valued graphons. We prove that the weighted graphs sampled from probability-graphons are close to their original model for the cut distance $\delta_{\square,\mathcal{F}}$, where $\mathcal{F} = (f_k)_{k\in\mathbb{N}}$ (with $f_0 = \mathbb{1}$) is a convergence determining sequence. In doing so, we prove analogues for probability-graphons of the first and second sampling lemmas for real-valued kernels and graphons (see [16] or [46, Sections 10.3 and 10.4]).

It would have been more natural to work in Sections 6 and 7 with the Kantorovitch–Rubinshtein norm or the Fortet–Mourier norm that both treats all test functions in a uniform manner. Unfortunately, the supremum in the definition of both of these norms does not behave well regarding the probabilities and expectations of graphs sampled from probability-graphons. We need in our proofs (and in particular that of the First Sampling Lemma 6.7 below) to consider simultaneously only a finite number of test functions in order to control the probability of failure for our stochastic bounds. The proof of the First Sampling Lemma 6.7 is based on applying the first sampling lemma for real-valued kernels to several one dimensional-valued projections of measure-valued kernels. The proof of the Second Sampling Lemma 6.12 is similar to the one for real-valued kernels and graphons.

6.1. $\mathcal{M}_1(\mathbf{Z})$-Graphs and weighted graphs

A graph $G = (V, E)$ is composed of a finite set of vertices $V(G) = V$, and a set of edges $E(G) = E$ which is a subset of $V \times V$ avoiding the diagonal. When its set of edges $E(G)$ is symmetric, we say that G is *symmetric* or *non-oriented*. We denote by $v(G) = |V(G)|$ the number of vertices of this graph, and by $e(G) = |E(G)|$ its number of edges.

Definition 6.1 ($\mathcal{X}$-graphs). — *Let $\mathcal{X}$ be a non-empty set. A* $\mathcal{X}$-graph *is a triplet $G = (V, E, \Phi)$ where (V, E) is a graph and $\Phi : E \to \mathcal{X}$ is a map that associates a decoration $x = \Phi(e) \in \mathcal{X}$ to each edge $e \in E$. When $\mathcal{X} = \mathbf{Z}$, we say that G is a* weighted graph.

Furthermore, the graph G is said to be symmetric *if (V, E) is a symmetric graph and if Φ is a symmetric function, i.e. for every edge $(x, y) \in E$, we have $(y, x) \in E$ and $\Phi(x, y) = \Phi(y, x)$.*

Remark 6.2 ($\mathcal{M}_1(\mathbf{Z})$-Graphs as probability-graphons). — Any labeled $\mathcal{M}_1(\mathbf{Z})$-graph G can be naturally represented as an $\mathcal{M}_1(\mathbf{Z})$-valued graphon, which we denote by W_G, in the following way. Let $G = (V, E, M)$ be a $\mathcal{M}_1(\mathbf{Z})$-graph, with $v(G) = n \in \mathbb{N}^*$. Denote by $V = [n] = \{1, \dots, n\}$ the vertices of G. Consider intervals of length $1/n$: for $1 \leqslant i \leqslant n$, let $J_i = ((i-1)/n, i/n]$. We then define the $\mathcal{M}_1(\mathbf{Z})$-valued graphon stepfunction W_G associated with the $\mathcal{M}_1(\mathbf{Z})$-graph G by:

$$\forall (i,j) \in E, \quad \forall (x,y) \in J_i \times J_j, \quad W_G(x,y;\mathrm{d}z) = \Phi(i,j)(\mathrm{d}z);$$

and $W_G(x,y;\mathrm{d}z)$ equals the Dirac mass at ∂ otherwise, where ∂ is an element of $\mathbf{Z}$ used as a cemetery point for missing edges in graphs.

In this section, we investigate weighted graphs sampled from probability-graphons. Hence, using the cemetery point argument in the remark above, we only consider complete graphs for the rest of this section.

Let d_{m} be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$. If G and H have the same vertex-set, the cut distance between them is defined as the cut distance between their associated graphons:

$$d_{\square,\mathrm{m}}(G,H) = d_{\square,\mathrm{m}}(W_G, W_H).$$

When G and H do not have the same vertex-sets, as the numbering of the vertices in Remark 6.2 is arbitrary, we must consider the unlabeled cut distance between them defined as the cut distance between their associated graphons:

$$\delta_{\square,\mathrm{m}}(G,H) = \delta_{\square,\mathrm{m}}(W_G, W_H).$$

Recall that when the distance d_{m} derives from a norm N_{m} on $\mathcal{M}_{\pm}(\mathbf{Z})$, Lemma 3.19 applies, and the cut distance $d_{\square,\mathrm{m}}(G,H)$ can be rewritten as a combinatorial optimization over whole steps.

Remark 6.3 (Weighted graphs as $\mathcal{M}_1(\mathbf{Z})$-graphs). — We will sometimes need to interpret a weighted graph G as a $\mathcal{M}_1(\mathbf{Z})$-graph where a weight x on an edge is replaced by δ_x the Dirac mass located at x.

Notation 6.4 (The real-weighted graph $G[f]$). — For a $\mathcal{M}_1(\mathbf{Z})$-graph (resp. weighted graph) G and a function $f \in C_b(\mathbf{Z})$, we denote by $G[f]$ the real-weighted graph with the same vertex set and edge set as G, and where the edge (i,j) has weight $\Phi_{G[f]}(i,j) = \Phi_G(i,j;f) = \int_{\mathbf{Z}} f(z)\Phi_G(i,j;\mathrm{d}z)$ (resp. $\Phi_{G[f]}(i,j) = f(\Phi_G(i,j))$), where Φ_G is the decoration of the $\mathcal{M}_1(\mathbf{Z})$-graph G. This notation is consistent when one identifies the decoration $z \in \mathbf{Z}$ for a $\mathbf{Z}$-graph with the decoration given by the Dirac mass at z for the corresponding $\mathcal{M}_1(\mathbf{Z})$-graph. It is also consistent with the notation in (3.7) for probability graphons.

6.2. W-random graphs

Let W be a probability-graphon, and $x = (x_1, \ldots, x_n)$, $n \in \mathbb{N}^*$, be a sequence of points from $[0,1]$. We define the $\mathcal{M}_1(\mathbf{Z})$-graph $\mathbb{H}(x, W)$ as the complete graph whose vertex set is $[n] = \{1, \ldots, n\}$, and with each edge (i, j) decorated by the probability measure $W(x_i, x_j; \mathrm{d}z)$.

Let H be any $\mathcal{M}_1(\mathbf{Z})$-graph. We can define from H a random weighted (directed) graph $\mathbb{G}(H)$ whose vertex set $V(H)$ and edge set $E(H)$ are the same as H, and with each edge (i, j) having a random weight $\beta_{i,j}$ distributed according to the probability distribution decorating the edge (i, j) in H, all the weights being independent from each other. For the special case where $H = \mathbb{H}(x, W)$, we simply note $\mathbb{G}(x, W) = \mathbb{G}(\mathbb{H}(x, W))$.

An important special case is when the sequence X is chosen at random: $X = (X_i)_{1 \leqslant i \leqslant n}$ where the X_i are independent and uniformly distributed on $[0,1]$. For this special case, we simply note $\mathbb{H}(n, W) = \mathbb{H}(X, W)$ and $\mathbb{G}(n, W) = \mathbb{G}(X, W)$, that are conditionally on $X = x$, distributed respectively as $\mathbb{H}(x, W)$ and $\mathbb{G}(x, W)$. The random graphs $\mathbb{H}(n, W)$ and $\mathbb{G}(n, W)$ are called W-random graphs.

Remark 6.5 (The case of symmetric graphons). — In the special case where W is a symmetric probability-graphon, the $\mathcal{M}_1(\mathbf{Z})$-graph $\mathbb{H}(x, W)$ is also symmetric. From a symmetric $\mathcal{M}_1(\mathbf{Z})$-graph H, the random weighted graph $\mathbb{G}(H)$ is not necessarily symmetric, but we can define a random symmetric weighted graph $\mathbb{G}^{\mathrm{sym}}(H)$ whose vertex set $V(H)$ and $E(H)$ are the same as H, and with independent weights $\beta_{i,j} = \beta_{j,i}$ on each edge $(i, j) = (j, i)$ distributed according to $\Phi_H(i, j; \cdot)$. For $H = \mathbb{H}(x, W)$ we simply note $\mathbb{G}^{\mathrm{sym}}(x, W)$ and $\mathbb{G}^{\mathrm{sym}}(n, W)$.

For a weighted graph G, and for $1 \leqslant k \leqslant v(G)$, we can define the random weighted graph $\mathbb{G}(k, G)$ as being the sub-graph of G induced by a uniform random subset of k distinct vertices from G. Then, upper bounding by the probability that a uniformly-chosen map $[k] \to V(G)$ is non-injective, we get the following bound on the total variation distance between the graphs obtained from G and its associated graphon W_G:

$$d_{\mathrm{var}}(\mathbb{G}(k, G), \mathbb{G}(k, W_G)) \leqslant \binom{k}{2} \frac{1}{v(G)},$$

where d_{var} is the total variation distance between probability measures.

6.3. Estimation of the distance by sampling

6.3.1. The first sampling lemma

In this subsection, we link sampling from graphons with the cut distance. This result is the equivalent of Lemma 10.6 in [46]. The main consequence of the following lemma is that the cut distance $d_{\square,\mathcal{F}}$ between two probability-graphons can be estimated by sampling.

Notation 6.6 (The random stepfunction W_X). — For a measure-valued kernel W (resp. a real-valued kernel w) and a vector $X = (X_i)_{1\leqslant i\leqslant k}$ composed of k independent random variables uniformly distributed over $[0,1]$, we denote by $W_X = W_{\mathbb{H}(k,W)}$ (resp. w_X) the random measure-valued (resp. real-valued) stepfunction with k steps of size $1/k$, and where the step (i,j) has value $W(X_i, X_j;\cdot)$ (resp. $w(X_i, X_j)$).

Lemma 6.7 (First Sampling Lemma). — *Let $\mathcal{F}$ be a convergence determining sequence. Let $k \in \mathbb{N}^*$, and $U, W \in \mathcal{W}_1$ be two probability-graphons, and let X be a random vector uniformly distributed over $[0,1]^k$. Then with probability at least $1 - 4k^{1/4}\,\mathrm{e}^{-\sqrt{k}/8}$, we have:*

$$\Big|\|U_X - W_X\|_{\square,\mathcal{F}} - \|U - W\|_{\square,\mathcal{F}}\Big| \leqslant \frac{11}{k^{1/4}}\cdot$$

An immediate consequence of Lemma 6.7 is that the decorated graphs with probability measures on their edges $\mathbb{H}(k,U)$ and $\mathbb{H}(k,W)$ can be coupled in order that $d_{\square,\mathcal{F}}(\mathbb{H}(k,U),\mathbb{H}(k,W))$ is close to $d_{\square,\mathcal{F}}(U,W)$ with high probability.

To prove the first sampling lemma, we first need to prove the following lemma which states that the cut norm $\|\cdot\|_{\square,\mathcal{F}}$ can be approximated by the maximum of the one-sided cut norm using a finite number of functions. Recall from Remark 3.24 the definition of the one-sided version of the cut norm $\|\cdot\|^+_{\square,\mathbb{R}}$.

Lemma 6.8 (Approximation bound with $\|\cdot\|_{\square,\mathcal{F}}$ and $\|\cdot\|^+_{\square,\mathbb{R}}$). — *Let $U, W \in \mathcal{W}_1$ and let $N \in \mathbb{N}$. For every $\varepsilon = (\varepsilon_n)_{1\leqslant n\leqslant N} \in \{\pm 1\}^N$, define $g_{N,\varepsilon} = \sum_{n=1}^N 2^{-n}\varepsilon_n f_n$. Then, we have:*

$$\|U - W\|_{\square,\mathcal{F}} - 2^{-N} \leqslant \max_{\varepsilon\in\{\pm1\}^N} \|(U-W)\,[g_{N,\varepsilon}]\|^+_{\square,\mathbb{R}} \leqslant \|U - W\|_{\square,\mathcal{F}}.$$

Proof. — First remark that for $n \in \mathbb{N}$, f_n takes values in $[0,1]$, and thus $U[f_n] - W[f_n]$ takes values in $[-1,1]$. Recall that $f_0 = \mathbb{1}$, and thus

$U[f_0] - W[f_0] \equiv 0$. Upper bounding integrals by 1 for indices $n > N$, we get:

$$\|U - W\|_{\square,\mathcal{F}} \leqslant \sup_{S,T\subset[0,1]} \sum_{n=1}^{N} 2^{-n} \left| \int_{S\times T} (U - W)[f_n](x, y)\, \mathrm{d}x\mathrm{d}y \right| + 2^{-N}.$$

And adding the non-negative terms for $n > N$, we get:

$$\sup_{S,T\subset[0,1]} \sum_{n=1}^{N} 2^{-n} \left| \int_{S\times T} (U - W)[f_n](x, y)\, \mathrm{d}x\mathrm{d}y \right| \leqslant \|U - W\|_{\square,\mathcal{F}}.$$

Using the same idea as in (3.12) and (3.13), we get:

$$\sup_{S,T\subset[0,1]} \sum_{n=1}^{N} 2^{-n} \left| \int_{S\times T} (U - W)[f_n](x, y)\, \mathrm{d}x\mathrm{d}y \right| = \max_{\varepsilon\in\{\pm1\}^N} \|(U - W)\,[g_{N,\varepsilon}]\|^{+}_{\square,\mathbb{R}}\,,$$

which concludes the proof. □

Proof of Lemma 6.7. — Remark that for $f \in C_b(\mathbf{Z})$ and $W \in \mathcal{W}_\pm$, we have $(W_X)[f] = (W[f])_X$, and we thus write $W[f]_X$ without any ambiguity.

Assume that $k \geqslant 2^4$ (otherwise the lower bound in the lemma is trivial). Set $N = \lceil \log_2(k^{1/4})\rceil$, so that $2^{-1}k^{-1/4} < 2^{-N} \leqslant k^{-1/4}$. Let $\varepsilon \in \{\pm1\}^N$. Remark that as the f_n take values in $[0, 1]$, the real-valued kernels $(U - W)[f_n]$ take values in $[-1, 1]$, and thus the real-valued kernel $(U - W)[g_{N,\varepsilon}]$ also take values in $[-1, 1]$. Applying [16, Theorem 4.6] to the real-valued kernel $(U - W)\,[g_{N,\varepsilon}]$, we get with probability at least $1 - 2\,\mathrm{e}^{-\sqrt{k}/8}$ that:

$$\left| \|(U - W)[g_{N,\varepsilon}]_X\|^{+}_{\square,\mathbb{R}} - \|(U - W)[g_{N,\varepsilon}]\|^{+}_{\square,\mathbb{R}} \right| \leqslant \frac{10}{k^{1/4}}, \tag{6.1}$$

where recall that $\|\cdot\|^{+}_{\square,\mathbb{R}}$ is the one-sided version of the cut norm for real-valued kernels defined in (3.11). Hence, with probability at least $1 - 2^{N+1}\,\mathrm{e}^{-\sqrt{k}/8} \geqslant 1 - 4k^{1/4}\,\mathrm{e}^{-\sqrt{k}/8}$, we have that the bounds in (6.1) holds for every $\varepsilon \in \{\pm1\}^N$ simultaneously; and when all of this holds, applying Lemma 6.8 to U, W and to U_X, W_X, we get:

$$\begin{aligned}\|U_X - W_X\|_{\square,\mathcal{F}} &\leqslant \max_{\varepsilon\in\{\pm1\}^N} \|(U - W)[g_{N,\varepsilon}]_X\|^{+}_{\square,\mathbb{R}} + 2^{-N}\\ &\leqslant \max_{\varepsilon\in\{\pm1\}^N} \|(U - W)[g_{N,\varepsilon}]\|^{+}_{\square,\mathbb{R}} + \frac{11}{k^{1/4}}\\ &\leqslant \|U - W\|_{\square,\mathcal{F}} + \frac{11}{k^{1/4}},\end{aligned}$$

and similarly:

$$\begin{aligned}\|U-W\|_{\square,\mathcal{F}} &\leqslant \max_{\varepsilon\in\{\pm 1\}^N} \|(U-W)[g_{N,\varepsilon}]\|^{+}_{\square,\mathbb{R}} + 2^{-N} \\ &\leqslant \max_{\varepsilon\in\{\pm 1\}^N} \|(U-W)[g_{N,\varepsilon}]_X\|^{+}_{\square,\mathbb{R}} + \frac{11}{k^{1/4}} \\ &\leqslant \|U_X - W_X\|_{\square,\mathcal{F}} + \frac{11}{k^{1/4}}.\end{aligned}$$

This concludes the proof. □

6.3.2. Approximation with random weighted graphs

As a consequence of the First Sampling Lemma 6.7, we get that the cut distance between the sampled graphs $\mathbb{H}(k,U)$ and $\mathbb{H}(k,W)$ (with the proper coupling) is close to the cut distance between the probability-graphons U and W. The following lemma states that if k is large enough, then $\mathbb{G}(k,W)$ is close to $\mathbb{H}(k,W)$ in the cut distance $d_{\square,\mathcal{F}}$, and thus the cut distance between the random weighted graphs $\mathbb{G}(k,U)$ and $\mathbb{G}(k,W)$ is also close to $d_{\square,\mathcal{F}}(U,W)$.

Recall from Section 6.2 the definition of the random weighted graph $\mathbb{G}(H)$ when H is an $\mathcal{M}_1(\mathbf{Z})$-graph. Following Remarks 6.3 and 6.2, we shall see the weighted graph $\mathbb{G}(H)$ as a $\mathcal{M}_1(\mathbf{Z})$-graph or even as a probability-graphon.

Lemma 6.9 (Bound in probability for $d_{\square,\mathcal{F}}(\mathbb{G}(H),H)$). — *For every $\mathcal{M}_1(\mathbf{Z})$-graph H with k vertices, and for every $\varepsilon \geqslant 10/\sqrt{k}$, we have:*

$$\mathbb{P}\Big(d_{\square,\mathcal{F}}(\mathbb{G}(H),H) > 2\varepsilon\Big) \leqslant e^{-\varepsilon^2 k^2}.$$

Remark 6.10 (Bound in expectation for $d_{\square,\mathcal{F}}(\mathbb{G}(H),H)$). — Recall that $d_{\square,\mathcal{F}}(\mathbb{G}(H),H) \leqslant 1$. Applying Lemma 6.9 with $\varepsilon = 10/\sqrt{k}$, we get the following bound on the expectation of $d_{\square,\mathcal{F}}(\mathbb{G}(H),H)$:

$$\mathbb{E}[d_{\square,\mathcal{F}}(\mathbb{G}(H),H)] \leqslant \frac{20}{\sqrt{k}} + \mathrm{e}^{-100k} < \frac{21}{\sqrt{k}}.$$

Proof of Lemma 6.9. — Let H and ε be as in the lemma. Assume that $\varepsilon \leqslant 1/2$ (otherwise the probability to bound in the lemma is null). To simplify the notations, denote by $G = \mathbb{G}(H)$ throughout this proof. Define $N = \lceil \log_2(\varepsilon^{-1}) \rceil$, so that $\sum_{n=N+1}^{\infty} 2^{-n} \leqslant \varepsilon$. Upper bounding by 1 the terms for $n > N$ in (3.14), we get for $U, W \in \mathcal{W}_1$:

$$d_{\square,\mathcal{F}}(U,W) \leqslant \sum_{n=1}^{N} 2^{-n} \|U[f_n] - W[f_n]\|_{\square,\mathbb{R}} + \ \varepsilon,$$

where recall that $\|\cdot\|_{\square,\mathbb{R}}$ is the cut norm for real-valued kernels defined in (3.11). Using this equation with the graphs G and H, we get:

$$\begin{aligned}\mathbb{P}(d_{\square,\mathcal{F}}(G,H) > 2\varepsilon) &\leqslant \mathbb{P}\left(\sum_{n=1}^{N} 2^{-n} d_{\square,\mathbb{R}}(G[f_n],H[f_n]) \ > \varepsilon\right) \\ &\leqslant \sum_{n=1}^{N} \mathbb{P}\left(d_{\square,\mathbb{R}}(G[f_n],H[f_n]) \ > \varepsilon\right), \end{aligned} \tag{6.2}$$

where $d_{\square,\mathbb{R}}$ denotes the cut distance associated to the cut norm $\|\cdot\|_{\square,\mathbb{R}}$ for real-valued graphons and kernels. Remark that for every $n \in \mathbb{N}$, $H[f_n]$ and $G[f_n]$ are real-weighted graphs with weights in $[0,1]$. Thus, by a straightforward adaptation of the proof of [46, Lemma 10.11], we get:

$$\forall n \in [N], \quad \mathbb{P}(d_{\square}(G[f_n],H[f_n]) \ > \varepsilon) \leqslant 2\cdot 4^k\, \mathrm{e}^{-2\varepsilon^2 k^2}. \tag{6.3}$$

Combining (6.2) and (6.3), we get for $\varepsilon > 10/\sqrt{k}$:

$$\mathbb{P}(d_{\square,\mathcal{F}}(G,H) > 2\varepsilon) \leqslant 2N4^k e^{-2\varepsilon^2 k^2} \leqslant \mathrm{e}^{-\varepsilon^2 k^2},$$

where the last bound derives from simple calculus. This concludes the proof. □

We can apply the First Sampling Lemma 6.7 along with Lemma 6.9 to get the following lemma, equivalent of the first sampling lemma for the random weighted graph $\mathbb{G}(k,W)$:

Corollary 6.11 (First Sampling Lemma for $\mathbb{G}(k,W)$). — *Let $U,W \in \mathcal{W}_1$ be two probability-graphons, and $k \in \mathbb{N}^*$. Then, we can couple the random weighted graphs $\mathbb{G}(k,U)$ and $\mathbb{G}(k,W)$ such that with probability at least $1-(4k^{1/4}+1)\,\mathrm{e}^{-\sqrt{k}/8}$, we have:*

$$\left|d_{\square,\mathcal{F}}(\mathbb{G}(k,U),\mathbb{G}(k,W)) - d_{\square,\mathcal{F}}(U,W)\right| \leqslant \frac{14}{k^{1/4}}\cdot$$

Proof. — Assume that $k \geqslant 14^4$ (otherwise the bound in the corollary is trivial). Then, we have with probability at least $1-4k^{1/4}\,\mathrm{e}^{-\sqrt{k}/8} - 2\,\mathrm{e}^{-100k}$

$> 1 - (4k^{1/4} + 1)\,\mathrm{e}^{-\sqrt{k}/8}$:

$$\begin{aligned}(6.4)\quad & \left|d_{\square,\mathcal{F}}(\mathbb{G}(k,U),\mathbb{G}(k,W)) - d_{\square,\mathcal{F}}(U,W)\right| \\ &\leqslant \left|d_{\square,\mathcal{F}}(\mathbb{G}(k,U),\mathbb{G}(k,W)) - d_{\square,\mathcal{F}}(\mathbb{H}(k,U),\mathbb{H}(k,W))\right| \\ &\quad + \left|d_{\square,\mathcal{F}}(\mathbb{H}(k,U),\mathbb{H}(k,W)) - d_{\square,\mathcal{F}}(U,W)\right| \\ &\leqslant d_{\square,\mathcal{F}}(\mathbb{G}(k,U),\mathbb{H}(k,U)) \\ &\quad + d_{\square,\mathcal{F}}(\mathbb{G}(k,W),\mathbb{H}(k,W)) + \frac{11}{k^{1/4}} \\ &\leqslant \frac{40}{\sqrt{k}} + \frac{11}{k^{1/4}} \\ &\leqslant \frac{14}{k^{1/4}},\end{aligned}$$

where we used the upper bound from the First Sampling Lemma 6.7 (which gives the coupling with the same random vector X to define both graphs $U_X = \mathbb{H}(k,U)$ and $W_X = \mathbb{H}(k,W)$) for the second inequality, the upper bound from Lemma 6.9 with $\varepsilon = 10/\sqrt{k}$ with both U and W for the third inequality, and that $\frac{1}{\sqrt{k}} \leqslant \frac{1}{14k^{1/4}}$ for the last inequality. □

6.4. The distance between a probability-graphon and its sample

In this section, we present the Second Sampling Lemma, that shows that a sampled $\mathcal{M}_1(\mathbf{Z})$-graph is close to its original probability-graphon with high probability. Note that we use the unlabeled cut distance $\delta_{\square,\mathcal{F}}$ rather than $d_{\square,\mathcal{F}}$ as the sample points are unordered. The bound on the distance is much weaker than the one in the First Sampling Lemma 6.7, but nevertheless goes to 0 as the sample size increases.

The proof is a straightforward adaptation of the proof of [46, Lemma 10.16] (replacing the weak regularity lemma and the first sampling lemma by their counterparts for probability-graphons, that is Lemmas 4.17 and 6.7; the sample concentration theorem for real-valued graphons can easily be adapted to probability-graphons).

Lemma 6.12 (Second Sampling Lemma). — *Let $\mathcal{F}$ be a convergence determining sequence. Let $W \in \widetilde{\mathcal{W}}_1$ be a probability-graphon and $k \in \mathbb{N}^*$. Then, with probability at least $1 - \exp(-k/(2\ln(k)))$ we have:*

$$\delta_{\square,\mathcal{F}}(\mathbb{H}(k,W),W) \leqslant \frac{21}{\sqrt{\ln(k)}} \qquad \text{and} \qquad \delta_{\square,\mathcal{F}}(\mathbb{G}(k,W),W) \leqslant \frac{22}{\sqrt{\ln(k)}}.$$

In the above lemma, the asymmetric random graph $\mathbb{G}(k, W)$ can be replaced by the symmetric random graph $\mathbb{G}^{\mathrm{sym}}(k, W)$ without changing the proof. Similarly, the results in Section 6.3.2 can be reformulated with symmetric random graphs $\mathbb{G}^{\mathrm{sym}}(k, W)$ and $\mathbb{G}^{\mathrm{sym}}(H)$ (but with a slight modification of the proof for Lemma 6.9 to symmetrize the random variable $X_{i,j}$ and with the upper bound $e^{-\varepsilon^2 k^2/2}$, see also [46, Lemma 10.11]).

As an immediate consequence of Lemma 6.12 and of the Borel–Cantelli lemma, we get the convergence of the sampled subgraphs for the cut distance $\delta_{\square,\mathcal{F}}$.

Theorem 6.13 (Convergence of sampled subgraphs). — *Let $\mathcal{F}$ be a convergence determining sequence. Let $W \in \widetilde{\mathcal{W}}_1$ be a probability-graphon. Then, a.s. the sequence of sampled subgraphs $(\mathbb{G}(k, W))_{k\in\mathbb{N}^*}$ converges to W for the cut distance $\delta_{\square,\mathcal{F}}$, and thus for any cut distance $\delta_{\square,\mathrm{m}}$ from Theorem 5.5.*

7. The Counting Lemmas and the topology of probability-graphons

In this section, we introduce the homomorphism densities for probability-graphons, and then we link those to the cut distance $\delta_{\square,\mathcal{F}}$ through the Counting Lemma and the Inverse Counting Lemma. Those results are analogous to the case of real-valued graphons, see [46, Chapter 7] for the definition of homomorphism densities and [46, Chapter 10] for the Counting Lemma and Inverse Counting Lemma. The main differences with [46] are: the decoration of the edges of the graphs with functions from $C_b(\mathbf{Z})$; the Counting Lemma for the decorations belonging only in the convergence determining sequence $\mathcal{F}$; the more technical proof of the Inverse Counting Lemma. Note that we need to work with $\delta_{\square,\mathcal{F}}$ here as the proof of the Inverse Counting Lemma relies on the second sampling Lemma 6.12.

Finally, we prove our main result, Theorem 7.11, which states that the topology on the space of probability-graphons, see Theorem 5.5, which can be induced by several choices of cut distances (e.g. $\delta_{\square,\mathrm{LP}}$, $\delta_{\square,\mathrm{KR}}$, $\delta_{\square,\mathrm{FM}}$ and $\delta_{\square,\mathcal{F}}$) coincides with the topology of convergence for all sampled subgraphs.

7.1. The homomorphism densities

In the case of non-weighted graphs, the homomorphism densities $t(F, G)$ allow to characterize a graph (up to twin-vertices expansion), and also allow to define a topology for real-valued graphons. In the case of weighted graphs and probability-graphons, we need to replace the absence/presence of edges (which is 0-1 valued) by test functions from $C_b(\mathbf{Z})$ decorating each edge.

In this section, we often need to fix the underlying (directed) graph structure $F = (V, E)$ (which may be incomplete) of a $C_b(\mathbf{Z})$-graph and to vary only the $C_b(\mathbf{Z})$-decorating functions $g = (g_e)_{e\in E}$, thus we will write $F^g = (V, E, g)$ for a $C_b(\mathbf{Z})$-graph. Moreover, when there exists a convergence determining sequence $\mathcal{F}$ such that $g_e \in \mathcal{F}$ for every edge $e \in E$, we say that F^g is a $\mathcal{F}$-graph and use the same notation conventions.

Definition 7.1 (Homomorphism density). — *We define the* homomorphism density *of a $C_b(\mathbf{Z})$-graph F^g in a signed measure-valued kernel $W \in \mathcal{W}_\pm$ as:*

$$t(F^g, W) = M_W^F(g) = \int_{[0,1]^{V(F)}} \prod_{(i,j)\in E(F)} W(x_i, x_j; g_{i,j}) \prod_{i\in V(F)} \mathrm{d}x_i. \tag{7.1}$$

Moreover, M_W^F defines a measure on $\mathbf{Z}^E$ (which we still denote by M_W^F) which is characterized by $M_W^F(\otimes_{e\in E} g_e) = M_W^F(g)$ for $g = (g_e)_{e\in E}$.

Remark 7.2 (Invariance under relabeling of homomorphism densities). Let $\varphi : [0,1] \to [0,1]$ be a measure-preserving map. As $\varphi^{\otimes k} : (x_1, \ldots, x_k) \mapsto (\varphi(x_1), \ldots, \varphi(x_k))$ is a measure-preserving map on $[0,1]^k$, applying the transfer formula (see (2.1)), we get that for every $C_b(\mathbf{Z})$-graph F^g and every signed measure-valued kernel $W \in \mathcal{W}_\pm$, we have $t(F^g, W^\varphi) = t(F^g, W)$. Thus $t(F^g, \cdot)$ can be extended to $\widetilde{\mathcal{W}}_\pm$.

Remark 7.3. — When $W \in \mathcal{W}_+$ is a measure-valued kernel, and F is the graph with two vertices and one edge, we get that $M_W^F = M_W$ the measure defined in (4.1).

Remark 7.4 (Adding missing edges to F). — When we work with probability-graphons, we can always assume the graph F to be complete, by adding the missing edges (i, j) and decorating them with the constant function $g_{(i,j)} = \mathbb{1}$.

For a finite weighted graph G, we define the *homomorphism density* of the $C_b(\mathbf{Z})$-graph F^g in G as $t(F^g, G) = t(F^g, W_G)$ (recall from Remark 6.2 the definition of W_G), that is:

$$t(F^g, G) = \frac{1}{v(G)^k} \sum_{(x_1,\cdots,x_k)\in V(G)^k} \prod_{(i,j)\in E(F)} g_{(i,j)}(\Phi_G(x_i, x_j)),$$

where $k = v(F)$ and $\Phi_G(x_i, x_j)$ is the weight of the directed edge from x_i to x_j.

7.2. The Counting Lemma

The following lemma links the homomorphism densities with the cut distance $\delta_{\square,\mathcal{F}}$ for some convergence determining sequence $\mathcal{F} = (f_n)_{n\in\mathbb{N}}$ (with $f_0 = \mathbb{1}$ and f_n takes values in $[0,1]$). This lemma is a generalization to probability-graphons of the Counting Lemma for real-valued graphons (see Lemmas 10.22 and 10.23 from [46]). Recall that by Remark 7.2, $t(F^g, \cdot)$ is defined on $\widetilde{\mathcal{W}}_\pm$.

Lemma 7.5 (Counting Lemma). — *Let $\mathcal{F} = (f_n)_{n\in\mathbb{N}}$ be a convergence determining sequence (with $f_0 = \mathbb{1}$ and f_n taking values in $[0,1]$). Let F^g be a $\mathcal{F}$-graph, and for every edge $e \in E(F)$, let $n_e \in \mathbb{N}$ be such that $g_e = f_{n_e}$. Then, for every probability-graphons $W, W' \in \widetilde{\mathcal{W}}_1$, we have:*

$$|t(F^g, W) - t(F^g, W')| \leqslant \left(\sum_{e\in E(F)} 2^{n_e} \right) \delta_{\square,\mathcal{F}}(W, W').$$

Remark 7.6 ($W \mapsto t(F^g, W)$ is Lipschitz). — The Lipschitz constant given by the lemma is too large to be useful in practical cases. Nevertheless, the homomorphism density function $W \mapsto t(F^g, W)$ is Lipschitz on the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$ equipped with the cut distance $\delta_{\square,\mathcal{F}}$.

Proof of Lemma 7.5. — To do this proof, we will apply Lemma 10.24 from [46], which applies to graphs F whose edges are decorated with (possibly different) real-valued graphons $w = (w_e : e \in E(F))$, and the associated homomorphism density is defined as

$$t(F, w) = \int_{[0,1]^{V(F)}} \prod_{(i,j)\in E(F)} w_e(x_i, x_j) \prod_{i\in V(F)} \mathrm{d}x_i. \tag{7.2}$$

Recall from (3.7) that for a probability-graphon $W \in \mathcal{W}_1$ and a function $f \in \mathcal{F}$ (which is $[0,1]$-valued by our definition of convergence determining sequences), we have that $W[f]$ is a real-valued graphon. Define the collections of real-valued graphons $w = (W[g_e] : e \in E(F))$ and $w' = (W'[g_e] : e \in E(F))$. Notice from (7.1) and (7.2) that we have $t(F, w) = t(F^g, W)$ and $t(F, w') = t(F^g, W')$. Applying [46, Lemma 10.24] to the graph F and edge-decorations w and w', we get:

$$|t(F^g, W) - t(F^g, W')| = |t(F, w) - t(F, w')| \leqslant \sum_{e \in E(F)} \|W[g_e] - W'[g_e]\|_{\square,\mathbb{R}},$$

where the norm $\|\cdot\|_{\square,\mathbb{R}}$ in the upper bound is the cut norm for real-valued graphons (see (3.11) for definition of this object). For $e \in E(F)$, by definition of the cut distance $d_{\square,\mathcal{F}}$ and using (3.10), we have:

$$\|W[g_e] - W'[g_e]\|_{\square,\mathbb{R}} \leqslant 2^{n_e}\, d_{\square,\mathcal{F}}(W, W').$$

Hence, combining all those upper bounds, we get the bound in the lemma but with $d_{\square,\mathcal{F}}$ instead of $\delta_{\square,\mathcal{F}}$. Since $t(F^g, \cdot)$ is invariant under relabeling by Remark 7.2, taking the infimum other all relabelings allows to replace $d_{\square,\mathcal{F}}$ by $\delta_{\square,\mathcal{F}}$ and to get the bound in the lemma. □

We have just seen that homomorphism densities defined using only functions from $\mathcal{F}$ are Lipschitz. We are going to see that the other homomorphism densities are nevertheless continuous.

Lemma 7.7 (Weak Counting Lemma). — *Let $\mathcal{F}$ be a convergence determining sequence (with $f_0 = \mathbb{1}$). Let $(W_n)_{n\in\mathbb{N}}$ and W be probability-graphons such that $\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$ for all $\mathcal{F}$-graphs F^g (which in particular the case if $\lim_{n\to\infty} \delta_{\square,\mathcal{F}}(W_n, W) = 0$ by the Counting Lemma 7.5). Then, for every $C_b(\mathbf{Z})$-graph F^g we have:*

$$t(F^g, W_n) \underset{n\to\infty}{\longrightarrow} t(F^g, W).$$

Proof. — Let $F = (V, E)$ be some fixed (directed) graph. By assumption, we have for all edge-decorations $g = (g_e)_{e\in E}$ in $\mathcal{F}$ that $\lim_{n\to\infty} M^F_{W_n}(\otimes_{e\in E} g_e) = M^F_W(\otimes_{e\in E} g_e)$ (see Definition 7.1). By [24, Chapter 3, Proposition 4.6], $\mathcal{F}^{\otimes E}$ is a (countable) convergence determining family on $\mathcal{M}_+(\mathbf{Z}^E)$. Thus, the sequence of measures $(M^F_{W_n})_{n\in\mathbb{N}}$ converges to M^F_W for the weak topology on $\mathcal{M}_+(\mathbf{Z}^E)$. And in particular, for every edge-decoration function $g = (g_e)_{e\in E}$ (here for every $e \in E$, $g_e \in C_b(\mathbf{Z})$ is arbitrary) we have $M^F_{W_n}(\otimes_{e\in E} g_e) = t(F^g, W_n) \to t(F^g, W) = M^F_W(\otimes_{e\in E} g_e)$ as $n \to \infty$. This being true for all choices of the graph F, it concludes the proof. □

7.3. The Inverse Counting Lemma

The goal of this subsection is to establish a converse to the Counting Lemma: if two probability-graphons are close in terms of homomorphism densities, then they are close w.r.t. the cut distance $\delta_{\square,\mathcal{F}}$.

Lemma 7.8 (Inverse Counting Lemma). — *Let $\mathcal{F} = (f_n)_{n\in\mathbb{N}}$ be a convergence determining sequence (with $f_0 = \mathbb{1}$ and f_n takes values in $[0,1]$). Let $U, W \in \widetilde{\mathcal{W}}_1$ be two probability-graphons, and let $k, n_0 \in \mathbb{N}^*$. Assume that we have $|t(F^g, U) - t(F^g, W)| \leqslant 2^{-k-n_0k^2}$ for every (complete) $C_b(\mathbf{Z})$-graph F^g with k vertices and such that the edge-decoration functions $g = (g_e)_{e\in E(F)}$ are products (without repetition) of the functions $(f_n)_{1\leqslant n\leqslant n_0}$ and $(1 - f_n)_{1\leqslant n\leqslant n_0}$. Then, we have:*

$$\delta_{\Box,\mathcal{F}}(U, W) \leqslant \frac{44}{\sqrt{\log(k)}} + 2^{-n_0}.$$

To prove Lemma 7.8, we first need to prove the special case where the space $\mathbf{Z}$ is finite.

Lemma 7.9 (Inverse Counting Lemma, case with finite space $\mathbf{Z}$). — *Assume that the space $\mathbf{Z}$ is finite with cardinality n_1, for simplicity say $\mathbf{Z} = [n_1]$. Define the indicator functions $f_n : z \mapsto \mathbb{1}_{\{z=n\}}$ for $n \in [n_1]$, in particular $\mathcal{H} = (f_n)_{1\leqslant n\leqslant n_1}$ is a finite convergence determining sequence. Let $U, W \in \widetilde{\mathcal{W}}_1$ be two probability-graphons, and let $k \in \mathbb{N}^*$. Assume that we have $|t(F^g, U) - t(F^g, W)| < 2^{-k-\log_2(n_1)k^2}$ for every (complete) $\mathcal{H}$-graph F^g with k vertices.*

Then, for any (possibly finite) convergence determining sequence $\mathcal{F}$, we have:

$$\delta_{\Box,\mathcal{F}}(U, W) \leqslant \frac{44}{\sqrt{\log(k)}}.$$

Abusing notations, we can identify a weight-value $n \in \mathbf{Z}$ with its indicator function f_n, and doing this identification for edge-decoration functions, we can identify a $\mathcal{F}$-graph F^g with its corresponding weighted graph. In particular, doing so we get $t(F^g, W) = \mathbb{P}(\mathbb{G}(k, W) = F^g)$ for every $\mathcal{F}$-graph F^g with k vertices. The proof of Lemma 7.9 is then a straightforward adaptation of the proof of [46, Lemma 10.31 and Lemma 10.32].

Proof of Lemma 7.8. — As the functions $(f_n)_{n\in\mathbb{N}}$ take value in $[0,1]$, for all φ measure-preserving map, for all $S, T \subset [0,1]$ measurable sets and for all $n \in \mathbb{N}$, we have:

$$\Big| U(S\times T; f_n) - W^{\varphi}(S\times T; f_n) \Big| \leqslant 1.$$

Using this bound, we get the following bound (recall that $f_0 = \mathbb{1}$):

(7.3)

$$\delta_{\Box,\mathcal{F}}(U, W) \leqslant \inf_{\varphi\in S_{[0,1]}} \sup_{S,T\subset[0,1]} \sum_{n=1}^{n_0} 2^{-n} \Big| U(S\times T; f_n) - W^{\varphi}(S\times T; f_n) \Big| + 2^{-n_0}.$$

Hence, for a point $z \in \mathbf{Z}$, the upper bound in (7.3) uses only the information given by $(f_n(z))_{n\in[n_0]}$. In order to discretize the space $[0,1]^{n_0}$, we replace a point $p = (p_1, \ldots, p_{n_0}) \in [0,1]^{n_0}$ by a random point $(Y_1, \ldots, Y_{n_0}) \in \{0,1\}^{n_0}$ where the Y_i are independent random variables with Bernoulli distribution of parameter p_i. This leads us to replace a $\mathcal{M}_1(\mathbf{Z})$-valued kernel W by the $\mathcal{M}_1(\{0,1\}^{n_0})$-valued kernel $\tilde{W}$ defined for all $(x,y) \in [0,1]^2$, and for all $s = (s_1, \ldots, s_{n_0}) \in \{0,1\}^{n_0}$ as:

$$\tilde{W}(x,y;\{s\}) = W(x,y;f^s) \qquad \text{where} \qquad f^s = \prod_{n=1}^{n_0} f_n^{s_n}(1-f_n)^{1-s_n}.$$

Fix some enumeration $(s^m)_{m\in[2^{n_0}]}$ of the points in $\{0,1\}^{n_0}$, and define the indicator functions $\tilde{h}_m : s \mapsto \mathbb{1}_{\{s=s^m\}}$ for $m \in [2^{n_0}]$, in particular $\tilde{\mathcal{H}} = (\tilde{h}_m)_{1\leqslant m\leqslant 2^{n_0}}$ is a finite convergence determining sequence on $\mathcal{M}_+(\{0,1\}^{n_0})$. Let $F^{\tilde{g}}$ be a $\tilde{\mathcal{H}}$-graph with vertex set $V(F) = [k]$, and for every edge $e \in E(F)$, let $m_e \in [2^{n_0}]$ be such that $\tilde{g}_e = \tilde{h}_{m_e}$. Define the edge-decoration functions $g = (g_e)_{e\in E(F)}$ for every edge $e \in E(F)$ as $g_e = f^{s^{m_e}}$, then we get:

$$t(F^{\tilde{g}}, \tilde{W}) = \int_{[0,1]^k} \prod_{(i,j)\in E(F)} \tilde{W}(x_i, x_j; \{s^{m_e}\}) \prod_{i=1}^{k} \mathrm{d}x_i = t(F^g, W).$$

Thus, the $\mathcal{M}_1(\{0,1\}^{n_0})$-valued graphons $\tilde{U}$ and $\tilde{W}$ inherit the bounds on the homomorphism densities: for every $\tilde{\mathcal{H}}$-graph $F^{\tilde{g}}$, we have $|t(F^{\tilde{g}}, \tilde{U}) - t(F^{\tilde{g}}, \tilde{W})| \leqslant 2^{-k-n_0k^2}$.

Define for all $n \in [n_0]$ the function $\tilde{f}_n : s \mapsto \mathbb{1}_{\{s_n=1\}}$, and let $\tilde{\mathcal{F}}$ be the concatenation of $(\tilde{f}_n)_{n\in[n_0]}$ and $\tilde{\mathcal{H}}$, in particular $\tilde{\mathcal{F}}$ is a finite convergence determining sequence on $\mathcal{M}_+(\{0,1\}^{n_0})$. Finally, as $\delta_{\square,\tilde{\mathcal{F}}}(\tilde{U}, \tilde{W})$ upper bounds the first term in the upper bound of (7.3), applying Lemma 7.9 with the finite space $\mathbf{Z} = \{0,1\}^{n_0}$ and $n_1 = 2^{n_0}$, the finite convergence determining sequences $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{H}}$, and the $\mathcal{M}_1(\{0,1\}^{n_0})$-valued graphons $\tilde{U}$ and $\tilde{W}$, we get:

$$\delta_{\square,\mathcal{F}}(U,W) \leqslant \frac{44}{\sqrt{\ln(k)}} + 2^{-n_0},$$

which concludes the proof. □

7.4. Subgraph sampling and the topology of probability-graphons

Thanks to the Weak Counting Lemma 7.7 and the Inverse Counting Lemma 7.8, we can formulate a new informative characterization of weak isomorphism, i.e. equality in the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$. Note that the propositions and the theorem in this subsection can in particular be applied to $\delta_{\square,\mathrm{m}}$ when d_{m} is a quasi-convex distance continuous w.r.t. the weak topology, as then $d_{\square,\mathrm{m}}$ is invariant, smooth, weakly regular and regular w.r.t. the stepping operator (see Proposition 4.13).

Proposition 7.10 (Characterization of equality for $\delta_{\square,\mathrm{m}}$). — *Let $U, W \in \mathcal{W}_1$ be two probability-graphons. The following properties are equivalent:*

(i) $\delta_{\square,\mathrm{m}}(U,W) = 0$ *for some (and hence for every) choice of the distance d_{m} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is (invariant) smooth.*
(ii) *There exist $\varphi, \psi \in \bar{S}_{[0,1]}$ such that $U^\varphi = W^\psi$ almost everywhere on $[0,1]^2$.*
(iii) $t(F^g, U) = t(F^g, W)$ *for all $C_b(\mathbf{Z})$-graphs F^g.*
(iv) $t(F^g, U) = t(F^g, W)$ *for all $\mathcal{F}$-graphs F^g.*

Proof. — The equivalence between Properties (i) and (ii) is a consequence of Proposition 3.18 on the cut distance. Remark 7.2 gives that Property (ii) implies Property (iii). It is clear that Property (iii) implies Property (iv). The Inverse Counting Lemma 7.8 with the Weak Counting Lemma 7.7 give that Property (iv) implies Property (i) (with $d_{\mathrm{m}} = d_{\mathcal{F}}$). Hence, we have the desired equivalence. □

Thanks to the Weak Counting Lemma 7.7 and the Inverse Counting Lemma 7.8, we get the following characterization of the topology induced by the cut distance $\delta_{\square,\mathrm{m}}$ on the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$ in terms of homomorphism densities

Theorem 7.11 (Characterization of the topology induced by $\delta_{\square,\mathrm{m}}$). — *Let $(W_n)_{n\in\mathbb{N}}$ and W be unlabeled probability-graphons from $\widetilde{\mathcal{W}}_1$. The following properties are equivalent:*

(i) $\lim_{n\to\infty} \delta_{\square,\mathrm{m}}(W_n, W) = 0$ *for some (and hence for every) choice of the distance d_{m} on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ and the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is (invariant) smooth, weakly regular and regular w.r.t. the stepping operator.*
(ii) $\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$ *for all $C_b(\mathbf{Z})$-graphs F^g.*
(iii) $\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$ *for all $\mathcal{F}$-graphs F^g.*
(iv) *For all $k \geqslant 2$, the sequence of sampled subgraphs $(\mathbb{G}(k, W_n))_{n\in\mathbb{N}}$ converges in distribution to $\mathbb{G}(k, W)$.*

In particular, the topology induced by the cut distance $\delta_{\square,\mathcal{F}}$ on the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$ coincides with the topology generated by the homomorphism densities functions $W \mapsto t(F^g, W)$ for all $C_b(\mathbf{Z})$-graphs F^g.

Proof. — By Theorem 5.5, convergence for $\delta_{\square,\mathcal{F}}$ is equivalent to convergence for $\delta_{\square,\mathrm{m}}$ for every choice of the distance d_m on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that d_m induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ and the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is (invariant) smooth, weakly regular and regular w.r.t. the stepping operator. Taking $d_\mathrm{m} = d_\mathrm{LP}$, the Weak Counting Lemma 7.7 gives that Property (i) implies Property (ii). It is clear that Property (ii) implies Property (iii). The Inverse Counting Lemma 7.8 with the Weak Counting Lemma 7.7 give that Property (iii) implies Property (i) (with $d_\mathrm{m} = d_\mathcal{F}$). Notice that when F is the complete graph with k vertices, M_W^F is the joint measure of all the edge-weights of the random graph $\mathbb{G}(k, W)$, and thus characterizes the distribution random graph $\mathbb{G}(k, W)$. Thus (recall Definition 7.1), Property (ii) and Property (iv) are equivalent. Hence, we have the desired equivalence. □

Question 7.12 (Do the distances $d_{\square,\mathcal{F}}$ all induce the same topology?). *Even though every distance $\delta_{\square,\mathcal{F}}$ generates the same topology on the space of unlabeled probability-graphons $\widetilde{\mathcal{W}}_1$, it is an open question whether or not it is also the case that every distance $d_{\square,\mathcal{F}}$ induces the same topology on the space of labeled probability-graphons $\mathcal{W}_1$.*

The following proposition states that to prove existence of a limit unlabeled probability-graphon it is enough to prove that there exists a convergence determining sequence $\mathcal{F}$ such that for every $\mathcal{F}$-graph F^g the homomorphism densities $t(F^g, \cdot)$ converge.

Proposition 7.13 (Existence of a limit unlabeled probability-graphon). *Let d_m be a distance on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ such that d_m induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ and the cut distance $d_{\square,\mathrm{m}}$ on $\mathcal{W}_1$ is (invariant) smooth, weakly regular and regular w.r.t. the stepping operator.*

Let $(W_n)_{n\in\mathbb{N}}$ be sequence of unlabeled probability-graphons in $\widetilde{\mathcal{W}}_1$ that is tight. Let $\mathcal{F}$ be a convergence determining sequence such that for every $\mathcal{F}$-graph F^g the sequence $(t(F^g, W_n))_{n\in\mathbb{N}}$ converges. Then, there exists an unlabeled probability-graphon $W \in \widetilde{\mathcal{W}}_1$ such that the sequence $(W_n)_{n\in\mathbb{N}}$ converges to W for $\delta_{\square,\mathrm{m}}$.

Proof. — Since the sequence $(W_n)_{n\in\mathbb{N}}$ is tight, by Theorem 5.1, there exists a subsequence $(W_{n_k})_{k\in\mathbb{N}}$ of the sequence $(W_n)_{n\in\mathbb{N}}$ that converges to some W for $\delta_{\square,\mathrm{m}}$. By Theorem 7.11, we have for every $\mathcal{F}$-graph F^g that $\lim_{k\to\infty} t(F^g, W_{n_k}) = t(F^g, W)$; and as we already know that the sequence $(t(F^g, W_n))_{n\in\mathbb{N}}$ converges, we have that $\lim_{n\to\infty} t(F^g, W_n) = t(F^g, W)$. Hence, by Theorem 7.11, we get that the sequence $(W_n)_{n\in\mathbb{N}}$ converges to W for $\delta_{\square,\mathrm{m}}$. $\square$

Remark 7.14. — For the special case $\mathbf{Z} = \{0, 1\}$, which is compact, we find back that convergence for real-valued graphons is characterized by the convergence of the homomorphism densities. Notice the tightness condition of Proposition 7.13 is automatically satisfied as $\mathbf{Z}$ is compact.

8. Proofs of Theorem 5.1 and Theorem 5.5

We start by proving a lemma that allows to construct a convergent subsequence and its limit kernel for a tight sequence of measure-valued kernels. This lemma is useful for the proofs of both Theorem 5.1 and Theorem 5.5. For the proof of Theorem 5.5, we will also need the convergence to hold simultaneously for two distances $\delta_\square$ and $\delta'_\square$. Recall from Definition 4.1 the definition of the stepfunction $W_\mathcal{P}$ for a signed measure-valued kernel W and a finite partition $\mathcal{P}$ of $[0, 1]$. For a finite partition $\mathcal{P}$ of $[0, 1]$, define its diameter as the smallest diameter of its sets, i.e. $\mathrm{diam}(\mathcal{P}) = \min_{S\in\mathcal{P}} \mathrm{diam}(S) = \min_{S\in\mathcal{P}} \sup_{x,y\in S} |x - y|$.

Lemma 8.1 (Convergence using given approximation partitions). — *Let d be an invariant smooth distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$). Let $(W_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) which is tight (resp. uniformly bounded and tight). Further assume that we are given, for every $n, k \in \mathbb{N}$, partitions $\mathcal{P}_{n,k}$ of $[0, 1]$, such that these partitions and the corresponding stepfunctions $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$ satisfy the following conditions:*

(i) the partition $\mathcal{P}_{n,k+1}$ is a refinement of $\mathcal{P}_{n,k}$,

(ii) $\mathrm{diam}(\mathcal{P}_{n,k}) \leqslant 2^{-k}$ and $|\mathcal{P}_{n,k}| = m_k$ depends only on k (and not on n),

(iii) $d(W_n, W_{n,k}) \leqslant 1/(k + 1)$.

Then, there exists a subsequence $(W_{n_\ell})_{\ell\in\mathbb{N}}$ of the sequence $(W_n)_{n\in\mathbb{N}}$ and a measure-valued kernel $W \in \mathcal{W}_1$ (resp. $W \in \mathcal{W}_+$ or $W \in \mathcal{W}_\pm$) such that $(W_{n_\ell})_{\ell\in\mathbb{N}}$ converges to W for $\delta_\square$.

Moreover, assume that d' is another invariant smooth distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) such that for every $n \in \mathbb{N}$ and $k \in \mathbb{N}$, $W_{n,k}$ also satisfies:

(iv) $d'(W_n, W_{n,k}) \leqslant 1/(k + 1)$.

Then, there exists a subsequence $(W_{n_\ell})_{\ell\in\mathbb{N}}$ of the sequence $(W_n)_{n\in\mathbb{N}}$ and a measure-valued kernel $W \in \mathcal{W}_1$ (resp. $W \in \mathcal{W}_+$ or $W \in \mathcal{W}_\pm$) such that $(W_{n_\ell})_{\ell\in\mathbb{N}}$ converges to the same measure-valued kernel W simultaneously both for $\delta_\Box$ and for $\delta'_\Box$, the cut distance associated with d'.

Proof. — We adapt here the general scheme from the proof of Theorem 9.23 in [46], but the argument for the convergence of the U_k, defined below, takes into account that measure-valued kernels are infinite-dimensional valued. We set (recall from (3.1) the definition of $\|\cdot\|_\infty$):

$$C = \sup_{n\in\mathbb{N}} \|W_n\|_\infty < +\infty.$$

The proof is divided into four steps.

Step 1: Without loss of generality, the partitions $\mathcal{P}_{n,k}$ are made of intervals. For every $n \in \mathbb{N}$, we can rearrange the points of $[0,1]$ by a measure-preserving map so that the partitions $\mathcal{P}_{n,k}$ are made of intervals, and we replace W_n by its rearranged version.

An argument similar to the next lemma is used in the proof in [46, Proof of Theorem 9.23] without any reference. So, we provide a proof and stress that diameters of the partitions shrinking to zero is an important assumption (see Remark 8.3 below). We say a measurable map φ from $[0,1]$ to $[0,1]$ is a.e. invertible if there exists a measurable function ψ from $[0,1]$ to $[0,1]$ such that $\varphi \circ \psi(x) = \psi \circ \varphi(x) = x$ for a.e. $x \in [0,1]$; and that it is bi-measurable if φ is measurable and for all Borel set $B \subset [0,1]$, $\varphi(B)$ is also a Borel set.

Lemma 8.2 (Kernel rearrangement with interval partitions). — *Let $(\mathcal{P}_k)_{k\in\mathbb{N}}$ be a refining sequence of finite partitions of $[0,1]$ whose diameter converges to zero. Then, there exist a measure-preserving map $\varphi \in \bar{S}_{[0,1]}$ which is bi-mesurable and a.e. invertible, and a refining sequence of partitions made of intervals $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$, and all set $S \in \mathcal{P}_k$ there exists a set $R \in \mathcal{Q}_k$ such that a.e. $\mathbb{1}_R = \mathbb{1}_{\varphi^{-1}(S)}$.*

In particular, if W is a signed measure-valued kernel, then for $U = W^\varphi$, we have that a.e. $U_{\mathcal{Q}_k} = (W^\varphi)_{\mathcal{Q}_k} = (W_{\mathcal{P}_k})^\varphi$ for all $k \in \mathbb{N}$.

Notice that, according to Remark 4.4, the sequence of refining partition $(\mathcal{P}_k)_{k\in\mathbb{N}}$, with a partition diameter converging to 0, separates points and thus generates the Borel σ-field of $[0,1]$.

Proof. — Consider the infinite Ulam–Harris tree $\mathcal{T}^\infty = \{u_1 \cdots u_k : k \in \mathbb{N}, u_1, \ldots, u_k \in \mathbb{N}^*\}$, where for $k = 0$ the empty word $u = \partial$ is called the root node of the tree; for a node $u = u_1 \cdots u_k \in \mathcal{T}^\infty$, we define its height as $h(u) = k$, and if $k > 0$ we define its parent node as $p(u) = u_1 \cdots u_{k-1}$

and we say that u is a child node of $p(u)$. We order vertices on the tree $\mathcal{T}^\infty$ with the lexicographical (total) order $<_{\text{lex}}$. As a first step, we construct a subtree $\mathcal{T} \subset \mathcal{T}^\infty$ that indexes the sets in the partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$, such that for every $k \in \mathbb{N}$, $\mathcal{P}_k = \{S_u : u \in \mathcal{T}, h(u) = k\}$, and such that if $S_v \subset S_u$ with $S_v \in \mathcal{P}_k$ and $S_u \in \mathcal{P}_{k-1}$, then $p(v) = u$.

Without loss of generality, we may assume that $\mathcal{P}_0 = \{[0,1]\}$, and we label its only set by the empty word ∂, and we set $S_\partial = [0,1]$. Then, suppose we have already labeled the sets from $\mathcal{P}_0, \ldots, \mathcal{P}_k$, and we proceed to label the sets from $\mathcal{P}_{k+1}$. Because the partition $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$, we can group the sets of $\mathcal{P}_{k+1}$ by their unique parent set from $\mathcal{P}_k$, i.e. for every $S_u \in \mathcal{P}_k$, let $\mathcal{O}_u = \{S \in \mathcal{P}_{k+1} : S \subset S_u\}$, then $S_u = \cup_{S\in\mathcal{O}_u} S$. For $S_u \in \mathcal{P}_k$, we fix an arbitrary enumeration of $\mathcal{O}_u = \{S^1, \ldots, S^\ell\}$ with $\ell = |\mathcal{O}_u|$, then label the set S^j by uj, and set $S^j = S_{uj}$; remark that the parent node of $w = uj$ is $p(w) = u$, and the height of node w is $h(w) = h(u) + 1 = k + 1$. Hence, we have labeled every set from $\mathcal{P}_{k+1}$. To finish the construction, we set $\mathcal{T} = \{u : \exists k \in \mathbb{N}, \exists S \in \mathcal{P}_k,\ S \text{ has label } u\}$.

We now proceed to construct a measure-preserving map ψ such that the image of every set S_u is a.e. equal to an interval, and such that those intervals are ordered w.r.t. to the order of their labels in $\mathcal{T}$.

Define the map $\sigma : [0,1] \to \mathcal{T}^\mathbb{N}$ by $\sigma(x) = (u^k(x))_{k\in\mathbb{N}} \in \mathcal{T}^\mathbb{N}$ where $u^k(x)$ is the only node of $\mathcal{T}$ with height k such that $x \in S_{u^k(x)}$ (and thus $u^{k+1}(x)$ is a child node of $u^k(x)$). Remark that if $u^{k_0}(x) <_{\text{lex}} u^{k_0}(y)$ for some $k_0 \in \mathbb{N}$, then $u^k(x) <_{\text{lex}} u^k(y)$ for every $k \geqslant k_0$. We extend naturally the total order $<_{\text{lex}}$ from $\mathcal{T}$ to a the total order on $\mathcal{T}^\mathbb{N}$: for $(u^k)_{k\in\mathbb{N}}, (v^k)_{k\in\mathbb{N}} \in \mathcal{T}^\mathbb{N}$, $(u^k)_{k\in\mathbb{N}} <_{\text{lex}} (v^k)_{k\in\mathbb{N}}$ if $u^{k_0} <_{\text{lex}} v^{k_0}$ where k_0 is the smallest k such that $u^k \neq v^k$.

For every $u \in \mathcal{T}$, define:

$$A^-(u) = \bigcup_{v <_{\text{lex}} u \,:\, h(v)=h(u)} S_v \quad \text{and} \quad A^+(u) = A^-(u) \cup S_u,$$

and then define $C^-(u) = \lambda(A^-(u))$ and $C^+(u) = \lambda(A^+(u))$. Now, define ψ as, for $x \in [0,1]$:

$$\psi(x) = \lambda(A^-(x)) \quad \text{where} \quad A^-(x) = \{y \in [0,1] : \sigma(y) <_{\text{lex}} \sigma(x)\} = \cup_{k\in\mathbb{N}} A^-(u^k(x)).$$

Moreover, as the sequence of partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ has a diameter that converges to zero, and thus separates points, the map σ is injective. Thus, we also have:

$$\psi(x) = \lambda(A^+(x)) \quad \text{where} \quad A^+(x) = \{y \in [0,1] : \sigma(y) \leqslant_{\text{lex}} \sigma(x)\} = A^-(x) \cup \{x\}.$$

Remark that both $A^-(x)$ and $A^+(x)$ are Borel measurable.

Remark that for every $k \in \mathbb{N}$, we have $A^-(u^k(x)) \subset A^-(x) \subset A^+(x) \subset A^+(u^k(x))$. In particular, for every $u \in \mathcal{T}$, we have $\psi(S_u) \subset [C^-(u), C^+(u)]$; however $\psi(S_u)$ is not necessarily an interval, but we shall see that $\lambda(\psi(S_u)) = C^+(u) - C^-(u)$, i.e. $\psi(S_u)$ is a.e. equal to $[C^-(u), C^+(u)]$. Remark that, as the sequence of partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ is refining, we get that $[C^-(u), C^+(u)] = \cup_{v\,:\,p(v)=u}[C^-(v), C^+(v)]$ for every $u \in \mathcal{T} \setminus \{\partial\}$.

As the diameter of the partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ converges to zero, we have the following alternative formula for ψ:

$$\psi(x) = \lim_{k\to\infty} C^-(u^k(x)) = \lim_{k\to\infty} C^+(u^k(x)).$$

For every $k \in \mathbb{N}$, the map $x \mapsto C^-(u^k(x))$ is a simple function (constant on each $S \in \mathcal{P}_k$ and takes finitely-many values), and thus ψ is measurable as a limit of measurable maps.

We outline the rest of the proof. We first prove that ψ is measure-preserving. Secondly, we prove that ψ is a.e. invertible and construct its a.e. inverse map φ. Thirdly, we prove that $(\varphi^{-1}(\mathcal{P}_k))_{k\in\mathbb{N}}$ is a refining sequence of partitions. And lastly, we approximate almost everywhere the sequence of partitions $(\varphi^{-1}(\mathcal{P}_k))_{k\in\mathbb{N}}$ by a sequence of refining partitions composed of intervals.

We now prove that ψ is measure preserving. Remark that $\psi(x)$ is a non-decreasing function of $\sigma(x)$ for the total relation order $<_{\text{lex}}$, i.e. $\psi(y) \leqslant \psi(x)$ if and only if $\sigma(y) \leqslant_{\text{lex}} \sigma(x)$. Hence, $\psi^{-1}([0, \psi(x)]) = \{y \in [0,1] : \sigma(y) \leqslant_{\text{lex}} \sigma(x)\}$, and we have:

$$\lambda(\psi^{-1}([0, \psi(x)])) = \lambda(\{y \in [0,1] : \sigma(y) \leqslant_{\text{lex}} \sigma(x)\}) = \psi(x).$$

Thus, to show that ψ is measure preserving we just need to show that $\psi([0,1])$ is dense in $[0,1]$. For every $u \in \mathcal{T}$, as $\psi(S_u) \subset [C^-(u), C^+(u)]$, we know that the interval $[C^-(u), C^+(u)]$ contains at least one point of the form $\psi(x)$. Remark that for all $k \in \mathbb{N}$, we have:

$$[0,1] = \cup_{u\in\mathcal{T}:h(u)=k}[C^-(u), C^+(u)].$$

Hence, as $\lambda([C^-(u), C^+(u)]) = \lambda(S_u) \leqslant \operatorname{diam}(\mathcal{P}_{h(u)})$ for every $u \in \mathcal{T}$, and as the diameter of the partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ converges to zero, we know that each interval of positive length contains a point of the form $\psi(x)$ for some $x \in [0,1]$, which implies that $\psi([0,1])$ is indeed dense in $[0,1]$.

We now prove that ψ is a.e. invertible and construct its a.e. inverse map φ. Without loss of generality, assume that there is no set S_u with null measure. Consider two distinct elements $x, y \in [0,1]$ such that $\sigma(x) <_{\text{lex}} \sigma(y)$. Assume that $\psi(x) = \psi(y)$, and let $N \in \mathbb{N}$ be the last index k such that $u^k(x) = u^k(y)$. Then, for every $k > N$, we have $u^k(x) <_{\text{lex}} u^k(y)$, which implies that $\psi(x) \leqslant C^+(u^k(x)) \leqslant C^-(u^k(y)) \leqslant \psi(y)$; and thus $\psi(x) = \psi(y) = C^+(u^k(x)) = C^-(u^k(y))$, which in turn implies that there is no node of $\mathcal{T}$ between $u^k(x)$ and $u^k(y)$. Remark that this situation is analogous to the terminating decimal versus repeating decimal situation. Hence, we proved that there is no node between $u^{N+1}(x)$ and $u^{N+1}(y)$ and that for every $k > N$, $u^{k+1}(x)$ is the right-most child of $u^k(x)$, and $u^{k+1}(y)$ is the left-most child of $u^k(y)$ (i.e. $u^{k+1}(x) = u^k(x)|\mathcal{O}_{u^k(x)}|$ and $u^{k+1}(y) = u^k(y)1$). Recall that the map σ is injective. Putting all of this together, we get that the set $\{(x,y) \in [0,1] : \psi(x) = \psi(y), x < y\}$ can be indexed by the nodes of $\mathcal{T}$, and is thus at most countable. The set $D = \{x \in [0,1] :\ \psi^{-1}(\{\psi(x)\}) \quad \text{is a singleton}\}$ is a Borel set as $[0,1] \setminus D$ is at most countable. Hence, the map ψ is injective on the subset $D \subset [0,1]$ with measure one, and as ψ is measure preserving, we get that $\psi(D)$ has measure one, and thus ψ is bijective from D to $\psi(D)$, that is, ψ is a.e. invertible. We construct the map φ as the inverse map of ψ for $x \in \psi(D)$ and $\varphi(x) = 0$ for $x \in [0,1] \setminus \psi(D)$. Without loss of generality, we assume that $0 \notin D$. Thus, φ is the a.e. inverse map of ψ.

We are left to prove that φ is bi-measurable and measure preserving. As we saw that each point $z \in [0,1]$ has a pre-image $\psi^{-1}(z) = \{x \in [0,1] : \psi(x) = z\}$ at most countable (indeed of cardinal at most 2), thus [51] insures that ψ is bi-measurable. Let $B \subset (0,1]$ be a Borel set. We have that $\varphi^{-1}(B) = \varphi^{-1}(B \cap D) = \psi(B \cap D)$ is a Borel set, where the first equality uses that $\varphi([0,1]) = D \cup \{0\}$, the second equality uses that ψ is the inverse of φ on D, and lastly we used that ψ is bi-measurable. We also have that $\varphi^{-1}(B \cup \{0\}) = \varphi^{-1}(B) \cup ([0,1] \setminus \psi(D))$ is a Borel set. Thus, we deduce that φ is measurable. We now prove that φ is bi-measurable. Let $B \subset [0,1]$ be a Borel set. We have that $\varphi(B)$ is equal to $\psi^{-1}(B)$ if $B \subset \psi(D)$ and $\psi^{-1}(B \cap \psi(D)) \cup \{0\}$ otherwise. In both cases, this is a Borel set as $\psi(D)$ is a Borel set since ψ is bi-measurable and D is a Borel set.

Moreover, we have:

$$\lambda(\varphi^{-1}(B)) = \lambda(\psi(B \cap D)) = \lambda(\psi^{-1}(\psi(B \cap D))) = \lambda(B \cap D) = \lambda(B),$$

where we used that $\varphi^{-1}(B) = \psi(B \cap D)$ for the first equality, that ψ is measure preserving for the second equality, that ψ is bijective from D to $\psi(D)$ for the third equality, and that D has measure one for the last equality. We also have:

$$\lambda(\varphi^{-1}(B \cup \{0\})) = \lambda(\varphi^{-1}(B)) + \lambda([0,1] \setminus \psi(D)) = \lambda(B) = \lambda(B \cup \{0\}),$$

where we used that $\varphi^{-1}(B) \subset \psi(D)$ and $[0,1] \setminus \psi(D)$ are disjoint sets for the first equality, that $\lambda(\varphi^{-1}(B)) = \lambda(B)$ and that $\psi(D)$ has measure one for the second equality. Hence, the map φ is measurable and measure preserving.

We now prove that $(\varphi^{-1}(\mathcal{P}_k))_{k\in\mathbb{N}}$ is a refining sequence of partitions. For $k \in \mathbb{N}$, as $\mathcal{P}_k$ is a finite partition of $[0,1]$, we have that $\varphi^{-1}(\mathcal{P}_k) = \{\varphi^{-1}(S_u) : u \in \mathcal{T}, h(u) = k\}$ is also a finite partition of $[0,1]$. Moreover, as $(\mathcal{P}_k)_{k\in\mathbb{N}}$ is a refining sequence of partitions, we get that the sequence of partitions $(\varphi^{-1}(\mathcal{P}_k))_{k\in\mathbb{N}}$ is also refining. Remark that the sets $\varphi^{-1}(S_u)$ are not necessarily intervals, they are intervals minus some at most countable sets (this is similar to the unit line minus the Cantor set).

To finish the proof, we are left to construct a refining sequence of partitions made of intervals $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ that agrees almost everywhere with the refining sequence of partitions $(\varphi^{-1}(\mathcal{P}_k))_{k\in\mathbb{N}}$. For $u \in \mathcal{T}$, define $R_u = [C^-(u), C^+(u))$ (and $R_u = [C^-(u), C^+(u)]$ if u is the unique node such that $v \leqslant_{\text{lex}} u$ for every $v \in \mathcal{T}$ with $h(v) = h(u)$). As ψ is measure preserving, and as $\psi(S_u) \subset [C^-(u), C^+(u)]$ with $\lambda(S_u) = C^+(u) - C^-(u)$, we get that $\lambda([C^-(u), C^+(u)] \setminus \psi(S_u)) = 0$. As φ is the a.e. inverse map of ψ, we have that a.e. $\mathbb{1}_{\varphi^{-1}(S_u)} = \mathbb{1}_{\psi(S_u)} = \mathbb{1}_{[C^-(u),C^+(u)]} = \mathbb{1}_{R_u}$, i.e. R_u agrees almost everywhere with $\varphi^{-1}(S_u)$. For $k \in \mathbb{N}$, define the finite partition $\mathcal{Q}_k = \{R_u : h(u) = k\}$. Then, by definition of the sets R_u, the sequence of partitions $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ is refining. This concludes the proof. □

Remark 8.3 (The shrinking diameter assumption is important). — Even if it is not stressed in [46, Proof of Theorem 9.23], the measure preserving map φ in Lemma 8.2, which is bi-measurable and a.e. invertible, cannot be obtained without any assumption on the refining sequence of partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ (in our case, we assumed that their diameter converges to zero). Indeed consider the sequence of partitions where for every $k \in \mathbb{N}$, $\mathcal{P}_k$ is composed of the sets:

$$S_{k,j} = [j2^{-k-1}, (j+1)2^{-k-1}) \ \cup\ [1/2 + j2^{-k-1}, 1/2 + (j+1)2^{-k-1}), \qquad 0 \leqslant j < 2^k,$$

i.e. $S_{k,j}$ is the union of two dyadic interval translated by $1/2$, (also add 1 to the set $S_{k,0}$ to get a complete partition). Then, for every $x \in [0,1/2)$, x and $x+1/2$ belong to the same set of $\mathcal{P}_k$ for every $k \in \mathbb{N}$; in particular the diameter of the partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ does not converge to zero. By contradiction, assume there exist a measure preserving map $\varphi \in \bar{S}_{[0,1]}$ and a sequence of interval partitions $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$ and all set $S_{k,j} \in \mathcal{P}_k$ with $0 \leqslant j < 2^k$, there exists a interval set $I_{k,j} \in \mathcal{Q}_k$ such that a.e. $\mathbb{1}_{I_{k,j}} = \mathbb{1}_{\varphi^{-1}(S_{k,j})}$. In particular, the set $I_{k,j}$ must be an interval of length 2^{-k}. Hence, $\mathcal{Q}_k$ is a dyadic partition with stepsize 2^{-k}, and thus the diameter of the partitions $(\mathcal{Q}_k)_{k\in\mathbb{N}}$ converges to zero. For every $x \in [0,1/2)$, we get that $\operatorname{diam}(\varphi^{-1}(\{x,x+1/2\})) \leqslant \operatorname{diam}(\mathcal{Q}_k) = 2^{-k}$ for all $k \in \mathbb{N}$; this implies that $\varphi^{-1}(\{x,x+1/2\})$ is a singleton, i.e. either $x \notin \varphi([0,1])$ or $x+1/2 \notin \varphi([0,1])$. Hence, as φ is bi-measurable, considering the measurable set $\mathrm{A} = \varphi([0,1])$, we have $\lambda\big([0,1/2)\cap\mathrm{A}\big) = \lambda\big([1/2,1)\setminus\mathrm{A}\big)$ and $\lambda\big([0,1/2)\setminus\mathrm{A}\big) = \lambda\big([1/2,1)\cap\mathrm{A}\big)$. As $\lambda\big([0,1/2)\big) = \lambda\big([0,1/2)\cap\mathrm{A}\big) + \lambda\big([0,1/2)\setminus\mathrm{A}\big) = 1/2$ because φ is measure preserving, we get that $\lambda\big(\mathrm{A}\big) = \lambda\big([0,1/2)\cap\mathrm{A}\big) + \lambda\big([1/2,1)\cap\mathrm{A}\big) = 1/2$, which contradicts the fact that φ is measure preserving.

Now, for every $n \in \mathbb{N}$, applying Lemma 8.2 to $(\mathcal{P}_{n,k})_{k\in\mathbb{N}}$ and W_n, we get a measure-preserving map φ_n and a refining sequence of partitions $(\mathcal{P}'_{n,k})_{k\in\mathbb{N}}$ made of intervals such that for all $k \in \mathbb{N}$, and all set $R \in \mathcal{P}'_{n,k}$ there exists a set $S \in \mathcal{P}_{n,k}$ such that a.e. $\mathbb{1}_R = \mathbb{1}_{\varphi_n^{-1}(S)}$. In particular, for all $k \in \mathbb{N}$, the sequence of partitions $(\mathcal{P}_{n,k})_{k\in\mathbb{N}}$ still satisfy (i)–(ii). Set $W'_n = W_n^{\varphi_n}$ and $W'_{n,k} = W_{n,k}^{\varphi_n}$ so that almost everywhere:

$$W'_{n,k} = \big((W_n)_{\mathcal{P}_{n,k}}\big)^{\varphi_n} = (W_n^{\varphi_n})_{\mathcal{P}'_{n,k}} = (W')_{\mathcal{P}'_{n,k}}.$$

As d and d' are invariant, we have for every $n,k \in \mathbb{N}$ that $d(W_n, W_{n,k}) = d(W'_n, W'_{n,k})$, and similarly for d'. This insures that the signed measure-valued kernels $(W'_n)_{n\in\mathbb{N}}$ and $(W'_{n,k})_{n\in\mathbb{N}}$, $k \in \mathbb{N}$, still satisfy (iii)–(iv). Recall that for a measure-valued kernel W and a measure-preserving map φ, $\delta_{\square,\mathrm{m}}(W, W^\varphi) = 0$. Hence, we can replace the signed measure-valued kernels $(W_n)_{n\in\mathbb{N}}$ and $(W_{n,k})_{n\in\mathbb{N}}$, $k \in \mathbb{N}$, by $(W'_n)_{n\in\mathbb{N}}$ and $(W'_{n,k})_{n\in\mathbb{N}}$, $k \in \mathbb{N}$, and assume that the partitions $\mathcal{P}_{n,k}$ are made of intervals.

Step 2: There exists a subsequence $(W_{n_\ell})_{\ell\in\mathbb{N}}$ such that for every $k \in \mathbb{N}$ and $\epsilon \in \{+,-\}$, the subsequence $(W^\epsilon_{n_\ell,k})_{\ell\in\mathbb{N}}$ weakly converges, as $\ell \to \infty$, almost everywhere to a limit, say U^ϵ_k which is a step-function adapted to a partition with m_k elements (some elements might be empty sets).

Fix some $k \in \mathbb{N}$. The stepfunctions $(W_{n,k} = (W_n)_{\mathcal{P}_{n,k}})_{n\in\mathbb{N}}$ all have the same number of steps m_k. For $n \in \mathbb{N}$, denote by $\mathcal{P}_{n,k} = \{S_{n,k,i} : 1 \leqslant i \leqslant$

$m_k\}$ the interval partition adapted to $W_{n,k}$ where the intervals are order according to the natural order on $[0,1]$ (note that some intervals might be empty, simply put them at the end). For $n \in \mathbb{N}$ and $1 \leqslant i \leqslant m_k$, let $\lambda(S_{n,k,i})$ denote the length of the interval $S_{n,k,i} \in \mathcal{P}_{n,k}$. As the lengths of steps take values in the compact set $[0,1]$, there exists a subsequence of indices $(n_\ell)_{\ell\in\mathbb{N}}$ such that for every $1 \leqslant i \leqslant m_k$, there exists $s_{k,i} \in [0,1]$ such that $\lim_{\ell\to\infty} \lambda(S_{n_\ell,k,i}) = s_{k,i}$. Denote by $\mathcal{P}_k = \{S_{k,i} : 1 \leqslant i \leqslant m_k\}$ the interval partition composed of m_k intervals where the i-th interval $S_{k,i}$ has length $s_{k,i}$ (note that some intervals might be empty). Up to a diagonal extraction, we can assume that the convergence holds for every $k \in \mathbb{N}$ simultaneously. Remark that for all $n, k \in \mathbb{N}$, the fact that $\mathcal{P}_{n,k+1}$ is a refinement of $\mathcal{P}_{n,k}$ can be simply restated as linear relations on the interval lengths $(\lambda(S_{n,k,i}))_{1\leqslant i\leqslant m_k}$ and $(\lambda(S_{n,k+1,i}))_{1\leqslant i\leqslant m_{k+1}}$. As linear relations are preserved when taking the limit, we get that the partition $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$ for all $k \in \mathbb{N}$. We assume from now on that $(W_n)_{n\in\mathbb{N}}$ and $(W_{n,k})_{n\in\mathbb{N}}$, $k \in \mathbb{N}$, are the corresponding subsequences.

For every $n \in \mathbb{N}$, we decompose $W_n = W_n^+ - W_n^-$ into its positive and negative kernel parts, see Lemma 3.3. For $n, k \in \mathbb{N}$ and $\epsilon \in \{+,-\}$, we define $W^\epsilon_{n,k} = (W^\epsilon_n)_{\mathcal{P}_{n,k}}$. In particular, remark that $W_{n,k} = W^+_{n,k} - W^-_{n,k}$ and for all $\ell \geqslant k$, that $W^\epsilon_{n,k} = (W^\epsilon_{n,\ell})_{\mathcal{P}_{n,k}}$. Let $\epsilon \in \{+,-\}$ and $1 \leqslant i, j \leqslant m_k$ such that $s_{k,i}s_{k,j} > 0$ be fixed. For every $n \in \mathbb{N}$, we have on $S_{n,k,i} \times S_{n,k,j}$ that $W^\epsilon_{n,k} = \mu^{i,j,\epsilon}_{n,k} \in \mathcal{M}_+(\mathbf{Z})$ with:

$$\mu^{i,j,\epsilon}_{n,k}(\cdot) = \frac{1}{\lambda(S_{n,k,i})\lambda(S_{n,k,j})} W^\epsilon_n(S_{n,k,i} \times S_{n,k,j}; \cdot).$$

We have that:

$$\|\mu^{i,j,\epsilon}_{n,k}\|_\infty \leqslant \|W_n\|_\infty \leqslant C.$$

This gives that the sequence $(\mu^{i,j,\epsilon}_{n,k})_{n\in\mathbb{N}}$ in $\mathcal{M}_\pm(\mathbf{Z})$ is bounded. We now prove it is tight. Let $\eta > 0$. As $\lim_{n\to\infty} \lambda(S_{n,k,\ell}) = s_{k,\ell} > 0$ for $\ell = i, j$, we deduce that there exists $c > 0$ such that for every $n \in \mathbb{N}$ large enough and $\ell = i, j$, we have $\lambda(S_{n,k,\ell}) > c$. Set $\eta' = c^2\eta$. As the sequence $(W_n)_{n\in\mathbb{N}}$ in $\widetilde{\mathcal{W}}_\pm$ is tight, there exists a compact set $K \subset \mathbf{Z}$ such that for every $n \in \mathbb{N}$, $M_{W_n}(K^c) \leqslant \eta'$. Hence, for every $n \in \mathbb{N}$ large enough, we have:

$$\mu^{i,j,\epsilon}_{n,k}(K^c) \leqslant \frac{1}{\lambda(S_{n,k,i})\lambda(S_{n,k,j})} M_{W_n}(K^c) \leqslant \eta.$$

This gives that the sequence $(\mu^{i,j,\epsilon}_{n,k})_{n\in\mathbb{N}}$ in $\mathcal{M}_+(\mathbf{Z})$ is bounded and tight, and thus by Lemma 2.8, it has a convergent subsequence. By diagonal extraction, we can assume there is a subsequence $(W_{n_\ell})_{\ell\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$, all $1 \leqslant i, j \leqslant m_k$ such that $s_{k,i}s_{k,j} > 0$, and all $\epsilon \in \{+,-\}$, the subsequence $(\mu_{n_\ell,k}^{i,j,\epsilon})_{\ell\in\mathbb{N}}$ weakly converges to a limit say $\mu_k^{i,j,\epsilon}$. Define the stepfunction $U_k^\epsilon \in \mathcal{W}_+$ adapted to the partition $\mathcal{P}_k$ which is equal to $\mu_k^{i,j,\epsilon}$ on $S_{k,i} \times S_{k,j}$ (if $s_{k,i}s_{k,j} = 0$, set $\mu_k^{i,j,\epsilon} = 0$). We have in particular obtained that, for every $k \in \mathbb{N}$, the subsequence $(W_{n_\ell,k}^\epsilon)_{\ell\in\mathbb{N}}$ weakly converges a.e. to U_k^ϵ which is a stepfunction adapted to a partition with m_k elements; this implies that the subsequence $(W_{n_\ell,k})_{\ell\in\mathbb{N}}$ also weakly converges a.e. to $U_k = U_k^+ - U_k^-$. We now assume that $(W_n)_{n\in\mathbb{N}}$ is such a subsequence. With this convention, notice that for all $k, n \in \mathbb{N}$ and $\epsilon \in \{+, -\}$:

$$\|U_k^\epsilon\|_\infty \leqslant \sup_{n\in\mathbb{N}} \|W_{n,k}^\epsilon\|_\infty \leqslant \sup_{n\in\mathbb{N}} \|W_n\|_\infty = C < +\infty. \tag{8.1}$$

Step 3: There exists a subsequence of $(U_k)_{k\in\mathbb{N}}$ which weakly converges to a limit $U \in \mathcal{W}_\pm$ almost everywhere on $[0,1]^2$. The proof of this step is postponed to the end. Without loss of generality we still write $(U_k)_{k\in\mathbb{N}}$ for this subsequence.

Step 4: We have $\lim_{n\to\infty} \delta_\square(U, W_n) = \lim_{n\to\infty} \delta'_\square(U, W_n) = 0$. Let $\eta > 0$. As the cut distances d is smooth, we deduce from Step 3, that for k large enough $d(U, U_k) \leqslant \eta$. By hypothesis (iii) on the sequence $(W_{n,k})_{n\in\mathbb{N}}$, we also have that for k large enough $d(W_n, W_{n,k}) \leqslant \eta$. For such large k, as by step 2 the sequence $(W_{n,k})_{n\in\mathbb{N}}$ weakly converges almost everywhere to U_k, and again as the cut distances d is smooth, there is a n_0 such that for every $n \geqslant n_0$, $d(U_k, W_{n,k}) \leqslant \eta$. Then for all $n \geqslant n_0$, we have:

$$\begin{aligned}\delta_\square(U, W_n) &\leqslant \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n)\\ &\leqslant d(U, U_k) + d(U_k, W_{n,k}) + d(W_{n,k}, W_n)\\ &\leqslant 3\eta.\end{aligned}$$

This gives that $\lim_{n\to\infty} \delta_\square(W_n, U) = 0$.

If we consider a second distance d' as in the lemma, then similarly $\lim_{n\to\infty} \delta'_\square(W_n, U) = 0$.

Proof of Step 3. Assume that the claim is true for measure-valued kernels. Then, if $(U_k)_{k\in\mathbb{N}}$ is a sequence of signed-measure valued kernels, applying the claim to $(U_k^\epsilon)_{k\in\mathbb{N}}$, for $\epsilon \in \{+, -\}$, we get a measure-valued $U^\epsilon \in \mathcal{W}_+$ such that the sequence $(U_k^\epsilon)_{k\in\mathbb{N}}$ weakly conveges a.e. to U^ϵ. Thus, the sequence $(U_k)_{k\in\mathbb{N}}$ weakly conveges a.e. to $U = U^+ - U^-$.

Hence, we are left to prove the claim for measure-valued kernels. The proof is divided in four steps. The first three steps also work for signed-measure valued kernels, but the last argument of step 3.d. only works for measures.

Step 3.a: The sequence $(U_k)_{k\in\mathbb{N}}$ inherit the tightness property from the sequence $(W_n)_{n\in\mathbb{N}}$. Let $\eta > 0$. Since the sequence $(W_n)_{n\in\mathbb{N}}$ is tight, there exists a compact set $K \subset \mathbf{Z}$ such that for every $n \in \mathbb{N}$, we have $M_{W_n}(K^c) \leqslant \eta$. Remark that:

$$M_{W_{n,k}} = \sum_{1\leqslant i,j\leqslant m_k} \lambda(S_{n,k,i})\lambda(S_{n,k,j})\mu_{n,k}^{i,j} = M_{W_n}$$
$$\text{and} \quad M_{U_k} = \sum_{1\leqslant i,j\leqslant m_k} s_{k,i}s_{k,j}\mu_k^{i,j}.$$

For all $k \in \mathbb{N}$ and $1 \leqslant i,j \leqslant m_k$, as the sequence $(\mu_{n,k}^{i,j})_{n\in\mathbb{N}}$ weakly converges to $\mu_k^{i,j}$, using [14, Theorem 2.7.4] with the open subset $K^c \subset \mathbf{Z}$, we get that $\mu_k^{i,j}(K^c) \leqslant \liminf_{n\to\infty} \mu_{n,k}^{i,j}(K^c)$. As $\lim_{n\to\infty} \lambda(S_{n,k,i}) = s_{k,i}$ for all $1 \leqslant i \leqslant m_k$, and summing those bounds, we get:

$$M_{U_k}(K^c) \leqslant \liminf_{n\to\infty} M_{W_{n,k}}(K^c) = \liminf_{n\to\infty} M_{W_n}(K^c) \leqslant \eta.$$

Consequently, the sequence $(U_k)_{k\in\mathbb{N}}$ is tight.

Step 3.b: Convergence of the measures $\hat{U}_k$ in $\mathcal{M}_+([0,1]^2 \times \mathbf{Z})$ defined for $k \in \mathbb{N}$ as:

$$\hat{U}_k(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = U_k(x, y; \mathrm{d}z)\, \lambda_2(\mathrm{d}x, \mathrm{d}y). \tag{8.2}$$

Since the sequence $(M_{U_k})_{k\in\mathbb{N}}$ in $\mathcal{M}_+(\mathbf{Z})$ is tight, for all $\eta > 0$, there exists a compact set $K \subset \mathbf{Z}$ such that for every $k \in \mathbb{N}$, $M_{U_k}(K^c) \leqslant \eta$; and thus $\hat{U}_k(\hat{K}^c) = M_{U_k}(K^c) \leqslant \eta$ where $\hat{K} = [0,1]^2 \times K$ is a compact subset of $[0,1]^2 \times \mathbf{Z}$, that is, the sequence $(\hat{U}_k)_{k\in\mathbb{N}}$ in $\mathcal{M}_+([0,1]^2 \times \mathbf{Z})$ is tight. The sequence $(\hat{U}_k)_{k\in\mathbb{N}}$ is also bounded as $\|\hat{U}_k\|_\infty \leqslant \|U_k\|_\infty \leqslant C$ thanks to (8.1). Hence, using Lemma 2.8, there exists a subsequence $(\hat{U}_{k_\ell})_{\ell\in\mathbb{N}}$ of the sequence $(\hat{U}_k)_{k\in\mathbb{N}}$ that converges to some measure, say $\hat{U}$, in $\mathcal{M}_+([0,1]^2 \times \mathbf{Z})$. Remark that, when considering the subsequence of indices $(k_\ell)_{\ell\in\mathbb{N}}$, the subsequences $(W_{n,k_\ell})_{\ell\in\mathbb{N}}$, $n \in \mathbb{N}$, still satisfy properties (i)-(iv) of Lemma 8.1, and for all $\ell \in \mathbb{N}$, the sequence $(W_{n,k_\ell})_{n\in\mathbb{N}}$ still weakly converges to U_{k_ℓ}. Without loss of generality, we now work with this subsequence and thus write k instead of k_ℓ.

Step 3.c: The measure $\hat{U}(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ can be disintegrated w.r.t. $\lambda_2(\mathrm{d}x, \mathrm{d}y)$ giving us an element of $\mathcal{W}_+$. To prove this, we need the following disintegration theorem for measures, see [38, Theorem 1.23] (stated in more the general framework of Borel spaces) which generalizes the disintegration theorem for probability measures [37, Theorem 6.3]. The notation $\mu \sim \nu$ for two measures μ and ν means that $\mu \ll \nu$ and $\nu \ll \mu$, where $\mu \ll \nu$ means that μ is absolutely continuous w.r.t. ν.

Lemma 8.4 (Disintegration theorem for measures, [38, Theorem 1.23]). *Let ρ be a measure on $S \times T$, where S is a measurable space and T a Polish space. Then there exist a measure $\nu \equiv \rho(\cdot \times T)$ on S and a probability kernel $\mu : S \to \mathcal{M}_1(T)$ such that $\rho = \nu \otimes \mu$ (i.e. $\rho(ds, dt) = \nu(ds)\mu(s; dt)$). Moreover, the measures $\mu_s = \mu(s; \cdot)$ are unique for ν-a.e. $s \in S$.*

Using Lemma 8.4 with $S = [0, 1]^2$ and $T = \mathbf{Z}$, we get that there exists a probability kernel U' in $\mathcal{W}_1$ such that:

$$\hat{U}(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = U'(x, y; \mathrm{d}z)\, \pi(\mathrm{d}x, \mathrm{d}y),$$

where $\pi = \hat{U}(\cdot \times \mathbf{Z})$ is a measure on $[0, 1]^2$.

We now need to prove that $\pi \ll \lambda_2$. By contradiction, assume this is false, then there exists a measurable set $A \in \mathcal{B}([0, 1]^2)$ such that $\lambda_2(A) = 0$ and $\pi(A) > 0$. As the measure $\int_A U'(x, y; \cdot)\, \pi(\mathrm{d}x, \mathrm{d}y)$ is not null, there exists $f \in C_b(\mathbf{Z})$ such that $\int_A U'(x, y; f)\, \pi(\mathrm{d}x, \mathrm{d}y) \neq 0$. As the sequence $(\hat{U}_k)_{k\in\mathbb{N}}$ weakly converges to $\hat{U}$ in $\mathcal{M}_+([0, 1]^2 \times \mathbf{Z})$ by step 3.b, we have that the sequence of measures $\hat{U}_k(\mathrm{d}x, \mathrm{d}y; f) = U_k(x, y; f)\, \lambda_2(\mathrm{d}x, \mathrm{d}y)$ weakly converges as $k \to \infty$ to $\hat{U}(\mathrm{d}x, \mathrm{d}y; f) = U'(x, y; f)\, \pi(\mathrm{d}x, \mathrm{d}y)$ in $\mathcal{M}_+([0, 1]^2)$. Moreover, as the maps $x, y \mapsto U_k(x, y; f)$ are uniformly bounded (by $\|f\|_\infty \|U_k\|_\infty \leqslant C\|f\|_\infty$, see (8.1)), they are also uniformly integrable (w.r.t. λ_2), and applying [12, Corollary 4.7.19] there exist a subsequence $(U_{k_\ell})_{\ell\in\mathbb{N}}$ and a bounded function g_f on $[0, 1]^2$ such that for every bounded measurable function $h \in L^\infty([0, 1]^2)$, we have:

$$\lim_{\ell\to\infty} \int U_{k_\ell}(x, y; f) h(x, y)\, \lambda_2(\mathrm{d}x, \mathrm{d}y) = \int g_f(x, y) h(x, y)\, \lambda_2(\mathrm{d}x, \mathrm{d}y).$$

In particular, the sequence of measures $(U_{k_\ell}(x, y; f)\, \lambda_2(\mathrm{d}x, \mathrm{d}y))_{\ell\in\mathbb{N}}$ weakly converges to the measure $g_f(x, y)\, \lambda_2(\mathrm{d}x, \mathrm{d}y)$, which imposes the equality between measures:

$$\hat{U}(\mathrm{d}x, \mathrm{d}y, f) = U'(x, y; f)\pi(\mathrm{d}x, \mathrm{d}y) = g_f(x, y)\, \lambda_2(\mathrm{d}x, \mathrm{d}y).$$

Hence, taking $h = \mathbb{1}_A$, we get:

$$\hat{U}(A, f) = \int_A U'(x, y; f)\pi(\mathrm{d}x, \mathrm{d}y) = \int_A g_f(x, y)\lambda_2(\mathrm{d}x, \mathrm{d}y) = 0,$$

which yields a contradiction. Consequently, the measure π is absolutely continuous w.r.t. λ_2, with density still denoted by π, and we set λ_2-a.e. on $[0, 1]^2$:

$$U(x, y; \mathrm{d}z) = \pi(x, y)\, U'(x, y; \mathrm{d}z) \quad \text{and thus} \quad \hat{U}(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = U(x, y; \mathrm{d}z)\, \lambda_2(\mathrm{d}x, \mathrm{d}y). \tag{8.3}$$

Step 3.d: The sequence $(U_k)_{k\in\mathbb{N}}$ weakly converges to U almost everywhere on $[0,1]^2$. Recall that by construction, the stepfunction U_k is adapted to the partition $\mathcal{P}_k$ defined in Step 2, and that $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$. A closer look at Step 2 yields that for all $\ell \geqslant k$, since $W_{n,k} = (W_{n,\ell})_{\mathcal{P}_{n,k}}$, we also get:

$$U_k = (U_\ell)_{\mathcal{P}_k}. \tag{8.4}$$

We prove (8.4) for $\ell = k+1$, the other cases follow by induction. As $\mathcal{P}_{n,k+1}$ is a refinement of $\mathcal{P}_{n,k}$, we already know that U_k and $(U_{k+1})_{\mathcal{P}_k}$ are both stepfunctions adapted to the finite partition $\mathcal{P}_k$. Thus, we only need to verify that U_k and $(U_{k+1})_{\mathcal{P}_k}$ take the same value on each step. For every $n \in \mathbb{N}$, the fact that $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$ implies that for all $1 \leqslant i,j \leqslant m_k$ such that $\lambda(S_{n,k,i})\lambda(S_{n,k,j}) > 0$, we have:

$$\mu_{i,j}^{n,k} = \sum_{i'\in I_i, j'\in I_j} \frac{\lambda(S_{n,k+1,i'})\lambda(S_{n,k+1,j'})}{\lambda(S_{n,k,i})\lambda(S_{n,k,j})} \mu_{i',j'}^{n,k+1},$$

and this equation is preserved when taking the limit $n \to \infty$, which gives us:

$$\mu_{i,j}^{k} = \sum_{i'\in I_i, j'\in I_j} \frac{s_{k+1,i'} s_{k+1,j'}}{s_{k,i} s_{k,j}} \mu_{i',j'}^{k+1}, \quad \text{for all } 1 \leqslant i,j \leqslant m_k \text{ such that } s_{k,i}s_{k,j} > 0.$$

This proves that the stepfunctions U_k and $(U_{k+1})_{\mathcal{P}_k}$ take the same value on each step $S_{k,i} \times S_{k,j}$ with positive size $s_{k,i}s_{k,j} > 0$ (on a step with null size $s_{k,i}s_{k,j} = 0$, U_k and $(U_{k+1})_{\mathcal{P}_k}$ are both equal to the null measure). This gives that $U_k = (U_{k+1})_{\mathcal{P}_k}$.

Let $f \in C_b(\mathbf{Z})$ be a bounded continuous function, and X, Y be independent uniform random variables on $[0,1]$. Then (8.4) and (8.1) imply that the sequence $N^f = (N_k^f = U_k(X,Y;f))_{k\in\mathbb{N}}$ is a martingale bounded by $C\|f\|_\infty$ for the filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$, where the σ-field $\mathcal{F}_k$ is generated by the events $\{X \in S_{k,i}\} \cap \{Y \in S_{k,j}\}$ for $1 \leqslant i,j \leqslant m_k$ and $S_{k,\ell} \in \mathcal{P}_k$. By the martingale convergence theorem, the martingale N^f is almost surely convergent, that is, the sequence $(U_k[f])_{k\in\mathbb{N}}$ converges λ_2-a.e. to a bounded measurable function u_f. Let $g : [0,1]^2 \to \mathbb{R}$ be a bounded measurable function. We get:

$$\begin{aligned}\int_{[0,1]^2} g(x,y)\, U(x,y;f)\, \lambda_2(\mathrm{d}x\mathrm{d}y) &= \int_{[0,1]^2\times\mathbf{Z}} g(x,y)\, f(z)\, \hat{U}(\mathrm{d}x,\mathrm{d}y,\mathrm{d}z)\\ &= \lim_{k\to\infty} \int_{[0,1]^2\times\mathbf{Z}} g(x,y)\, f(z)\, \hat{U}_k(\mathrm{d}x,\mathrm{d}y,\mathrm{d}z)\\ &= \lim_{k\to\infty} \mathbb{E}\left[g(X,Y)U_k(X,Y;f)\right]\\ &= \int_{[0,1]^2} g(x,y)\, u_f(x,y)\, \lambda_2(\mathrm{d}x\mathrm{d}y),\end{aligned}$$

where we used the definition (8.3) of U for the first equality, that $(\hat{U}_k)_{k\in\mathbb{N}}$ weakly converges to $\hat{U}$ for the second, the definition (8.2) of $\hat{U}_k$ for the third, and the convergence of the martingale N^f for the last. Since g is arbitrary, we deduce that λ_2-a.e. $U(\cdot,\cdot;f) = u_f$ and thus that the sequence $(U_k[f])_{k\in\mathbb{N}}$ converges λ_2-a.e. to $U[f]$. Applying this result for all $f \in \mathcal{F} = (f_m)_{m\in\mathbb{N}}$ a convergence determining sequence (with the convention $f_0 = \mathbb{1}$), we deduce that the sequence $(U_k)_{k\in\mathbb{N}}$ weakly converges to U almost everywhere on $[0,1]^2$. Recall from Section 2 that convergence determining sequences exist only for measures and not for signed measures in general, this is why we worked with measures in Step 3. This ends the proof of Step 3, and thus ends the proof of the lemma. □

We are now ready to prove Theorem 5.1.

Proof of Theorem 5.1. — We first prove Point (i) on $\mathcal{W}_\pm$ (the proof on $\mathcal{W}_1$ is similar). Since the distance d is weakly regular and the sequence $(W_n)_{n\in\mathbb{N}}$ is uniformly bounded and tight in $\mathcal{W}_\pm$, we can construct inductively for every $n \in \mathbb{N}$ a sequence $(\mathcal{P}_{n,k})_{k\in\mathbb{N}}$ of partitions of $[0,1]$ such that hypothesis (i)-(iii) of Lemma 8.1 are satisfied: $\mathcal{P}_{n,k+1}$ being obtained by applying the weak regularity property (see Definition 4.10-(i)) with starting partition $\mathcal{Q}_{n,k} = \mathcal{P}_{n,k} \wedge \mathcal{D}_k$, where $\mathcal{D}_k$ is the dyadic partition with stepsize $2^{-(k+1)}$. (We may assume that the partitions $P_{n,k}$ for all $n \in \mathbb{N}$ have the same size m_k by adding empty sets.) Then as d is also invariant and smooth on $\mathcal{W}_\pm$, the first part of Lemma 8.1 directly gives Point (i).

Before proving Point (ii), we first need to prove the following lemma.

Lemma 8.5 (Compactness theorem for $\mathcal{W}_\mathcal{M}$). — *Let d be an invariant, smooth and weakly regular distance on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$). Let $\mathcal{M}$ be a convex and weakly closed subset of $\mathcal{M}_1(\mathbf{Z})$ (resp. $\mathcal{M}_+(\mathbf{Z})$ or $\mathcal{M}_\pm(\mathbf{Z})$). Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{M}$-valued kernels which is tight and uniformly bounded. Then, $(W_n)_{n\in\mathbb{N}}$ has a subsequence that converges for $\delta_\square$ to some $\mathcal{M}$-valued kernel.*

Proof. — First remark that, as $\mathcal{M}$ is convex, the image of $\mathcal{W}_{\mathcal{M}}$ by the stepping operator $W \mapsto W_{\mathcal{P}}$, where $\mathcal{P}$ is a finite partition of $[0,1]$, is a subset of $\mathcal{W}_{\mathcal{M}}$. Hence, a close look at the proof of Lemma 8.1 (the partitions are constructed as in the proof of Point (i) from Theorem 5.1), and using the notation therein, shows that, up to taking subsequences, one can take the stepping kernels $W_{n,k}$ and U_k in $\mathcal{W}_{\mathcal{M}}$, such that $(U_k)_{k\in\mathbb{N}}$ weakly converges to U a.e. and the subsequence $(W_{n_\ell})_{\ell\in\mathbb{N}}$ converges to U w.r.t. $\delta_\square$. Since $U_k(x,y;\cdot) \in \mathcal{M}$ weakly converges to $U(x,y;\cdot)$ for almost every $x,y \in [0,1]$ and since $\mathcal{M}$ is weakly closed (and thus sequentially weakly closed), we deduce that $U(x,y;\cdot)$ belongs to $\mathcal{M}$ for almost every $x,y \in [0,1]$. This means that $U \in \mathcal{W}_{\mathcal{M}}$. $\square$

We prove Point (ii) for $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$ (the proof for $\mathcal{M} \subset \mathcal{M}_1(\mathbf{Z})$ is identical). The fact that $\mathcal{W}_{\mathcal{M}}$ and $\widetilde{\mathcal{W}}_{\mathcal{M}}$ are convex is clear as $\mathcal{M}$ is convex. Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of elements of $\widetilde{\mathcal{W}}_{\mathcal{M}}$. Since $\mathcal{M}$ is convex, we deduce that $(M_{W_n})_{n\in\mathbb{N}}$ is a sequence in $\mathcal{M}$. As $\mathcal{M}$ is sequentially compact for the weak topology, $\mathcal{M}$ is tight and bounded by Lemma 2.8, and thus the sequence $(W_n)_{n\in\mathbb{N}}$ is tight and uniformly bounded (recall Definition 4.7). Hence, using Lemma 8.5, we get that from any sequence in $\widetilde{\mathcal{W}}_{\mathcal{M}}$, we can extract a subsequence which converges for $\delta_\square$ to an element in $\widetilde{\mathcal{W}}_{\mathcal{M}}$. This implies that $(\widetilde{\mathcal{W}}_{\mathcal{M}}, \delta_\square)$ is compact.

Point (iii) is a direct consequence of Point (ii) as if $\mathbf{Z}$ is compact, so is $\mathcal{M}_1(\mathbf{Z})$. $\square$

Proof of Point (iii) from Proposition 5.2. — We prove Point (iii). The fact that $\mathcal{W}_{\mathcal{M}}$ and $\widetilde{\mathcal{W}}_{\mathcal{M}}$ are convex is clear as $\mathcal{M}$ is convex. To prove that $\widetilde{\mathcal{W}}_{\mathcal{M}}$ is closed, we consider a sequence $(W_n)_{n\in\mathbb{N}}$ in $\widetilde{\mathcal{W}}_{\mathcal{M}}$ that converges for $\delta_{\square,\mathrm{m}}$ to some $W \in \widetilde{\mathcal{W}}_\pm$. As $(W_n)_{n\in\mathbb{N}}$ is a Cauchy sequence for $\delta_{\square,\mathrm{m}}$, by Lemma 4.9, $(M_{W_n})_{n\in\mathbb{N}}$ is a Cauchy sequence for d_m and thus is tight. Hence, $(W_n)_{n\in\mathbb{N}}$ is uniformly bounded and tight. Applying Lemma 8.5, there exists a subsequence $(W_{n_k})_{k\in\mathbb{N}}$ of the sequence $(W_n)_{n\in\mathbb{N}}$ which converges for $\delta_{\square,\mathrm{m}}$ to some $\mathcal{M}$-valued kernel $U \in \widetilde{\mathcal{W}}_{\mathcal{M}}$. But as a subsequence, $(W_{n_k})_{k\in\mathbb{N}}$ must also converge for $\delta_{\square,\mathrm{m}}$ to W. This implies that $W = U$ is a $\mathcal{M}$-valued kernel. $\square$

In order to prove Theorem 5.5, we first prove a lemma that allows to construct the partitions needed to use Lemma 8.1.

LEMMA 8.6 (Construction of partitions for two distances). — *Let d and d' be two distances on $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) which are invariant, smooth, weakly regular and regular w.r.t. the stepping operator (see Definitions 3.10 and 4.10). Let $(W_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{W}_1$ (resp. $\mathcal{W}_+$ or $\mathcal{W}_\pm$) which is* *tight (resp. uniformly bounded and tight). Then, there exists sequences* $(\mathcal{P}_{n,k})_{k\in\mathbb{N}}$, $n \in \mathbb{N}$, *of partitions of* $[0,1]$ *such that hypothesis (i)–(iv) of Lemma 8.1 are satisfied.*

Proof. — We prove the result on $\mathcal{W}_\pm$ (the proof on $\mathcal{W}_1$ and $\mathcal{W}_+$ is similar). To simplify notations, write $d^1 = d$ and $d^2 = d'$. We proceed by induction on $k \in \mathbb{N} \cup \{-1\}$. For every $n \in \mathbb{N}$, set $\mathcal{P}_{n,-1} = \{[0,1]\}$ the trivial partition with size 1. Let $k \in \mathbb{N}$ and assume that we have already constructed partitions $(\mathcal{P}_{n,k-1})_{n\in\mathbb{N}}$ that have the same size m_{k-1}. Now we proceed to construct partitions $(\mathcal{P}_{n,k})_{n\in\mathbb{N}}$ that satisfy hypothesis (i)-(iv).

Set $C = \sup_{n\in\mathbb{N}} \|W_n\|_\infty$, which is finite as the sequence $(W_n)_{n\in\mathbb{N}}$ is uniformly bounded. As d^i, with $i = 1, 2$, are regular w.r.t. the stepping operator, there exists a finite constant $C_0 > 0$ such that for every $W, U \in \mathcal{W}_\pm$, with $\|W\|_\infty \leqslant C$ and $\|U\|_\infty \leqslant C$, and U a stepfunction adapted to a finite partition $\mathcal{Q}$:

$$d^i(W, W_\mathcal{Q}) \leqslant C_0\, d^i(W, U). \tag{8.5}$$

Set $\varepsilon = 1/C_0(k+1)$. Since d^i, with $i = 1, 2$, are weakly regular and the sequence $(W_n)_{n\in\mathbb{N}}$ is tight and uniformly bounded, there exists $r_k \in \mathbb{N}^*$, such that for every $n \in \mathbb{N}$, there exists a partition $\mathcal{R}^i_{n,k}$ of $[0,1]$ that refines $\mathcal{Q}_{n,k} = \mathcal{P}_{n,k-1} \wedge \mathcal{D}_k$, where $\mathcal{D}_k$ is the dyadic partition with stepsize 2^{-k}, such that:

$$|\mathcal{R}^i_{n,k}| \leqslant r_k|\mathcal{Q}_{n,k}| \leqslant 2^k r_k|\mathcal{P}_{n,k-1}| \quad \text{and} \quad d^i\Big(W_n, (W_n)_{\mathcal{R}^i_{n,k}}\Big) \leqslant \varepsilon. \tag{8.6}$$

(Indeed, a close look at the proof shows that $\mathcal{P}_{n,k-1}$ refines $\mathcal{D}_{k-1}$ by construction, thus $\mathcal{Q}_{n,k}$ cuts each set of $\mathcal{P}_{n,k-1}$ in at most 2 sets, and we get $|\mathcal{Q}_{n,k}| \leqslant 2|\mathcal{P}_{n,k-1}|$.) Now, let $\mathcal{P}_{n,k}$ be the common refinement of $\mathcal{R}^1_{n,k}$ and $\mathcal{R}^2_{n,k}$; it is a refinement of $\mathcal{P}_{n,k-1}$, has diameter at most 2^{-k} and size:

$$|\mathcal{P}_{n,k}| \leqslant 2^{2k} r_k^2 |\mathcal{P}_{n,k-1}|^2 = 2^{2k} r_k^2 m_{k-1}^2.$$

If necessary, by completing $\mathcal{P}_{n,k}$ with null sets, we may assume that $|\mathcal{P}_{n,k}| = m_k$, where $m_k = 2^{2k} r_k^2 m_{k-1}^2$. As $(W_n)_{\mathcal{R}^i_{n,k}}$ is a stepfunction adapted to the partition $\mathcal{P}_{n,k}$, we deduce from (8.5) and (8.6) that for $i = 1, 2$ and $n \in \mathbb{N}$:

$$d^i(W_n, (W_n)_{\mathcal{P}_{n,k}}) \leqslant C_0\, d^i\Big(W_n, (W_n)_{\mathcal{R}^i_{n,k}}\Big) \leqslant C_0\, \varepsilon = \frac{1}{k+1}\cdot$$

Hence, for every $n \in \mathbb{N}$, the partition $\mathcal{P}_{n,k}$ satisfies the hypothesis (i)-(iv) of Lemma 8.1. Thus, the induction is complete. □

Proof of Theorem 5.5. — Let d_m and $d_{\mathrm{m}'}$ be as in Theorem 5.5.

Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of probability-graphons that converges to some $W \in \widetilde{\mathcal{W}}_1$ for $\delta_{\square,\mathrm{m}}$. By Lemma 4.9, the sequence of probability measure $(M_{W_n})_{n\in\mathbb{N}}$ converges to M_W for the distance d_{m}. As d_{m} induces the weak topology on $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, we have that the sequence $(M_{W_n})_{n\in\mathbb{N}}$ is tight, and thus the sequence $(W_n)_{n\in\mathbb{N}}$ is also tight (recall Definition 4.7). The sequence $(W_n)_{n\in\mathbb{N}}$ is also uniformly bounded as a sequence in $\widetilde{\mathcal{W}}_1$. Applying Lemma 8.6 with the distances $d = d_{\square,\mathrm{m}}$ and $d' = d_{\square,\mathrm{m}'}$, which are invariant, smooth, weakly regular and regular w.r.t. the stepping operator, we get sequences of partitions $(\mathcal{P}_{n,k})_{k\in\mathbb{N}}$, $n \in \mathbb{N}$, that satisfy hypothesis (i)-(iv) of Lemma 8.1. We then deduce from the last part of Lemma 8.1 that any subsequence of $(W_n)_{n\in\mathbb{N}}$ has a further subsequence which converges to the same limit for both $\delta_{\square,\mathrm{m}}$ and $\delta_{\square,\mathrm{m}'}$, this limit must then be W. This implies that the sequence $(W_n)_{n\in\mathbb{N}}$ converges to W for $\delta_{\square,\mathrm{m}'}$.

The role of d_{m} and $d_{\mathrm{m}'}$ being symmetric, we conclude that the distances $\delta_{\square,\mathrm{m}}$ and $\delta_{\square,\mathrm{m}'}$ induce the same topology on $\widetilde{\mathcal{W}}_1$. □

9. Index of notations

Measures

- $(\mathbf{Z}, \mathcal{O}_{\mathbf{Z}})$ a topological Polish space (p. 37)
- $\mathcal{B}(\mathbf{Z})$ the Borel σ-field induced by $\mathcal{O}_{\mathbf{Z}}$ (p. 37)
- $C_b(\mathbf{Z})$ the set of continuous bounded real-valued functions on $\mathbf{Z}$ (p. 37)
- *measure* = positive measure (p. 36)
- $\mathcal{M}_{\pm}(\mathbf{Z})$ the set of signed measures on $\mathbf{Z}$ (p. 37)
- $\mathcal{M}_{+}(\mathbf{Z})$ the set of measures on $\mathbf{Z}$ (p. 37)
- $\mathcal{M}_{1}(\mathbf{Z})$ the set of probability measures on $\mathbf{Z}$ (p. 37)
- $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$ the set of sub-probability measures on $\mathbf{Z}$, i.e. measures with total mass at most 1 (p. 37)
- μ^+, μ^- the positive and negative parts of μ from its Hahn–Jordan decomposition (p. 37)
- $|\mu| = \mu_+ + \mu_-$ the total variation measure of μ (p. 37)
- $\|\mu\|_\infty = |\mu|(\mathbf{Z})$ the total mass of μ (p. 37)
- d_{m} a distance on either $\mathcal{M}_{\leqslant 1}(\mathbf{Z})$, $\mathcal{M}_{+}(\mathbf{Z})$ or $\mathcal{M}_{\pm}(\mathbf{Z})$ (p. 45)
- N_{m} a norm on $\mathcal{M}_{\pm}(\mathbf{Z})$ (p. 45)
- d_{LP} the Lévy–Prokhorov distance (p. 55)
- $\|\cdot\|_{\mathrm{KR}}$ the Kantorovitch–Rubinshtein norm (p. 55)

- $\|\cdot\|_{\mathrm{FM}}$ the Fortet–Mourier norm (p. 55)
- $\|\cdot\|_{\mathcal{F}}$ the norm based on a convergence determining sequence $\mathcal{F}$ (p. 56)

Relabelings and partitions

- $S_{[0,1]}$ the set of bijective measure-preserving maps from $([0,1],\lambda)$ to itself (p. 36)
- $\bar{S}_{[0,1]}$ the set of measure-preserving maps from $([0,1],\lambda)$ to itself (p. 36)
- $|\mathcal{P}|$ the number of sets in the finite partition $\mathcal{P}$ (p. 44)

Kernels and graphons spaces

- $\mathcal{W}_1$ the set of probability-graphons (p. 42)
- $\mathcal{W}_+$ the set of measure-valued kernels (p. 42)
- $\mathcal{W}_\pm$ the set of signed measure-valued kernels (p. 42)
- $\mathcal{W}_\mathcal{M}$ the set of $\mathcal{M}$-valued kernels with $\mathcal{M} \subset \mathcal{M}_\pm(\mathbf{Z})$ (p. 43)
- $\widetilde{\mathcal{W}}_1$ the set of unlabeled probability-graphons (p. 50)
- $\widetilde{\mathcal{W}}_+$ the set of unlabeled measure-valued kernels (p. 50)
- $\widetilde{\mathcal{W}}_\pm$ the set of unlabeled signed measure-valued kernels (p. 50)
- $\widetilde{\mathcal{W}}_\mathcal{M}$ the set of unlabeled $\mathcal{M}$-valued kernels (p. 50)

Kernels and graphons

- W^+ and W^- the positive and negative part of $W \in \mathcal{W}_\pm$ (p. 43)
- $|W| = W^+ + W^-$ (p. 43)
- $W(A;\cdot) = \int_A W(x,y;\cdot)\ \mathrm{d}x\mathrm{d}y$ for $A \subset [0,1]^2$ (p. 45)
- $W[f](x,y) = W(x,y;f)$ for $f \in C_b(\mathbf{Z})$ (p. 53)
- $W_\mathcal{P}$ the stepping of W w.r.t. a partition $\mathcal{P}$ (p. 59)
- $\|W\|_\infty := \sup_{x,y\in[0,1]} \|W(x,y;\cdot)\|_\infty$ (p. 42)
- $M_W(\mathrm{d}z) = |W|([0,1]^2;\mathrm{d}z)$ (p. 61)
- W_G the probability-graphon associated to a $\mathcal{M}_1(\mathbf{Z})$-graph or a weighted graph G (p. 80)
- $\mathbb{H}(k,W)$ the $\mathcal{M}_1(\mathbf{Z})$-graph with k vertices sampled from $W \in \mathcal{W}_1$ (p. 82)
- $\mathbb{G}(k,W)$ the $\mathcal{M}_1(\mathbf{Z})$-graph with k vertices sampled from $W \in \mathcal{W}_1$ (p. 82)
- F^g a finite graph whose edges are decorated with functions in $C_b(\mathbf{Z})$ (p. 89)
- $t(F^g,W) = M_W^F(g)$ the homomorphism density of F^g in W (p. 89)

Distances/norms on graphon spaces

- $d_{\square,\mathrm{m}}$ the cut distance associated to d_m (p. 45)
- $N_{\square,\mathrm{m}}$ the cut norm associated to N_m (p. 45)

Definitions

Acknowledgement

We thank Pierre-André Zitt for some helpful discussion in particular on Lemma 3.20. We also thank the referees for their careful reading and for their helpful suggestions which improved the paper.

BIBLIOGRAPHY

[1] R. Albert, I. Albert & G. L. Nakarado, Structural Vulnerability of the North American Power Grid, *Physical Review E* **69** (2004), no. 2, article no. 025103, https://doi.org/10.1103/PhysRevE.69.025103.

[2] R. Albert & A.-L. Barabási, Statistical Mechanics of Complex Networks, *Reviews of Modern Physics* **74** (2002), no. 1, p. 47–97, https://doi.org/10.1103/RevModPhys.74.47.

[3] G. Aletti & G. Naldi, Opinion Dynamics on Graphon: The Piecewise Constant Case, *Applied Mathematics Letters* **133** (2022), article no. 108227, https://doi.org/10.1016/j.aml.2022.108227.

[4] L. Alonso, J. A. Méndez-Bermúdez, A. González-Meléndrez & Y. Moreno, Weighted Random-Geometric and Random-Rectangular Graphs: Spectral and Eigenfunction Properties of the Adjacency Matrix, *Journal of Complex Networks* **6** (2018), no. 5, p. 753–766, https://doi.org/10.1093/comnet/cnx053.

[5] H. Amini, M. Draief & M. Lelarge, Flooding in Weighted Sparse Random Graphs, *SIAM Journal on Discrete Mathematics* **27** (2013), no. 1, p. 1–26, https://doi.org/10.1137/120865021.

[6] H. Amini & M. Lelarge, The Diameter of Weighted Random Graphs, *The Annals of Applied Probability* **25** (2015), no. 3, p. 1686–1727, https://doi.org/10.1214/14-AAP1034.

[7] E. Archer, A. Nachmias & M. Shalev, The GHP Scaling Limit of Uniform Spanning Trees in High Dimensions, *Communications in Mathematical Physics* **405** (2024), no. 3, article no. 73, https://doi.org/10.1007/s00220-023-04923-2.

[8] S. Athreya, S. Pal, R. Somani & R. Tripathi, Path Convergence of Markov Chains on Large Graphs, 2023, https://arxiv.org/abs/2308.09214.

[9] N. Ayi & N. P. Duteil, Graph Limit for Interacting Particle Systems on Weighted Random Graphs, 2023, https://arxiv.org/abs/2307.12801.

[10] S. Bhamidi, R. Van Der Hofstad & G. Hooghiemstra, First Passage Percolation on Random Graphs with Finite Mean Degrees, *The Annals of Applied Probability* **20** (2010), no. 5, p. 1907–1965, https://doi.org/10.1214/09-AAP666.

[11] P. Billingsley, *Convergence of Probability Measures*, second edition ed., Wiley, New York, 1999.

[12] V. I. Bogachev, *Measure Theory Vol. 1*, vol. 1, Springer, Berlin ; New York, 2007.

[13] ———, *Measure Theory Vol. 2*, vol. 2, Springer, Berlin ; New York, 2007.

[14] ———, *Weak Convergence of Measures*, Mathematical Surveys and Monographs, no. volume 234, American Mathematical Society, Providence, Rhode Island, 2018.

[15] B. Bollobás, S. Janson & O. Riordan, The Cut Metric, Random Graphs, and Branching Processes, *Journal of Statistical Physics* **140** (2010), no. 2, p. 289–335, https://doi.org/10.1007/s10955-010-9982-z.

[16] C. Borgs, J. Chayes, L. Lovász, V. Sós & K. Vesztergombi, Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing, *Advances in Mathematics* **219** (2008), no. 6, p. 1801–1851, https://doi.org/10.1016/j.aim.2008.07.008.

[17] C. Borgs & J. Chayes, Graphons: A Nonparametric Method to Model, Estimate, and Design Algorithms for Massive Networks, in *Proceedings of the 2017 ACM Conference on Economics and Computation* (Cambridge Massachusetts USA), ACM, 2017, p. 665–672, https://doi.org/10.1145/3033274.3084097.

[18] C. Borgs, J. Chayes, L. Lovász, V. Sós & K. Vesztergombi, Convergent Sequences of Dense Graphs II. Multiway Cuts and Statistical Physics, *Annals of Mathematics* **176** (2012), no. 1, p. 151–219, https://doi.org/10.4007/annals.2012.176.1.2.

[19] S. Bornholdt & H. G. Schuster (eds.), *Handbook of Graphs and Networks: From the Genome to the Internet*, 1 ed., Wiley, 2002.

[20] J.-F. Delmas, D. Dronnier & P.-A. Zitt, An Infinite-Dimensional Metapopulation SIS Model, *Journal of Differential Equations* **313** (2022), p. 1–53, https://doi.org/10.1016/j.jde.2021.12.024.

[21] M. Draief & L. Massoulié, *Epidemics and Rumours in Complex Networks*, London Mathematical Society Lecture Note Series, no. 369, Cambridge University Press, Cambridge ; New York, 2010.

[22] L. Dubins & D. Freedman, Measurable Sets of Measures, *Pacific Journal of Mathematics* **14** (1964), no. 4, p. 1211–1222, https://doi.org/10.2140/pjm.1964.14.1211.

[23] R. Engelking, *General topology*, rev. and completed ed ed., Sigma series in pure mathematics, no. 6, Heldermann, Berlin, 1989.

[24] S. N. Ethier & T. G. Kurtz, *Markov processes: characterization and convergence*, vol. 282, John Wiley & Sons, 2009.

[25] V. Falgas-Ravry, K. O'Connell, J. Strömberg & A. Uzzell, Multicolour Containers and the Entropy of Decorated Graph Limits, 2016, https://arxiv.org/abs/1607.08152.

[26] M. Freedman, L. Lovász & A. Schrijver, Reflection Positivity, Rank Connectivity, and Homomorphism of Graphs, *Journal of the American Mathematical Society* **20** (2006), no. 1, p. 37–51, https://doi.org/10.1090/S0894-0347-06-00529-7.

[27] A. Frieze & R. Kannan, Quick Approximation to Matrices and Applications, *Combinatorica* **19** (1999), no. 2, p. 175–220, https://doi.org/10.1007/s004930050052.

[28] F. Garbe, R. Hancock, J. Hladký & M. Sharifzadeh, Limits of Latin Squares, *Discrete analysis* **2023** (2023), article no. 8, https://doi.org/10.19086/da.83253.

[29] D. Garlaschelli, The Weighted Random Graph Model, *New Journal of Physics* **11** (2009), no. 7, article no. 073005, https://doi.org/10.1088/1367-2630/11/7/073005.

[30] J. Hladký, A. Nachmias & T. Tran, The Local Limit of the Uniform Spanning Tree on Dense Graphs, *Journal of Statistical Physics* **173** (2018), no. 3-4, p. 502–545, https://doi.org/10.1007/s10955-017-1933-5.

[31] J. Hladký & G. Viswanathan, Random Minimum Spanning Tree and Dense Graph Limits, 2023, https://arxiv.org/abs/2310.11705.

[32] Y. Hu, X. Wei, J. Yan & H. Zhang, Graphon Mean-Field Control for Cooperative Multi-Agent Reinforcement Learning, *Journal of the Franklin Institute* (2023), https://doi.org/10.1016/j.jfranklin.2023.09.002.

[33] A. Humphreys & S. Simpson, Separable Banach Space Theory Needs Strong Set Existence Axioms, *Transactions of the American Mathematical Society* **348** (1996), no. 10, p. 4231–4255, https://doi.org/10.1090/S0002-9947-96-01725-4.

[34] T. R. Hurd & J. P. Gleeson, On Watts' Cascade Model with Random Link Weights, *Journal of Complex Networks* **1** (2013), no. 1, p. 25–43, https://doi.org/10.1093/comnet/cnt003.

[35] S. Janson, Graphons, Cut Norm and Distance, Couplings and Rearrangements, *New York Journal of Mathematics* **4** (2013), https://arxiv.org/abs/1009.2376.

[36] V. Jog & P.-L. Loh, Recovering Communities in Weighted Stochastic Block Models, in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (Monticello, IL), IEEE, 2015, p. 1308–1315, https://doi.org/10.1109/ALLERTON.2015.7447159.

[37] O. Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002, xx+638 pages.

[38] ———, *Random Measures, Theory and Applications*, Probability Theory and Stochastic Modelling, vol. 77, Springer International Publishing, Cham, 2017.

[39] N. Keriven & S. Vaiter, What Functions Can Graph Neural Networks Compute on Random Graphs? The Role of Positional Encoding, in *Advances in Neural Information Processing Systems*, vol. 36, Curran Associates, Inc., 2023, p. 11823–11849, https://proceedings.neurips.cc/paper_files/paper/2023/file/271ec4d1a9ff5e6b81a6e21d38b1ba96-Paper-Conference.pdf.

[40] I. Z. Kiss, J. C. Miller & P. L. Simon, *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, Interdisciplinary Applied Mathematics, vol. 46, Springer International Publishing, Cham, 2017.

[41] I. Kolossváry & B. Ráth, Multigraph Limits and Exchangeability, *Acta Mathematica Hungarica* **130** (2011), no. 1-2, p. 1–34, https://doi.org/10.1007/s10474-010-0037-3.

[42] D. Kunszenti-Kovács, L. Lovász & B. Szegedy, Multigraph Limits, Unbounded Kernels, and Banach Space Decorated Graphs, *Journal of Functional Analysis* **282** (2022), no. 2, article no. 109284, https://doi.org/10.1016/j.jfa.2021.109284.

[43] D. Lacker & A. Soret, A Label-State Formulation of Stochastic Graphon Games and Approximate Equilibria on Large Networks, *Mathematics of Operations Research* (2022), article no. moor.2022.1329, https://doi.org/10.1287/moor.2022.1329.

[44] M. Lelarge, L. Massoulie & J. Xu, Reconstruction in the Labeled Stochastic Block Model, in *2013 IEEE Information Theory Workshop (ITW)* (Sevilla, Spain), IEEE, 2013, p. 1–5, https://doi.org/10.1109/ITW.2013.6691264.

[45] T. G. Lewis & T. G. Lewis, *Network Science: Theory and Practice*, Wiley, Hoboken, NJ, 2009.

[46] L. Lovász, *Large networks and graph limits*, vol. 60, American Mathematical Society, Colloquium Publications, 2012.

[47] L. Lovász & B. Szegedy, Limits of Dense Graph Sequences, *Journal of Combinatorial Theory, Series B* **96** (2006), no. 6, p. 933–957, https://doi.org/10.1016/j.jctb.2006.05.002.

[48] ———, Szemerédi's Lemma for the Analyst, *GAFA Geometric And Functional Analysis* **17** (2007), no. 1, p. 252–270, https://doi.org/10.1007/s00039-007-0599-6.

[49] ———, Limits of Compact Decorated Graphs, 2010, https://arxiv.org/abs/1010.5155.

[50] M. E. J. Newman, The Structure and Function of Complex Networks, *SIAM Review* **45** (2003), no. 2, p. 167–256, https://doi.org/10.1137/S003614450342480.

[51] R. Purves, Bimeasurable functions, *Fund. Math.* **58** (1966), p. 149–157, https://doi.org/10.4064/fm-58-2-149-157.

[52] V. S. Varadarajan, Weak Convergence of Measures on Separable Metric Spaces, *Sankhyā: The Indian Journal of Statistics (1933-1960)* **19** (1958), no. 1/2, p. 15–22, http://www.jstor.org/stable/25048364.

[53] J. Xu, L. Massoulié & M. Lelarge, Edge Label Inference in Generalized Stochastic Block Models: from Spectral Theory to Impossibility Results, in *Proceedings of The 27th Conference on Learning Theory* (Barcelona, Spain) (M. F. Balcan, V. Feldman & C. Szepesvári, eds.), Proceedings of Machine Learning Research, vol. 35, PMLR, 13–15 Jun 2014, p. 903–920, https://proceedings.mlr.press/v35/xu14.html.

[54] M. Xu, V. Jog & P.-L. Loh, Optimal Rates for Community Estimation in the Weighted Stochastic Block Model, *The Annals of Statistics* **48** (2020), no. 1, p. 183–204, https://doi.org/10.1214/18-AOS1797.

[55] S.-Y. Yun & A. Proutiere, Optimal Cluster Recovery in the Labeled Stochastic Block Model, in *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon & R. Garnett, eds.), vol. 29, Curran Associates, Inc., 2016, https://proceedings.neurips.cc/paper_files/paper/2016/file/a8849b052492b5106526b2331e526138-Paper.pdf.

Manuscript received 22nd December 2023,
accepted 19th February 2025.

Romain ABRAHAM
Institut Denis Poisson
Université d'Orléans, Université de Tours, CNRS
France
romain.abraham@univ-orleans.fr

Jean-François DELMAS
CERMICS
Ecole des Ponts
France
jean-francois.delmas@enpc.fr

Julien WEIBEL
Institut Denis Poisson
Université d'Orléans, Université de Tours, CNRS
France
and CERMICS

Ecole des Ponts
France
julien.weibel@normalesup.org

Innovations in Graph Theory is a member of the Mersenne Center for Open Publishing
ISSN: 3050-743X