GRAPHONS, CUT NORM AND DISTANCE, COUPLINGS AND REARRANGEMENTS

SVANTE JANSON

Abstract. We give a survey of basic results on the cut norm and cut metric for graphons (and sometimes more general kernels), with emphasis on the equivalence problem. The main results are not new, but we add various technical complements, and a new proof of the uniqueness theorem by Borgs, Chayes and Lovász. We allow graphons on general probability spaces whenever possible. We also give some new results for {0,1}-valued graphons and for pure graphons.

1. Introduction

In the recent theory of *graph limits*, introduced by Lovász and Szegedy [48] and further developed by e.g. Borgs, Chayes, Lovász, Sós and Vesztergombi [14, 15], a prominent role is played by *graphons*. These are symmetric measurable functions $W : \Omega^2 \to [0,1]$, where, in general, Ω is an arbitrary probability space. The basic fact is that every graph limit can be represented by a graphon (where we further may choose $\Omega = [0,1]$ if we like); however, such representations of graph limits are far from unique, see e.g., [12, 13, 14, 24, 48]. (This representation is essentially equivalent to the representation by Aldous and Hoover of exchangeable arrays of random variables, see [43] for details of this representation and [4, 24] for the connection, which is summarized in Appendix D.) See Appendix B for a very brief summary.

It turns out that for studying both convergence and equivalence of graphons, a key tool is the *cut metric* [14]. The purpose of this paper is to give a survey over basic, and often elementary, facts on the cut norm and cut metric. Most results in this paper are not new, even when we do not give a specific reference. (Most results are in at least one of [12, 13, 14, 24, 48].) However, the results are sometimes difficult to find in the literature, since they are spread out over several papers, with somewhat different versions of the definitions and assumptions; moreover, some elementary results have only been given implicitly and without proof before. Hence we try to collect the results and proofs here, and state them in as general forms as we find convenient. For example, we allow general probability spaces whenever possible. We thus add various technical complements to previous results. We also give some new results, including some results on $\{0,1\}$-valued graphons

Date: 10 September, 2010; revised 31 August 2012.
2000 *Mathematics Subject Classification.* 05C99, 28A20.

in Section 10, and some results on pure graphons leading to a new proof of the uniqueness theorem by Borgs, Chayes and Lovász [13] in Section 9.

We include below for convenience some standard facts from measure theory, sometimes repeating standard arguments. Some general references (from different points of view) are [6, 19, 42, 54].

Remark 1.1. The basic idea of graph limits has been generalized to limits of many other finite combinatorial objects such as weighted graphs, directed graphs, multigraphs, bipartite graphs, hypergraphs, posets and permutations, see for example [4, 14, 15, 24, 26, 36, 40, 44, 51, 52]. Many results below extend in a straightforward way to such extensions, but for simplicity we leave such extensions to the reader and concentrate on the standard case.

2. The setting

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. (We will usually denote this space simply by Ω or (Ω, μ), with $\mathcal{F}$ and perhaps μ being clear from the context.) Often we take Ω to be $[0,1]$ (or $(0,1]$) with $\mu = \lambda$, the Lebesgue measure; this is sometimes convenient, and it is often possible to reduce to this case; in fact, in several papers on graph limits only this case is considered for convenience. (See [38] for a general representation theorem.) However, it is also often convenient to consider other Ω, and we will here be general and allow arbitrary probability spaces.

Nevertheless, we will often consider $[0,1]$ or $(0,1]$. Except when we explicitly say otherwise, we will always assume that these spaces are equipped with the Borel σ-field $\mathcal{B}$ and the Lebesgue measure, which we denote by λ. (We denote the Lebesgue σ-field by $\mathcal{L}$; we will occasionally use it instead of $\mathcal{B}$, but not without saying so. Recall that $\mathcal{L}$ is the completion of $\mathcal{B}$, see e.g. [19].)

Remark 2.1. Our default use of $\mathcal{B}$ is important when we consider mappings into $[0,1]$, but for functions defined on $[0,1]$ or $[0,1]^2$, it often does not matter whether we use $\mathcal{B}$ or $\mathcal{L}$, since every $\mathcal{L}$-measurable function is a.e. equal to a $\mathcal{B}$-measurable one. In fact, it is sometimes more convenient to use $\mathcal{L}$.

In a few cases, we will need some technical assumptions on Ω. We refer to Appendix A for the definitions of *atomless*, *Borel* and *Lebesgue* probability spaces.

We will study functions on Ω^2, and various (semi)metrics on such functions. Of course, Ω^2 is itself a probability space, equipped with the product measure $\mu^2 := \mu \times \mu$ and the product σ-field (or its completion; this makes no difference for our purposes).

Remark 2.2. The definitions and many results can be extended to functions of Ω^r for arbitrary $r \geq 2$, which is the setting for hypergraph limits; see e.g. [10] and [26].

All subsets and all functions on Ω or Ω^2 that we consider will tacitly be assumed to be measurable. We will usually identify functions that are

a.e. equal. This also means that functions only have to be defined a.e. (In particular, this means that it does not make any significant difference if we replace $\mathcal{F}$ by its completion; for example, on $[0,1]$ and $[0,1]^2$, with Lebesgue measure, it does not matter whether we consider Borel or Lebesgue measurable functions, cf. Remark 2.1. Moreover, in this case it does not matter whether we take $[0,1]$, $(0,1]$ or $(0,1)$.)

The natural domain of definition for the various metrics we consider is $L^1(\Omega^2)$, but we are really mainly interested in some subclasses.

Definition 2.3. A *kernel* on Ω is an integrable, symmetric function $W : \Omega^2 \to [0,\infty)$.

A *standard kernel* or *graphon* on Ω is a (measurable) symmetric function $W : \Omega^2 \to [0,1]$.

We let $\mathcal{W} = \mathcal{W}(\Omega)$ denote the set of all graphons on a given Ω.

We are mainly interested in the graphons (standard kernels), since they correspond to graph limits. We use kernels when we find it more natural to state results in this generality, but we will often consider just graphons for convenience, leaving possible extensions to the reader.

Warning. The terminology varies between different authors and papers. *Kernel* and *graphon* are used more or less interchangeably, with somewhat different definitions in different papers. (This includes my own papers, where again there is no consistency.) Apart from the two cases in the definition above, one sometimes considers the intermediate case of arbitrary bounded symmetric functions $\Omega^2 \to [0,\infty)$. Moreover, sometimes one considers W with arbitrary values in $\mathbb{R}$, and not just $W \geq 0$; for simplicity, we will not consider this case here. (Extensions to these cases are typically straightforward when they are possible.)

Remark 2.4. For consistency we here require W to be measurable for the product σ-field $\mathcal{F} \times \mathcal{F}$, but it makes no essential difference if we only require W to be measurable for the completion of $\mathcal{F} \times \mathcal{F}$, since every kernel of the latter type is a.e. equal to an $\mathcal{F} \times \mathcal{F}$-measurable kernel.

Remark 2.5. A kernel is said to be *Borel* if it is defined on a Borel space, and *Lebesguian* if it is defined on a Lebesgue space, see Appendix A for definitions. We sometimes have to restrict to such special kernels (which include all common examples). Note that the difference between Borel and Lebesguian kernels is very minor: A Lebesgue probability space is the same as the completion of a Borel probability space. Hence, if W is a Borel kernel defined on some (Borel) space $(\Omega, \mathcal{F}, \mu)$, then W can also be regarded as a Lebesguian kernel defined on $(\Omega, \widehat{\mathcal{F}}, \mu)$, where $\widehat{\mathcal{F}}$ is the completion of $\mathcal{F}$ (for μ). Conversely, if W is a Lesbeguian kernel defined on $(\Omega, \mathcal{F}, \mu)$, then $\mathcal{F}$ is the completion of a sub-σ-field $\mathcal{F}_0$ such that $(\Omega, \mathcal{F}_0, \mu)$ is a Borel space. Hence W is a.e. equal to some $\mathcal{F}_0 \times \mathcal{F}_0$-measurable function W_0, which we may be assume to be symmetric and with values in $[0,1]$; thus $W = W_0$ a.e. where W_0 is a Borel kernel. Consequently, up to a.e. equivalence, the

classes of Borel and Lebesgue kernels are the same, and it is a matter of taste which version we choose when we introduce one of these restrictions. Cf. Remark 2.1.

Remark 2.6. The definitions and results can be extended to the non-symmetric case, considering instead of $\mathcal{W}(\Omega)$ the set of arbitrary (measurable) functions $\Omega^2 \to [0,1]$ or, more generally, $\Omega_1 \times \Omega_2 \to [0,1]$. Such functions (*bigraphons*) appear in the graph limit theory for bipartite graphs, see e.g. [24] and [51].

Example 2.7. Let G be a (simple, undirected) graph. Then G defines naturally a graphon W_G, which forms a link between graphs and graphons and is central in the graph limit theory, see e.g. [14]. In fact, there are two natural versions, which we denote by W_G^{V} and W_G^{I}.

For the first version, we regard the vertex set V of G as a probability space with each vertex having equal probability $1/|G|$. We define the graphon $W_G^{\mathsf{V}} : V^2 \to [0,1]$ on this probability space by

$$W_G^{\mathsf{V}}(u,v) = \begin{cases} 1 & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

In other words, W_G^{V} equals (up to notation) the adjacency matrix of G.

For the second version we choose the probability space $\Omega = (0,1]$. Let $n := |G|$ and partition $(0,1]$ into n intervals $I_{in} := (\frac{i-1}{n}, \frac{i}{n}]$. We assume that the vertices of G are labelled $1,\dots,n$ (or, equivalently, that $V = \{1,\dots,n\}$), and define

$$W_G^{\mathsf{I}}(x,y) := W_G^{\mathsf{V}}(i,j) \quad \text{if } x \in I_{in},\ y \in I_{jn}. \tag{2.2}$$

The graphons W_G^{V} and W_G^{I} are equivalent in the sense defined below, see Example 6.8. Usually it does not matter which version we choose, and we let W_G denote any of them when the choice is irrelevant.

3. Step functions

Recall that a function f on Ω is *simple* or a *step function* if there is a finite partition $\Omega = \bigcup_{i=1}^n A_i$ of Ω such that f is constant on each A_i. Similarly, we say that a function W on Ω^2 is a *step function* if there is a finite partition $\Omega = \bigcup_{i=1}^n A_i$ of Ω such that W is constant on each $A_i \times A_j$. Step functions are also said to be of *finite type*. If W is a kernel or graphon that also is a step function, we call it a *step kernel* or *step graphon*.

When necessary, we may be more specific and say, for example, that W is a *$\mathcal{P}$-step function*, where $\mathcal{P}$ is the partition $\{A_i\}$ above, or an *n-step function*, when the number of parts A_i is (at most) n.

Step kernels (and graphons) are important mainly as a technical tool, see several proofs below. However, they can also be studied for their own sake; see Lovász and Sós [47], which can be seen as a study of step graphons, although the results are stated in terms of the corresponding graph limits and convergent sequences of graphs.

Remark 3.1. Note that being a step function on Ω^2 is stronger than being a simple function on that space, which means constant on the sets of some arbitrary partition of Ω^2; it is important that we use product sets in the definition of a step function on Ω^2. See also Example 5.3 below.

Warning. Some authors use different terminology. For example, when studying functions on $[0,1]$, step functions are sometimes defined as functions constant on some finite set of intervals partitioning $[0,1]$, i.e., the parts A_i are required to be intervals. We make no such assumption.

4. The cut norm

For functions in $L^1(\Omega^2)$ we have the usual L^1 norm

$$\|W\|_1 := \int_{\Omega^2} |W| \,\mathrm{d}\mu^2 \tag{4.1}$$

and the corresponding metric $\|W_1 - W_2\|_1$.

For the graph limit theory, it turns out that another norm is more important. This is the *cut norm* $\|W\|_\square$ of W, which was introduced for a different purpose by Frieze and Kannan [29], and given a central role in the graph limit theory by Borgs, Chayes, Lovász, Sós and Vesztergombi [14]. (Its history actually goes back much further. For functions on $[0,1]^2$, the version in (4.3) is the same as the *Fréchet variation* of the corresponding distribution function $F(x,y) := \int_0^x \int_0^y W$, see Fréchet [28]; more generally, $\|W\|_{\square,2}$ equals the Fréchet variation of the bimeasure on Ω^2 corresponding to W. See further e.g. Littlewood [46] (where also the discrete version is considered), Clarkson and Adams [18] and Morse [53], and in particular Blei [7] with further references.)

There are several versions of the cut norm, equivalent within constant factors. Following [29] and [14], for $W \in L^1(\Omega^2)$ we define

$$\|W\|_{\square,1} := \sup_{S,T} \Bigl| \int_{S\times T} W(x,y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \Bigr|, \tag{4.2}$$

where the supremum is taken over all pairs of measurable subsets of Ω. Alternatively, one can take

$$\|W\|_{\square,2} := \sup_{\|f\|_\infty, \|g\|_\infty \le 1} \Bigl| \int_{\Omega^2} W(x,y) f(x) g(y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \Bigr|, \tag{4.3}$$

taking the supremum over all (real-valued) functions f and g with values in $[-1,1]$. (We let $\|f\|_\infty$ denote the norm in L^∞ of f, i.e., the essential supremum of $|f|$.) It is easily seen that in taking the supremum in (4.3) one can restrict to functions f and g taking only the values ± 1. Note that (4.2) is equivalent to (4.3) with the supremum taken over only f and g with values in $\{0,1\}$ (i.e., indicator functions); it follows that

$$\|W\|_{\square,1} \le \|W\|_{\square,2} \le 4\|W\|_{\square,1}. \tag{4.4}$$

Thus the two norms $\|\cdot\|_{\square,1}$ and $\|\cdot\|_{\square,2}$ are equivalent, and it will almost never matter which one we use. We shall write $\|\cdot\|_\square$ for either norm, when the choice of definition does not matter. For further, equivalent, versions of the cut norm, see Appendix E.

We usually do not indicate Ω or μ explicitly in the notation; when necessary we may add them as subscripts and write, for example, $\|\cdot\|_{\square,\Omega,\mu}$ or $\|\cdot\|_{\square,\Omega,\mu,1}$.

Remark 4.1. Similarly, it is easily seen that (4.2) is equivalent to (4.3) with the supremum taken over only f and g with values in $[0,1]$.

One advantage of the version $\|\cdot\|_{\square,2}$ is the simple "Banach module" property: For any bounded functions h and k on Ω,

$$\|h(x)k(y)W(x,y)\|_{\square,2} \le \|h\|_\infty \|k\|_\infty \|W\|_{\square,2}. \tag{4.5}$$

A similar advantage is seen in Lemma 4.5 below. (In both cases, using $\|\cdot\|_{\square,1}$ would introduce some constants.) On the other hand, $\|\cdot\|_{\square,1}$ is perhaps more natural, and probably more familiar, in combinatorics.

Note that for either definition of the cut norm we have

$$\Bigl|\int_{\Omega^2} W\Bigr| \le \|W\|_\square \le \|W\|_1. \tag{4.6}$$

Remark 4.2. The definition (4.3) is natural for a functional analyst. This norm is the dual of the projective tensor product norm in $L^\infty(\Omega)\hat{\otimes}L^\infty(\Omega)$, and is thus the injective tensor product norm in $L^1(\Omega)\check{\otimes}L^1(\Omega)$; equivalently, it is equal to the operator norm of the corresponding integral operator $L^\infty(\Omega) \to L^1(\Omega)$. This contrasts nicely to the L^1 norm on Ω^2, which is the projective tensor product norm in $L^1(\Omega)\hat{\otimes}L^1(\Omega)$. (See e.g. [59].)

Remark 4.3. We may similarly define the cut norm of functions defined on a product of two different spaces.

Remark 4.4. The one-dimensional version of the cut norm coincides with the L^1 norm. This is exact for $\|\cdot\|_{\square,2}$: If f is any integrable function of Ω, then

$$\|f\|_1 = \sup_{\|g\|_\infty\le 1}\Bigl|\int_\Omega f(x)g(x)\,\mathrm{d}\mu(x)\Bigr|. \tag{4.7}$$

For the one-dimensional version of $\|\cdot\|_{\square,1}$, we may in analogy with (4.4) lose a factor 2; we omit the details.

We define the *marginals* of a function $W \in L^1(\Omega^2)$ by

$$\overline{W}^{(1)}(x) := \int_\Omega W(x,y)\,\mathrm{d}\mu(y), \tag{4.8}$$

$$\overline{W}^{(2)}(y) := \int_\Omega W(x,y)\,\mathrm{d}\mu(x). \tag{4.9}$$

It a well-known consequence of Fubini's theorem that $\|\overline{W}^{(1)}\|_{L^1(\Omega)} \le \|W\|_{L^1(\Omega^2)}$ for any $W \in L^1(\Omega^2)$. This extends to the cut norm on Ω^2,

even though this norm is weaker. This is stated in the next lemma, which can be seen as a consequence of Remark 4.4 and the fact taking marginals (in any product, and in any dimension) does not increase the cut norm.

Lemma 4.5. *If* $W \in L^1(\Omega^2)$*, then* $\|\overline{W}^{(1)}\|_{L^1(\Omega)}, \|\overline{W}^{(2)}\|_{L^1(\Omega)} \le \|W\|_{\square,2}$.

Proof. By symmetry, it suffices to consider $\overline{W}^{(1)}$. If $f \in L^\infty(\Omega)$, then

$$\int_\Omega \overline{W}^{(1)}(x) f(x)\,\mathrm{d}\mu(x) = \int_{\Omega^2} W(x,y) f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$$

and the result follows from (4.3), letting $g(y) = 1$ and taking the supremum over all f with $\|f\|_\infty \le 1$, using (4.7). (Or simply taking $f(x)$ equal to the sign of $\overline{W}^{(1)}(x)$.) □

Remark 4.6. It is a standard fact that the step functions are dense in $L^1(\Omega)$ and $L^1(\Omega^2)$. As a consequence, they are dense also in the cut norm in these spaces.

We finally note that the cut norm really is a norm if we, as usual, identify functions that are equal a.e.

Lemma 4.7. *If* $W \in L^1(\Omega^2)$*, then* $\|W\|_\square = 0 \iff W = 0$ *a.e.*

Proof. Suppose that $\|W\|_\square = 0$. Thus $\int_{S\times T} W(x,y) = 0$ for all subsets $S, T \subseteq \Omega$. It follows that $\int_{\Omega^2} W(x,y) f(x,y) = 0$ for every step function f on Ω^2.

Let g be any function on Ω^2 with $\|g\|_\infty \le 1$. Since step functions are dense in $L^1(\Omega^2)$, there exists a sequence g_n of step functions such that $g_n \to g$ in $L^1(\Omega^2)$; by considering a subsequence we may further assume that $g_n \to g$ a.e., and by truncating each g_n at ± 1 that $|g_n| \le 1$. By dominated convergence, $\int_{\Omega^2} W g_n \to \int_{\Omega^2} W g$, but each $\int_{\Omega^2} W g_n = 0$ since g_n is a step function; hence $\int_{\Omega^2} W g = 0$. If we choose $g := \operatorname{sgn}(W)$, this shows that $\int_{\Omega^2} |W| = 0$, and thus $W = 0$ a.e. □

5. Pull-backs and rearrangements

Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two probability spaces.

A mapping $\varphi : \Omega_1 \to \Omega_2$ is *measure-preserving* if it is measurable and $\mu_1(\varphi^{-1}(A)) = \mu_2(A)$ for every $A \in \mathcal{F}_2$ (i.e., for every measurable $A \subseteq \Omega_2$).

A mapping $\varphi : \Omega_1 \to \Omega_2$ is a *measure-preserving bijection* if φ is a bijection of Ω_1 onto Ω_2, and both φ and φ^{-1} are measure-preserving. (In other words, φ is an isomorphism between the measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ in category theory sense.) Equivalently, φ is a measure-preserving bijection if and only if it is a bijection that is measure-preserving, and further φ^{-1} is measurable (and then automatically measure-preserving). Note that if Ω_1 and Ω_2 are Borel spaces, then measurability of φ^{-1} is automatic by Theorem A.6, so it suffices to check that φ is a bijection and measure-preserving.

Note that if $\varphi : \Omega_1 \to \Omega_2$ is a measurable mapping, then $\varphi \otimes \varphi : \Omega_1^2 \to \Omega_2^2$ defined by $\varphi \otimes \varphi(x, y) = (\varphi(x), \varphi(y))$ is a measurable mapping, and if φ is measure-preserving or a measure-preserving bijection, then so is $\varphi \otimes \varphi$.

We define, for any functions f on Ω_2 and W on Ω_2^2, the *pull-backs*

$$f^\varphi(x) := f(\varphi(x)), \tag{5.1}$$

$$W^\varphi(x, y) := W(\varphi(x), \varphi(y)); \tag{5.2}$$

these are functions on Ω_1 and Ω_1^2, respectively.

We will only consider measure-preserving φ. In the special case that φ is a measure-preserving bijection, we say that f^φ and W^φ are *rearrangements* of f and W. (However, we will not assume that φ is injective or bijective unless we say so explicitly.) We further say that W' is an *a.e. rearrangement* of W if $W' = W^\varphi$ a.e. where W^φ is a rearrangement of W. Note that the relation "W_1 is a rearrangement of W_2" is symmetric and, moreover, an equivalence relation, and similarly for a.e. rearrangements.

Remark 5.1. Note that if W is symmetric, then W^φ is too by (5.2); recall that this is the case we really are interested in.

If we want to study general W, for example in connection with bipartite graphs as mentioned in Remark 2.6, it is often more natural to allow different maps φ_1 and φ_2 acting on the two coordinates.

Remark 5.2. Instead of measure-preserving bijections, it may be convenient to consider *measure-preserving almost bijections*, which are mappings φ that are measure-preserving bijections $\Omega_1 \setminus N_1 \to \Omega_2 \setminus N_2$ for some null sets N_1 and N_2. This makes essentially no difference below, and we leave the details to the reader. (See Theorem 8.6(vii) for a situation where almost bijections occur.)

Example 5.3. A kernel is a step kernel if and only if it is a pull-back W^φ of some kernel defined on a finite probability space. (The same holds for general functions $\Omega^2 \to \mathbb{R}$. Recall that step functions are the same as functions of finite type.)

Remark 5.4. We take here the point of view that (Ω_1, μ_1) and (Ω_2, μ_2) are given probability spaces, and we consider suitable maps between them. A closely related idea is to take a probability space (Ω_1, μ_1) and a measurable space Ω_2 (without any particular measure). A measurable map $\varphi : \Omega_1 \to \Omega_2$ then maps the measure μ_1 to a measure μ_1^φ on Ω_2 given by $\mu_1^\varphi(A) := \mu_1(\varphi^{-1}(A))$ for all $A \subseteq \Omega_2$. Note that μ_1^φ is the unique measure on Ω_2 that makes φ measure-preserving. This well-known construction (called *push-forward*) can be seen as a dual to the pull-back above; note that measures map forward, from Ω_1 to Ω_2, while functions map backward, from Ω_2 to Ω_1.

Note that, on the contrary, given a measurable map $\varphi : \Omega_1 \to \Omega_2$ between two measurable spaces, and a probability measure μ_2 on Ω_2, there is in general no measure μ_1 on Ω_1 which makes φ measure-preserving. This is a source of some of the technical difficulties in the theory.

It is easy to see that the norms defined above are invariant under rearrangements, and more generally under pull-backs by measure-preserving maps:

Lemma 5.5. *If φ is measure-preserving, then, taking the norms in the respective spaces, for any $f \in L^1(\Omega)$ and $W \in L^1(\Omega^2)$,*

$$\|f^\varphi\|_1 = \|f\|_1, \qquad \|W^\varphi\|_1 = \|W\|_1, \tag{5.3}$$

$$\|W^\varphi\|_\square = \|W\|_\square. \tag{5.4}$$

Proof. The equalities (5.3) are standard.

The cut norm equality (5.4) is obvious if φ is a measurable bijection. In general, it seems simplest to first assume that W is a step function, so that W is constant on each $A_i \times A_j$ for some partition $\Omega_2 = \bigcup_1^n A_i$, say $W = w_{ij}$ on $A_i \times A_j$. Then $A_i' := \varphi^{-1}(A_i)$ defines a partition of Ω_1, and W^φ is a step function constant on each $A_i' \times A_j'$, and equal to w_{ij} there.

Consider first $\|W\|_{\square,2}$. In the definition (4.3), we may replace f by its average on each A_i (i.e., by its conditional expectation given the partition) without changing the integral, and similarly for g. This shows that it is enough to consider f and g that are constant on each A_i, and we find

$$\|W\|_{\square,2} = \sup\Bigl|\sum_{i,j} w_{ij} a_i b_j \mu_2(A_i)\mu_2(A_j)\Bigr|, \tag{5.5}$$

taking the supremum over all real numbers a_i and b_j with $|a_i|, |b_j| \le 1$. Since $\mu_1(A_i') = \mu_2(A_i)$, the same argument shows that $\|W^\varphi\|_{\square,2}$ is given by the same quantity, and thus (5.4) holds in this case.

For $\|\cdot\|_{\square,1}$ we argue for step functions in exactly the same way, using Remark 4.1 and taking $a_i, b_j \in [0,1]$ in (5.5).

For a general W, let $\varepsilon > 0$ and let W_1 be a step function on Ω_2^2 such that $\|W - W_1\|_1 < \varepsilon$. Then

$$\|W - W_1\|_\square \le \|W - W_1\|_1 < \varepsilon.$$

Further, $\|W_1^\varphi\|_\square = \|W_1\|_\square$ by what we just have shown, and

$$\|W^\varphi - W_1^\varphi\|_\square \le \|W^\varphi - W_1^\varphi\|_1 = \|(W - W_1)^\varphi\|_1 = \|W - W_1\|_1 < \varepsilon.$$

The result $\|W^\varphi\|_\square = \|W\|_\square$ follows by some applications of the triangle inequality. $\square$

However, the distances $\|W_1 - W_2\|_1$ and $\|W_1 - W_2\|_\square$ between two kernels are, in general, not invariant under rearrangements of just one of the kernels, since, in general, $\|W - W^\varphi\| \neq 0$ for a kernel W on a space Ω and a measure-preserving bijection $\varphi : \Omega \to \Omega$. In the graph limit theory, we need a metric space where all rearrangements are equivalent (and thus have distance 0 to each other); we obtain this by taking the infimum over rearrangements.

Given two kernels W_1, W_2 on $[0,1]$, the *cut metric* of Borgs, Chayes, Lovász, Sós and Vesztergombi [14] may be defined by

$$\delta_\square(W_1, W_2) = \inf_\varphi \|W_1 - W_2^\varphi\|_\square, \tag{5.6}$$

taking the infimum over all measure-preserving bijection $\varphi : [0,1] \to [0,1]$; in other words, over all rearrangements W_2^φ of W_2. (If we wish to specify which version of the cut norm is involved, we write $\delta_{\square,1}$ or $\delta_{\square,2}$.) Borgs, Chayes, Lovász, Sós and Vesztergombi [14] showed that for kernels on $[0,1]$, there are several equivalent definitions of $\delta_\square$, see Theorem 6.9 below. For general probability spaces Ω, we have to use couplings between different kernels instead of rearrangements, see the following section; it then further is irrelevant whether the kernels are defined on the same probability space or not.

On the other hand, if we restrict ourselves to $[0,1]$, we can do with a special simple case of rearrangements. Following Borgs, Chayes, Lovász, Sós and Vesztergombi [14], we define an *n-step interval permutation* to be the map $\widetilde{\sigma}$ defined for a permutation σ of $\{1,\dots,n\}$ by taking the partition $(0,1] = \bigcup I_{in}$ with $I_{in} := ((i-1)/n, i/n]$ and mapping each I_{in} by translation to $I_{\sigma(i),n}$. (For completeness we also let $\widetilde{\sigma}(0) = 0$.) Evidently, $\widetilde{\sigma}$ is a measure-preserving bijection $[0,1] \to [0,1]$. We shall see in Theorem 6.9 below that it suffices to use such interval permutations in (5.6).

Example 5.6. To see one problem caused by using (5.6) for kernels on a general probability space, let Ω be the two-point space $\{1,2\}$, and let $\mu\{1\} = \frac{1}{2} - \varepsilon$, $\mu\{2\} = \frac{1}{2} + \varepsilon$, for some small $\varepsilon > 0$. Let $W_1(x,y) := \mathbf{1}\{x = y = 1\}$ and $W_2(x,y) := \mathbf{1}\{x = y = 2\}$. On this probability space there is no measure-preserving bijection except the identity, so (5.6) yields $\|W_1 - W_2\|_\square = (\frac{1}{2} + \varepsilon)^2 > \frac{1}{4}$, while the coupling definition (6.1) below yields $\delta_\square(W_1, W_2) = 2\varepsilon$.

6. Couplings and the cut metric

Given two probability spaces (Ω_1, μ_1), (Ω_2, μ_2), a *coupling* of these spaces is a pair of measure preserving maps $\varphi_i : \Omega \to \Omega_i$, $i = 1,2$, defined on a common (but arbitrary) probability space (Ω, μ).

Remark 6.1. Couplings are more common in the context of two random variables, say X_1 and X_2. These are often real-valued, but may more generally take values in any measurable spaces Ω_1 and Ω_2. A coupling of X_1 and X_2 then is a pair (X_1', X_2') of random variables defined on a common probability space such that $X_1' \overset{\mathrm{d}}{=} X_1$ and $X_2' \overset{\mathrm{d}}{=} X_2$. This is the same as a coupling of the two probability spaces (Ω_1, μ_1) and (Ω_2, μ_2) according to our definition above, where μ_1 is the distribution of X_1 and μ_2 the distribution of X_2.

The general definition of the cut metric, for kernels defined on arbitrary probability spaces (possibly different ones), is as follows.

Given kernels W_i on Ω_i, $i = 1,2$, or more generally any functions $W_i \in L^1(\Omega_i^2)$, we define the *cut metric* by

$$\delta_\square(W_1, W_2) = \inf \|W_1^{\varphi_1} - W_2^{\varphi_2}\|_\square, \tag{6.1}$$

where the infimum is taken over all couplings $(\varphi_1, \varphi_2) : \Omega \to (\Omega_1, \Omega_2)$ of Ω_1 and Ω_2 (with Ω arbitrary), and $W_i^{\varphi_i}$ is the pull-back defined in (5.2).

We similarly define

$$\delta_1(W_1, W_2) = \inf \|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{L^1(\Omega^2)}, \tag{6.2}$$

again taking the infimum over all couplings (φ_1, φ_2) of Ω_1 and Ω_2.

Remark 6.2. It is not obvious that the definition (6.1) agrees with (5.6) for kernels on $[0, 1]$, but, as shown in [14], this is the case; see Theorem 6.9 below. Note, in somewhat greater generality, that if W_1 and W_2 are kernels of probability spaces Ω_1 and Ω_2, and $\varphi : \Omega_1 \to \Omega_2$ is measure-preserving, then (ι, φ) is a coupling defined on Ω_1. (We let here and below ι denote the identity map in any space.) Hence, we always have $\delta_\square(W_1, W_2) \le \|W_1 - W_2^{\varphi}\|_\square$.

Note that $\delta_\square$ and δ_1 really are pseudometrics rather than metrics, since $\delta_\square(W_1, W_2) = 0$ and $\delta_1(W_1, W_2) = 0$ in many cases with $W_1 \ne W_2$, for example if $W_1 = W_2^{\varphi}$ for a measure preserving φ (use the coupling (ι, φ) in (6.1), see Remark 6.2). Nevertheless, it is customary to call this pseudometric the *cut metric*. We will return to the important problem of when $\delta_\square(W_1, W_2) = 0$ in Section 8.

It is obvious from the definition (6.1) that $\delta_\square$ and δ_1 are non-negative and symmetric, and $\delta_\square(W, W) = \delta_1(W, W) = 0$ for every W. It is less obvious that they really are subadditive, i.e., that the triangle inequality holds, so we give a detailed proof in Lemma 6.5 below.

A coupling (φ_1, φ_2) of two probability spaces (Ω_1, μ_1) and (Ω_2, μ_2), with φ_1, φ_2 defined on (Ω, μ), defines a map $\Phi := (\varphi_1, \varphi_2) : \Omega \to \Omega_1 \times \Omega_2$, which induces a unique measure $\widetilde{\mu}$ on $\Omega_1 \times \Omega_2$ such that $\Phi : (\Omega, \mu) \to (\Omega_1 \times \Omega_2, \widetilde{\mu})$ is measure-preserving (see Remark 5.4). Let $\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i$ be the projection; then $\varphi_i = \pi_i \circ \Phi$, $i = 1, 2$. Note that if $A \subseteq \Omega_i$, then

$$\widetilde{\mu}\left(\pi_i^{-1}(A)\right) = \mu\left(\Phi^{-1}\left(\pi_i^{-1}(A)\right)\right) = \mu\left(\varphi_i^{-1}(A)\right) = \mu_i(A),$$

since Φ and φ_i are measure-preserving; thus $\pi_i : (\Omega_1 \times \Omega_2, \widetilde{\mu}) \to (\Omega_i, \mu_i)$ is measure-preserving. Hence, (π_1, π_2) is a coupling of (Ω_1, μ_1) and (Ω_2, μ_2). If $W_i \in L^1(\Omega_i^2)$, then $W_i^{\varphi_i} = (W_i^{\pi_i})^{\Phi}$ and thus, using (5.4),

$$\|W_1^{\varphi_1} - W_2^{\varphi_2}\|_\square = \|(W_1^{\pi_1} - W_2^{\pi_2})^{\Phi}\|_\square = \|W_1^{\pi_1} - W_2^{\pi_2}\|_\square. \tag{6.3}$$

Consequently, in (6.1) it suffices to consider couplings of the type (π_1, π_2) defined on $(\Omega_1 \times \Omega_2, \widetilde{\mu})$, where $\widetilde{\mu}$ is a probability measure such that π_1 and π_2 are measure-preserving, i.e., such that $\widetilde{\mu}$ has the correct marginals μ_1 and μ_2.

Before proving the triangle inequality, we prove a technical lemma and a partial result.

Lemma 6.3. *Let Ω_1 and Ω_2 be probability spaces and let $W_1 \in L^1(\Omega_1^2)$ and $W_2 \in L^1(\Omega_2^2)$ be step functions with corresponding partitions $\Omega_1 = \bigcup_{i=1}^I A_i$ and $\Omega_2 = \bigcup_{j=1}^J B_j$. If (φ_1, φ_2) and (φ_1', φ_2') are two couplings of Ω_1 and Ω_2,*

defined on (Ω,μ) *and* (Ω',μ') *respectively, such that* $\mu\bigl(\varphi_1^{-1}(A_i)\cap\varphi_2^{-1}(B_j)\bigr) = \mu'\bigl(\varphi_1'^{-1}(A_i)\cap\varphi_2'^{-1}(B_j)\bigr)$ *for every* i *and* j*, then* $\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_{\square,\mu} = \|W_1^{\varphi_1'}-W_2^{\varphi_2'}\|_{\square,\mu'}$ *and, similarly,* $\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_{1,\mu} = \|W_1^{\varphi_1'}-W_2^{\varphi_2'}\|_{1,\mu'}$.

Proof. Recall that $\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_\square$ is given by (4.3), in case of $\|\cdot\|_{\square,1}$ further assuming $f,g\ge 0$, see Remark 4.1. Since $W_1^{\varphi_1}-W_2^{\varphi_2}$ is constant on each set $C_{ij} := \varphi_1^{-1}(A_i)\cap\varphi_2^{-1}(B_j)$, we may as in the proof of Lemma 5.5 average f and g in (4.3) over each such set, so it suffices to consider f and g that are constant on each set C_{ij}. Consequently, if $W_1 = u_{ik}$ on $A_i\times A_k$ and $W_2 = v_{jl}$ on $B_j\times B_l$, then

$$\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_{\square,\mu} = \max_{(f_{ij}),(g_{kl})}\biggl|\sum_{i,j,k,l}\mu(C_{ij})\mu(C_{kl})(u_{ik}-v_{jl})f_{ij}g_{kl}\biggr|, \tag{6.4}$$

taking the maximum over all arrays (f_{ij}) and (g_{kl}) of numbers in $[0,1]$ for $\|\cdot\|_{\square,1}$ and in $[-1,1]$ for $\|\cdot\|_{\square,2}$. This depends on the coupling only through the numbers $\mu(C_{ij})$, and the result follows.

For the L^1 norm we have immediately, with the same notation,

$$\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_{1,\mu} = \sum_{i,j,k,l}\mu(C_{ij})\mu(C_{kl})\bigl|u_{ik}-v_{jl}\bigr|,$$

and the result follows. □

Lemma 6.4. *Let* Ω_1 *and* Ω_2 *be probability spaces and* $W_1,W_1'\in L^1(\Omega_1^2)$ *and* $W_2\in L^1(\Omega_2^2)$*. Then* $\delta_\square(W_1,W_2)\le\delta_\square(W_1',W_2)+\|W_1-W_1'\|_\square$ *and, similarly,* $\delta_1(W_1,W_2)\le\delta_1(W_1',W_2)+\|W_1-W_1'\|_1$.

Proof. Let (φ_1,φ_2) be a coupling of Ω_1 and Ω_2. Then, using Lemma 5.5,

$$\begin{aligned}\delta_\square(W_1,W_2) &\le \|W_1^{\varphi_1}-W_2^{\varphi_2}\|_\square \le \|(W_1')^{\varphi_1}-W_2^{\varphi_2}\|_\square + \|W_1^{\varphi_1}-(W_1')^{\varphi_1}\|_\square\\ &= \|(W_1')^{\varphi_1}-W_2^{\varphi_2}\|_\square + \|(W_1-W_1')^{\varphi_1}\|_\square\\ &= \|(W_1')^{\varphi_1}-W_2^{\varphi_2}\|_\square + \|W_1-W_1'\|_\square.\end{aligned}$$

The result for $\delta_\square$ follows by taking the infimum over all couplings. The proof for δ_1 is the same. □

Lemma 6.5. *Let, for* $i=1,2,3$*,* Ω_i *be a probability space and* $W_i\in L^1(\Omega_i^2)$*. Then* $\delta_\square(W_1,W_3)\le\delta_\square(W_1,W_2)+\delta_\square(W_2,W_3)$ *and, similarly,* $\delta_1(W_1,W_3)\le\delta_1(W_1,W_2)+\delta_1(W_2,W_3)$*. Hence* $\delta_\square$ *and* δ_1 *are (pseudo)metrics.*

Proof. Roughly speaking, given a coupling of Ω_1 and Ω_2 and another coupling of Ω_2 and Ω_3, we want to couple the couplings so that we can compare pull-backs of W_1 and W_3. This simple idea, unfortunately, leads to technical difficulties in general, but it works easily if, for example, the spaces are finite. We use therefore an approximation argument with step functions which essentially reduces to the finite case.

Thus, suppose first that W_1,W_2,W_3 are step functions with corresponding partitions $\Omega_1=\bigcup_{i=1}^I A_i$, $\Omega_2=\bigcup_{j=1}^J B_j$, $\Omega_3=\bigcup_{k=1}^K C_k$, and assume for

simplicity that $\mu_1(A_i)$, $\mu_2(B_j)$ and $\mu_3(C_k)$ are non-zero for all i,j,k. (μ_ℓ denotes the measure on Ω_ℓ.)

We consider $\delta_\square$; the proof for δ_1 is the same. Let $\varepsilon > 0$. By the definition of $\delta_\square$ and the comments just made (see (6.3)), there exist measures μ' on $\Omega_1 \times \Omega_2$ and μ'' on $\Omega_2 \times \Omega_3$, with marginals μ_ℓ on Ω_ℓ, such that

$$\|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\mu'} < \delta_\square(W_1, W_2) + \varepsilon, \tag{6.5}$$

$$\|W_2^{\pi_2} - W_3^{\pi_3}\|_{\square,\mu''} < \delta_\square(W_2, W_3) + \varepsilon. \tag{6.6}$$

(We abuse notation a little by letting π_ℓ denote the projection onto Ω_ℓ from any product space.)

Define a measure μ on $\Omega_1 \times \Omega_2 \times \Omega_3$ by, for $E \subseteq \Omega_1 \times \Omega_2 \times \Omega_3$,

$$\mu(E) := \sum_{i,j,k} \frac{\mu'(A_i \times B_j)\mu''(B_j \times C_k)}{\mu_2(B_j)} \cdot \frac{\mu_1 \times \mu_2 \times \mu_3(E \cap (A_i \times B_j \times C_k))}{\mu_1(A_i)\mu_2(B_j)\mu_3(C_k)}.$$

We have $\mu_1(A_i) = \sum_j \mu'(A_i \times B_j)$, $\mu_2(B_j) = \sum_i \mu'(A_i \times B_j) = \sum_k \mu''(B_j \times C_k)$, and $\mu_3(C_k) = \sum_j \mu''(B_j \times C_k)$. It follows that the three mappings $\pi_\ell : (\Omega_1 \times \Omega_2 \times \Omega_3, \mu) \to (\Omega_\ell, \mu_\ell)$ are measure-preserving since, for example, if $F \subseteq \Omega_1$, then $\pi_1^{-1}(F) = F \times \Omega_2 \times \Omega_3$ and

$$\begin{aligned}
\mu\big(\pi_1^{-1}(F)\big) &= \mu(F \times \Omega_2 \times \Omega_3) \\
&= \sum_{i,j,k} \frac{\mu'(A_i \times B_j)\mu''(B_j \times C_k)}{\mu_2(B_j)} \cdot \frac{\mu_1 \times \mu_2 \times \mu_3\big((F \cap A_i) \times B_j \times C_k\big)}{\mu_1(A_i)\mu_2(B_j)\mu_3(C_k)} \\
&= \sum_{i,j,k} \frac{\mu'(A_i \times B_j)\mu''(B_j \times C_k)}{\mu_2(B_j)} \cdot \frac{\mu_1(F \cap A_i)}{\mu_1(A_i)} \\
&= \sum_{i,j} \mu'(A_i \times B_j)\frac{\mu_1(F \cap A_i)}{\mu_1(A_i)} = \sum_i \mu_1(F \cap A_i) = \mu_1(F).
\end{aligned}$$

In particular, μ is a probability measure.

The projections $\pi_{12} : \Omega_1 \times \Omega_2 \times \Omega_3 \to \Omega_1 \times \Omega_2$ and $\pi_{23} : \Omega_1 \times \Omega_2 \times \Omega_3 \to \Omega_2 \times \Omega_3$ map μ to measures $\widetilde{\mu}'$ on $\Omega_1 \times \Omega_2$ and $\widetilde{\mu}''$ on $\Omega_2 \times \Omega_3$. We have, for any i and j,

$$\begin{aligned}
\widetilde{\mu}'(A_i \times B_j) &= \mu(\pi_{12}^{-1}(A_i \times B_j)) = \mu(A_i \times B_j \times \Omega_3) \\
&= \sum_k \frac{\mu'(A_i \times B_j)\mu''(B_j \times C_k)}{\mu_2(B_j)} \cdot \frac{\mu_1 \times \mu_2 \times \mu_3(A_i \times B_j \times C_k)}{\mu_1(A_i)\mu_2(B_j)\mu_3(C_k)} \\
&= \sum_k \frac{\mu'(A_i \times B_j)\mu''(B_j \times C_k)}{\mu_2(B_j)} = \mu'(A_i \times B_j).
\end{aligned}$$

Hence, by Lemma 6.3,

$$\|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\Omega_1\times\Omega_2,\widetilde{\mu}'} = \|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\Omega_1\times\Omega_2,\mu'}. \tag{6.7}$$

Further, since $\pi_{12} : (\Omega_1 \times \Omega_2 \times \Omega_3, \mu) \to (\Omega_1 \times \Omega_2, \widetilde{\mu}')$ is measure-preserving, Lemma 5.5 implies that (recall our generic use of π_ℓ)

$$\|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\Omega_1\times\Omega_2\times\Omega_3,\mu} = \|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\Omega_1\times\Omega_2,\widetilde{\mu}'}. \tag{6.8}$$

Combining (6.5), (6.7) and (6.8), we find

$$\|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\mu} < \delta_\square(W_1, W_2) + \varepsilon. \tag{6.9}$$

Similarly,

$$\|W_2^{\pi_2} - W_3^{\pi_3}\|_{\square,\mu} < \delta_\square(W_2, W_3) + \varepsilon. \tag{6.10}$$

We have reached our goal of finding suitable couplings on the same space, viz. $(\Omega_1 \times \Omega_2 \times \Omega_3, \mu)$, and we can now use the triangle inequality for $\|\cdot\|_\square$ and deduce

$$\begin{aligned}\delta_\square(W_1, W_3) &\le \|W_1^{\pi_1} - W_3^{\pi_3}\|_{\square,\mu} \le \|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square,\mu} + \|W_2^{\pi_2} - W_3^{\pi_3}\|_{\square,\mu} \\ &< \delta_\square(W_1, W_2) + \delta_\square(W_2, W_3) + 2\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this implies the desired inequality $\delta_\square(W_1, W_3) \le \delta_\square(W_1, W_2) + \delta_\square(W_2, W_3)$ in the case of step functions.

In general, we approximate first each W_ℓ by a step function W'_ℓ such that $\|W_\ell - W'_\ell\|_{\square,\Omega_\ell} < \varepsilon$. (We may assume, as we did above, that all sets in the partition have positive measures by removing any null sets in them, redefining W'_ℓ on a null set.) The result for step functions together with several applications of Lemma 6.4 yield

$$\begin{aligned}\delta_\square(W_1, W_3) &\le \delta_\square(W'_1, W'_3) + 2\varepsilon \le \delta_\square(W'_1, W'_2) + \delta_\square(W'_2, W'_3) + 2\varepsilon \\ &\le \delta_\square(W_1, W_2) + \delta_\square(W_2, W_3) + 6\varepsilon.\end{aligned}$$

The result $\delta_\square(W_1, W_2) \le \delta_\square(W_1, W_2) + \delta_\square(W_2, W_3)$ follows. □

Corollary 6.6. *Let, for $i = 1, 2, 3$, Ω_i be a probability space and $W_i \in L^1(\Omega_i^2)$. If $\delta_\square(W_1, W_2) = 0$, then $\delta_\square(W_1, W_3) = \delta_\square(W_2, W_3)$. (The same result holds for δ_1.)* □

Consider the class $\mathcal{W}^* := \bigcup_\Omega \mathcal{W}(\Omega)$ of all graphons (on any probability space). We define a relation $\cong$ on this class (or on the even larger class $\bigcup_\Omega L^1(\Omega^2)$) by

$$W_1 \cong W_2 \quad \text{if} \quad \delta_\square(W_1, W_2) = 0. \tag{6.11}$$

Corollary 6.6 shows that this is an equivalence relation, and that $\delta_\square$ is a true metric on the quotient space $\widehat{\mathcal{W}} := \mathcal{W}^*/\cong$. We say that two graphons W_1, W_2 are *equivalent* if $W_1 \cong W_2$, i.e., if $\delta_\square(W_1, W_2) = 0$. (We will see in Theorem 8.10 below that $\delta_1(W_1, W_2) = 0$ defines the same equivalence relation.)

It is a central fact in the graph limit theory [14] that this quotient space $\widehat{\mathcal{W}} := \mathcal{W}^*/\cong$ is homeomorphic to (and thus can be identified with) the set of graph limits; moreover, the metric space $(\widehat{\mathcal{W}}, \delta_\square)$ is compact. (See also [24].) The compactness is closely related to Szemerédi's regularity lemma, see [49].

We will always regard $\widehat{\mathcal{W}}$ as a compact metric space equipped with the metric $\delta_\square$, except a few times when we explicitly use δ_1 instead. Note that δ_1 is a larger metric and thus gives a stronger topology. In particular, $(\widehat{\mathcal{W}}, \delta_1)$ is *not* compact. (See Pikhurko [55] and Section 10. For example, if G_n is the random graph $G(n, 1/2)$, then W_{G_n} a.s. converges to $1/2$ in $\delta_\square$, but not in δ_1.)

Example 6.7. If $W : \Omega^2 \to [0, 1]$ is any graphon (or kernel) on a probability space Ω, and $\varphi : \Omega' \to \Omega$ is a measure-preserving map, then, as remarked above, W is equivalent to its pull-back W^φ.

Example 6.8. Let G be a graph with vertex set $V = \{1, \dots, n\}$, and consider the graphons W_G^{V} and W_G^{I} defined in Example 2.7. Let $\varphi : (0, 1] \to V$ be the map $x \mapsto \lceil nx \rceil$. Then φ is measure-preserving and (2.2) defines W_G^{I} as the pull-back $(W_G^{\mathsf{V}})^\varphi$. Hence $W_G^{\mathsf{V}} \cong W_G^{\mathsf{I}}$.

We can now prove, following [14], that the definition (5.6) agrees with our definition (6.1) of the cut metric for $[0, 1]$, and more generally for any atomless Borel spaces. We include several related versions; note that (i) is (6.1) and (v) is (5.6).

Theorem 6.9. *Let W_1 and W_2 be two kernels defined on probability spaces (Ω_1, μ_1) and (Ω_2, μ_2), respectively. Then the following are the same, and thus all define $\delta_\square(W_1, W_2)$.*

(i) *For any Ω_1 and Ω_2,*
$$\inf_{\varphi_1, \varphi_2} \|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{\square, \Omega, \mu},$$
where the infimum is over all couplings (pairs of measure-preserving maps) $\varphi_1 : (\Omega, \mu) \to (\Omega_1, \mu_1)$ and $\varphi_2 : (\Omega, \mu) \to (\Omega_2, \mu_2)$.

(ii) *For any Ω_1 and Ω_2,*
$$\inf_{\mu} \|W_1^{\pi_1} - W_2^{\pi_2}\|_{\square, \Omega_1 \times \Omega_2, \mu},$$
where $\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i$ is the projection and the infimum is over all measures μ on $\Omega_1 \times \Omega_2$ having marginals μ_1 and μ_2.

(iii) *For any Ω_1 and Ω_2, for $\delta_{\square,2}$,*
$$\inf_{\mu} \sup_{\|f\|_\infty, \|g\|_\infty \le 1} \Big| \int_{(\Omega_1 \times \Omega_2)^2} \big(W_1(x_1, y_1) - W_2(x_2, y_2)\big) \cdot f(x_1, x_2) g(y_1, y_2) \, \mathrm{d}\mu(x_1, x_2) \, \mathrm{d}\mu(y_1, y_2) \Big|,$$
taking the infimum over all measures μ on $\Omega_1 \times \Omega_2$ having marginals μ_1 and μ_2; for $\delta_{\square,1}$ we further restrict to $f, g \ge 0$.

(iv) *Provided Ω_1 and Ω_2 are atomless Borel spaces,*
$$\inf_{\varphi} \|W_1 - W_2^\varphi\|_\square,$$
where the infimum is over all measure-preserving $\varphi : \Omega_1 \to \Omega_2$.

(v) *Provided* Ω_1 *and* Ω_2 *are atomless Borel spaces,*

$$\inf_{\varphi} \|W_1 - W_2^{\varphi}\|_{\square},$$

where the infimum is over all measure-preserving bijections $\varphi : \Omega_1 \to \Omega_2$*, i.e., over all rearrangements of* W_2 *defined on* Ω_1*.*

(vi) *Provided* $\Omega_1 = \Omega_2 = [0,1]$*,*

$$\inf_{\widetilde{\sigma}} \|W_1 - W_2^{\widetilde{\sigma}}\|_{\square},$$

where the infimum is over all interval permutations $\widetilde{\sigma} : [0,1] \to [0,1]$*, defined by permutations* σ *of* $\{1,\dots,n\}$ *with* n *arbitrary.*

Proof. (i) $\iff$ (ii). We have shown this in (6.3) and the accompanying argument.

(ii) $\iff$ (iii). Directly from the definition (4.3) (using Remark 4.1 for $\delta_{\square,1}$), writing $x = (x_1, x_2)$ and $y = (y_1, y_2)$. (The expression in (iii) is just writing the definition explicitly in this case.)

For (iv) and (v), we first note that by Theorem A.7, Ω_1 and Ω_2 are isomorphic to $[0,1]$ (equipped with Lebesgue measure), i.e., there are measure-preserving bijections $\psi_j : [0,1] \to \Omega_j$. It is evident that we may use these maps to transfer the problem to the pull-backs $W_1^{\psi_1}$ and $W_2^{\psi_2}$ on $[0,1]$. In other words, we may in (iv) and (v) assume that $\Omega_1 = \Omega_2 = [0,1]$.

In this case, denote the quantities in (i)–(vi) by $\delta_{(\mathrm{i})}, \dots, \delta_{(\mathrm{vi})}$. We have $\delta_{(\mathrm{i})} \le \delta_{(\mathrm{iv})} \le \delta_{(\mathrm{v})} \le \delta_{(\mathrm{vi})}$, since we take infima over smaller and smaller sets of maps. Further, we have shown that $\delta_{(\mathrm{i})} = \delta_{(\mathrm{ii})} = \delta_{(\mathrm{iii})}$. To complete the proof, it thus suffices to show that $\delta_{(\mathrm{vi})} \le \delta_{(\mathrm{ii})}$.

Let $\varepsilon > 0$ and let I_{iN} denote the interval $((i-1)/N, i/N]$, for $1 \le i \le N$. The set of step functions $W : [0,1]^2 \to \mathbb{R}$ that correspond to partitions of $[0,1]$ (or rather $(0,1]$, but the difference does not matter here) into m equally long intervals $I_{1m}, \dots, I_{mm}$ for $m = 1, 2, \dots$, is a dense subset of $L^1([0,1]^2)$. Hence we may choose $m > 0$ and two such step functions W_1' and W_2' so that $\|W_i - W_i'\|_1 < \varepsilon$, $i = 1, 2$. (We may first obtain such W_i' with different m_1 and m_2, but we may then replace both by $m := m_1 m_2$.) By Lemma 6.4 and its proof, which applies to all the versions $\delta_{(\mathrm{i})}, \dots, \delta_{(\mathrm{vi})}$, we have

$$\delta_*(W_1, W_2) - 2\varepsilon \le \delta_*(W_1', W_2') \le \delta_*(W_1, W_2) + 2\varepsilon \tag{6.12}$$

for every $* = (\mathrm{i}), \dots, (\mathrm{vi})$.

Choose a probability measure μ on $\Omega_1 \times \Omega_2 = [0,1]^2$ such that $\|W_1'^{\pi_1} - W_2'^{\pi_2}\|_{\square} < \delta_{(\mathrm{ii})}(W_1', W_2') + \varepsilon$. We may evaluate this cut norm by (6.4) (replacing W_i by W_i') and as asserted in Lemma 6.3, the cut norm depends only on the numbers $\mu(C_{ij})$, where now $C_{ij} := \pi_1^{-1}(I_{im}) \cap \pi_2^{-1}(I_{jm}) = I_{im} \times I_{jm}$, so we may assume that the coupling measure μ on $[0,1]^2$ on each square C_{ij} equals a constant factor λ_{ij} times the Lebesgue measure. (Hence, $\mu(C_{ij}) = \lambda_{ij}/m^2$.) We adjust these factors so that every $\mu(C_{ij})$ is rational; we may do this so that the marginals still are correct, i.e., for every i

and j,

$$\sum_l \mu(C_{il}) = \sum_l \mu(C_{lj}) = \frac{1}{m}. \tag{6.13}$$

The adjustment will change cut norm in (6.4) by an arbitrary small amount, so we can do this and still have $\|W_1'^{\pi_1} - W_2'^{\pi_2}\|_\square < \delta_{(\mathrm{ii})}(W_1', W_2') + \varepsilon$.

We now have $\mu(C_{ij}) = a_{ij}/N$ for some integers N and a_{ij}, $1 \le i, j \le m$. Let $b := N/m$. By (6.13), for every i and j,

$$\sum_j a_{ij} = \sum_i a_{ij} = \frac{N}{m} = b. \tag{6.14}$$

Hence, b is an integer, and thus every interval I_{im} is a union $\bigcup_{k=b(i-1)+1}^{bi} I_{kN}$ of b intervals I_{kN} of length $1/N$. By (6.14), we may construct a permutation σ of $\{1, \dots, N\}$ such that σ maps exactly a_{ij} of the indices $k \in [b(i-1)+1, bi]$ into $[b(j-1)+1, bj]$, for all i, j. Hence, $\lambda(I_{im} \cap \widetilde{\sigma}^{-1}(I_{jm})) = a_{ij}/N = \mu(C_{ij})$. Thus, Lemma 6.3 applies to the couplings (π_1, π_2) and $(\iota, \widetilde{\sigma})$ (defined on $[0,1]$); hence,

$$\delta_{(\mathrm{vi})}(W_1', W_2') \le \|W_1' - W_2'^{\widetilde{\sigma}}\|_\square = \|W_1'^{\pi_1} - W_2'^{\pi_2}\|_\square < \delta_{(\mathrm{ii})}(W_1', W_2') + \varepsilon.$$

Finally, (6.12) yields $\delta_{(\mathrm{vi})}(W_1, W_2) < \delta_{(\mathrm{ii})}(W_1, W_2) + 5\varepsilon$, and the result follows since ε is arbitrary. $\square$

Remark 6.10. On spaces with atoms, the quantities $\delta_{(\mathrm{iv})}$ and $\delta_{(\mathrm{v})}$ defined in (iv) and (v) are in general different from $\delta_\square$, see Example 5.6. (In this case, they are larger than $\delta_\square$, see Remark 6.2.) Furthermore, for two general probability spaces Ω_1 and Ω_2, it is possible that there are no measure-preserving maps $\Omega_1 \to \Omega_2$ at all, in which case the definitions in (iv) and (v) are not appropriate; and even if we may interpret $\delta_{(\mathrm{iv})}$ as a default value 1 (for graphons; for general kernels we would have to use ∞), in such cases, $\delta_{(\mathrm{iv})}$ is not even symmetric in general. (For an example, modify Example 5.6 by replacing W_2 by a pull-back defined on $[0,1]$; then there are measure-preserving maps $\Omega_2 \to \Omega_1$ but not conversely. We have $\delta_{(\mathrm{iv})}(W_2, W_1) = 2\varepsilon < \delta_{(\mathrm{iv})}(W_1, W_2) = 1$.)

Remark 6.11. In (iv), it suffices that Ω_1 and Ω_2 are Borel spaces such that Ω_1 is atomless. To see this, replace W_2 by its pull-back W_2^π defined on the atomless Borel space $\widetilde{\Omega}_2 := \Omega_2 \times [0,1]$, where π is the projection onto Ω_2.

Remark 6.12. (iv) and (v) hold also for atomless Lebesgue spaces, since then, for $\ell = 1, 2$, $\Omega_\ell = (\Omega_\ell, \mathcal{F}_\ell, \mu_\ell)$ is the completion of some Borel space $\Omega_\ell^0 = (\Omega_\ell, \mathcal{F}_\ell^0, \mu_\ell)$, and we may replace W_ℓ by a kernel W_ℓ^0 on Ω_ℓ^0 with $W_\ell = W_\ell^0$ a.e., cf. Remark 2.5; note that every measure-preserving map $\varphi : \Omega_1^0 \to \Omega_2^0$ also is measure-preserving $\Omega_1 \to \Omega_2$.

Remark 6.13. An obvious analogue of Theorem 6.9 holds for δ_1. (In (iii), the integral is $\int_{(\Omega_1 \times \Omega_2)^2} \bigl|W_1(x_1, y_1) - W_2(x_2, y_2)\bigr| \,\mathrm{d}\mu(x_1, x_2) \,\mathrm{d}\mu(y_1, y_2)$, and there are no f and g.) See Pikhurko [55] for some parts and related results.

Remark 6.14. In probabilistic notation, see Remark 6.1, (iii) can be written as

$$\inf_{(X_1',X_2')} \sup_{\|f\|_\infty,\|g\|_\infty\le 1} \Bigl|\mathbb{E}\Bigl(\bigl(W_1(X_1',X_1'')-W_2(X_2',X_2'')\bigr)f(X_1',X_2')g(X_1'',X_2'')\Bigr)\Bigr|,$$

where the infimum is taken over all couplings (X_1',X_2') of two random variables X_1 and X_2 such that X_ℓ is Ω_ℓ-valued and has distribution μ_ℓ, and (X_1'',X_2'') is an independent copy of (X_1',X_2').

Corollary 6.15. *Let Ω be an atomless Borel spaces, e.g. $[0,1]$, and let W be a graphon on Ω. Then the equivalence class of all graphons on Ω equivalent to W equals the closure of the orbit of W under measure-preserving bijections (or maps); i.e.,*

$$\{W'\in\mathcal{W}(\Omega): W'\cong W\} = \overline{\{W^\varphi:\varphi\in S_{\mathrm{mp}}\}} = \overline{\{W^\varphi:\varphi\in S_{\mathrm{mpb}}\}},$$

where S_{mp} is the set of all measure-preserving $\varphi:\Omega\to\Omega$, and S_{mpb} is the subset of all measure-preserving bijections. The closure may here be taken either for the cut norm or for the L^1 norm.

Proof. For the closure in cut norm, this follows from Theorem 6.9(iv) and (v). By Remark 6.13, the same holds for the closure in L^1 norm and the equivalence class $\{W'\in\mathcal{W}(\Omega):\delta_1(W',W)=0\}$. However, by Theorem 8.10 below, δ_1 and $\delta_\square$ define the same equivalence classes. $\square$

Remark 6.10 shows that the (equivalent) definitions in Theorem 6.9(i)–(iii) are the only ones useful for general probability spaces. Another advantage of them is that, as shown by Bollobás and Riordan [12], the infima are attained, at least for Borel spaces. (This is not true in general for the versions in (iv)–(vi), not even in the special case when the infimum is 0, see Example 8.1 below.)

Theorem 6.16. *Let W_1 and W_2 be two kernels defined on Borel probability spaces (Ω_1,μ_1) and (Ω_2,μ_2), respectively. Then the infima in Theorem 6.9*(i)–(iii) *are attained. In other words, there exists a probability measure μ on $\Omega_1\times\Omega_2$ with marginals μ_1 and μ_2, and thus a corresponding coupling (π_1,π_2), such that $\delta_\square(W_1,W_2)=\|W_1^{\pi_1}-W_2^{\pi_2}\|_{\square,\Omega_1\times\Omega_2,\mu}$.*

Proof. By Theorem A.4 and Remark A.5, every Borel measurable space is either countable or isomorphic to the Cantor cube $\mathcal{C}:=\{0,1\}^\infty$. Hence, we may without loss of generality assume that each of the two spaces Ω_ℓ (where, as in the rest of the proof, $\ell=1,2$) is either a finite set, the countable set $\{0\}\cup\{1/n:n\in\mathbb{N}\}$ or $\mathcal{C}$, equipped with some probability measure μ_ℓ. Note that in every case Ω_ℓ is a compact metric space.

For $\ell_1,\ell_2\in\{1,2\}$, Let $\mathcal{A}(\Omega_{\ell_1}\times\Omega_{\ell_2})$ be the set of all step functions on $\Omega_{\ell_1}\times\Omega_{\ell_2}$ corresponding to partitions $\Omega_{\ell_m}=\bigcup_i A_{im}$ where every part A_{im} is clopen (closed and open) in Ω_{ℓ_m}, $m=1,2$. (We extend here the definition of step functions on $\Omega\times\Omega$ to products of two different spaces in the natural way.) For the spaces we consider, $\mathcal{A}(\Omega_{\ell_1}\times\Omega_{\ell_2})$ is dense in $L^1(\Omega_{\ell_1}\times\Omega_{\ell_2},\mu)$

for any probability measure μ on the product. (This is the reason why we replaced $[0,1]$ by the totally disconnected space $\mathcal{C}$. It is possible to use $[0,1]$ instead, with minor modifications, see [12].)

Denote the integral in Theorem 6.9(iii) by $\Phi(W_1, W_2, f, g, \mu)$. By Theorem 6.9, there exist probability measures ν_n on $\Omega_1 \times \Omega_2$ such that

$$\sup_{\|f\|_\infty, \|g\|_\infty \le 1} |\Phi(W_1, W_2, f, g, \nu_n)| < \delta_\square(W_1, W_2) + 1/n. \tag{6.15}$$

(For $\delta_{\square,1}$, we tacitly assume that $f, g \ge 0$.)

Since Ω_1 and Ω_2 are compact metric spaces, $\Omega_1 \times \Omega_2$ is too. Hence, the set of probability measures on $\Omega_1 \times \Omega_2$ is compact and metrizable (see [6]), so there exists a subsequence of (ν_n) that converges (in the usual weak topology) to some probability measure ν on $\Omega_1 \times \Omega_2$. We consider in the sequel this subsequence only.

Let $\varepsilon > 0$. By the remarks above, we may find $W'_\ell \in \mathcal{A}(\Omega_\ell^2)$ with $\|W_\ell - W'_\ell\|_{L^1(\Omega_\ell \times \Omega_\ell)} < \varepsilon$, and hence, assuming $\|f\|_\infty, \|g\|_\infty \le 1$,

$$\begin{aligned} |\Phi(W_1, W_2, f, g, \nu)| &\le |\Phi(W'_1, W'_2, f, g, \nu)| + \|W_1 - W'_1\|_\square + \|W_2 - W'_2\|_\square \\ &\le |\Phi(W'_1, W'_2, f, g, \nu)| + 2\varepsilon \end{aligned} \tag{6.16}$$

and similarly, for every n and every f, g with $\|f\|_\infty, \|g\|_\infty \le 1$,

$$|\Phi(W'_1, W'_2, f, g, \nu_n)| \le |\Phi(W_1, W_2, f, g, \nu_n)| + 2\varepsilon. \tag{6.17}$$

Since W'_1 and W'_2 are step functions, they are bounded, so there exists some M with $\|W'_\ell\|_\infty \le M$. For any f and g with $\|f\|_\infty, \|g\|_\infty \le 1$, we may similarly find f' and g' in $\mathcal{A}(\Omega_1 \times \Omega_2)$, with $\|f\|_\infty, \|g\|_\infty \le 1$, such that $\|f - f'\|_{L^1(\nu)}, \|g - g'\|_{L^1(\nu)} \le \varepsilon/M$. It follows that

$$|\Phi(W'_1, W'_2, f, g, \nu)| \le |\Phi(W'_1, W'_2, f', g', \nu)| + 4\varepsilon. \tag{6.18}$$

Since W'_1, W'_2, f', g' all are step functions in the sets $\mathcal{A}$, the integral $\Phi(W'_1, W'_2, f', g', \mu)$ can be written as a linear combination of integrals

$$\int_{A_i \times B_j \times A_k \times B_m} \mathrm{d}\mu(x_1, x_2)\, \mathrm{d}\mu(y_1, y_2) = \mu(A_i \times B_j)\mu(A_k \times B_m),$$

where further the sets A_i, B_j, A_k, B_m are clopen. Hence each term, and thus $\Phi(W'_1, W'_2, f', g', \mu)$, is a continuous functional of μ; consequently,

$$\Phi(W'_1, W'_2, f', g', \nu_n) \to \Phi(W'_1, W'_2, f', g', \nu). \tag{6.19}$$

By (6.17) and (6.15),

$$|\Phi(W'_1, W'_2, f', g', \nu_n)| < \delta_\square(W_1, W_2) + 1/n + 2\varepsilon, \tag{6.20}$$

and thus (6.19) yields

$$|\Phi(W'_1, W'_2, f', g', \nu)| \le \delta_\square(W_1, W_2) + 2\varepsilon \tag{6.21}$$

and, by (6.16) and (6.18),

$$|\Phi(W_1, W_2, f, g, \nu)| \le |\Phi(W'_1, W'_2, f', g', \nu)| + 6\varepsilon \le \delta_\square(W_1, W_2) + 8\varepsilon. \tag{6.22}$$

Since ε is arbitrary, we thus obtain $|\Phi(W_1, W_2, f, g, \nu)| \le \delta_\square(W_1, W_2)$, and

$$\|W_1 - W_2\|_{\square,\nu} = \sup_{\|f\|_\infty, \|g\|_\infty \le 1} |\Phi(W_1, W_2, f, g, \nu)| \le \delta_\square(W_1, W_2), \tag{6.23}$$

which shows that equality is attained in Theorem 6.9(ii)–(iii) by ν, and in Theorem 6.9(i) by the coupling (π_1, π_2) defined on $(\Omega_1 \times \Omega_2, \nu)$. $\square$

The assumption that the spaces are Borel (or Lebesgue, see Remark 6.12) really is essential here, even when the infimum $\delta_\square(W_1, W_2) = 0$; in Example 8.13 we will see an example of two equivalent kernels such that none of the infima in Theorem 6.9 is attained.

7. Representation on $[0, 1]$

As said in the introduction, many papers consider only kernels or graphons on $[0, 1] = ([0, 1], \lambda)$. This is justified by the fact that every kernel [graphon] is equivalent to such a kernel [graphon]. (See [38] for a generalization.)

Theorem 7.1. *Every kernel [graphon] on a probability space $(\Omega, \mathcal{F}, \mu)$ is equivalent to a kernel [graphon] on $([0, 1], \lambda)$.*

Corollary 7.2. *The quotient space $\widehat{\mathcal{W}} := \bigcup_\Omega \mathcal{W}(\Omega)/\cong$, which as said above can be identified with the space of graph limits, can as well be defined $\widehat{\mathcal{W}} := \mathcal{W}([0, 1])/\cong$.*

Before proving Theorem 7.1, we prove a partial result.

Lemma 7.3. *Every kernel [graphon] on a probability space $(\Omega, \mathcal{F}, \mu)$ is a pull-back of a kernel [graphon] on some Borel probability space.*

Proof. Let $W : \Omega^2 \to [0, \infty)$ be a kernel. Since W is measurable, each set $E_r := \{(x, y) : W(x, y) < r\}$, where $r \in \mathbb{R}$, belongs to $\mathcal{F} \times \mathcal{F}$, and it follows that there exists a countable subset $\mathcal{A}_r \subseteq \mathcal{F}$ such that $E_r \in \mathcal{F}(\mathcal{A}_r) \times \mathcal{F}(\mathcal{A}_r)$, where $\mathcal{F}(\mathcal{A}_r)$ is the σ-field generated by $\mathcal{A}_r$. Hence, if $\mathcal{F}_0$ is the σ-field generated by the countable set $\mathcal{A} := \bigcup_{r\in\mathbb{Q}} \mathcal{A}_r$, then $\mathcal{F}_0 \subseteq \mathcal{F}$ and W is $\mathcal{F}_0 \times \mathcal{F}_0$-measurable.

List the elements of $\mathcal{A}$ as $\{A_1, A_2, \dots\}$. (If $\mathcal{A}$ is finite, we for convenience repeat some element.) Let $\mathcal{C} := \{0, 1\}^\infty$ be the Cantor cube (see Remark A.5) and define a map $\varphi : \Omega \to \mathcal{C} := \{0, 1\}^\infty$ by $\varphi(x) = (\mathbf{1}\{x \in A_i\})_{i=1}^\infty$. Let ν be the probability measure on $\mathcal{C}$ that makes $\varphi : \Omega \to \mathcal{C}$ measure-preserving, see Remark 5.4.

The σ-field on Ω generated by φ equals $\mathcal{F}_0$, and thus the σ-field on $\Omega \times \Omega$ generated by $(\varphi, \varphi) : \Omega^2 \to \mathcal{C}^2$ equals $\mathcal{F}_0 \times \mathcal{F}_0$. Since W is measurable for this σ-field, W equals $V \circ (\varphi, \varphi) = V^\varphi$ for some measurable $V : \mathcal{C}^2 \to [0, \infty)$. Since W is symmetric, we may here replace $V(x, y)$ by $\frac{1}{2}\big(V(x, y) + V(y, x)\big)$ and thus assume that also V is symmetric. Hence, V is a kernel on $\mathcal{C}$ and $W = V^\varphi$. If W is a graphon, we may assume that $V : \Omega^2 \to [0, 1]$, and thus V too is a graphon. This proves the result with the Borel probability space $(\mathcal{C}, \nu)$. $\square$

Proof of Theorem 7.1. Let W be a kernel on some probability space Ω. By Lemma 7.3, $W \cong V$ for some kernel V on a Borel probability space (Ω_1, ν). (With $\Omega_1 = \mathcal{C}$ in the proof above.) If ν is atomless, the result follows by Theorem A.7. In general, let $\Omega_2 := \Omega \times [0,1]$, with product measure ν_2. The projection $\pi : \Omega_2 \to \Omega_1$ is measure-preserving, so $V \cong V_2 := V^\pi$. Moreover, (Ω_2, ν_2) is an atomless Borel probability space, so by Theorem A.7 there exists a measure-preserving bijection $\psi : [0,1] \to \Omega_2$. Hence $U := V_2^\psi$ is a kernel [graphon] on $[0,1]$ and $U \cong V_2 \cong V \cong W$. $\square$

Remark 7.4. If μ is atomless, we may by Lemma A.1 find an increasing family of sets $B_r \subseteq \Omega$, $r \in [0,1]$, such that $\mu(B_r) = r$. In the construction in the proof of Lemma 7.3, we may add each B_r with rational r to the family $\mathcal{A}$. Then the measure ν on $\mathcal{C}$ is atomless, because if x were an atom, then $E := \varphi^{-1}\{x\}$ would be a subset of Ω with $\mu(E) > 0$ such that for each rational r, either $E \subseteq B_r$ or $E \cap B_r = \emptyset$, but this leads to a contradiction as in the proof of Lemma A.3. Consequently, we then can use Theorem A.7 directly to find a measure-preserving bijection $\psi : [0,1] \to \mathcal{C}$, and a kernel $U := V^\psi$ on $[0,1]$ such that $W = V^\varphi = U^{\psi^{-1}\circ\varphi}$. Consequently, every kernel on an atomless probability space (Ω, μ) is a pull-back of a kernel on $[0,1]$, which combines and improves Lemma 7.3 and Theorem 7.1 in this case. (Conversely, by Lemma A.3, no kernel on a space (Ω, μ) with atoms is a pull-back of a kernel on $[0,1]$.)

8. Equivalence

We have seen that if W_1 and W_2 are two kernels on some probability spaces Ω_1 and Ω_2, and $W_1 = W_2^\varphi$ (or just $W_1 = W_2^\varphi$ a.e.) for some measure-preserving $\varphi : \Omega_1 \to \Omega_2$, then $W_1 \cong W_2$. The converse does not hold, as shown by the following standard examples [13].

Example 8.1. Let $\varphi : [0,1] \to [0,1]$ be given by $\varphi(x) = 2x \mod 1$. Take $W_1(x,y) = xy$, and $W_2 := W_1^\varphi$. Then W_1 and W_2 are graphons on $[0,1]$, and $\delta_\square(W_1, W_2) = 0$. However, there is no measure-preserving $\psi : [0,1] \to [0,1]$ such that $W_1 = W_2^\psi$ a.e., and as a consequence, the infima in Theorem 6.9(iv)–(vi) are not attained. (See Lemma 4.7.) In fact, if such a ψ existed, then $W_1 = (W_1^\varphi)^\psi = W_1^{\varphi\circ\psi}$ a.e., which implies (e.g. by considering the marginal $\int_0^1 W(x,y)\,\mathrm{d}y = x/2$) that $\varphi(\psi(x)) = x$ a.e., and thus $\psi(x) \in \{x/2, x/2 + 1/2\}$ a.e. However, if $E := \psi^{-1}([0,1/2])$, it follows that for any a and b with $0 < a < b < 1$, $E \cap [a,b] = \psi^{-1}([a/2, b/2])$, so $\lambda(E \cap [a,b])/(b-a) = 1/2$. In particular, for every $x \in (0,1)$, the density $\lim_{\varepsilon\to 0} \lambda(E \cap (x-\varepsilon, x+\varepsilon))/2\varepsilon = 1/2$. On the other hand, by the Lebesgue density theorem, this density is 1 for a.e. $x \in E$ and 0 for a.e. $x \notin E$, a contradiction.

Example 8.2. More generally, let $\varphi : [0,1] \to [0,1]$ be given by $\varphi_n(x) = nx \mod 1$, and define $W_n := W^{\varphi_n}$ with the same W as in Example 8.1. If

$W_n = W_m^\psi$ a.e., then $m\psi(x) \equiv nx \mod 1$ a.e. Let $E := \psi^{-1}(0, 1/m)$. Then, for $0 < a < b < 1/m$, $\psi^{-1}([a,b]) = \bigcup_{j=0}^{n-1} E \cap ([ma/n, mb/n] + j/n)$ (a.e.) and thus, if ψ is measure-preserving,

$$b - a = \lambda\left(\psi^{-1}([a,b])\right) = \sum_{j=0}^{n-1} \lambda\left(E \cap \left[\frac{ma+j}{n}, \frac{mb+j}{n}\right]\right).$$

Divide by $b-a$, take $a = x - \varepsilon$ and $b = x + \varepsilon$, and let $\varepsilon \to 0$. The Lebesgue differentiation theorem implies that for a.e. $x \in (0, 1/m)$,

$$1 = \frac{m}{n} \sum_{j=0}^{n-1} \mathbf{1}\Big\{\frac{mx+j}{n} \in E\Big\}.$$

Since the sum is an integer for each x, this implies that n is a multiple of m. Conversely, if $n = m\ell$ for an integer ℓ, then $\varphi_n = \varphi_m \circ \varphi_\ell$, and thus $W_n = (W^{\varphi_m})^{\varphi_\ell} = W_m^{\varphi_\ell}$. Consequently, all W_n are equivalent (being pull-backs of W_1), and there exists a measure-preserving $\psi : [0,1] \to [0,1]$ such that $W_n = W_m^\psi$ a.e. if and only if n is a multiple of m. In particular, W_2 and W_3 are equivalent, but neither of them is a pull-back of the other.

However, equivalence is characterized by sequences of pull-backs. We begin with a simple result.

Theorem 8.3. *Let W' and W'' be kernels defined of probability spaces Ω' and Ω''. Then $W' \cong W''$, i.e. $\delta_\square(W', W'') = 0$, if and only if there exists a finite sequence of kernels W_i defined on probability spaces Ω_i, $i = 0, \dots, n$, with $W_0 = W'$ and $W_n = W''$, such that for each $i \ge 1$, either $W_{i-1} = W_i^{\varphi_i}$ a.e. for some measure-preserving $\varphi_i : \Omega_{i-1} \to \Omega_i$, or $W_i = W_{i-1}^{\psi_i}$ a.e. for some measure-preserving $\psi_i : \Omega_i \to \Omega_{i-1}$.*

Proof. Suppose that $W' \cong W''$. We show that we can construct such a sequence with $n = 4$. We thus take $W_0 := W'$ and $W_4 := W''$. By Lemma 7.3, we can find W_1 and W_3 on Borel probability spaces Ω_1 and Ω_3 such that $W_0 = W_1^{\varphi_1}$ and $W_4 = W_3^{\psi_4}$ for some measure-preserving φ_1 and ψ_4. Then $W_1 \cong W_0 \cong W_4 \cong W_3$, so $\delta_\square(W_1, W_3) = 0$. By Theorem 6.16, there exists a probability measure μ on $\Omega_1 \times \Omega_3$ such that $\|W_1^{\pi_1} - W_3^{\pi_3}\|_{\square, \Omega_1 \times \Omega_3, \mu} = 0$, where π_i is the projection onto Ω_i. Thus, by Lemma 4.7, $W_1^{\pi_1} = W_3^{\pi_3}$ a.e. Hence, we can take $\Omega_2 := (\Omega_1 \times \Omega_3, \mu)$, $\psi_2 := \pi_2$, $\varphi_3 := \pi_3$ and $W_2 := W_1^{\psi_2} = W_1^{\pi_2}$.

The converse is obvious by Example 6.7 and Corollary 6.6. $\square$

Example 8.2 shows that we cannot in general do with a single pull-back in Theorem 8.3. However, we can always do with a chain of length 2 in Theorem 8.3. In fact, Borgs, Chayes and Lovász [13] proved the following, more precise and much more difficult, result. (We will not use this theorem later; the simpler Theorem 8.3 is sufficient for our applications.)

Theorem 8.4. *Let W_1 and W_2 be kernels defined of probability spaces Ω_1 and Ω_2. Then $W_1 \cong W_2$, i.e. $\delta_\square(W_1, W_2) = 0$, if and only if there exists a kernel W on some probability space Ω and measure-preserving maps $\varphi_j : \Omega_j \to \Omega$ such that $W_j = W^{\varphi_j}$ a.e., $j = 1, 2$.*

We can always take Ω to be a Borel space. If Ω_1 and Ω_2 are atomless, we may take $\Omega = [0, 1]$.

Proof. It suffices to prove the theorem for graphons W_1 and W_2; the general case follows easily by considering the transformations $W_1/(1 + W_1)$ and $W_2/(1 + W_2)$.

We give a proof in Section 9. (Except for the final statement, which is shown below.) See also Borgs, Chayes and Lovász [13] for the long and technical original proof. In their formulation, the space Ω is constructed as a Lebesgue space, and the maps φ_j are only assumed to be measurable from the completions $(\Omega_j, \widehat{\mathcal{F}}_j, \mu_j)$ of Ω_j to Ω. However, this is easily seen to be equivalent: If Ω is such a Lebesgue space, then $\Omega = (\Omega, \mathcal{F}, \mu)$ is the completion of some Borel space $\Omega_0 = (\Omega, \mathcal{F}_0, \mu)$. We may replace W by an a.e. equal kernel that is $\mathcal{F}_0 \times \mathcal{F}_0$-measurable, i.e., a kernel on the Borel space Ω_0. Further, since every Borel measurable space is isomorphic to a Borel subset of $[0, 1]$, see Theorem A.4, the map $\varphi_j : \Omega_j \to \Omega_0$ which is $\widehat{\mathcal{F}}_j$-measurable, is a.e. equal to an $\mathcal{F}_j$-measurable map φ'_j. Replacing Ω by Ω_0 and φ_j by φ'_j, we obtain the result as stated above, with Ω Borel.

For the final statement, suppose that Ω_1 and Ω_2 are atomless, and let W, φ_j and Ω be as in the first part of the theorem, with Ω Borel. Suppose that Ω has atoms, i.e., points $a \in \Omega$ with $\mu\{a\} > 0$. Replace each such point a by a set I_a which is a copy of the interval $[0, \mu\{a\}]$ (with Borel σ-field and Lebesgue measure), and let Ω' be the resulting Borel probability space. There is an obvious map $\pi : \Omega' \to \Omega$, mapping each I_a to a and being the identity elsewhere, and we let $W' := W^\pi$. For each atom a, and $j = 1, 2$, let $A_{aj} := \varphi_j^{-1}(a) \subseteq \Omega_j$. Then A_{aj} is an atomless measurable space, and by Lemma A.2 (and scaling), there is a measure-preserving map $A_{aj} \to I_a$. Combining these maps and the original φ_j, we find a measure-preserving map $\varphi'_j : \Omega_j \to \Omega'$ such that $\varphi_j = \pi \circ \varphi'_j$, and thus $W_j = W^{\varphi_j} = (W')^{\varphi'_j}$ a.e. Finally, Ω' is an atomless Borel probability space, and may thus be replaced by $[0, 1]$ by Theorem A.7. $\square$

Remark 8.5. With a Lebesgue space Ω, it is both natural and necessary to consider maps $\Omega_j \to \Omega$ that are measurable with respect to the completion of Ω_j, as done in [13]. For example, if $\Omega_1 = \Omega_2 = [0, 1]$ with the Borel σ-field and $W_1(x, y) = W_2(x, y) = xy$, we can take $\Omega = [0, 1]$ and $\varphi_j = \iota$, but if we equip Ω with the Lebesgue σ-field, then φ_j is not measurable $\Omega_j \to \Omega$ (and cannot be modified on a null set to become measurable). This is just a trivial technicality that is no real problem, and as seen in the proof above, it can be avoided by using Borel spaces.

Theorem 8.4 says that a pair of equivalent graphons always are pull-backs of a single graphon. We may also try to go in the opposite direction and try to find a common pull-back of two equivalent graphons. As shown by Borgs, Chayes and Lovász [13], this is not always possible, see Example 8.13 below, but it is possible for graphons defined on Borel or Lebesgue spaces. We state this, in several versions, in the next theorem, together with conditions under which W_1 is a pull-back or rearrangement of W_2. (Recall that Example 8.2 shows that this does not hold in general, not even for a nice Borel space like $[0,1]$.)

If W is kernel defined on a probability space Ω, we say following [13] that $x_1, x_2 \in \Omega$ are *twins* (for W) if $W(x_1, y) = W(x_2, y)$ for a.e. $y \in \Omega$. We say that W is *almost twinfree* if there exists a null set $N \subset \Omega$ such that there are no twins $x_1, x_2 \in \Omega \setminus N$ with $x_1 \neq x_2$.

Various parts of the following theorem are given, at least for the standard case of graphons on $\Omega_1 = \Omega_2 = [0,1]$, in Diaconis and Janson [24] (as a consequence of Hoover's equivalence theorem for representations of exchangeable arrays [43, Theorem 7.28]), Bollobás and Riordan [12], and Borgs, Chayes and Lovász [13]. A similar theorem in the related case of partial orders is given in [40].

Theorem 8.6. *Let W_1 and W_2 be kernels defined on Borel probability spaces (Ω_1, μ_1) and (Ω_2, μ_2). Then the following are equivalent.*

(i) $W_1 \cong W_2$.

(ii) *There exist a coupling (φ_1, φ_2), i.e., two measure preserving maps $\varphi_j : \Omega \to \Omega_j$, $j = 1, 2$, for some probability space Ω, such that $W_1^{\varphi_1} = W_2^{\varphi_2}$ a.e., i.e., $W_1\bigl(\varphi_1(x), \varphi_1(y)\bigr) = W_2\bigl(\varphi_2(x), \varphi_2(y)\bigr)$ a.e.*

(iii) *There exist measure preserving maps $\varphi_j : [0,1] \to \Omega_j$, $j = 1, 2$, such that $W_1^{\varphi_1} = W_2^{\varphi_2}$ a.e., i.e., $W_1\bigl(\varphi_1(x), \varphi_1(y)\bigr) = W_2\bigl(\varphi_2(x), \varphi_2(y)\bigr)$ a.e. on $[0,1]^2$.*

(iv) *There exists a measure-preserving map $\psi : \Omega_1 \times [0,1] \to \Omega_2$ such that $W_1^{\pi_1} = W_2^{\psi}$ a.e., where $\pi_1 : \Omega_1 \times [0,1] \to \Omega_1$ is the projection, i.e., $W_1(x, y) = W_2\bigl(\psi(x, t_1), \psi(y, t_2)\bigr)$ for a.e. $x, y \in \Omega_1$ and $t_1, t_2 \in [0,1]$.*

(v) *There exists a probability measure μ on $\Omega_1 \times \Omega_2$ with marginals μ_1 and μ_2 such that $W_1^{\pi_1} = W_2^{\pi_2}$ a.e. on $(\Omega_1 \times \Omega_2)^2$, i.e., $W_1(x_1, y_1) = W_2(x_2, y_2)$ for μ-a.e. $(x_1, x_2), (y_1, y_2) \in \Omega_1 \times \Omega_2$.*

If W_2 is almost twinfree, then these are also equivalent to:

(vi) *There exists a measure preserving map $\varphi : \Omega_1 \to \Omega_2$ such that $W_1 = W_2^{\varphi}$ a.e., i.e. $W_1(x, y) = W_2\bigl(\varphi(x), \varphi(y)\bigr)$ a.e. on Ω_1^2.*

If both W_1 and W_2 are almost twinfree, then these are also equivalent to:

(vii) *There exists a measure preserving map $\varphi : \Omega_1 \to \Omega_2$ such that φ is a bimeasurable bijection of $\Omega_1 \setminus N_1$ onto $\Omega_2 \setminus N_2$ for some null sets $N_1 \subset \Omega_1$ and $N_2 \subset \Omega_2$, and $W_1 = W_2^{\varphi}$ a.e., i.e. $W_1(x, y) =$*

$W_2\bigl(\varphi(x),\varphi(y)\bigr)$ a.e. on Ω_1^2. If further (Ω_2,μ_2) is atomless, for example if $\Omega_2=[0,1]$, then we may take $N_1=N_2=\emptyset$, so W_1 is a rearrangement of W_2 and vice versa.

The same results hold if Ω_1 and Ω_2 are Lebesgue spaces, provided in (iii) *$[0,1]$ is equipped with the Lebesgue σ-field, and in* (iv) *$\Omega_1\times[0,1]$ has the completed σ-field.*

Proof. We assume that Ω_1 and Ω_2 are Borel spaces. The Lebesgue space case follows immediately from this case by replacing W_1 and W_2 by (a.e. equal) Borel kernels, see Remark 2.5.

We may also, when convenient, assume that W_1 and W_2 are graphons by using again the transformations $W_1/(1+W_1)$ and $W_2/(1+W_2)$.

First note that any of (iii)–(vii) is a special case of (ii), and that (ii) implies $W_1\cong W_1^{\varphi_1}\cong W_2^{\varphi_2}\cong W_2$; thus any of (ii)–(vii) implies (i). We turn to the converses.

(i) $\implies$ (ii),(v): Assume $W_1\cong W_2$, i.e., $\delta_\square(W_1,W_2)=0$. First, by Theorem 6.16, there exists a coupling (φ_1,φ_2) such that $\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_\square=\delta_\square(W_1,W_2)=0$, and thus, by Lemma 4.7, $W_1^{\varphi_1}=W_2^{\varphi_2}$ a.e. Consequently, (ii) holds. Moreover, by the same theorem and Theorem 6.9(ii), we may take this coupling (φ_1,φ_2) as the projections (π_1,π_2) for a suitable measure μ on $\Omega_1\times\Omega_2$, which shows (v).

(v) $\implies$ (iii): Since $(\Omega_1\times\Omega_2,\mu)$ is a Borel probability space, Theorem A.9 shows that there exists a measure-preserving map $\psi:[0,1]\to\Omega_1\times\Omega_2$, and then $(\pi_1\circ\psi,\pi_2\circ\psi)$ is a coupling defined on $\Omega=[0,1]$, which shows (iii). (Alternatively, (i) $\implies$ (iii) follows also easily by Theorem A.9 from the special case $\Omega_1=\Omega_2=[0,1]$ showed in [24].)

(i) $\implies$ (iv): By Theorem A.9, there exist measure preserving maps $\gamma_j:[0,1]\to\Omega_j$, $j=1,2$. Then $W_1^{\gamma_1}$ and $W_2^{\gamma_2}$ are kernels on $[0,1]$, and $W_1^{\gamma_1}\cong W_1\cong W_2\cong W_2^{\gamma_2}$. The equivalence (i) $\iff$ (iv) was shown (for graphons, which suffices as remarked above) in [24] in the special case $\Omega_1=\Omega_2=[0,1]$, based on [43, Theorem 7.28], and thus (iv) holds for $W_1^{\gamma_1}$ and $W_2^{\gamma_2}$. In other words, there exists a measure preserving function $h:[0,1]^2\to[0,1]$ such that $W_1^{\gamma_1}(x,y)=W_2^{\gamma_2}\bigl(h(x,z_1),h(y,z_2)\bigr)$ for a.e. $x,y,z_1,z_2\in[0,1]$. By Lemma 8.9 below (applied to (Ω_1,μ_1) and γ_1), there exists a measure preserving map $\alpha:\Omega_1\times[0,1]\to[0,1]$ such that $\gamma_1(\alpha(s,u))=s$ a.e. Hence, for a.e. $x,y\in\Omega_1$ and $u_1,u_2,z_1,z_2\in[0,1]$,

$$\begin{aligned}W_1(x,y)&=W_1\bigl(\gamma_1\circ\alpha(x,u_1),\gamma_1\circ\alpha(y,u_2)\bigr)=W_1^{\gamma_1}\bigl(\alpha(x,u_1),\alpha(y,u_2)\bigr)\\&=W_2^{\gamma_2}\bigl(h(\alpha(x,u_1),z_1),h(\alpha(y,u_2),z_2)\bigr)\\&=W_2\bigl(\gamma_2\circ h(\alpha(x,u_1),z_1),\gamma_2\circ h(\alpha(y,u_2),z_2)\bigr).\end{aligned}$$

Finally, let $\beta=(\beta_1,\beta_2)$ be a measure preserving map $[0,1]\to[0,1]^2$, and define $\psi(x,t):=\gamma_2\circ h\bigl(\alpha(x,\beta_1(t)),\beta_2(t)\bigr)$.

(iv) $\implies$ (vi): Since, for a.e. x,y,t_1,t_2,t_1',

$$W_2\bigl(\psi(x,t_1),\psi(y,t_2)\bigr)=W_1(x,y)=W_2\bigl(\psi(x,t_1'),\psi(y,t_2)\bigr)$$

and ψ is measure preserving, it follows that for a.e. x, t_1, t_1', $\psi(x,t_1)$ and $\psi(x,t_1')$ are twins for W_2. If W_2 is almost twin-free, with exceptional null set N, then further $\psi(x,t_1), \psi(x,t_1') \notin N$ for a.e. x, t_1, t_1', since ψ is measure preserving, and consequently $\psi(x,t_1) = \psi(x,t_1')$ for a.e. x, t_1, t_1'. It follows that we can choose a fixed t_1' (almost every choice will do) such that $\psi(x,t) = \psi(x,t_1')$ for a.e. x, t. Define $\varphi(x) := \psi(x,t_1')$. Then $\psi(x,t) = \varphi(x)$ for a.e. x, t, which in particular implies that φ is measure preserving, and (iv) yields $W_1(x,y) = W_2\bigl(\varphi(x), \varphi(y)\bigr)$ a.e.

(vi) $\implies$ (vii): Let $N' \subset \Omega_1$ be a null set such that if $x \notin N'$, then $W_1(x,y) = W_2(\varphi(x), \varphi(y))$ for a.e. $y \in \Omega_1$. If $x, x' \in \Omega_1 \setminus N'$ and $\varphi(x) = \varphi(x')$, then x and x' are twins for W_1. Consequently, if W_1 is almost twinfree with exceptional null set N'', then φ is injective on $\Omega_1 \setminus N_1$ with $N_1 := N' \cup N''$. Since $\Omega_1 \setminus N_1$ and Ω_2 are Borel spaces, Theorem A.6 shows that the injective map $\varphi : \Omega_1 \setminus N_1 \to \Omega_2$ has measurable range and is a bimeasurable bijection $\varphi : \Omega_1 \setminus N_1 \to \Omega_2 \setminus N_2$ for some measurable set $N_2 \subset \Omega_2$. Since φ is measure preserving, $\mu_2(N_2) = 0$.

If Ω_2 has no atoms, we may take an uncountable null set $N_2' \subset \Omega_2 \setminus N_2$. Let $N_1' := \varphi^{-1}(N_2')$. Then $N_1 \cup N_1'$ and $N_2 \cup N_2'$ are uncountable Borel spaces so they are isomorphic and there is a bimeasurable bijection $\eta : N_1 \cup N_1' \to N_2 \cup N_2'$. Redefine φ on $N_1 \cup N_1'$ so that $\varphi = \eta$ there; then φ becomes a bijection $\Omega_1 \to \Omega_2$. $\square$

Remark 8.7. A probabilistic reformulation of (ii), along the lines of Remark 6.1, is that there exists a coupling (X,Y) of random variables with the distributions μ_1 on Ω_1 and μ_2 on Ω_2, such that if (X',Y') is an independent copy of (X,Y), then $W_1(X,X') = W_2(Y,Y')$ a.s. Similarly, (v) says that there exists a distribution (i.e., probability measure) μ on $\Omega_1 \times \Omega_2$ with marginals μ_1 and μ_2 such that if (X,Y) and (X',Y') are independent with the same distribution μ, then $W_1(X,X') = W_2(Y,Y')$ a.s. [13].

Remark 8.8. In (iv), the seemingly superfluous variables t_1 and t_2 act as extra randomization; (iv) thus yields a kind of "randomized pull-back" using a "randomized measure-preserving map" ψ, even when no suitable measure-preserving map as in (vi) exists. It is an instructive exercise to see how this works for Example 8.1; we leave this to the reader.

The proof above uses the following consequence of the transfer theorem [42, Theorem 6.10].

Lemma 8.9. *Suppose that (Ω,μ) is a Borel probability space and that $\gamma : [0,1] \to \Omega$ is a measure preserving function. Then there exists a measure preserving function $\alpha : \Omega \times [0,1] \to [0,1]$ such that $\gamma\bigl(\alpha(s,y)\bigr) = s$ for $\mu \times \lambda$-a.e. $(s,y) \in \Omega \times [0,1]$.*

Proof. Let $\eta : [0,1] \to [0,1]$ and $\tilde\xi : \Omega \to \Omega$ be the identity maps $\eta(x) = x$, $\tilde\xi(s) = s$, and let $\xi = \gamma : [0,1] \to \Omega$. Then (ξ,η) is a pair of random variables, defined on the probability space $([0,1],\lambda)$, with values in Ω and

$[0,1]$, respectively; further, $\tilde\xi$ is a random variable defined on (Ω,μ) with $\tilde\xi \overset{\mathrm{d}}{=} \xi$. By the transfer theorem [42, Theorem 6.10], there exists a measurable function $\alpha : \Omega\times[0,1] \to [0,1]$ such that if $\tilde\eta(s,y) := \alpha(\tilde\xi(s),y) = \alpha(s,y)$, then $(\tilde\xi,\tilde\eta)$ is a pair of random variables defined on $\Omega\times[0,1]$ with $(\tilde\xi,\tilde\eta) \overset{\mathrm{d}}{=} (\xi,\eta)$. Since $\xi = \gamma(\eta)$, this implies $\tilde\xi = \gamma(\tilde\eta)$ a.e., and thus $s = \tilde\xi(s) = \gamma\bigl(\alpha(s,y)\bigr)$ a.e. $\square$

There are several other, quite different, characterizations of equivalence. We give several important conditions from [13], [14] and [24] that use the homomorphism densities $t(F,W)$ and the random graphs $G(n,W)$ defined in Appendix C and Appendix D.

Theorem 8.10. *Let W and W' be two graphons (possibly defined on different probability spaces). Then the following are equivalent:*

(i) $W \cong W'$.
(ii) $\delta_\square(W,W') = 0$.
(iii) $\delta_1(W,W') = 0$.
(iv) $t(F,W) = t(F,W')$ *for every simple graph* F.
(v) $t(F,W) = t(F,W')$ *for every loopless multigraph* F.
(vi) *The random graphs* $G(n,W)$ *and* $G(n,W')$ *have the same distribution for every finite* n.
(vii) *The infinite random graphs* $G(\infty,W)$ *and* $G(\infty,W')$ *have the same distribution.*

Proof. (i) $\iff$ (ii): This is just our definition of $\cong$.

(ii) $\implies$ (iii): If $\delta_\square(W,W') = 0$, let $W_0,\dots,W_n$ be a chain of graphons as in Theorem 8.3 (or Theorem 8.4). We have $\delta_1(W_i,W_i^\varphi) = 0$ for any pull-back of a graphon W_i, and thus $\delta_1(W_{i-1},W_i) = 0$ for every $i \ge 1$. Hence $\delta_1(W,W') = 0$ by the triangle inequality Lemma 6.5.

(iii) $\implies$ (ii): Trivial.

(ii) $\implies$ (v): This is immediate from (C.1) for a pull-back, and the general case follows again by Theorem 8.3.

(v) $\implies$ (iv): Trivial.

(iv) $\iff$ (vi): The distribution of $G(n,W)$ is determined by the family $\{t(F,W) : |F| \le n\}$ of homomorphism densities for all (simple) graphs F with $|F| \le n$, and conversely, cf. Remark D.1.

(vi) $\iff$ (vii): The distribution of $G(\infty,W)$ is determined by the family of distributions of the restrictions $G(\infty,W)|_{[n]}$ to the first n vertices, for $n \ge 1$, and conversely. However, $G(\infty,W)|_{[n]} = G(n,W)$. See [24] for details.

(iv) $\implies$ (ii): See [14] or, for a different proof, [13]. (This is highly non-trivial.) Alternatively, (vii) $\implies$ (ii) follows from [43, Theorem 7.28], see [24, Proof of Theorem 7.1]. $\square$

Remark 8.11. One of the central results in [14] is that, for graphons $W_1,W_2,\dots$ and W, $\delta_\square(W_n,W) \to 0$ if and only if $t(F,W_n) \to t(F,W)$ for

every (simple) graph F. (Taking $W_n = W_{G_n}$ for a sequence of graphs with $|G_n| \to \infty$, this says in particular that $G_n \to W \iff t(F, G_n) \to t(F, W)$ for every graph F, see Appendix B.) As pointed out in [12], this equivalence is equivalent to the corresponding equivalence (ii) $\iff$ (iv) in Theorem 8.10.

One way to see this is to define a new semimetric on the class $\mathcal{W}^*$ of graphons by

$$\delta_{\mathtt{t}}(W, W') := \sum_{n=1}^{\infty} 2^{-n} |t(F_n, W) - t(F_n, W')|,$$

where $F_1, F_2, \dots$ is some (arbitrary but fixed) enumeration of all unlabelled (simple) graphons. By Theorem 8.10, $\delta_{\mathtt{t}}(W, W') = 0 \iff W \cong W' \iff \delta_\square(W, W') = 0$, so $\delta_{\mathtt{t}}$ is, just as $\delta_\square$, a metric on the quotient space $\widehat{\mathcal{W}}$. Moreover, the easy result Lemma C.2 that each $W \mapsto t(F_n, W)$ is continuous for $\delta_\square$ implies that $\delta_{\mathtt{t}}$ is continuous on $(\widehat{\mathcal{W}}, \delta_\square)$, so the topology on $\widehat{\mathcal{W}}$ defined by $\delta_{\mathtt{t}}$ is weaker than the topology defined by $\delta_\square$. (Equivalently, the identity map $(\widehat{\mathcal{W}}, \delta_\square) \to (\widehat{\mathcal{W}}, \delta_{\mathtt{t}})$ is continuous.) However, since $(\widehat{\mathcal{W}}, \delta_\square)$ is compact, this implies that topologies are the same, i.e., that the metrics $\delta_\square$ and $\delta_{\mathtt{t}}$ are equivalent on $\widehat{\mathcal{W}}$, which is the result we want. (This argument in [12] is essentially the same, but stated somewhat differently. Another equivalent version is to consider the mapping $\widehat{\mathcal{W}} \to [0,1]^\infty$ given by $W \mapsto (t(F_n, W))_{n=1}^\infty$, see [24]; this map is continuous and, by Theorem 8.10, injective, so again by compactness it is a homeomorphism onto some subset.) Note the importance of the compactness of $\widehat{\mathcal{W}}$ in these arguments.

The *distribution* of a kernel W defined on a probability space (Ω, μ) is the distribution of W regarded as a random variable defined on Ω^2, i.e., the push-forward of μ^2 by W, or equivalently the probability measure on $\mathbb{R}$ that makes $W : \Omega^2 \to \mathbb{R}$ measure-preserving, see Remark 5.4.

Corollary 8.12. *If W_1 and W_2 are two equivalent graphons, defined on two probability spaces Ω_1 and Ω_2, then W_1 and W_2 have the same distributions. In particular, $\int_{\Omega_1^2} W_1^k = \int_{\Omega_2^2} W_2^k$ for every $k \geq 1$.*

Proof. The conclusion obviously holds if W_1 is a.e. equal to a pull-back W_2^φ of W_2, or conversely. The general case follows by Theorem 8.3 (or Theorem 8.4) and transitivity. Alternatively, we may use Theorem 8.10 and observe that $\int W_\ell^k = t(M_k, W_\ell)$ if M_k is the multigraph consisting of k parallel edges, see Example C.1. $\square$

Note that, for any $k > 1$, $W \to \int_{\Omega^2} W^k$ is *not* continuous in the cut norm, see Example C.3.

Finally we give, as promised above, the counter-example by Borgs, Chayes and Lovász [13], showing that the condition that the spaces are Borel (or Lebesgue) is needed in Theorem 8.6.

Example 8.13. Let $A \subseteq [0,1]$ be a non-measurable set such that the outer measure $\lambda^*(A) = 1$ and the inner measure $\lambda_*(A) = 0$. (Equivalently, every

measurable set contained in A or in its complement has measure 0.) Let $\mathcal{L}_A := \{B \cap A : B \in \mathcal{L}\}$, the *trace* of the Lebesgue σ-field on A. Then the outer Lebesgue measure λ^* is a probability measure on $(A, \mathcal{L}_A)$, and the injection $\iota : A \to [0,1]$ is measure-preserving. (See e.g. [19, Exercises 1.5.8 and 1.5.11].)

Let $W(x,y) := xy$. W is a graphon on $[0,1]$, so its pull-back $W_1 := W^\iota$, which equals the restriction of W to $A \times A$, is a graphon on $\Omega_1 := (A, \mathcal{L}_A, \lambda^*)$, and $W_1 \cong W$. The complement $A^\mathsf{c} := [0,1] \setminus A$ satisfies the same condition as A, so we may also define $\Omega_2 := (A^\mathsf{c}, \mathcal{L}_{A^\mathsf{c}}, \lambda^*)$, and let $W_2 \cong W$ be the restriction of W to $A^\mathsf{c} \times A^\mathsf{c}$.

Then $W_1 \cong W \cong W_2$, so $W_1 \cong W_2$. However, suppose that (φ_1, φ_2) is a coupling of Ω_1 and Ω_2, defined on some space Ω, such that $W_1^{\varphi_1} = W_2^{\varphi_2}$ a.e. Then the marginal $\overline{W_1^{\varphi_1}}^{(1)}$ equals the pull-back $(\overline{W_1}^{(1)})^{\varphi_1}$ of the marginal $\overline{W}_1^{(1)} : \Omega_1 \to [0,1]$, but the marginal of W_1 is

$$\overline{W}_1^{(1)}(x) = \overline{W}^{(1)}(x) := \int_0^1 W(x,y)\,\mathrm{d}y = x/2;$$

hence $\overline{W_1^{\varphi_1}}^{(1)}(x) = \varphi_1(x)/2$ for all $x \in \Omega$. Similarly, $\overline{W_2^{\varphi_2}}^{(1)}(x) = \varphi_2(x)/2$ for all $x \in \Omega$. Our assumption $W_1^{\varphi_1} = W_2^{\varphi_2}$ a.e. implies that the marginals are equal a.e., and thus $\varphi_1(x)/2 = \varphi_2(x)/2$ a.e.; consequently, $\varphi_1(x) = \varphi_2(x)$ for a.e. $x \in \Omega$. This is a contradiction since for every x, $\varphi_1(x) \in A$ while $\varphi_2(x) \in A^\mathsf{c}$.

Consequently, for every coupling (φ_1, φ_2) of Ω_1 and Ω_2 we have $W_1^{\varphi_1} \neq W_2^{\varphi_2}$ on a set of positive measure and thus $\|W_1^{\varphi_1} - W_2^{\varphi_2}\|_\square > 0$ by Lemma 4.7. Hence, the infima in Theorem 6.9(i)–(iii) are not attained, and none of Theorem 8.6(ii)–(vii) holds, although (i) does.

9. Pure graphons and a canonical version of a graphon

We present here a way to select an essentially unique, canonical choice of graphon among all equivalent graphons corresponding to a graph limit; more precisely we construct a graphon that is determined uniquely up to a.e. rearrangements. This construction is based on Lovász and Szegedy [51], although formulated somewhat differently. This will also lead to a new proof of Theorem 8.4, a proof which we find simpler than the original one.

For convenience, we consider only graphons, although the construction extends to general kernels with very few modifications.

Let W be a graphon on a probability space $(\Omega, \mathcal{F}, \mu)$. For each $x \in \Omega$, the section W_x is defined by

$$W_x(y) := W(x,y), \qquad y \in \Omega. \tag{9.1}$$

Thus W_x is a measurable function $\Omega \to [0,1]$, and in particular $W_x \in L^1(\Omega, \mathcal{F}, \mu)$.

Let $\psi_W : \Omega \to L^1(\Omega, \mathcal{F}, \mu)$ be the map defined by $\psi_W(x) := W_x$. By a standard monotone class argument (using e.g. the version of the monotone

class theorem in [37, Theorem A.1]), see also [25, Lemma III.11.16], $\psi_W : \Omega \to L^1(\Omega, \mathcal{F}, \mu)$ is measurable.

Let μ_W be the push-forward μ^{ψ_W} of μ by ψ_W, i.e., the probability measure on $L^1(\Omega, \mathcal{F}, \mu)$ that makes $\psi_W : (\Omega, \mu) \to (L^1(\Omega, \mathcal{F}, \mu), \mu_W)$ measure-preserving, see Remark 5.4; explicitly,

$$\mu_W(A) = \mu(\psi_W^{-1}(A)), \qquad A \subseteq L^1(\Omega, \mathcal{F}, \mu). \tag{9.2}$$

Further, let Ω_W be the support of μ_W, i.e.,

$$\Omega_W := \{f \in L^1(\Omega, \mathcal{F}, \mu) : \mu_W(U) > 0 \text{ for every neighbourhood } U \text{ of } f\}. \tag{9.3}$$

Ω_W is a subset of $L^1(\Omega, \mathcal{F}, \mu)$, and we equip it with the induced metric, given by the norm in $L^1(\Omega, \mathcal{F}, \mu)$, and the Borel σ-field generated by the metric topology.

Theorem 9.1. (i) *Ω_W is a complete separable metric space, μ_W is a probability measure on Ω_W and $\psi_W(x) \in \Omega_W$ for μ-a.e. $x \in \Omega$. We can thus regard ψ_W as a mapping $\Omega \to \Omega_W$ (defined a.e.); then $\psi_W : (\Omega, \mu) \to (\Omega_W, \mu_W)$ is measure-preserving.*

(ii) *μ_W has full support on Ω_W, i.e., if $U \subseteq \Omega_W$ is open and non-empty, then $\mu_W(U) > 0$.*

(iii) *The range of ψ_W is dense in Ω_W. More precisely, $\psi_W(\Omega) \cap \Omega_W = \{W_x : x \in \Omega\} \cap \Omega_W$ is a dense subset of Ω_W.*

(iv) $\Omega_W \subseteq \{f \in L^1(\Omega, \mathcal{F}, \mu) : 0 \le f \le 1 \text{ a.e.}\}$.

(v) *There exists a graphon $\widehat{W}$ on (Ω_W, μ_W) such that the pull-back $\widehat{W}^{\psi_W} = W$ a.e.; this graphon $\widehat{W}$ is unique up to a.e. equality. In particular, $W \cong \widehat{W}$.*

Proof. Recall first that $L^1(\Omega, \mathcal{F}, \mu)$ is a Banach space, and thus a complete metric space. In many cases, $L^1(\Omega, \mathcal{F}, \mu)$ is separable (for example if $\Omega = [0, 1]$ or another Borel space); however, there are cases when $L^1(\Omega, \mathcal{F}, \mu)$ is non-separable, see Appendix G, and in order to be completely general, we have to include some technical details on separability below; these can be ignored when Ω is a Borel space (and at the first reading).

Recall also that if B is a Banach space, then $L^1(\Omega, \mathcal{F}, \mu; B)$ is the Banach space of functions $f : \Omega \to B$ that are measurable and essentially separably valued, i.e., there exists a separable subspace $B_1 \subseteq B$ such that $f(x) \in B_1$ for a.e. x, and further $\int_\Omega \|f\|_B \, \mathrm{d}\mu < \infty$, see e.g. [25, Chapter III, in particular Section III.6] or the summary in [37, Appendix C]. (Note that [25] uses a definition of measurability which implicitly includes essential separability, see [25, Lemma III.6.9].)

Returning to our setting, we remarked above that $\psi_W : \Omega \to L^1(\Omega, \mathcal{F}, \mu)$ is measurable; furthermore it is bounded since $\|\psi(x)\|_{L^1} = \|W_x\|_{L^1} \le 1$, and a monotone class argument (again using e.g. [37, Theorem A.1]) shows that ψ_W is separably valued; thus $\psi_W \in L^1\bigl(\Omega, \mathcal{F}, \mu; L^1(\Omega, \mathcal{F}, \mu)\bigr)$. In fact, see [25, III.11.16–17], the mapping $W \mapsto \psi_W$ extends to $L^1(\Omega \times \Omega, \mu \times \mu)$,

and more generally to $L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2)$ for a product of any two probability spaces (or, more generally, σ-finite measure spaces), and this yields an isometric isomorphism

$$L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2) \cong L^1\bigl(\Omega_1, \mathcal{F}_1, \mu_1; L^1(\Omega_2, \mathcal{F}_2, \mu_2)\bigr). \tag{9.4}$$

As just said, ψ_W is separably valued, i.e., there exists a separable subspace $B_1 \subseteq B := L^1(\Omega, \mathcal{F}, \mu)$ such that $\psi_W(x) \in B_1$ for all $x \in \Omega$. We may replace B_1 by $\overline{B}_1$, and we may thus assume that B_1 is a closed subspace of B, and thus a Banach space. Then $\mu_W(B \setminus B_1) = \mu(\psi_W^{-1}(B \setminus B_1)) = \mu(\emptyset) = 0$, and it follows from (9.3) that $\Omega_W = \operatorname{supp}(\mu_W) \subseteq B_1$ and, more precisely,

$$\Omega_W := \{f \in B_1 : \mu_W(U) > 0 \text{ for every open } U \subseteq B_1 \text{ with } f \in U\}. \tag{9.5}$$

Let $\mathcal{A}$ be the family of all open subsets U of B_1 such that $\mu_W(U) = 0$. Then (9.5) shows that $\Omega_W = B_1 \setminus \bigcup_{U \in \mathcal{A}} U$. The union $\bigcup_{U \in \mathcal{A}} U$ is open, so this shows that Ω_W is a closed subset of B_1, and thus a complete separable metric space as asserted. Moreover, since B_1 is separable, this union equals the union of some countable subfamily; hence $\mu_W\bigl(\bigcup_{U \in \mathcal{A}} U\bigr) = 0$ and $\mu_W(\Omega_W) = 1$, so μ_W is a probability measure on Ω_W.

By the definition of μ_W, $\mu_W(\Omega_W) = \mu\{x : \psi_W(x) \in \Omega_W\}$, so this also shows that $\psi_W(x) \in \Omega_W$ for μ-a.e. x. Thus we can modify ψ_W on a null set in Ω so that $\psi_W : \Omega \to \Omega_W$, and then ψ_W is measure-preserving by the definition of μ_W.

This proves (i). Next, if U is an open subset of Ω_W with $\mu_W(U) = 0$, then $U = V \cap \Omega_W$ for some open $V \subset B_1$. Then $\mu_W(V) = \mu_W(U) = 0$, and thus $V \in \mathcal{A}$, so $V \subseteq B_1 \setminus \Omega_W$ and $U = V \cap \Omega_W = \emptyset$, which proves (ii).

If $U \subseteq \Omega_W$ is open and nonempty, then $\mu_W(U) > 0$ by (ii) and thus $\psi_W^{-1}(U) \neq \emptyset$ by (9.2); hence $U \cap \psi(\Omega_W) \neq \emptyset$, which shows (iii).

The set $Q := \{f \in L^1(\Omega, \mathcal{F}, \mu) : 0 \le f \le 1 \text{ a.e.}\}$ is a closed subset of $L^1(\Omega, \mathcal{F}, \mu)$. Since $W_x(y) = W(x, y) \in [0, 1]$ for every x and y, it follows that $\psi(x) \in Q$ for every x, and (iii) implies (iv).

To show (v), let $\mathcal{F}_W \subseteq \mathcal{F}$ be the σ-field on Ω induced by ψ_W, i.e.,

$$\mathcal{F}_W := \bigl\{\psi_W^{-1}(A) : A \subseteq L^1(\Omega, \mathcal{F}, \mu) \text{ is measurable}\bigr\}. \tag{9.6}$$

By definition, ψ_W is measurable $(\Omega, \mathcal{F}_W, \mu) \to L^1(\Omega, \mathcal{F}, \mu)$, so we can regard ψ_W as an element of, using (9.4),

$$L^1\bigl(\Omega, \mathcal{F}_W, \mu; L^1(\Omega, \mathcal{F}, \mu)\bigr) \cong L^1(\Omega \times \Omega, \mathcal{F}_W \times \mathcal{F}, \mu \times \mu). \tag{9.7}$$

This shows the existence of $W_1 \in L^1(\Omega \times \Omega, \mathcal{F}_W \times \mathcal{F}, \mu \times \mu)$ such that $W_1 = W$ a.e. Consequently, the conditional expectation

$$\mathbb{E}(W \mid \mathcal{F}_W \times \mathcal{F}) = \mathbb{E}(W_1 \mid \mathcal{F}_W \times \mathcal{F}) = W_1 = W \quad \text{a.e.}$$

By symmetry, also $\mathbb{E}(W \mid \mathcal{F} \times \mathcal{F}_W) = W$ a.e., and thus

$$\mathbb{E}(W \mid \mathcal{F}_W \times \mathcal{F}_W) = \mathbb{E}(\mathbb{E}(W \mid \mathcal{F}_W \times \mathcal{F}) \mid \mathcal{F} \times \mathcal{F}_W) = W \quad \text{a.e.}$$

Hence, $W = W_2$ a.e., where $W_2 : \Omega^2 \to [0, 1]$ is $\mathcal{F}_W \times \mathcal{F}_W$-measurable, which implies that $W_2 = \widetilde{W}^{\psi_W}$ for some measurable $\widetilde{W} : \Omega_W^2 \to [0, 1]$; we

can symmetrize $\widetilde{W}$ to obtain the desired graphon $\widehat{W}(x,y) := (\widetilde{W}(x,y) + \widetilde{W}(y,x))/2$.

If $\widehat{W}_1 : \Omega_W \to [0,1]$ is another graphon such that $\widehat{W}_1^{\psi_W} = W = \widehat{W}^{\psi_W}$ $\mu \times \mu$-a.e., then $\widehat{W}_1 = \widehat{W}$ $\mu_W \times \mu_W$-a.e., by the definitions of pull-back and μ_W.

Finally, $W \cong \widehat{W}$ since W is a.e. equal to a pull-back of $\widehat{W}$. $\square$

Remark 9.2. Sine Ω_W is a complete separable metric space, the probability space (Ω_W, μ_W) is a Borel space, see Appendix A.2.

Following [51], but using our notations, we make the following definition:

Definition 9.3. A graphon W on Ω is *pure* if the mapping ψ_W is a bijection $\Omega \to \Omega_W$.

Note that ψ_W is injective $\iff$ W is twinfree (see Section 8). It follows easily that a graphon is pure if and only if it is twinfree and the metric $r(x,y) := \|W(x,\cdot) - W(y,\cdot)\|_{L^1}$ on Ω is complete, and further μ has full support in the metric space (Ω, r). (Then, automatically, (Ω, r) is separable.) See further [51].

Remark 9.4. Let W, W' be graphons on the same probability space Ω with $W' = W$ a.e. Then $W_x(y) = W'_x(y)$ for a.e. y, for a.e. x; in other words, $\psi_W(x) = \psi_{W'}(x)$ for a.e. x. Consequently, $\mu_W = \mu_{W'}$, and thus also $\Omega_W = \Omega_{W'}$. We have $\widehat{W'}^{\psi_W} = \widehat{W'}^{\psi_{W'}} = W' = W$ a.e., and thus $\widehat{W'} = \widehat{W}$ a.e. by the uniqueness statement in Theorem 9.1.

Note that if W is pure and $W' = W$ a.e., then W' is not necessarily pure; however, W' is pure if $\mu\{y : W(x,y) \neq W'(x,y)\} = 0$ for every x (and not just for almost every x).

We let $\widehat{W}$ denote the graphon constructed in Theorem 9.1. Note that $\widehat{W}$ is defined only up to a.e. equivalence, so we have some freedom in choosing $\widehat{W}$. We will show (Lemma 9.6) that there is a choice of $\widehat{W}$ that is a pure graphon.

Lemma 9.5. *Let W_1 and W_2 be two graphons defined on probability spaces (Ω_1, μ_1) and (Ω_2, μ_2), respectively, and $\varphi : \Omega_1 \to \Omega_2$ is a measure-preserving mapping such that $W_1 = W_2^{\varphi}$ a.e. Then the pull-back map $\varphi^* : f \mapsto f^{\varphi}$ is an isometric measure-preserving bijection of Ω_{W_2} onto Ω_{W_1}, and $\widehat{W}_1^{\varphi^*} = \widehat{W}_2$ a.e.*

Proof. By Remark 9.4, we may replace W_1 by W_2^{φ} and thus assume $W_1 = W_2^{\varphi}$ everywhere and not just a.e.

Since φ is measure-preserving, φ^* is an isometric injection $L^1(\Omega_2, \mu_2) \to L^1(\Omega_1, \mu_1)$.

If $x \in \Omega_1$, then the composition $\varphi^* \circ \psi_{W_2} \circ \varphi$ maps $x \in \Omega_1$ to

$$\varphi^*\bigl(\psi_{W_2}(\varphi(x))\bigr) = \varphi^*\bigl(W_{2,\varphi(x)}\bigr) = \bigl(W_{2,\varphi(x)}\bigr)^{\varphi},$$

which is the mapping

$$x' \mapsto \bigl(W_{2,\varphi(x)}\bigr)^{\varphi}(x') = W_2\bigl(\varphi(x),\varphi(x')\bigr) = W_2^{\varphi}(x,x') = W_1(x,x') = W_{1,x}(x'),$$

and thus

$$\varphi^*\bigl(\psi_{W_2}(\varphi(x))\bigr) = W_{1,x} = \psi_{W_1}(x). \tag{9.8}$$

In other words, $\varphi^* \circ \psi_{W_2} \circ \varphi = \psi_{W_1}$. Since ψ_{W_1} , ψ_{W_2} and φ are measure-preserving, it follows that for any $A \subseteq L^1(\Omega_1,\mu_1)$,

$$\begin{aligned}\mu_{W_2}\bigl((\varphi^*)^{-1}(A)\bigr) &= \mu_2\bigl(\psi_{W_2}^{-1}\bigl((\varphi^*)^{-1}(A)\bigr)\bigr) = \mu_1\bigl(\varphi^{-1}\bigl(\psi_{W_2}^{-1}\bigl((\varphi^*)^{-1}(A)\bigr)\bigr)\bigr)\\ &= \mu_1\bigl(\psi_{W_1}^{-1}(A)\bigr) = \mu_{W_1}(A).\end{aligned} \tag{9.9}$$

Hence, $\varphi^* : L^1(\Omega_2,\mu_2) \to L^1(\Omega_1,\mu_1)$ is measure-preserving.

In particular, (9.9) shows that $\mu_{W_2}\bigl((\varphi^*)^{-1}(\Omega_{W_1})\bigr) = \mu_{W_1}(\Omega_{W_1}) = 1$. Moreover, $(\varphi^*)^{-1}(\Omega_{W_1})$ is closed in $L^1(\Omega_2,\mu_2)$ since Ω_{W_1} is closed in $L^1(\Omega_1,\mu_1)$ and φ^* is continuous, and thus it follows, see (9.3), that

$$\Omega_{W_2} = \operatorname{supp}(\mu_{W_2}) \subseteq (\varphi^*)^{-1}(\Omega_{W_1}).$$

In other words, $\varphi^* : \Omega_{W_2} \to \Omega_{W_1}$.

Next, since φ^* is an isometry and Ω_{W_2} is a complete metric space (by Theorem 9.1(i)), $\varphi^*(\Omega_{W_2})$ is a complete subset of the metric space Ω_{W_1}, and thus $\varphi^*(\Omega_{W_2})$ is closed. By (9.9),

$$\mu_{W_1}\bigl(\varphi^*(\Omega_{W_2})\bigr) = \mu_{W_2}\bigl((\varphi^*)^{-1}(\varphi^*(\Omega_{W_2}))\bigr) = \mu_{W_2}\bigl(\Omega_{W_2}\bigr) = 1.$$

Thus by (9.3) again (or by Theorem 9.1(ii)),

$$\Omega_{W_1} = \operatorname{supp}(\mu_{W_1}) \subseteq \varphi^*(\Omega_{W_2}).$$

Hence φ^* is a bijection $\Omega_{W_2} \to \Omega_{W_1}$.

Finally, by (9.8), a.e. on $\Omega_1 \times \Omega_1$,

$$\bigl(\bigl(\widehat{W}_1^{\varphi^*}\bigr)^{\psi_{W_2}}\bigr)^{\varphi} = \bigl(\widehat{W}_1\bigr)^{\varphi^*\circ\psi_{W_2}\circ\varphi} = \bigl(\widehat{W}_1\bigr)^{\psi_{W_1}} = W_1 = \bigl(W_2\bigr)^{\varphi},$$

and thus $\bigl(\widehat{W}_1^{\varphi^*}\bigr)^{\psi_{W_2}} = W_2$ a.e. on $\Omega_2 \times \Omega_2$. Consequently, $\widehat{W}_1^{\varphi^*} = \widehat{W}_2$ a.e. by the uniquness statement in Theorem 9.1(v). □

Lemma 9.6. *For any graphon W, $\widehat{W}$ in Theorem 9.1 can be chosen to be a pure graphon on Ω_W.*

Proof. Let W be defined on Ω. The construction in Theorem 9.1 yields the graphon $\widehat{W}$ defined on $\Omega_W \subseteq L^1(\Omega,\mathcal{F},\mu)$. We repeat the construnction, starting with $\widehat{W}$ on Ω_W, and obtain the graphon $\widehat{\widehat{W}}$ on $\Omega_{\widehat{W}}$, where $\Omega_{\widehat{W}} \subseteq L^1(\Omega_W,\mu_W)$. Since $\psi_W : \Omega \to \Omega_W$ is measure-preserving and $\widehat{W}^{\psi_W} = W$ a.e. by Theorem 9.1, it follows by Lemma 9.5 that ψ_W^* is an isometric bijection of $\Omega_{\widehat{W}}$ onto Ω_W; thus $(\psi_W^*)^{-1}$ is a bijection $\Omega_W \to \Omega_{\widehat{W}}$.

We will show that we can modify $\widehat{W}$ on a null set so that $\psi_{\widehat{W}} = (\psi_W^*)^{-1}$.

For a.e. $x \in \Omega$, we have $\psi_W(x) \in \Omega_W \subseteq L^1(\Omega, \mathcal{F}, \mu)$ and then $\psi_{\widehat{W}}(\psi_W(x))$ is by (9.1) the function in $L^1(\Omega_W, \mu_W)$ given by

$$\psi_{\widehat{W}}(\psi_W(x))(g) = \widehat{W}\big(\psi_W(x), g\big), \qquad g \in \Omega_W.$$

Consequently, the pull-back map ψ_W^* in Lemma 9.5 maps this to the function on Ω given by, for a.e. x and y,

$$\begin{aligned}\psi_W^*(\psi_{\widehat{W}}(\psi_W(x)))(y) &= \psi_{\widehat{W}}\big(\psi_W(x)\big)\big(\psi_W(y)\big) = \widehat{W}\big(\psi_W(x), \psi_W(y)\big)\\ &= \widehat{W}^{\psi_W}(x,y) = W(x,y) = \psi_W(x)(y);\end{aligned}$$

thus $\psi_W^*(\psi_{\widehat{W}}(\psi_W(x))) = \psi_W(x)$ for a.e. $x \in \Omega$.

Let

$$A := \{f \in \Omega_W : \psi_W^*(\psi_{\widehat{W}}(f)) = f\}.$$

We have just shown that $\mu\{x : \psi_W(x) \in A\} = 1$, so by (9.2) $\mu_W(A) = 1$. Thus, $\psi_W^*(\psi_{\widehat{W}}) : \Omega_W \to \Omega_W$ equals the identity map a.e.

Since ψ_W^* is a bijection, $\psi_{\widehat{W}} = (\psi_W^*)^{-1}$ on $A \subseteq \Omega_W$. The idea is to modify $\psi_{\widehat{W}}$ on the null set $\Omega_W \setminus A$ such that this equality holds everywhere. The space $\Omega_{\widehat{W}}$ is included in a separable subspace $B_1 \subseteq L^1(\Omega_W, \mu_W)$, and by Lemma G.1, there exists a measurable evaluation map $\Phi : B_1 \times \Omega_W \to \mathbb{R}$ such that $\Phi(F, g) = F(g)$ for every $F \in B_1$ and μ_W-a.e. $g \in \Omega_W$. Define

$$H(f,g) := \Phi\big((\psi_W^*)^{-1}(f), g\big), \qquad f, g \in \Omega_W;$$

then $H : \Omega_W \times \Omega_W \to \mathbb{R}$ is measurable and for every $f \in \Omega_W$,

$$H(f,g) = (\psi_W^*)^{-1}(f)(g), \qquad \text{for a.e. } g \in \Omega_W. \tag{9.10}$$

For every $f \in \Omega_W$, $(\psi_W^*)^{-1}(f) \in \Omega_{\widehat{W}}$, so by (9.10) and Theorem 9.1(iv), $0 \le H(f,g) \le 1$ for a.e. $g \in \Omega_W$. Let $\overline{H}(f,g) := \min\big\{\max\{H(f,g), 0\}, 1\big\} \in [0,1]$. Then, for every $f \in \Omega_W$, by (9.10),

$$\overline{H}(f,g) = H(f,g) = (\psi_W^*)^{-1}(f)(g), \qquad \text{for a.e. } g \in \Omega_W. \tag{9.11}$$

Thus $\overline{H}$ has the desired sections. We define a graphon $\widehat{W}_1$ on Ω_W by

$$\widehat{W}_1(x,y) := \begin{cases} \widehat{W}(f,g), & f, g \in A;\\ \overline{H}(f,g), & f \notin A, g \in A;\\ \overline{H}(g,f), & f \in A, g \notin A;\\ 0, & f, g \notin A. \end{cases} \tag{9.12}$$

Then $\widehat{W}_1 = \widehat{W}$ a.e., because $\mu(A) = 1$, so we may replace $\widehat{W}$ by $\widehat{W}_1$ in Theorem 9.1(v). Moreover, if $f \in A$ then $\widehat{W}_1(f,g) = \widehat{W}(f,g)$ for a.e. g, and thus $\psi_{\widehat{W}_1}(f) = \psi_{\widehat{W}}(f) = (\psi_W^*)^{-1}(f)$.

If $f \notin A$, then $\widehat{W}_1(f,g) = \overline{H}(f,g) = (\psi_W^*)^{-1}(f)(g)$ for a.e. g by (9.12) and (9.11), and thus $\psi_{\widehat{W}_1}(f) = (\psi_W^*)^{-1}(f)$ in this case too.

Consequently, $\psi_{\widehat{W}_1} = (\psi_W^*)^{-1}$ is a bijection $\Omega_W \to \Omega_{\widehat{W}} = \Omega_{\widehat{W}_1}$, where the final equality is by Remark 9.4. $\square$

Theorem 9.7. *Two graphons W_1 and W_2 are equivalent if and only if $\widehat{W_1}$ is an a.e. rearrangement of $\widehat{W_2}$ by a measure-preserving bijection $\Omega_{W_1} \to \Omega_{W_2}$ that further can be taken to be an isometry.*

In other words, $W_1 \cong W_2$ if and only if there is an isometric measure-preserving bijection $\varphi : \Omega_{W_1} \to \Omega_{W_2}$ such that $\widehat{W_2}^{\varphi} = \widehat{W_1}$ a.e.

Proof. Consider the class $\mathcal{W}_{\mathsf{m}}$ of all graphons that are defined on a probability space that is also a metric space. Define $W_1 \equiv W_2$ if $W_1, W_2 \in \mathcal{W}_{\mathsf{m}}$ and W_1 is a.e. equal to a rearrangement of W_2 by an isometric measure-preserving bijection; this is an equivalence relation on $\mathcal{W}_{\mathsf{m}}$. Lemma 9.5 shows that if W_1 is a pullback of W_2, then $\widehat{W_1} \equiv \widehat{W_2}$.

If $W_1 \cong W_2$, then Theorem 8.3 yields a chain of pullbacks linking W_1 and W_2, and thus $\widehat{W_1} \equiv \widehat{W_2}$.

Conversely, if $\widehat{W_1}$ equals a rearrangement of $\widehat{W_2}$ a.e., then $W_1 \cong \widehat{W_1} \cong \widehat{W_2} \cong W_2$ by Theorem 9.1(v). □

Corollary 9.8. *If W_1 and W_2 are equivalent graphons, then Ω_{W_1} and Ω_{W_2} are isometric metric spaces.* □

Theorem 9.9. *Every graphon is equivalent to a pure graphon. Two pure graphons W_1 and W_2 are equivalent if and only if they are a.e. rearrangements of each other.*

Proof. If W is a graphon, then $W \cong \widehat{W}$ for a pure graphon $\widehat{W}$ by Theorem 9.1(v) and Lemma 9.6.

If W_1 is pure, then ψ_{W_1} is a bijection so W_1 is an a.e. rearrangement of $\widehat{W_1}$ by Theorem 9.1(v). The same applies to W_2, and if further $W_1 \cong W_2$, then $\widehat{W_1} \cong \widehat{W_2}$ and Theorem 9.7 yields that $\widehat{W_1}$ is an a.e. rearrangement of $\widehat{W_2}$. Since being an a.e. rearrangement is an equivalence relation, this shows that W_1 is an a.e. rearrangement of W_2, and conversely. □

Proof of Theorem 8.4. Suppose that $W_1 \cong W_2$. By Theorem 9.7, $\widehat{W_1}$ is an a.e. rearrangement of $\widehat{W_2}$. Further, W_1 is a.e. equal to a pull-back of $\widehat{W_1}$ by Theorem 9.1, so by composition, W_1 is a.e. equal to a pull-back of $\widehat{W_2}$, and so is W_2 by Theorem 9.1 again. This proves Theorem 8.4 (except the last sentence, which was shown in Section 8) for graphons, which suffices as remarked earlier. Note that $\widehat{W_2}$ is defined on Ω_{W_2}, which by Remark 9.2 is a Borel space. □

By Corollary 9.8, every graph limit Γ, i.e. every element of the quotient space $\widehat{\mathcal{W}}$, defines a complete separable metric space Ω_W by taking any graphon W that represents Γ; this metric space is uniquely defined up to isometry. Hence metric and topological properties of Ω_W are invariants of graph limits. See Lovász and Szegedy [51] for some relations between such properties of Ω_W and combinatorial properties of the graph limit; it would be interesting to find further such results.

Example 9.10. A trivial example is that a graph limit is of finite type, i.e. it can be represented by a step graphon, if and only if Ω_W is a finite set, see Example 5.3. Theorems 9.7 and 9.1 imply that every graphon equivalent to a step graphon is a.e. equal to a step graphon.

Theorem 9.9 shows that we can regard pure graphons as the canonical choices among all graphons representing a given graph limit. By considering only pure graphons, equivalence boils down to a.e. equality and rearrangements, and every graphon W has a pure version constructed as $\widehat{W}$. This is theoretically pleasing. (Nevertheless, for many applications it is more convenient to use other graphons, for example defined on $[0,1]$, regardless of whether they are pure or not.)

Remark 9.11. In this section, we have used mappings into $L^1(\Omega,\mu)$ and have constructed Ω_W as a subset of $L^1(\Omega,\mu)$. We could just as well use $L^2(\Omega,\mu)$, or $L^p(\Omega,\mu)$ for any $p\in[1,\infty)$. In fact, by Theorem 9.1(iv) $\Omega_W\subset L^p(\Omega,\mu)$ and the different L^p-metrics are equivalent on Ω_W by Hölder's inequality; it follows easily that the construction above yields the same space Ω_W for any L^p with $1\le p<\infty$, with a different but equivalent metric and thus the same topology.

9.1. **The weak topology on Ω_W.** Since $\Omega_W\subset L^2(\Omega,\mu)$, the inner product $\langle f,g\rangle:=\int_\Omega fg\,\mathrm{d}\mu$ is defined and continuous on $\Omega_W\times\Omega_W$. We define further, again following Lovász and Szegedy [51] in principle but not in all details,

$$r_{W\circ W}(f,g):=\int_{\Omega_W}|\langle f-g,h\rangle|\,\mathrm{d}\mu_W(h),\qquad f,g\in\Omega_W. \tag{9.13}$$

Since $\psi_W:\Omega\to\Omega_W$ is measure-preserving, and $\psi_W(x)=W_x$, this can also be written

$$r_{W\circ W}(f,g)=\int_\Omega|\langle f-g,W_x\rangle|\,\mathrm{d}\mu(x). \tag{9.14}$$

Since $\langle f,g\rangle$ is continuous and bounded on $\Omega_W\times\Omega_W$, it follows by dominated convergence that $r_{W\circ W}(f,g)$ is continuous on $\Omega_W\times\Omega_W$. We will soon see (in Theorem 9.13) that it is a metric. We let r_W denote the original metric on Ω_W, i.e. the L^1-norm:

$$r_W(f,g):=\|f-g\|_{L^1}:=\int_\Omega|f(x)-g(x)|\,\mathrm{d}\mu(x). \tag{9.15}$$

We have $0\le W_x\le 1$ and thus $|\langle f-g,W_x\rangle|\le\|f-g\|_{L^1}$, so by (9.14) and (9.15),

$$r_{W\circ W}(f,g)\le r_W(f,g). \tag{9.16}$$

Remark 9.12. If W is pure, so ψ_W is a bijection, then $r_{W\circ W}$ induces a metric on Ω which we also denoted by $r_{W\circ W}$; explicitly,

$$\begin{aligned} r_{W\circ W}(x,y) := r_{W\circ W}(W_x,W_y) &= \int_\Omega |\langle W_x - W_y, W_z\rangle| \,\mathrm{d}\mu(z) \\ &= \int_\Omega \biggl|\int_\Omega \bigl(W(x,u)-W(y,u)\bigr)W(z,u)\,\mathrm{d}\mu(u)\biggr| \,\mathrm{d}\mu(z) \\ &= \int_\Omega \bigl|W\circ W(x,z) - W\circ W(y,z)\bigr| \,\mathrm{d}\mu(z) \\ &= \|W\circ W(x,\cdot) - W\circ W(y,\cdot)\|_{L^1(\Omega,\mu)} \end{aligned}$$

where $W\circ W(x,y) := \int_\Omega W(x,u)W(u,y)\,\mathrm{d}\mu(u)$. Thus, if T_W is the integral operator with kernel W, then $W\circ W$ is the kernel of the integral operator $T_W\circ T_W$, which explains the notation.

Recall that the *weak topology* $\sigma = \sigma_{L^\infty}$ on $L^1(\Omega,\mu)$ is the topology generated by the linear functionals $f\mapsto \langle f,h\rangle = \int_\Omega fh\,\mathrm{d}\mu$ for $h\in L^\infty(\Omega,\mu)$. In general, if X and Y are two subsets of $L^1(\Omega)$ such that $\int_\Omega |fh| < \infty$ when $f\in X$ and $h\in Y$, let (X,σ_Y) denote X with the weak topology generated by the linear functionals $f\mapsto\langle f,h\rangle$, $h\in Y$. Since the elements of Ω_W are uniformly bounded functions by Theorem 9.1(iv), it is well-known, see Lemma F.1(i), that the weak topology on Ω_W also is generated by $f\mapsto\langle f,h\rangle$ for $h\in L^1(\Omega,\mu)$ (this is the *weak* topology* on $L^\infty(\Omega,\mu)$ restricted to Ω_W), or by the subset $h\in L^p(\Omega,\mu)$ (this is the weak topology on $L^q(\Omega,\mu)$, where $1/p+1/q=1$, restricted to the subset Ω_W, cf. Remark 9.11). Thus,

$$(\Omega_W,\sigma) = (\Omega_W,\sigma_{L^\infty}) = (\Omega_W,\sigma_{L^1}) = (\Omega_W,\sigma_{L^2}). \tag{9.17}$$

We let $\overline{\Omega_W}^\sigma$ be the closure of Ω_W in $L^1(\Omega,\mu)$ in the weak topology. It follows by Theorem 9.1(iv) that $\overline{\Omega_W}^\sigma \subseteq \{f\in L^1(\Omega,\mu) : 0\le f\le 1 \text{ a.e.}\}$, and thus, by Lemma F.1(i) again,

$$(\overline{\Omega_W}^\sigma,\sigma) = (\overline{\Omega_W}^\sigma,\sigma_{L^\infty}) = (\overline{\Omega_W}^\sigma,\sigma_{L^1}) = (\overline{\Omega_W}^\sigma,\sigma_{L^2}). \tag{9.18}$$

Moreover, the weak closure $\overline{\Omega_W}^\sigma$ is the same as the weak closure of Ω_W in $L^p(\Omega,\mu)$ for any $p<\infty$, and the weak* closure in $L^\infty(\Omega,\mu)$.

Recall also that two metrics r_1 and r_2 on the same space are *equivalent* if they induce the same topology, i.e., if $r_1(x_n,x)\to 0 \iff r_2(x_n,x)\to 0$ for any point x and sequence (x_n) in the space; the metrics are *uniformly equivalent* if $r_1(x_n,y_n)\to 0 \iff r_2(x_n,y_n)\to 0$ for any sequences (x_n) and (y_n).

Lovász and Szegedy [51] showed essentially the following.

Theorem 9.13. (i) *$r_{W\circ W}$ is a metric on Ω_W and it defines the weak topology σ on Ω_W. The same holds on the weak closure $\overline{\Omega_W}^\sigma$.*

(ii) *The metric space $(\overline{\Omega_W}^\sigma, r_{W\circ W})$ is compact. Thus $(\overline{\Omega_W}^\sigma, r_{W\circ W})$ is the completion of $(\Omega_W, r_{W\circ W})$. In particular, $\overline{\Omega_W}^\sigma = \Omega_W$ if and only if $(\Omega_W, r_{W\circ W})$ is complete.*

(iii) *The inequality* $r_W \geq r_{W\circ W}$ *holds, and thus the identity mapping* $(\Omega_W, r_W) \to (\Omega_W, r_{W\circ W})$ *is uniformly continuous.*

(iv) Ω_W *is compact if and only if the metrics* r_W *and* $r_{W\circ W}$ *are equivalent on* Ω_W *and further* Ω_W *is weakly closed,* $\overline{\Omega_W}^{\sigma} = \Omega_W$.

(v) *The metrics* r_W *and* $r_{W\circ W}$ *are uniformly equivalent on* Ω_W *if and only if* Ω_W *is compact for the norm topology given by* r_W.

It seems more difficult to characterize when r_W and $r_{W\circ W}$ are equivalent on Ω_W, see Examples 9.16–9.17 below.

Before proving the theorem, we introduce more notation. Using the fact that $\Omega_W \subset L^2(\Omega, \mu)$ (cf. Remark 9.11), let A be the closed linear span of Ω_W in $L^2(\Omega, \mu)$, and let B be the unit ball of A; thus $\Omega_W \subseteq B$. We extend the definition (9.13) of $r_{W\circ W}$ to all $f, g \in A$.

Lemma 9.14. (i) $r_{W\circ W}$ *is a metric on* A.

(ii) *The metric* $r_{W\circ W}$ *defines the weak topology* σ_{L^2} *on* B. *In other words,* $(B, r_{W\circ W}) = (B, \sigma_{L^2})$ *as topological spaces.*

(iii) *The metric space* $(B, r_{W\circ W})$ *is compact.*

Proof. (i): Symmetry and the triangle inequality are immediate from the definition (9.13). Suppose that $r_{W\circ W}(f, g) = 0$ for some $f, g \in A$. Since $h \mapsto |\langle f - g, h\rangle|$ is continuous on Ω_W, and its integral $r_{W\circ W}(f, g)$ is 0, it follows from Theorem 9.1(ii) that $\langle f - g, h\rangle = 0$ for every $h \in \Omega_W$. The set $\{h \in L^2 : \langle f - g, h\rangle = 0\}$ is a closed linear subspace of $L^2(\Omega, \mu)$, and thus it contains A; i.e. $\langle f - g, h\rangle = 0$ for every $h \in A$. In particular,

$$\int_\Omega |f - g|^2 \, \mathrm{d}\mu = \langle f - g, f - g\rangle = 0.$$

Thus $f - g = 0$ a.e., i.e. $f = g$ in $A \subseteq L^2$. Hence $r_{W\circ W}$ is a metric.

(ii): If $h \in L^2(\Omega)$ and $h_2 \in A$ is the orthogonal projection of h, then $\langle f, h\rangle = \langle f, h_2\rangle$ for every $f \in A$. Consequently, $\sigma_{L^2} = \sigma_A$ on A.

Let D be a countable dense subset of Ω_W; then D is total in A, and thus A is a separable Hilbert space. It is a standard fact that the unit ball B of A with the weak topology $\sigma_A = \sigma_{L^2}$ then is a compact metric space. (It is compact by the Banach–Alaoglu theorem [25, Theorem V.4.2], and metric by [25, Theorem V.5.1]. Explicitly, $\sigma_A = \sigma_D$ on B, by the same argument as in the proof of Lemma F.1, and if $D = \{h_1, h_2, \dots\}$, we can define a metric on $(B, \sigma_{L^2}) = (B, \sigma_A)$ by $d(f, g) := \sum_i 2^{-i} |\langle f - g, h_i\rangle|$.)

We next show that the identity map $(B, \sigma_{L^2}) \to (B, r_{W\circ W})$ is continuous. Since, as just shown, (B, σ_{L^2}) is metrizable, it suffices to consider sequential continuity. Thus assume that $f_n, f \in B$ and $f_n \to f$ in σ_{L^2}. Then $\langle f_n - f, h\rangle \to 0$ as $n \to \infty$ for every $h \in \Omega_W \subset L^2$, and thus $r_{W\circ W}(f_n, f) \to 0$ by (9.13) and dominated convergence.

The identity map $(B, \sigma_{L^2}) \to (B, r_{W\circ W})$ is thus a continuous bijection of a compact space onto Hausdorff space, and it is thus a homeomorphism. Consequently, $(B, \sigma_{L^2}) = (B, r_{W\circ W})$.

(iii): A consequence of (ii) and its proof, where we showed that (B, σ_{L^2}) is compact. $\square$

Proof of Theorem 9.13. (i): Since $\Omega_W \subseteq B \subset A$, it follows by Lemma 9.14 that $\overline{\Omega_W}^\sigma \subseteq B$. Hence, using Lemma 9.14 again and (9.18), $r_{W\circ W}$ is a metric on $\overline{\Omega_W}^\sigma$ and $(\overline{\Omega_W}^\sigma, \sigma) = (\overline{\Omega_W}^\sigma, \sigma_{L^2}) = (\overline{\Omega_W}^\sigma, r_{W\circ W})$.

(ii): An immediate consequence of Lemma 9.14, together with standard facts on compact and complete metric spaces (see e.g. [27, Section 4.3]).

(iii): This is just (9.16).

(iv): If (Ω_W, r_W) is compact, then the identity mapping $(\Omega_W, r_W) \to (\Omega_W, r_{W\circ W})$, which is continuous by (iii), is a homeomorphism. The metrics are thus equivalent. Furthermore, $(\Omega_W, \sigma) = (\Omega_W, r_{W\circ W})$ is compact and thus closed in the weak topology on $L^1(\Omega, \mu)$.

Conversely, if $\Omega_W = \overline{\Omega_W}^\sigma$, then $(\Omega_W, r_{W\circ W})$ is compact by (ii), and if further the metrics are equivalent, then (Ω_W, r_W) is compact too.

(v): If (Ω_W, r_W) is compact, then the metrics r_W and $r_{W\circ W}$ on Ω_W are equivalent as seen in the proof of (iv). Moreover, as is easily seen (e.g. [27, Theorem 4.3.32]), two equivalent metrics on a compact metric space are uniformly equivalent.

Conversely, if r_W and $r_{W\circ W}$ are uniformly equivalent, then $(\Omega_W, r_{W\circ W})$ is a complete metric space, since (Ω_W, r_W) is by Theorem 9.1; hence $\overline{\Omega_W}^\sigma = \Omega_W$ by (ii), and thus Ω_W is compact by (ii) again. $\square$

The following analogue of Corollary 9.8 shows that also the metric space $(\Omega_W, r_{W\circ W})$ and its completion, the compact metric space $(\overline{\Omega_W}^\sigma, r_{W\circ W})$, are invariants of graph limits.

Theorem 9.15. *If W_1 and W_2 are equivalent graphons, then $(\Omega_{W_1}, r_{W_1\circ W_1})$ and $(\Omega_{W_2}, r_{W_2\circ W_2})$ are isometric metric spaces, and so are the compact metric spaces $(\overline{\Omega_{W_1}}^\sigma, r_{W_1\circ W_1})$ and $(\overline{\Omega_{W_2}}^\sigma, r_{W_2\circ W_2})$.*

Proof. By Theorem 8.3, it suffices to prove this in the case when W_1 is a pull-back of W_2 as in Lemma 9.5. In this case, for any $f, g \in \Omega_{W_2}$, using (9.13) and the fact that $f \mapsto f^\varphi$ is a measure-preserving bijection of Ω_{W_2} onto Ω_{W_1} by Lemma 9.5,

$$\begin{aligned} r_{W_1\circ W_1}(f^\varphi, g^\varphi) &= \int_{\Omega_{W_1}} \langle f^\varphi - g^\varphi, h\rangle \,\mathrm{d}\mu_{W_1}(h) \\ &= \int_{\Omega_{W_2}} \langle f^\varphi - g^\varphi, k^\varphi\rangle \,\mathrm{d}\mu_{W_2}(k) \\ &= \int_{\Omega_{W_2}} \langle f - g, k\rangle \,\mathrm{d}\mu_{W_2}(k) \\ &= r_{W_2\circ W_2}(f, g); \end{aligned}$$

thus the bijection $f \mapsto f^\varphi$ is an isometry also $(\Omega_{W_2}, r_{W_2\circ W_2}) \to (\Omega_{W_1}, r_{W_1\circ W_1})$. This extends to an isometric bijection of $(\overline{\Omega_{W_2}}^\sigma, r_{W_2\circ W_2})$ onto $(\overline{\Omega_{W_1}}^\sigma, r_{W_1\circ W_1})$ by Theorem 9.13(ii). $\square$

We say that a graphon W is *compact* if Ω_W is a compact metric space with the standard L^1 metric r_W, and *weakly compact* if $(\Omega_W, \sigma) = (\Omega_W, r_{W\circ W})$ is compact. By Corollary 9.8 and Theorem 9.15, the same then holds for every equivalent graphon, so we may say that a graph limit is [weakly] compact if some, and thus any, representing graphon is [weakly] compact.

Not every graphon is compact. Moreover, this can happen both with (Ω_W, σ) compact and (Ω_W, σ) non-compact, as shown by the following examples (inspired by a similar example in [51]). Note that exactly one of the two conditions in Theorem 9.13(iv) fails in each of the two examples.

Example 9.16. Let $\Omega := \{0,1\}^\infty = \bigl\{x = (x_i)_0^\infty : x_i \in \{0,1\}\bigr\}$ (the Cantor cube, which is homeomorphic to the Cantor set) with the product measure $\mu := \nu^\infty$, where $\nu\{0\} = \nu\{1\} = 1/2$. We write $\Omega = \Omega_0 \cup \Omega_1$, where $\Omega_j := \{x \in \Omega : x_0 = j\}$. Note that there is a measure-preserving map $[0,1] \to \Omega$ given by the binary expansion, so the examples below can be translated to examples on $[0,1]$ by taking pull-backs.

If F is a function $\Omega_0 \times \Omega_1 \to [-1,1]$, we define a graphon W on Ω by

$$W(x,y) := \begin{cases} \frac12 + \frac12 F(x,y), & x \in \Omega_0,\ y \in \Omega_1; \\ \frac12 + \frac12 F(y,x), & x \in \Omega_1,\ y \in \Omega_0; \\ \frac12, & x,y \in \Omega_0 \text{ or } x,y \in \Omega_1. \end{cases}$$

Define $F_x(y) = F(x,y)$ for $x \in \Omega_0$, $y \in \Omega_1$ and $\bar F_x(y) = F(y,x)$ for $x \in \Omega_1$, $y \in \Omega_0$; thus $F_x \in L^1(\Omega_1)$ for $x \in \Omega_0$ and $\bar F_x \in L^1(\Omega_0)$ for $x \in \Omega_1$.

Regard $L^1(\Omega_0)$ and $L^1(\Omega_1)$ as subspaces of $L^1(\Omega)$ in the obvious way (extending functions by 0). Define the maps $\Phi_0 : \Omega_0 \to L^1(\Omega_1)$ and $\Phi_1 : \Omega_1 \to L^1(\Omega_0)$ by $\Phi_0(x) = F_x$ and $\Phi_1(x) = \bar F_x$; then

$$\psi_W(x) = W_x = \tfrac12 + \tfrac12 \Phi_j(x), \qquad \text{for } x \in \Omega_j. \tag{9.19}$$

Let μ_j be the push-forward of μ by Φ_j; this is a measure on $L^1(\Omega_{1-j}) \subset L^1(\Omega)$ with total mass $1/2$. Let $X_j \subset L^1(\Omega_{1-j}) \subset L^1(\Omega)$ be the support of μ_j. It follows from (9.19) that the map $f \mapsto \frac12 + \frac12 f$ is measure-preserving $(L^1(\Omega), \mu_0 + \mu_1) \to (L^1(\Omega), \mu_W)$, and thus

$$\Omega_W = \bigl\{\tfrac12 + \tfrac12 f : f \in X_0 \cup X_1\bigr\}. \tag{9.20}$$

Define the functions $h_i : \Omega_1 \to \{-1,1\}$ by $h_i(x) = 2x_i - 1$, where $x \mapsto x_i$ is the i:th coordinate function. Let $\ell(x) := \inf\{i : x_i = 1\}$ (defined a.e. on Ω) and take

$$F(x,y) := h_{\ell(x)}(y). \tag{9.21}$$

(Thus $W(x,y) = y_{\ell(x)}$ for $x \in \Omega_0$, $y \in \Omega_1$.)

Then $\{F_x : x \in \Omega_0\} = \{h_i : i \ge 1\}$. The induced measure μ_0 on $L^1(\Omega_1)$ is thus a discrete measure with atoms h_i (each with positive measure), so

$$X_0 = \operatorname{supp} \mu_0 = \overline{\{h_i : i \ge 1\}} = \{h_i : i \ge 1\},$$

since $\{h_i\}$ is closed in $L^1(\Omega_1)$ because $\|h_i - h_j\|_{L^1(\Omega_1)} = 1/2$ when $i \ne j$.

If $y, z \in \Omega_1$, then

$$\|\bar{F}_y - \bar{F}_z\|_{L^1} = \int_{\Omega_0} |F(x,y) - F(x,z)| \, \mathrm{d}\mu(x) = \sum_{i=1}^{\infty} 2^{-i-1} |y_i - z_i|.$$

It is easily seen that this is a metric on Ω_1 which defines the product topology. Hence $\Phi_1 : y \mapsto \bar{F}_y$ is a homeomorphism of Ω_1 onto $\{\bar{F}_y : y \in \Omega_1\} \subset L^1(\Omega_0)$, and consequently $\{\bar{F}_y : y \in \Omega_1\}$ is a compact subset of L^1. Since further $\Phi_1 : (\Omega_1, \mu) \to (L^1(\Omega), \mu_1)$ is measure-preserving, and μ has full support on Ω_1, it follows that $X_1 = \operatorname{supp} \mu_1 = \{\bar{F}_y : y \in \Omega_1\} \cong \Omega_1 \cong \Omega$ (where $\cong$ denotes homeomorphisms.) Note also that X_0 and X_1 are disjoint; in fact, they have distance 1 in L^1.

It follows that $\Omega_W \cong X_0 \cup X_1 \cong \mathbb{N} \cup \Omega$, i.e., Ω_W is homeomeorphic to the disjoint union of the Cantor cube (or Cantor set) and a sequence of discrete points. Thus Ω_W is not compact.

With the weak topology σ, we have $(X_1, \sigma) = (X_1, r_W)$ because (X_1, r_W) is compact. Moreover, the sequence (h_i) is orthonormal in $L^1(\Omega_1, 2\mu)$ (for convenience normalizing the measure on Ω_1), and thus $h_i \to 0$ weakly in L^2 as $i \to \infty$. It follows by Lemma 9.14(ii) that $r_{W \circ W}(h_i, 0) \to 0$. For the corresponding elements $g_i := \frac{1}{2} + \frac{1}{2} h_i \in \Omega$, see (9.20), we have $r_{W \circ W}(g_i, \frac{1}{2}) \to 0$. It follows that (Ω_W, σ) consists of a compact set homeomorpic to Ω, and a sequence (g_i) converging to $\frac{1}{2}$. Since $\frac{1}{2} \notin \Omega_W$, it follows that $(\Omega_W, \sigma) = (\Omega_W, r_{W \circ W})$ is not compact; moreover, the identity map $(\Omega_W, r_W) \to (\Omega_W, r_{W \circ W})$ is a homeomorphism, so $(\Omega_W, \sigma) = (\Omega_W, r_W) \cong \mathbb{N} \cup \Omega$. Thus r_W and $r_{W \circ W}$ are equivalent on Ω_W but not uniformly equivalent. (Just as $\{1, 2, \dots\}$ and $\{1, 1/2, 1/3, \dots\}$, both with the usual metric on $\mathbb{R}$, are equivalent but not uniformly so.)

The weak closure $\overline{\Omega_W}^{\sigma} = \Omega_W \cup \{\frac{1}{2}\}$ is the one-point compactification of Ω_W.

Example 9.17. We modify the preceding example by taking, instead of (9.21),

$$F(x,y) := \begin{cases} 0, & \text{if } x_1 = x_2 = 1, \\ h_{\ell(x)}(y) & \text{otherwise.} \end{cases} \tag{9.22}$$

The only significant difference from the preceding example is that now X_0 also contains the function 0, and Ω_W thus the function $\frac{1}{2}$; note that $h_i \to 0$ weakly and thus in $r_{W \circ W}$ but not in r_W. In the norm topology. $X_0 = \{g_i\} \cup \{\frac{1}{2}\}$ is still an infinite discrete set, and thus $\Omega_W \cong X_0 \cup X_1 \cong \mathbb{N} \cup \Omega$ as in Example 9.16. (We have added one isolated point to Ω_W.)

In the weak topology, however, X_0 now consists of a convergent sequence and its limit point, and thus (X_0, σ) is compact and homeomorphic to the one-point compactification $\overline{\mathbb{N}}$ of $\mathbb{N}$ (or, equivalently, to $\{1/n : n \in \mathbb{N}\} \cup \{0\}$ with the usual topology). Thus $(\Omega_W, \sigma) \cong X_0 \cup X_1 \cong \overline{\mathbb{N}} \cup \Omega$. (Compared to Example 9.16, we have added the point at infinity in the one-point compactification.) In particular, $(\Omega_W, r_{W \circ W}) = (\Omega_W, \sigma)$ is compact but (Ω_W, r_W)

is not, and the two topologies are different so the metrics are not equivalent. The weak closure $\overline{\Omega_W}^{\sigma} = \Omega_W$.

10. Random-free graphons

Lovász and Szegedy [51] have studied the class of graph limits represented by $\{0,1\}$-valued graphons (and the corresponding graph properties); with a slight variation of their terminology we call such graphons and graph limits *random-free* (a reason for the name is given in Remark D.2):

Definition 10.1. A *random-free* graphon is a graphon W with values in $\{0,1\}$ a.e.

By Corollary 8.12, every graphon equivalent to a random-free graphon is random-free. Note that every graphon W_G defined by a graph as in Example 2.7 is random-free. (A reason for the name random-free is given in Remark D.2.)

Example 10.2. It is shown by Diaconis, Holmes and Janson [22] that every graph limit that is a limit of a sequence of threshold graphs can be represented by a graphon that is random-free (and has a monotonicity property, studied further in [11]). Hence every representing graphon is random-free, i.e., if G_n are threshold graphs and W is a graphon such that $G_n \to W$, then W is random-free.

Example 10.3. It is shown by Diaconis, Holmes and Janson [23] that every graph limit that is a limit of a sequence of interval graphs can be represented by the graphon $W(x,y) := \mathbf{1}\{x \cap y \neq \emptyset\}$ on the space $\Omega := \{[a,b] : 0 \le a \le b \le 1\}$ of all closed subintervals of $[0,1]$, equipped with some Borel probability measure μ. (Note that Ω and W are fixed, but μ varies.) Hence every graphon representing an interval graph limit is random-free. (This includes the threshold graph limits in Example 10.2 as a subset. The explicit representations in [22] and [23] are different, however.)

Lemma 10.4. *Let W be a graphon. Then the following are equivalent.*

(i) *W is random-free.*
(ii) $\int_{\Omega^2} W(1-W) = 0$.
(iii) $\int_{\Omega^2} W^2 = \int_{\Omega^2} W$.

Proof. This is trivial, noting that W is random-free if and only if $W(1-W) = 0$ a.e., and that $W(1-W) \ge 0$ for every graphon. □

Recall that $W \mapsto \int_{\Omega^2} W^2$ is not continuous for $\delta_\square$, see Example C.3; we therefore cannot conclude that the set of random-free graphons is closed. In fact, it is not; on the contrary, this set is dense in the space of all graphons.

Lemma 10.5. *The set of random-free graphons is dense in the space of all graphons. In other words, given any graphon W, on any probability space Ω, there exists a sequence of random-free graphons W_n such that $\delta_\square(W_n, W) \to 0$.*

Proof. By Remark B.2, there exists a sequence (G_n) of graphs such that $\delta_\square(W_{G_n}, W) \to 0$. Each W_{G_n} is random-free. $\square$

In contrast, the set is closed in the stronger metric δ_1. (See Pikhurko [55].)

Lemma 10.6. *The set of random-free graphons is closed in the space of all graphons equipped with the metric δ_1. In other words, if W and W_n are graphons, on any probability spaces, such that $\delta_1(W_n, W) \to 0$, and every W_n is random-free, then W is random-free.*

Proof. Let $F(x) := x(1-x)$. Then $F : [0,1] \to [0,1]$ and $|F'(x)| \le 1$ so $|F(x)-F(y)| \le |x-y|$ for $x, y \in [0,1]$. It follows easily that if W_n and W are graphons with $\delta_1(W_n, W) \to 0$, then $\delta_1(F(W_n), F(W)) \le \delta_1(W_n, W) \to 0$ and $|\int F(W_n) - \int F(W)| \to 0$. Since W_n is random-free, $\int F(W_n) = 0$ by Lemma 10.4 for each n, and thus $\int F(W) = 0$. By Lemma 10.4 again, this shows that W is random free. $\square$

We continue to investigate the metric δ_1 in connection with random-free graphons.

Lemma 10.7. *Let W_1 and W_2 be graphons on a probability space Ω, and let W_1' be a random-free n-step graphon on the same space. Then*

$$\|W_1 - W_2\|_{L^1(\Omega^2)} \le n^2 \|W_1 - W_2\|_\square + 2\|W_1 - W_1'\|_{L^1(\Omega^2)}. \tag{10.1}$$

Proof. Let $\{A_i\}_1^n$ be a partition of Ω such that W_1' is constant 0 or 1 on each $A_i \times A_j$.

If $W_1' = 0$ on $A_i \times A_j$, then

$$\begin{aligned}\iint_{A_i\times A_j} |W_1' - W_2| = \iint_{A_i\times A_j} W_2 &\le \iint_{A_i\times A_j} W_1 + \|W_1 - W_2\|_\square \\ &= \iint_{A_i\times A_j} |W_1 - W_1'| + \|W_1 - W_2\|_\square.\end{aligned}$$

If $W_1' = 1$ on $A_i \times A_j$, then

$$\begin{aligned}\iint_{A_i\times A_j} |W_1' - W_2| = \iint_{A_i\times A_j} (1 - W_2) &\le \iint_{A_i\times A_j} (1 - W_1) + \|W_1 - W_2\|_\square \\ &= \iint_{A_i\times A_j} |W_1 - W_1'| + \|W_1 - W_2\|_\square.\end{aligned}$$

Thus, in both cases $\iint_{A_i\times A_j} |W_1' - W_2| \le \iint_{A_i\times A_j} |W_1 - W_1'| + \|W_1 - W_2\|_\square$, and summing over all i and j yields

$$\|W_1' - W_2\|_{L^1} \le \|W_1 - W_1'\|_{L^1} + n^2 \|W_1 - W_2\|_\square.$$

The result follows by $\|W_1 - W_2\|_{L^1} \le \|W_1 - W_1'\|_{L^1} + \|W_1' - W_2\|_{L^1}$. $\square$

Remark 10.8. In particular, if W_1 is a random-free n-step graphon and W_2 an arbitrary graphon on the same probability space, then

$$\|W_1 - W_2\|_{L^1} \le n^2 \|W_1 - W_2\|_\square. \tag{10.2}$$

The constant n^2 in Lemma 10.7 and (10.2) is good enough for our purposes, but it is not the best possible, and it may easily be improved. In fact, an inspection of the proof shows that if we let $a_{ij} := \int_{A_i \times A_j} (W_1 - W_2)$, then we have simply estimated $|a_{ij}| \le \|W_1 - W_2\|_\square$ and thus $\sum_{i,j} |a_{ij}| \le n^2 \|W_1 - W_2\|_\square$. To obtain a better estimate, we use an inequality by Littlewood [46], see also [7] and [60, §6.2], which yields

$$\sum_i \Big(\sum_j |a_{ij}|^2 \Big)^{1/2} \le \sqrt{3} \sup_{\varepsilon_i, \varepsilon'_j = \pm 1} \Big| \sum_i \sum_j \varepsilon_i \varepsilon'_j a_{ij} \Big| \le \sqrt{3}\, \|W_1 - W_2\|_{\square,2}. \tag{10.3}$$

Consequently, by the Cauchy–Schwarz inequality,

$$\sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \le \sum_{i=1}^n n^{1/2} \Big(\sum_{j=1}^n |a_{ij}|^2 \Big)^{1/2} \le \sqrt{3}\, \sqrt{n}\, \|W_1 - W_2\|_{\square,2}, \tag{10.4}$$

which shows that n^2 in (10.1) and (10.2) can be replaced by $\sqrt{3n}$. Furthermore, the constant $\sqrt{3}$, which is implicit in [46], has been improved to $\sqrt{2}$ by Szarek [58]. (Szarek actually proved that $\sqrt{2}$ is the sharp constant in Khinchin's inequality, which implies Littlewood's, see [7]. See also [32] for related results.) Consequently, n^2 in (10.1) and (10.2) can be replaced by $\sqrt{2n}$.

This is, within a numerical constant, the best constant in these inequalities, as shown by the following examples which all yield a lower bound of order $n^{1/2}$.

Example 10.9. Let W be a symmetric Hadamard matrix of order n (i.e., a matrix with ± 1 entries and all rows ortogonal); such matrices exists at least if $n = 2^k$ for some k. (Take tensor powers of $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$.) We have $W = W_+ - W_-$ where $W_\pm$ are graphons on $[n]$. (We equip $[n]$ with the uniform probability distribution.)

Then $\|W\|_{L^1} = 1$, since $|W| = 1$. In order to estimate $\|W\|_{\square,2}$, let f and g be two functions $[n] \to [-1, 1]$, see the definition (4.3). Write $W = (w_{ij})_{i,j=1}^n$ and change notation to $a_i = f(i)$, $b_j = g(j)$; thus $|a_i|, |b_j| \le 1$.

Since W is a Hadamard matrix, the normalized matrix $n^{-1/2} W$ is orthogonal, and is thus an isometry as an operator in $\mathbb{R}^n$ (with the usual Euclidean norm); hence, W has norm $\sqrt{n}$. Consequently,

$$\int_{[n]^2} W(x,y) f(x) g(y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) = n^{-2} \sum_{i,j=1}^n a_i w_{ij} b_j$$

$$\le n^{-2}\sqrt{n}\left(\sum_{i=1}^n a_i^2\right)^{1/2}\left(\sum_{i=1}^n b_j^2\right)^{1/2} \le n^{-1/2}.$$

Hence $\|W\|_{\square,2} \le n^{-1/2}$ while $\|W\|_{L^1} = 1$, and thus the best constant in (10.2) is at least $\sqrt{n}$ for n such that a symmetric Hadamard matrix exists, and hence at least $\sqrt{n/2}$ for any n. See further [46] and [60, §6.3].

Example 10.10. Let q be a prime power with $q \equiv 1 \pmod 4$ and consider the Paley graph P_q, see [8, Section 13.2]; the vertex set of P_q is the finite field $\mathbb{F}_q$ and there is an edge xy if $x-y$ is a square in $\mathbb{F}_q$. Let $W_1 := W_{P_q}$ and $W_2 = 1/2$; then W_1 is a random-free q-step graphon, and $\|W_1 - W_2\|_{L^1} = 1/2$, since $W_1 - W_2 = \pm 1/2$ everywhere. By [8, Theorem 13.13] (and its proof, or Lemma E.1 below), $\|W_1 - W_2\|_{\square,1} = O(q^{-1/2})$. Hence, the constant in (10.2) is at least $\Omega(q^{1/2})$, for $n = q$ of this type. Since primes of the type $4k+1$ are dense in the natural numbers, it follows again that the constant is $\Omega(n^{1/2})$ for all n.

Example 10.11. We can use a random graph $G = G(n, 1/2)$ and let $W_1 := W_G$ and, again, $W_2 := 1/2$. Thus $\|W_1 - W_2\|_{L^1} = 1/2$. (Note that the Payley graph in Example 10.10 is an example of a quasirandom graph, so the two examples are related.)

We use for convenience the version $\|\ \|_{\square,4}$ of the cutnorm in Appendix E. If $S, T \subset [n]$ are disjoint, then $n^2 \int_{S\times T} W_G$ is the number of edges between S and T, and has thus a binomial distribution $\mathrm{Bi}(st, 1/2)$ where $s := |S|$ and $t := |T|$. Hence, a Chernoff bound [41, Remark 2.5] shows that, for any $c > 0$,

$$\begin{aligned}\mathbb{P}\Big(\Big|\int_{S\times T}(W_1 - W_2)\Big| > cn^{-1/2}\Big) &= \mathbb{P}\Big(\Big|n^2\int_{S\times T}(W_G - \mathbb{E}\,W_G)\Big| > cn^{3/2}\Big)\\ &\le 2\exp\Big(-\frac{2(cn^{3/2})^2}{st}\Big) \le 2\exp\Big(-\frac{2c^2n^3}{n^2/4}\Big) = 2\exp\Big(-8c^2n\Big),\end{aligned}$$

since $st \le s(n-s) \le n^2/4$. There are 3^n pairs S, T of disjoint subsets, and thus

$$\mathbb{P}\big(\|W_1 - W_2\|_{\square,4} > cn^{-1/2}\big) \le 2\cdot 3^n \exp\big(-8c^2n\big) = 2\exp\big((\log 3 - 8c^2)n\big).$$

Consequently, choosing for simplicity $c = 1$, so $8c^2 > \log 3$, with high probability $\|W_1 - W_2\|_{\square,4} \le n^{1/2}$, and thus by Lemma E.2

$$\|W_1 - W_2\|_{\square,1} \le 4n^{-1/2} = 8n^{-1/2}\|W_1 - W_2\|_{L^1},$$

showing that the best constant in (10.2) is at least $\frac{1}{8}n^{1/2}$ (for $\|\ \|_{\square,1}$).

Pikhurko [55] showed that convergence to a graphon W in δ_1 is equivalent to convergence in $\delta_\square$ if W is random-free, but not otherwise. To show this, we first show the corresponding result for $\|\ \|_\square$.

Lemma 10.12. *Let W and $W_1, W_2, \dots$ be graphons on a probability space Ω, and assume that W is random-free. Then $\|W_n - W\|_\square \to 0$ as $n \to \infty$ if and only if $\|W_n - W\|_{L^1(\Omega^2)} \to 0$.*

Proof. Assume $\|W_n - W\|_\square \to 0$. W is the indicator $\mathbf{1}_A$ of a measurable set $A \subseteq \Omega^2$. Any such set can be approximated in measure by a finite disjoint union of rectangle sets $\bigcup_i A_i \times B_i$, and we may assume that this set is symmetric since A is; in other words, given any $\varepsilon > 0$, there exists a $\{0,1\}$-valued step graphon W' such that $\|W - W'\|_{L^1} < \varepsilon$. Let the corresponding partition have $N = N(\varepsilon)$ parts. Lemma 10.7 then yields

$$\|W - W_n\|_{L^1} \le N^2 \|W - W_n\|_\square + 2\varepsilon \to 2\varepsilon$$

as $n \to \infty$. Hence, $\limsup_{n\to\infty} \|W - W_n\|_{L^1} = 0$.

The converse is obvious. □

Theorem 10.13 ([55])**.** *Let W and $W_1, W_2, \dots$ be graphons defined on some probability spaces, and assume that W is random-free. Then $\delta_\square(W_n, W) \to 0$ as $n \to \infty$ if and only if $\delta_1(W_n, W) \to 0$.*

Proof. Assume that $\delta_\square(W, W_n) \to 0$. By replacing the graphons by equivalent ones, we may by Theorem 7.1 assume that all graphons are defined on $[0,1]$. By Theorem 6.9, we may then find measure-preserving bijections $\varphi_n : [0,1] \to [0,1]$ such that $\|W - W_n^{\varphi_n}\|_\square < \delta_\square(W, W_n) + 1/n \to 0$. Hence, $\|W - W_n^{\varphi_n}\|_{L^1(\Omega^2)} \to 0$ by Lemma 10.12, and thus $\delta_1(W, W_n) \to 0$.

The converse is obvious. □

As said above, it is here necessary that W is random-free [55], which is seen from the following theorem.

Theorem 10.14. *Let W be a graphon. Then W is random-free if and only if $\delta_1(W_{G_n}, W) \to 0$ for some sequence of graphs G_n.*

Proof. There exists a sequence of graphs G_n with $\delta_\square(W_{G_n}, W) \to 0$ by Remark B.2. If W is random-free, then $\delta_1(W_{G_n}, W) \to 0$ by Theorem 10.13.

The converse follows by Lemma 10.6, since each W_{G_n} is random-free. □

Theorem 10.15. *Let W be a graphon. Then the following are equivalent.*

(i) *W is random-free.*
(ii) *$\int W_n^2 \to \int W^2$ whenever (W_n) is a sequence of graphons such that $\delta_\square(W_n, W) \to 0$.*
(iii) *$t(F, W_n) \to t(F, W)$ for every multigraph F whenever (W_n) is a sequence of graphons such that $\delta_\square(W_n, W) \to 0$.*

Proof. (i) $\implies$ (iii): If W is random-free and $\delta_\square(W_n, W) \to 0$, then Theorem 10.13 yields $\delta_1(W_n, W) \to 0$, and thus $t(F, W_n) \to t(F, W)$ for every multigraph F by Lemma C.4.

(iii) $\implies$ (ii): Immediate by taking F to be a double edge, see Example C.1.

(ii) $\implies$ (i): Take a sequence of graphs G_n such that $G_n \to W$, see Remark B.2; thus $\delta_\square(W_{G_n}, W) \to 0$. Hence $\int W_{G_n} \to \int W$. Further, every W_{G_n} is $\{0,1\}$-valued, so $W_{G_n}^2 = W_{G_n}$; hence

$$\int W_{G_n}^2 = \int W_{G_n} \to \int W.$$

If (ii) holds, then also $\int W_{G_n}^2 \to \int W^2$. Hence $\int W^2 = \int W$, so W is random-free by Lemma 10.4. $\square$

Finally, we mention two characterizations of random-free graphons in terms of the finite or infinite random graph $G(n, W)$ defined in Appendix D. First the finite case and entropy.

Theorem 10.16. *Let W be a graphon. Then W is random-free if and only if the entropy $\mathcal{E}(G(n, W)) = o(n^2)$ as $n \to \infty$.*

Proof. This is an immediate consequence of Theorem D.5, since $h \geq 0$ and thus the right-hand side of (D.1) vanishes if and only $h(W(x, y)) = 0$ a.e., which is equivalent to $W(x, y) \in \{0, 1\}$ a.e. $\square$

Problem 10.17. We may, as in [2, (15.30)], ask for the exact growth rate of $\mathcal{E}(G(n, W))$ for a random-free graphon W. It is easily seen that if W is a step graphon, then $\mathcal{E}(G(n, W)) = O(n)$; we conjecture that the converse holds too. As another example, for the "half graphon" $W(x, y) = \mathbf{1}\{x + y > 1\}$ on $[0, 1]$, it can be shown (e.g. using [22, Corollary 6.6]) that $\mathcal{E}(G(n, W)) = n \log n + O(n)$.

We represent the infinite random graph $G(\infty, W)$ by the family of indicator variables $J_{ij} := \mathbf{1}\{ij \text{ is an edge}\}$, $1 \leq i < j \leq \infty$. We define the *shell* σ-field (or *big tail* σ-field) to be the intersection

$$\mathcal{S} := \bigcap_{n=1}^{\infty} \sigma\{J_{ij} : i < j,\, j \geq n\} \tag{10.5}$$

of the σ-fields generated by all J_{ij} where at least one index is "big". Recall that a random variable is *a.s. $\mathcal{S}$-measurable* (or *essentially $\mathcal{S}$-measurable*) if it is a.s. equal to an $\mathcal{S}$-measurable variable; equivalently, it is measurable for the completion $\widehat{\mathcal{S}}$ of $\mathcal{S}$.

Theorem 10.18. *The following are equivalent for a graphon W:*

(i) *W is random-free.*
(ii) *The infinite random graph $G(\infty, W)$ is a.s. $\mathcal{S}$-measurable.*
(iii) *The indicator $J_{12} := \mathbf{1}\{12$ is an edge in $G(\infty, W)\}$ is a.s. $\mathcal{S}$-measurable.*

Proof. This is the symmetric version of [1, Proposition 3.6], see also [2, (14.15) and p. 133] and [21, (4.9)]. Since the details for the symmetric case are not given in these references, we give some of them for completeness.

First, note that we can write the definition of $G(\infty, W)$ in Appendix D as

$$J_{ij} = \mathbf{1}\{\xi_{ij} \le W(X_i, X_j)\}, \tag{10.6}$$

where ξ_{ij}, for $1 \le i < j$, and X_i, for $i \ge 1$, all are independent, and X_i has distribution μ on Ω while ξ_{ij} is uniform on $[0,1]$.

(i) $\implies$ (iii): If W is random-free, then (10.6) simplifies to $J_{ij} = W(X_i, X_j)$. Consider the array $(J_{2i-1,2j})_{i,j=1}^{\infty} = (W(X_{2i-1}, X_{2j})_{i,j=1}^{\infty}$, where the first index is odd and the second even; this is a separately exchangeable array, and by [1, Proposition 3.6] (or as a simple consequence of [43, Proposition 7.31]), it is a.s. $\mathcal{S}'$-measurable for the shell σ-field of this array. Since $\mathcal{S}' \subseteq \mathcal{S}$, (iii) follows.

(iii) $\iff$ (ii): $\mathcal{S}$ is invariant under finite permutations, so the exchangeability implies that every J_{ij} is $\mathcal{S}$-measurable if J_{12} is. The converse is trivial.

(iii) $\implies$ (i): It follows from (10.5) and (10.6) that ξ_{12} is independent of $\mathcal{S}$. If (iii) holds, then J_{12} is thus independent of ξ_{12}, which by (10.6) implies that $J_{12} = \mathbb{E}(J_{12} \mid X_1, X_2) = W(X_1, X_2)$ a.s., so W is $\{0,1\}$-valued a.e. $\square$

Appendix A. Special probability spaces

A.1. **Atoms.** An *atom* in a probability space (Ω, μ) is a subset A with $\mu(A) > 0$ such that every subset $B \subseteq A$ satisfies $\mu(B) = 0$ or $\mu(B) = \mu(A)$.

We say that Ω is *atomless* if there are no atoms.

Lemma A.1. *If (Ω, μ) is an atomless probability space, then there exists a family $(A_r)_{r\in[0,1]}$ of measurable sets such that $\mu(A_r) = r$ for every $r \in [0,1]$, and further $A_r \subseteq A_s$ if $r < s$ (i.e., the family is increasing).*

Proof. Consider families $(A_r)_{r\in E}$ with these properties, defined on some arbitrary subset E of $[0,1]$. By Zorn's lemma, there exists a maximal family; we claim that then $E = [0,1]$. In fact, $0, 1 \in E$, since we otherwise could enlarge the family by defining $A_0 = \emptyset$ or $A_1 = \Omega$. Further, E is closed, since otherwise there would exists $r \notin E$ and a sequence $r_n \in E$ such that either $r_n \nearrow r$ or $r_n \searrow r$; in the first case we can define $A_r := \bigcup_n A_{r_n}$, and in the second case $A_r := \bigcap_n A_{r_n}$. Finally, if $E \ne [0,1]$, the complement $[0,1] \setminus E$ thus is open, and thus a disjoint union of open intervals. Let (a,b) be one of these intervals. Then $a, b \in E$, and $A_b \setminus A_a$ is a set of measure $b - a > 0$. Since μ is atomless, there exists a subset $C \subseteq A_b \setminus A_a$ with $0 < \mu(C) < b-a$, but in this case, the family could be extended by $A_{a+\mu(C)} := A \cup C$, so we again contradict the maximality of the family. Hence $E = [0,1]$, which completes the proof. $\square$

We also give a reformulation in terms of a map to $[0,1]$.

Lemma A.2. *If (Ω, μ) is an atomless probability space, then there exists a measure-preserving map $\varphi : \Omega \to [0,1]$.*

Proof. Let $(A_r)_r$ be as in Lemma A.1, and define $\varphi(x) := \inf\{r \in [0,1] : x \in A_r\}$ (assuming as we may that $A_1 = \Omega$). $\square$

Lemma A.3. *If $\varphi : \Omega_1 \to \Omega_2$ is measure-preserving and Ω_2 is atomless, then Ω_1 is atomless too.*

Proof. Let $(A_r)_{r\in[0,1]}$ be a family of subsets of Ω_2 with the properties in Lemma A.1. Then $B_r := \varphi^{-1}(A_r)$ defines a family of subsets of Ω_1 with the same properties. Suppose that $A \subseteq \Omega_1$ is an atom. Then, for each r, $\mu(A \cap B_r) = 0$ or $\mu(A)$. Let $r_0 := \sup\{r : \mu(A \cap B_r) = 0\}$, and take any $r_- < r_0$ and $r_+ > r_0$. (If $r_0 = 0$, take $r_- = 0$, and if $r_0 = 1$, take $r_+ = 1$.) Then $\mu(A \cap B_{r_-}) = 0$ and $\mu(A \cap B_{r_+}) = \mu(A)$, so

$$\begin{aligned}\mu(A) &= \mu(A \cap B_{r_+}) - \mu(A \cap B_{r_-}) = \mu(A \cap (B_{r_+} \setminus B_{r_-}))\\ &\le \mu(B_{r_+} \setminus B_{r_-}) = r_+ - r_-.\end{aligned}$$

This is a contradiction, since $\mu(A) > 0$ while $r_+ - r_-$ can be arbitrarily small. $\square$

In the opposite direction, there are typically many measure-preserving maps from an atomless space Ω_1 into a space with atoms. Simple examples are the trivial map onto a one-point space, and the indicator function of a subset $B \subseteq \Omega_1$ seen as a map $(\Omega_1, \mu) \to (\{0,1\}, \nu)$, which is measure-preserving if $\nu\{1\} = \mu(B)$.

A.2. **Borel spaces.** To define Borel spaces, it is simplest to begin with measurable spaces, without any particular measures.

We say that two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are *isomorphic* if there is a bimeasurable bijection $\varphi : \Omega \to \Omega'$, i.e., a bijection such that both φ and φ^{-1} are measurable. (Similarly, two probability spaces $(\Omega, \mathcal{F}, \mu)$ and $(\Omega', \mathcal{F}', \mu')$ are isomorphic if there exists a bimeasurable bijection that further is measure-preserving.)

A measurable space is *Borel* (also called *standard* [19] or *Lusin* [20]) if it is isomorphic to a Borel subset of a Polish space (i.e., a complete metric space) with its Borel σ-field. A probability space $(\Omega, \mathcal{F}, \mu)$ is *Borel* if $(\Omega, \mathcal{F})$ is a Borel measurable space; equivalently, if it is isomorphic to a Borel subset of a Polish space equipped with a Borel measure.

In fact, we do not need arbitrary Polish spaces here; the following theorem shows that it suffices to consider subsets of $[0,1]$. We tacitly assume that $[0,1]$ and other Polish spaces are equipped with their Borel σ-fields.

Theorem A.4. *The following are equivalent for a measurable space $(\Omega, \mathcal{F})$, and thus each property characterizes Borel measurable spaces.*

(i) *$(\Omega, \mathcal{F})$ is isomorphic to a Borel subset of a Polish space.*
(ii) *$(\Omega, \mathcal{F})$ is isomorphic to a Polish space.*
(iii) *$(\Omega, \mathcal{F})$ is isomorphic to a Borel subset of $[0,1]$.*
(iv) *$(\Omega, \mathcal{F})$ is either countable (with all subsets measurable), or isomorphic to $[0,1]$.*

For a proof, see e.g. [19, Theorem 8.3.6] or [54, Theorem I.2.12]. An essentially equivalent statement is that any two Borel measurable spaces with the same cardinality are isomorphic.

Hence, a Borel probability space is either countable or isomorphic to $[0,1]$ equipped with some Borel probability measure. Consequently we can, when dealing with Borel spaces, restrict ourselves to $[0,1]$ without much loss of generality (the countable case is typically simple), but for applications it is convenient to allow general Borel spaces.

Remark A.5. Another simple Borel space is the *Cantor cube* $\mathcal{C} := \{0,1\}^\infty$ (which up to homeomorphism is the same as the usual *Cantor set*); this is a compact metric space, and thus a Polish space. Since $\mathcal{C}$ is uncountable, it is by Theorem A.4 isomorphic to $[0,1]$ as measurable spaces; consequently we may replace $[0,1]$ by $\mathcal{C}$ in Theorem A.4.

One important property of Borel spaces is the following theorem by Kuratowski, showing that a measurable bijection is bimeasurable, and thus an isomorphism.

Theorem A.6. *Let Ω and Ω' be Borel measurable spaces. If $f : \Omega \to \Omega'$ is a bijection that is measurable, then $f^{-1} : \Omega' \to \Omega$ is measurable, and thus f is an isomorphism.*

More generally, if $f : \Omega \to \Omega'$ is a measurable injection, then the image $f(\Omega)$ is a measurable subset of Ω' and f is an isomorphism of Ω onto $f(\Omega)$.

For a proof, see e.g. [19, Proposition 8.3.5 and Theorem 8.3.7]; see also further results in [19, Sections 8.3 and 8.6].

Let us now add measures to the spaces. There is a version of Theorem A.4 for probability spaces. For simplicity we begin with the atomless case. Recall that λ denotes the Lebegue measure.

Theorem A.7. *If (Ω,μ) is an atomless Borel probability space, then there exists a measure-preserving bijection of (Ω,μ) onto $([0,1],\lambda)$.*

In other words, all atomless Borel probability space are isomorphic, as measure spaces.

Proof. Since Ω is atomless, every point has measure 0 and thus every countable subset has measure 0; in particular, Ω cannot be countable. By Theorem A.4(iv), there exists a bimeasurable bijection φ_1 of Ω onto $[0,1]$. This maps the measure μ onto some Borel measure ν on $[0,1]$.

Since ν has no atoms, $x \mapsto \nu([0,x])$ is a continuous non-decreasing map of $[0,1]$ onto itself. We let $\psi : [0,1] \to [0,1]$ be its right-continuous inverse defined by

$$\psi(t) := \sup\bigl\{x \in [0,1] : \nu([0,x]) \le t\bigr\}. \tag{A.1}$$

Then $\nu([0,\psi(t)]) = t$ for every $t \in [0,1]$, which implies that ψ is strictly increasing. Hence, ψ is injective and measurable, and by Theorem A.6, ψ is a bimeasurable bijection of $[0,1]$ onto some Borel subset $B := \psi([0,1])$.

It follows from (A.1) that, for all $s,t \in [0,1]$, $\psi(t) \geq s \iff \nu([0,s]) \leq t$, and thus $\psi^{-1}([0,s)) = [0,\nu([0,s]))$. Hence,

$$\lambda\bigl(\psi^{-1}([0,s))\bigr) = \nu\bigl([0,s]\bigr) = \nu\bigl([0,s)\bigr), \qquad s \in [0,1],$$

which implies that $\lambda^\psi = \nu$ (see Remark 5.4 for the notation), i.e., that $\psi : ([0,1],\lambda) \to ([0,1],\nu)$ is measure-preserving.

Consequently, ψ is a measure-preserving bijection $\psi : ([0,1],\lambda) \to (B,\nu)$. Choose an uncountable null set $N \subseteq [0,1]$ (for example the Cantor set). Then $N' := \psi(N)$ is an uncountable null set in (B,ν). The restriction of ψ to $[0,1] \setminus N$ is a measure-preserving bijection onto $B \setminus N'$. Further, N and $N' \cup B^{\mathsf{c}}$, where $B^{\mathsf{c}} := [0,1] \setminus B$, are both uncountable Borel subsets of $[0,1]$, and thus by Theorem A.4, both are isomorphic as measurable spaces to $[0,1]$, and thus to each other. Hence there exists a measurable bijection $\psi_1 : N \to N' \cup B^{\mathsf{c}}$.

Define $\psi_2 : [0,1] \to [0,1]$ by $\psi_2(x) = \psi(x)$ when $x \notin N$ and $\psi_2(x) = \psi_1(x)$ when $x \in N$. Then ψ is a measure-preserving bijection $([0,1],\lambda) \to ([0,1],\nu)$. Consequently, $\psi_2^{-1} \circ \varphi$ is a measure-preserving bijection of (Ω,μ) onto $([0,1],\lambda)$. $\square$

It is easy to handle atoms too. An atom in a Borel probability space is, up to a null set, just a single point with a point mass; hence, a Borel space is atomless if and only if it has no point masses, i.e. no point with positive measure. In any Borel probability space there is at most a countable number of point masses, and removing them we obtain an atomless Borel measure space. This leads to the following characterization.

Theorem A.8. *A probability space is Borel if and only if it is isomorphic, by a measure-preserving bijection, to one of the following spaces.*

(i) *A countable set $\mathcal{D} = \{x_i\}_{i=1}^n$ (where $n \leq \infty$), with all subsets measurable and the discrete measure given by $\mu(A) = \sum_{i:x_i \in A} p_i$, for some $p_i \geq 0$. (Necessarily $\sum_i p_i = 1$.)*

(ii) *The disjoint union $\mathcal{D} \cup N$, where $\mathcal{D}$ is as in* (i) *and N is a null set given by any given uncountable Borel measurable space equipped with zero measure. (We may choose for example $N = [0,1]$ with zero measure, or the Cantor set with λ, which vanishes there.)*

(iii) *The disjoint union of a closed interval $([0,r],\lambda)$ with $0 < r \leq 1$ and a countable set $\mathcal{D}$ as in* (i) *(possibly empty); in this case $r + \sum_i p_i = 1$, and we may further assume that each $p_i > 0$.*

Proof. If (Ω,μ) is a Borel probability space, let $D := \{x \in \Omega : \mu\{x\} > 0\}$ and $\Omega' := D^{\mathsf{c}} = \Omega \setminus D$. Then D is countable, and (Ω',μ) is atomless. Let $r := \mu(\Omega')$. If $r > 0$, then by Theorem A.4 and a scaling, (Ω',μ) is isomorphic to $[0,r]$, which yields (iii). If $r = 0$, then Ω' is a null set. If further Ω' is uncountable, then $(\Omega',\mu) = (\Omega',0)$ is isomorphic to $(N,0)$ for any uncountable Borel space by Theorem A.4(iv), which yields (ii). Finally, if Ω' is countable, then Ω is countable and (i) holds.

The converse is obvious. □

Theorem A.9. *If Ω is a Borel probability space, then there is a measure-preserving map $[0,1] \to \Omega$.*

Proof. It suffices to show this for the spaces in Theorem A.8(i)–(iii), and for these it is easy to construct explicit maps. (For each $x \in \mathcal{D}$, map a suitable interval of length $\mu\{x\}$ to x; in (iii), map $[0,r]$ onto itself by the identity map.) □

A.3. **Lebesgue spaces.** A *Lebesgue probability space* is a probability space that is the completion of a Borel probability space; equivalently (see Theorem A.4), it is isomorphic to a Polish space (or, equivalently, a Borel subset of a Polish space) equipped with the completion of a Borel measure.

Theorem A.8 leads directly to the following characterization.

Theorem A.10. *A probability space is Lebesgue if and only if it is isomorphic, by a measure-preserving bijection, to one of the spaces given in Theorem A.8, with the modifications that in* (ii) *all subsets of N are measurable (with measure* 0*), and in* (iii) *the interval $[0,r]$ is equipped with the Lebesgue σ-field $\mathcal{L}$.* □

In other words, every Lebesgue probability space is, possibly ignoring a null sets, isomorphic to either a countable discrete space, an interval $([0,r],\mathcal{L},\lambda)$, or a disjoint union of an interval and a countable discrete part.

Corollary A.11. *An atomless Lebesgue space is isomorphic to* $([0,1],\mathcal{L},\lambda)$.

Proof. Immediate from either Theorem A.10 or Theorem A.4. □

Remark A.12. Lebesgue spaces were introduced by Rohlin [57] by a different, intrinsic, definition, see also Haezendonck [34]. The equivalence to the definition above follows from [57, §2.4] or [34, Remark 2, p. 250].

Appendix B. Graph limits

As said in the introduction, graph limits were introduced by Lovász and Szegedy [48] and further developed by Borgs, Chayes, Lovász, Sós and Vesztergombi [14, 15]. The central idea in graph limit theory is to assign limits to (some) sequences G_n of (unlabelled) graphs with $|G_n| \to \infty$. Part of the importance of this notion is the fact that several different definitions of convergence turn out to be equivalent. One definition is the following, which has the advantage that it easily is adapted to many other situations such as hypergraphs, bipartite graphs, directed graphs, compactly decorated graphs and posets, see [4, 14, 15, 24, 26, 36, 40, 44, 51, 52].

For each $k \le |G_n|$, let $G_n[k]$ be the random induced subgraph of G_n with k vertices obtained by selecting k (distinct) vertices $v_1,\dots,v_k \in G_n$ at random (uniformly); we regard $G_n[k]$ as a labelled graph with the vertices labelled $1,\dots,k$; equivalently, we regard $G_n[k]$ as a graph with vertex set $\{1,\dots,k\}$.

Definition B.1. A sequence of graphs (G_n) with $|G_n| \to \infty$ converges if for each fixed k, the distribution of the random graph $G_n[k]$ converges as $n \to \infty$.

In other words, for each k and each labelled graph G with $|G| = k$, we require that $\lim_{n\to\infty} \mathbb{P}(G_n[k] = G)$ exists.

Given this notion of convergence, graph limits can be defined abstractly, as equivalence classes of convergent sequences of graphs. Equivalently, one can easily introduce a metric on the set of unlabelled finite graphs such that the convergent sequences become the Cauchy sequences in the metric, and then construct the completion of this metric space.

It turns out that the space of limits can be identified with the quotient space $\widehat{\mathcal{W}} := \bigcup_\Omega \mathcal{W}(\Omega)/\cong$ defined in Section 6, see Lovász and Szegedy [48] and Borgs, Chayes, Lovász, Sós and Vesztergombi [14]. In other words, every graph limit is represented by a graphon, but non-uniquely, since every equivalent graphon represents the same graph limit. (Conversely, non-equivalent graphons represent different graph limits.)

Moreover, convergence to graph limits can be described by the cut metric. If (G_n) is a sequence of graphs with $|G_n| \to \infty$, and W is a graphon, then G_n converges to the graph limit represented by W if and only if $\delta_\square(W_{G_n}, W) \to 0$, where W_{G_n} is as in Example 2.7. In this case we also say that (G_n) converges to W, and write $G_n \to W$ (remembering the non-uniqueness of W).

Remark B.2. In particular, for every graphon W, there exist sequences of graphs (G_n) such that $G_n \to W$. (One construction of such G_n is the random construction in Appendix D below.)

Convergence to graph limits can also be described by the homomorphism densities defined in Appendix C: $G_n \to W$ if and only if $t(F, G_n) \to t(F, W)$ for every simple graph F.

For details and many other results, see Borgs, Chayes, Lovász, Sós and Vesztergombi [14]; for further aspects, see e.g. Austin [4], Bollobás and Riordan [12], Borgs, Chayes, Lovász, Sós and Vesztergombi [15], Diaconis and Janson [24], Lovász and Szegedy [51], and Appendix D below.

Appendix C. Homomorphism densities

Define, following [14] and [48], for a graphon (or, more generally, any bounded symmetric function) $W : \Omega^2 \to [0, 1]$ and a simple graph F vith vertex set $V(F)$ and edge set $E(F)$, the *the homomorphism density*

$$t(F, W) := \int_{\Omega^{V(F)}} \prod_{ij\in E(F)} W(x_i, x_j)\, \mathrm{d}\mu(x_1) \cdots \mathrm{d}\mu(x_{|F|}). \tag{C.1}$$

If X_i are i.i.d. random variables with values in Ω and distribution μ, we can write (C.1) as

$$t(F,W) := \mathbb{E} \prod_{ij\in E(F)} W(X_i,X_j). \tag{C.2}$$

The homomorphism densities can be defined for graphs too by $t(F,G) := t(F,W_G)$. It is easily seen that $t(F,G)$ is the proportion of maps $V(F) \to V(G)$ that are graph homomorphisms (or, equivalently, the probability that a random map $V(F) \to V(G)$ is a graph homomorphism. (This explains the name homomorphism density.)

The homomorhism denisities have a central place in the graph limit theory. In particular, as shown in [14], $G_n \to W$ if and only if $t(F,G_n) \to t(F,W)$ for every simple graph F.

The definition (C.1) makes sense also for loopless multigraphs F, where we allow repeated edges. (Loops are not allowed, since we want $t(F,W) = t(F,W')$ when $W = W'$ a.e., and this rules out a factor $W(x_i,x_i)$ in (C.1).)

Example C.1. Let M_k be the multigraph with 2 vertices connected by k parallel edges. Then $t(M_k,F) = \int_{\Omega^2} W^k$.

We have seen in Theorem 8.10 that $t(F,W) = t(F,W')$ when $W \cong W'$, for every multigraph F. In other words, the mapping $W \mapsto t(F,W)$ yields a well-defined mapping on the quotient space $\widehat{\mathcal{W}} := \mathcal{W}^*/\cong$, which is the same as the space of graph limits, see Appendix B.

Lemma C.2. *The mapping $W \mapsto t(F,W)$ is continuous on $(\widehat{\mathcal{W}},\delta_\square)$ if and only if F is a simple graph.*

In other words, if $\delta_\square(W_n,W) \to 0$, then $t(F,W_n) \to t(F,W)$ for every simple graph F. However, if F is a multigraph with parallel edges, then $\delta_\square(W',W) = 0$ implies $t(F,W') = t(F,W)$, but $\delta_\square(W_n,W) \to 0$ does not imply $t(F,W_n) \to t(F,W)$.

Proof. It is easy to see that $W \mapsto t(F,W)$ is continuous in $\delta_\square$ for every simple F, see [14] or [48]; more precisely, for any graphons W and W',

$$|t(F,W) - t(F,W')| \le |E(F)|\,\delta_\square(W,W'). \tag{C.3}$$

For the converse, suppose that the loopless multigraph F is not simple, and let F' be the simple graph obtained by identifying parallel edges in F. Thus $V(F') = V(F)$, but $|E(F')| < |E(F)|$.

Let W be the constant graphon $1/2$ defined on $[0,1]$, and let G_n be a sequence of graphs such that $G_n \to W$. (See Remark B.2. Such sequences are known as quasirandom, see [48]. For example, G_n can be a realization of the random graph $G(n,1/2)$, see Appendix D.)

Let W_{G_n} be the graphon corresponding to G_n as in Example 2.7; we thus have $\delta_\square(W_{G_n},W) \to 0$. On the other hand, W_{G_n} is $\{0,1\}$-valued, and thus $t(F,W_{G_n}) = t(F',W_{G_n})$ by (C.1). Hence, using the already proved part of

the lemma for F',

$$t(F, W_{G_n}) = t(F', W_{G_n}) \to t(F', W) = 2^{-|E(F')|} > 2^{-|E(F)|} = t(F, W). \quad \square$$

Example C.3. In particular, $W \mapsto t(K_2, W) = \int_{\Omega^2} W$ is continuous in the cut metric, but $W \mapsto t(M_2, W) = \int_{\Omega^2} W^2$, see Example C.1, is not. More generally, $W \mapsto \int_{\Omega^2} W^k$ is not continuous for any $k > 1$.

If we use the stronger metric δ_1, we have continuity for multigraphs too. (This metric is, however, much less useful.)

Lemma C.4. *The mapping $W \mapsto t(F, W)$ is continuous on $(\widehat{\mathcal{W}}, \delta_1)$ for every loopless multigraph.*

We omit the easy proof, similar to the proof for $\delta_\square$ and simple graphs in [14] or [48].

Appendix D. Graphons and random graphs

Let W be a graphon, defined on some probability space Ω. For $1 \le n \le \infty$, let $[n] = \{i \in \mathbb{N} : i \le n\}$; thus $[n] = \{1, \dots, n\}$ if n is finite and $[\infty] = \mathbb{N}$. We define a random graph $G(n, W)$ with vertex set $[n]$ by first taking an i.i.d. sequence $\{X_i\}_{i=1}^n$ of random points in Ω with the distribution μ, and then, given this sequence, letting ij be an edge in $G(n, W)$ with probability $W(X_i, X_j)$; for a given sequence $(X_i)_i$, this is done independently for all pairs $(i, j) \in [n]^2$ with $i < j$. (I.e., we first sample $X_1, X_2, \dots$ at random, and then toss a biased coin for each possible edge.)

The random graphs $G(n, W)$ thus generalize the standard random graphs $G(n, p)$ obtained by taking $W = p$ constant. Note that we may construct $G(n, W)$ for all n by first constructing $G(\infty, W)$ and then taking the subgraph induced by the first n vertices.

This construction was introduced in graph limit theory in [48] and [14]. (For other uses, see e.g. [9] and [21].)

Remark D.1. If F is a labelled graph, then the homomorphism density $t(F, W)$ in (C.1) equals the probability that F is a labelled subgraph of $G(\infty, W)$ (or of $G(n, W)$ for any $n \ge |F|$).

In particular, this shows that the family $\big(t(F, W)\big)_F$ and the distribution of $G(\infty, W)$ determine each other; see further Theorem 8.10 and [24].

Remark D.2. If W is a random-free graphon, i.e., $W(x, y) \in \{0, 1\}$ a.e., then the construction of $G(n, W)$ simplifies. We sample i.i.d. $X_1, X_2, \dots$ as before, and draw an edge ij if and only if $W(X_i, X_j) = 1$; thus the second random step in the construction disappears. (This explains the name "random-free"; of course, $G(n, W)$ still is random, but it is now a deterministic function of the random X_i.)

The infinite random graph $G(\infty, W)$ is an exchangeable random graph, i.e., its distribution is invariant under permutations of the vertices, and every exchangeable random graph is a mixture of such graphs, i.e., it can

be obtained by this construction with a random W. This is an instance of the representation theorem for exchangeable arrays by Aldous [1] and Hoover [35], see also Kallenberg [43]. Moreover, by Theorem 8.10, if W' is another graphon, then $G(\infty, W)$ and $G(\infty, W')$ have the same distribution if and only if $W \cong W'$. Consequently, the mapping $W \mapsto G(\infty, W)$ gives a bijection between the set $\widehat{\mathcal{W}} = \mathcal{W}^*/\cong$ of equivalence classes of graphons and a subset $\widehat{\mathcal{X}}$ of the set $\mathcal{X}$ of distributions of exchangeable infinite random graphs; this subset $\widehat{\mathcal{X}}$ is easily charaterized in several different ways, for example as follows.

Lemma D.3. *For an exchangeable infinite random graph $\overline{G}$, the following are equivalent, and thus all characterize $\mathcal{L}(\overline{G}) \in \widehat{\mathcal{X}}$.*

(i) $\overline{G} \overset{\mathrm{d}}{=} G(\infty, W)$ *for some graphon* W.
(ii) *The distribution* $\mathcal{L}(\overline{G})$ *is an extreme point in* $\mathcal{X}$.
(iii) $\overline{G}$ *is ergodic: every property that is (a.s.) invariant under left-shift (i.e., delete vertex* 1 *and its edges and relabel the remaining vertices* $i \mapsto i-1$*) has probability* 0 *or* 1.
(iv) *Every property of* $\overline{G}$ *that is (a.s.) invariant under finite permutations of the vertices has probability* 0 *or* 1.
(v) *For any two disjoint subsets of vertices* V_1 *and* V_2*, the induced subgraphs* $\overline{G}|_{V_1}$ *and* $\overline{G}|_{V_2}$ *are independent.*

Proof. See [24] and [43]. □

D.1. **Graph limits and random graphs.** There is also a simple connection between graph limits and exchangeable infinite random graphs. By Definition B.1, if (G_n) is a convergent sequence of graphs with $|G_n| \to \infty$, then for each k there exists a random graph $G[k]$ on the vertex set $[k]$ such that $G_n[k] \xrightarrow{\mathrm{d}} G[k]$. The distributions of $G[k]$ for different k are consistent, so by Kolmogorov's extension theorem, there exists a random infinite graph $\overline{G}$ on $[\infty]$ such that $G[k] \overset{\mathrm{d}}{=} \overline{G}|_{[k]}$, i.e., $G_n[k] \xrightarrow{\mathrm{d}} \overline{G}|_{[k]}$. Each $G_n[k]$ has an exchangeable distribution, and thus so has each $G[k]$; consequently, $\overline{G}$ is an exchangeable infinite random graph; furthermore, it is easily seen that $\overline{G}$ satisfies Lemma D.3(v), and thus its distribution belongs to $\widehat{\mathcal{X}}$. Thus every graph limit can be represented by an exchangeable infinite random graph with distribution in $\widehat{\mathcal{X}}$. Conversely, if $\overline{G}$ is any exchangeable infinite random graph with a distribution in $\widehat{\mathcal{X}}$, then the induced subgraphs $G_n := \overline{G}|_{[n]}$ a.s. satisfy $G_n[k] \xrightarrow{\mathrm{d}} \overline{G}|_{[k]}$ for every k, as can be seen from the limit theorem for reverse martingales [24] or directly [48]; thus the sequence (G_n) converges a.s., and its limit is represented by the infinite random graph $\overline{G}$.

This yields a bijection between the set of graph limits and the set $\widehat{\mathcal{X}}$, characterized in Lemma D.3, of distributions of exchangeable infinite random graphs.

This connection between graph limits and (distributions of) exchangeable infinite random graphs combines with the connection above between (equivalence classes of) graphons and (distributions of) exchangeable infinite random graphs to prove the central fact stated in Appendix B that there is a bijection between graph limits and equivalence classes of graphons; see further [4], [24], [44], [50].

In particular, for any graphon W, we have a.s. $G(n,W) \to W$ as $n \to \infty$, in the sense of Appendix B [48], cf. Remark B.2.

Remark D.4. This method of proving the connection between graph limits and graphons through the use of exchangeable infinite random graphs as an intermediary generalizes immediately to several extensions of the theory, and it may be used to find the correct analogue of graphons in new situations. See for example [4] (hypergraphs) and [24] (bipartite graphs, directed graphs).

Another example is compact decorated graphs [52], which are graphs with edges labelled by elements of a fixed second-countable compact space (i.e., a compact metrizable space [27, Theorem 4.2.8]) $\mathcal{K}$; this includes several interesting cases. $\mathcal{K}$-decorated graph limits are defined as in Definition B.1, now with $\mathcal{K}$-decorated graphs. The arguments sketched above show that there is a bijection between $\mathcal{K}$-decorated graph limits and distributions of exchangeable $\mathcal{K}$-decorated infinite random graphs satisfying the properties in Lemma D.3, and a further bijections to equivalence classes of graphons, where the graphons now take their values in the space $\mathcal{P}(\mathcal{K})$ of Borel probability measures on $\mathcal{K}$. (The representation theorem in [42] yields a representation where the label of ij is $f(X_i, X_j, \xi_{ij})$ for some fixed function $f : [0,1]^3 \to \mathcal{X}$ with X_i and ξ_{jk} uniform on $[0,1]$ and independent of each other; it is easily seen that this leads to an equivalent representation by $\mathcal{P}(\mathcal{K})$-valued graphons $W : [0,1]^2 \to \mathcal{P}(\mathcal{K})$.) For a different proof, see [52]. Many results in Sections 6–8 above extend to this case, but we leave that to the reader.

In fact, the arguments above on the equivalences work for any Polish space $\mathcal{K}$, also non-compact; however, compactness implies that the resulting space of decorated graph limits is compact, which is important for some results.

D.2. **Entropy.** If we regard $G(n,W)$ as a labelled random graph, we may identify it with the collection $(J_{ij})_{i<j}$ of the $\binom{n}{2}$ edge indicators $J_{ij} := \mathbf{1}\{ij \text{ is an edge}\}$, $1 \le i < j \in [n]$. For finite n, $G(n,W)$ is thus a discrete random variable with $2^{\binom{n}{2}}$ possible outcomes. Recall that for any discrete random variable Z, with outcomes (in any space) having probabilities $p_1, p_2, \dots$, say, its *entropy* $\mathcal{E}(Z)$ is defined by

$$\mathcal{E}(Z) := -\sum_i p_i \log p_i.$$

We also write $\mathcal{E}(Z_1, \dots, Z_n)$ for the entropy of a vector $(Z_1, \dots, Z_n)$, and $\mathcal{E}(Z \mid Z')$ for the entropy of the conditioned random variable $(Z \mid Z')$.

The following asymptotic calculation of the entropy of $G(n,W)$ is a special case of the symmetric version of the formula in [2, Remarks, p. 146]. Let

$$h(p) := -p\log p - (1-p)\log(1-p), \qquad p\in[0,1];$$

thus the entropy of a $\{0,1\}$-valued random variable $Z\in\mathrm{Be}(p)$ is $h(p)$. Note that h is continuous on $[0,1]$ with $0\le h(p)\le \log 2$ and $h(0)=h(1)=0$.

Theorem D.5. *Let W be a graphon, defined on a probability space (Ω,μ). Then, as $n\to\infty$,*

$$\frac{\mathcal{E}(G(n,W))}{\binom{n}{2}} \to \iint_{\Omega^2} h\bigl(W(x,y)\bigr)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y). \tag{D.1}$$

Proof. If we condition on $X_1,\dots,X_n$, then J_{ij} are independent and each $J_{ij}\in\mathrm{Be}(p_{ij})$ with $p_{ij}=W(X_i,X_j)$. Thus, using in the calculations here and below some simple standard results on entropy,

$$\begin{aligned}\mathcal{E}\bigl(G(n,W)\mid X_1,\dots,X_n\bigr) &= \sum_{i<j}\mathcal{E}\bigl(J_{ij}\mid X_1,\dots,X_n\bigr) = \sum_{i<j}\mathcal{E}\bigl(\mathrm{Be}(p_{ij})\bigr)\\ &= \sum_{i<j} h(p_{ij}) = \sum_{i<j} h(W(X_i,X_j)).\end{aligned}$$

Hence,

$$\begin{aligned}\mathcal{E}\bigl(G(n,W)\bigr) &\ge \mathbb{E}\,\mathcal{E}\bigl(G(n,W)\mid X_1,\dots,X_n\bigr) = \mathbb{E}\sum_{i<j} h(W(X_i,X_j))\\ &= \binom{n}{2}\iint_{\Omega^2} h\bigl(W(x,y)\bigr)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).\end{aligned}$$

Thus the left-hand side of (D.1) is greater than or equal to the right-hand side for every $n\ge 2$.

To obtain a corresponding upper bound, we for convenience assume that $\Omega=(0,1]$, as we may by Theorem 7.1 (noting that $\iint h(W)$ is preserved by pull-backs, and thus by equivalence, see Theorem 8.3).

Fix an integer m and let $M_i := \lceil mX_i\rceil$. Thus $M_i=k \iff X_i\in I_{km}$. We have

$$\begin{aligned}\mathcal{E}\bigl(G(n,W)\bigr) &\le \mathcal{E}\bigl(G(n,W),M_1,\dots,M_n\bigr)\\ &= \mathcal{E}(M_1,\dots,M_n) + \mathbb{E}\bigl(\mathcal{E}\bigl(G(n,W)\mid M_1,\dots,M_n\bigr)\bigr).\end{aligned} \tag{D.2}$$

Since $M_1,\dots,M_n$ are independent and uniformly distributed on $\{1,\dots,m\}$,

$$\mathcal{E}(M_1,\dots,M_n) = \sum_{i=1}^n \mathcal{E}(M_i) = n\log m. \tag{D.3}$$

Moreover,

$$\mathcal{E}\bigl(G(n,W)\mid M_1,\dots,M_n\bigr) \le \sum_{i<j}\mathcal{E}\bigl(J_{ij}\mid M_1,\dots,M_n\bigr) = \sum_{i<j}\mathcal{E}\bigl(J_{ij}\mid M_i,M_j\bigr). \tag{D.4}$$

Define, for $k, l = 1, \dots, m$,

$$w_m(k,l) := \mathbb{E}\big(W(X_1, X_2) \mid M_1 = k,\, M_2 = l\big) = m^2 \int_{I_{km}} \int_{I_{lm}} W(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

the average of W over $I_{km} \times I_{lm}$, and let

$$W_m(x,y) := w_m(k,l) \quad \text{if } x \in I_{km},\, y \in I_{lm}.$$

Thus $W_m(X_i, X_j)$ equals the conditional expectation $\mathbb{E}\big(W(X_1, X_2) \mid M_1, M_2\big)$.

Given $M_i = k$ and $M_j = l$,

$$\mathbb{P}(J_{ij} = 1) = \mathbb{E}\big(W(X_1, X_2) \mid M_1 = k,\, M_2 = l\big) = w_m(k,l),$$

and thus

$$\mathcal{E}\big(J_{ij} \mid M_i = k,\, M_j = l\big) = h\big(w_m(k,l)\big).$$

Consequently,

$$\mathbb{E}\big(\mathcal{E}\big(J_{ij} \mid M_i,\, M_j\big)\big) = m^{-2} \sum_{k,l=1}^{m} h\big(w_m(k,l)\big) = \iint_{[0,1]^2} h\big(W_m(x,y)\big)\,\mathrm{d}x\,\mathrm{d}y. \tag{D.5}$$

Combining (D.2)–(D.5), we obtain

$$\mathcal{E}(G(n,W)) \le n \log m + \binom{n}{2} \iint_{[0,1]^2} h\big(W_m(x,y)\big)\,\mathrm{d}x\,\mathrm{d}y.$$

and thus, for every $m \ge 1$,

$$\limsup_{n\to\infty} \binom{n}{2}^{-1} \mathcal{E}(G(n,W)) \le \iint_{[0,1]^2} h\big(W_m(x,y)\big)\,\mathrm{d}x\,\mathrm{d}y.$$

Now let $m \to \infty$. Then $W_m(x,y) \to W(x,y)$ a.e., and thus the right-hand side tends to $\iint h(W)$ by dominated convergence. □

Appendix E. Other versions of the cut norm

There are several other versions of the cut norm that are equivalent to the versions in (4.2) and (4.3) within constant factors or, in Subsection E.3, at least in a weaker sense.

E.1. **Restrictions on the pairs of subsets.** First, we may restrict the subsets S and T of Ω in (4.2) in various ways. Borgs, Chayes, Lovász, Sós and Vesztergombi [14, Section 7] give three versions where it is assumed that, respectively, $S = T$, S and T are disjoint, and S and T are the complements of each other, i.e.,

$$\|W\|_{\square,3} := \sup_S \Big| \int_{S\times S} W(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \Big|, \tag{E.1}$$

$$\|W\|_{\square,4} := \sup_{S\cap T=\emptyset} \Big| \int_{S\times T} W(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \Big|, \tag{E.2}$$

$$\|W\|_{\square,5} := \sup_S \Big| \int_{S\times S^{\mathsf{c}}} W(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \Big|. \tag{E.3}$$

These have natural combinatorial interpretations for graphs as follows. For a graph G with vertex set V and edge set E, we define, for $A, B \subseteq V$,

$$e(A,B) = e_G(A,B) := \bigl|\bigl\{(x,y) \in A \times B : \{x,y\} \in E\bigr\}\bigr|; \tag{E.4}$$

we also write $e_G(A) := e_G(A,A)$. (Thus, if A and B are disjoint, then $e(A,B)$ is the number of edges between A and B. On the other hand, $e(A)$ is twice the number of edges in A.)

Lemma E.1. *Let G_1 and G_2 be two graphs on the same vertex set V, and let $n := |V|$. Then, for both versions W_G^{V} and W_G^{I},*

$$\|W_{G_1} - W_{G_2}\|_{\square,3} = n^{-2} \max_{A \subseteq V} \bigl|e_{G_1}(A) - e_{G_2}(A)\bigr|, \tag{E.5}$$

$$\|W_{G_1} - W_{G_2}\|_{\square,4} = n^{-2} \max_{A \cap B = \emptyset} \bigl|e_{G_1}(A,B) - e_{G_2}(A,B)\bigr|, \tag{E.6}$$

$$\|W_{G_1} - W_{G_2}\|_{\square,5} = n^{-2} \max_{A \subseteq V} \bigl|e_{G_1}(A,A^{\mathsf{c}}) - e_{G_2}(A,A^{\mathsf{c}})\bigr|. \tag{E.7}$$

In particular, $\|W_{G_1} - W_{G_2}\|_{\square,5}$ measures directly the maximal difference in size of cuts in G_1 and G_2, which explains the name "cut norm".

Proof. For W_G^{V} this is immediate, since for every $S, T \subseteq \Omega = V$ we have $\int_{S\times T} W_{G_\ell}^{\mathsf{V}} = n^{-2} e_{G_\ell}(S,T)$.

For W_G^{I}, let $\bigl(a_{ij}^{(\ell)}\bigr)_{ij}$ be the adjacency matrix of G_ℓ, so $a_{ij}^{(\ell)} := \mathbf{1}\bigl\{\{i,j\} \in E(G_\ell)\bigr\}$. If $S, T \subseteq [0,1]$, let $s_i := \lambda(S \cap I_{in})$, $t_j := \lambda(T \cap I_{jn})$. Then

$$\int_{S\times T} \bigl(W_{G_1} - W_{G_2}\bigr) = \Bigl|\sum_{i,j=1}^n s_i t_j \bigl(a_{ij}^{(1)} - a_{ij}^{(2)}\bigr)\Bigr|. \tag{E.8}$$

It follows that

$$\|W_{G_1} - W_{G_2}\|_{\square,3} = \sup_{0 \le s_i \le 1/n} \Bigl|\sum_{i,j=1}^n s_i s_j \bigl(a_{ij}^{(1)} - a_{ij}^{(2)}\bigr)\Bigr|, \tag{E.9}$$

$$\|W_{G_1} - W_{G_2}\|_{\square,4} = \sup_{\substack{0 \le s_i \le 1/n \\ 0 \le u_j \le 1}} \Bigl|\sum_{i,j=1}^n s_i u_j (1 - s_j) \bigl(a_{ij}^{(1)} - a_{ij}^{(2)}\bigr)\Bigr|, \tag{E.10}$$

$$\|W_{G_1} - W_{G_2}\|_{\square,5} = \sup_{0 \le s_i \le 1/n} \Bigl|\sum_{i,j=1}^n s_i (1 - s_j) \bigl(a_{ij}^{(1)} - a_{ij}^{(2)}\bigr)\Bigr|. \tag{E.11}$$

Since $a_{ii}^{(1)} = a_{ii}^{(2)} = 0$, the diagonal terms in these sums vanish, and thus the sums are affine functions of each s_i and (for $\|\cdot\|_{\square,4}$) u_i. Hence, the suprema are attained when all s_i are either 0 or $1/n$, and u_i 0 or 1, i.e., when S and T are unions $S = \bigcup_{i\in A} I_{in}$ and $T = \bigcup_{j \in B} I_{jn}$ for some $A, B \subseteq V$, but then $\int_{S\times T} W_{G_\ell}^{\mathsf{I}} = n^{-2} e_{G_\ell}(A,B)$, so we obtain the same result as for W_G^{V}. $\square$

Lemma E.2 ([14])**.** *If Ω is atomless and $W \in L^1(\Omega^2)$ is symmetric, then the norms $\|W\|_{\square,i}$, $i = 1, ..., 5$, are equivalent. More precisely,*

$$\|W\|_{\square,1} \le \|W\|_{\square,2} \le 4\|W\|_{\square,1} \qquad \textit{for all } \Omega \textit{ and } W; \tag{E.12}$$

$$\frac{1}{2}\|W\|_{\square,1} \le \|W\|_{\square,3} \le \|W\|_{\square,1} \qquad \textit{if } W \textit{ is symmetric;} \tag{E.13}$$

$$\frac{1}{4}\|W\|_{\square,1} \le \|W\|_{\square,4} \le \|W\|_{\square,1} \qquad \textit{if } \Omega \textit{ is atomless;} \tag{E.14}$$

$$\frac{2}{3}\|W\|_{\square,4} \le \|W\|_{\square,5} \le \|W\|_{\square,4} \qquad \textit{if } W \textit{ is symmetric.} \tag{E.15}$$

Proof. The inequalities (E.12) were given in (4.4). For the others, the right-hand sides are trivial.

For the left-hand sides, let $W(S,T) := \int_{S\times T} W$. Then (E.13) follows from $W(S,T) = W(T,S)$ and

$$\begin{aligned} W(S,T) + W(T,S) = W(S\cup T, S\cup T) + W(S\cap T, S\cap T) \\ - W(S\setminus T, S\setminus T) - W(T\setminus S, T\setminus S). \end{aligned}$$

For (E.14) we randomize. Let $(A_i)_{i=1}^n$ be a partition of Ω with $\mu(A_i) = 1/n$ for each i (such partitions exist when Ω is atomless as a consequence of Lemma A.1), let I be a random subset of $\{1,\dots,n\}$ defined by including each element with probability $1/2$, independently of each other, and define a random subset B of Ω by $B := \bigcup_{i\in I} A_i$. Then, for any $S,T \subseteq \Omega$, $\mathbb{E}\bigl|W(S\cap B, T\setminus B)\bigr| \le \|W\|_{\square,4}$. Moreover,

$$\begin{aligned} \mathbb{E}\, W(S\cap B, T\setminus B) &= \sum_{i\ne j} \frac{1}{4} W(S\cap A_i, T\cap A_j) \\ &= \frac{1}{4} W(S,T) - \frac{1}{4}\sum_i W(S\cap A_i, T\cap A_i). \end{aligned}$$

The last sum is the integral of W over a subset of Ω^2 of measure $1/n$, so it tends to 0 as $n\to\infty$. Consequently, $\frac{1}{4}W(S,T) \le \|W\|_{\square,4}$, and (E.14) follows.

For (E.15), assume that $S\cap T = \emptyset$. Let $R := (S\cup T)^{\mathsf{c}}$. The result follows from

$$W(S,T) + W(T,S) = W(S, T\cup R) + W(T, S\cup R) - W(S\cup T, R). \qquad \square$$

Remark E.3. Some restrictions are necessary in Lemma E.2. For example, if W is anti-symmetric ($W(x,y) = -W(y,x)$), then $\|W\|_{\square,3} = 0$, so (E.13) does not hold for arbitrary W. More generally, if $\widetilde{W}(x,y) := \frac{1}{2}(W(x,y) + W(y,x))$ is the symmetrization of W, then $\|\cdot\|_{\square,3}$ never distinguishes between W and $\widetilde{W}$, so $\|\cdot\|_{\square,3}$ is appropriate only for symmetric W.

Similarly, if Ω has an atom A and $W(x,y) := \mathbf{1}\{x,y\in A\}$, then $\|W\|_{\square,4} = 0$ and (E.14) does not hold. Hence, in general $\|\cdot\|_{\square,4}$ and $\|\cdot\|_{\square,5}$ are not appropriate for spaces with atoms. (However, they work well also for W_G^{V} for graphs G, because $W_G^{\mathsf{V}}(x,x) = 0$ for every x, see Lemma E.1 and its proof.)

If W is anti-symmetric and the marginal $\int_\Omega W(x,y)\,\mathrm{d}\mu(y)=0$, then

$$\int_{S\times S^{\mathsf{c}}} W = \int_{S\times\Omega} W - \int_{S\times S} W = 0 \tag{E.16}$$

for every S, so $\|W\|_{\square,5}=0$ and (E.15) does not hold (unless $W=0$ a.e.). For example, we can take $W(x,y)=\sin(2\pi(x-y))$ on $[0,1]$, or take $\Omega=\{1,2,3\}$ with $\mu(i)=1/3$ for each $i\in\Omega$, and $W(i,j)\in\{-1,0,1\}$ with $W(i,j)\equiv i-j \pmod 3$. (In fact, if Ω is atomless, then $\|W\|_{\square,5}=0$ if and only if W is anti-symmetric and its marginals vanish a.e. To see this, note that if $\|W(x,y)\|_{\square,5}=0$, then $\|W(y,x)\|_{\square,5}=0$ as well, and thus $\|\widetilde{W}\|_{\square,5}=0$. By Lemma E.2, then $\|\widetilde{W}\|_{\square,2}=0$ and thus $\widetilde{W}=0$ a.e. By (E.16), $\int_S \overline{W}^{(2)}=\int_{S\times\Omega}W=0$ for every $S\subseteq\Omega$, and thus $\overline{W}^{(2)}=0$ a.e.) Cf. [39, Section 9].

Remark E.4. If $W\cong W'$, then $\|W\|_{\square,3}=\|W'\|_{\square,3}$; this is easily seen first for pull-backs by the argument in the proof of Lemma 5.5, and then in general by Theorem 8.3. The same holds for $\|\cdot\|_{\square,4}$ and $\|\cdot\|_{\square,5}$ provided W and W' are defined on atomless spaces, using also a randomization argument similar to the one in the proof of Lemma E.2. However, this is not true in general for spaces with atoms. For a trivial example, let $W=1$ on $[0,1]$ and $W'=1$ on one-point space; then $\|W'\|_{\square,4}=\|W'\|_{\square,5}=0$.

Remark E.5. The constants in (E.12)–(E.15) are best possible. Examples with equality in the left or right inequalities are given by the following matrices, interpreted as functions on $[0,1]^2$, with each row or column in an $n\times n$-matrix corresponding to an interval I_{in} of length $1/n$ (we could use a space Ω with n points, but we want Ω to be atomless):

(E.12): (1), $\begin{pmatrix}1&-1\\-1&1\end{pmatrix}$;

(E.13): $\begin{pmatrix}-1&0&1\\0&1&0\\1&0&-1\end{pmatrix}$, (1);

(E.14): (1), $\begin{pmatrix}1&-1\\-1&1\end{pmatrix}$;

(E.15): $\begin{pmatrix}0&3&-1\\3&0&-1\\-1&-1&0\end{pmatrix}$, (1).

E.2. **Complex and Hilbert space valued functions.** Another set of versions of the cut norm use (4.3) but consider other sets of functions f and g. For example, we may take the supremum over all complex-valued functions f and g with $|f|,|g|\le 1$, i.e.

$$\|W\|_{\square,\mathbb{C}} := \sup_{\substack{f,g:\Omega\to\mathbb{C}\\ \|f\|_\infty,\|g\|_\infty\le 1}} \Bigl|\int_{\Omega^2} W(x,y)f(x)g(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\Bigr|. \tag{E.17}$$

It is easily seen that $\|W\|_{\square,\mathbb{C}}\le 2\|W\|_{\square,2}$, which can be improved to [45]

$$\|W\|_{\square,2}\le\|W\|_{\square,\mathbb{C}}\le\sqrt{2}\|W\|_{\square,2}, \tag{E.18}$$

which is best possible. (For an example, consider a two-point space $\Omega=\{1,2\}$ with $\mu\{1\}=\mu\{2\}=1/2$, and let $W_1(x,y)=1/2$ and $W_2(x,y)=$

$\mathbf{1}\{x = y = 1\}$. Then $\|W_1 - W_2\|_{\square,2} = 1/4$ but $\|W_1 - W_2\|_{\square,\mathbb{C}} = \sqrt{2}/4$, obtained by taking $f = g = (1, \mathrm{i})$ in (E.17).)

An interesting version is to allow f and g to take values in the unit ball of an arbitrary Hilbert space H and define

$$\|W\|_{\square,\mathsf{H}} := \sup_{\substack{f,g:\Omega\to H\\ \|f\|_\infty,\|g\|_\infty\le 1}} \Big|\int_{\Omega^2} W(x,y)\langle f(x), g(y)\rangle \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\Big|. \tag{E.19}$$

(Since we only consider real W, it is easy to see that it does not matter whether we allow real or complex Hilbert spaces in (E.19).) In this case, the equivalence with $\|W\|_{\square,2}$ is a form of the famous *Grothendieck's inequality* [31], which says that

$$\|W\|_{\square,2} \le \|W\|_{\square,\mathsf{H}} \le K_G\|W\|_{\square,2}, \tag{E.20}$$

where the constant K_G, the real Grothendieck constant, is known to satisfy $\pi/2 \le K_G \le \pi/(2\log(1+\sqrt{2})) \approx 1.78221$ [45]. (The lower bound is improved in an unpublished manuscript [56].) We also have $\|W\|_{\square,\mathbb{C}} \le \|W\|_{\square,\mathsf{H}} \le K_G^{\mathbb{C}}\|W\|_{\square,\mathbb{C}}$, where $K_G^{\mathbb{C}}$ is the complex Grothendieck constant, known to satisfy $4/\pi \le K_G^{\mathbb{C}} < 1.40491$ [33]. Moreover, $\|W\|_{\square,\mathbb{C}}$ is obtained by taking only a fixed Hilbert space of dimension 2 in (E.19).

See [3] for an algorithmic use of the version $\|\cdot\|_{\square,\mathsf{H}}$ of the cut norm and Grothendieck's inequality.

E.3. **Other operator norms.** If W is a kernel on Ω, then it defines an integral operator $T_W : f \mapsto \int_\Omega W(x,y)f(y)\,\mathrm{d}\mu(y)$ (for suitable f). We have already noted in Remark 4.2 that $\|\cdot\|_{\square,2}$ is the operator norm of T_W as an operator $L^\infty(\Omega) \to L^1(\Omega)$, but we may also consider other spaces.

Let, for $1 \le p, q \le \infty$, $\|T\|_{p,q}$ denote the norm of T as an operator $L^p \to L^q$.

Lemma E.6. *If $|W| \le 1$, then for all $p, q \in [1,\infty]$,*

$$\|W\|_{\square,2} = \|T_W\|_{\infty,1} \le \|T_W\|_{p,q} \le \sqrt{2}\|W\|_{\square,2}^{\min(1-1/p,1/q)}.$$

Consequently, for any fixed $p > 1$ and $q < \infty$, if $W_1, W_2, \dots$ and W are graphons defined on the same space Ω, then $\|W_n - W\|_\square \to 0$ if and only if $\|T_{W_n} - T_W\|_{p,q} \to 0$.

Proof. We know that $\|W\|_{\square,2} = \|T_W\|_{\infty,1}$. Moreover, for any probability space, the inclusions $L^\infty \subseteq L^p$ and $L^p \subseteq L^1$ have norm 1, and thus $\|T\|_{\infty,1} \le \|T\|_{p,q}$ for any operator T.

Let $\theta := \min(1 - 1/p,\, 1/q)$, so $1 - \theta := \max(1/p,\, 1 - 1/q)$, and define $p_0, q_0 \in [1,\infty]$ by $1/p = (1-\theta)/p_0$ and $1 - 1/q = (1-\theta)(1 - 1/q_0)$. Further, let $p_1 = \infty$ and $q_1 = 1$. Then $(1/p,\, 1/q) = (1-\theta)(1/p_0,\, 1/q_0) + \theta(1/p_1,\, 1/q_1)$, and it follows from the Riesz–Thorin interpolation theorem (see e.g. [5, Theorem 1.1.1]) that, provided we work with complex L^p spaces,

$$\|T_W\|_{p,q} \le \|T_W\|_{p_0,q_0}^{1-\theta}\|T_W\|_{p_1,q_1}^{\theta}.$$

By (E.17) and (E.18),

$$\|T_W\|_{p_1,q_1} = \|T_W\|_{\infty,1} = \|W\|_{\square,\mathbb{C}} \le \sqrt{2}\|W\|_{\square,2},$$

and the assumption $|W| \le 1$ implies $\|T_W\|_{p_0,q_0} \le \|T_W\|_{1,\infty} \le \|W\|_\infty \le 1$. The result follows. $\square$

We consider the case $p = q = 2$ further, i.e., we regard T_W as an operator on the Hilbert space $L^2(\Omega)$. If W is bounded (or, more generally, in $L^2(\Omega^2)$), then T_W is bounded on L^2; it is further compact (and Hilbert–Schmidt) and selfadjoint (because W is symmetric). Hence T_W has a sequence of eigenvalues (λ_n). We define, for $1 \le p < \infty$, the Schatten S_p-norm of T_W to be

$$\|T_W\|_{S_p} := \|(\lambda_n)\|_{\ell^p} = \Bigl(\sum_n |\lambda_n|^p\Bigr)^{1/p}. \tag{E.21}$$

(See e.g. [30], where also the non-selfadjoint case is treated.) It is well-known that for $p = 2$, $\|\cdot\|_{S_2}$ equals the Hilbert–Schmidt norm and thus

$$\|T_W\|_{S_2} = \|W\|_{L^2(\Omega^2)}. \tag{E.22}$$

If $p = 2k$ is an even integer ≥ 4, then (E.21) yields

$$\|T_W\|_{S_{2k}}^{2k} = \sum_n \lambda_n^{2k} = \operatorname{Tr}\bigl(T_W^{2k}\bigr) = t(C_{2k}, W), \tag{E.23}$$

where the graph C_{2k} is the cycle of length $2k$.

Lemma E.7.

(i) *For $2 < p < \infty$, if $|W| \le 1$, then*

$$\|W\|_{\square,2} = \|T_W\|_{\infty,1} \le \|T_W\|_{2,2} \le \|T_W\|_{S_p} \le \sqrt{2}\|W\|_{\square,2}^{1/2-1/p}.$$

Consequently, for any fixed $p > 2$, if $W_1, W_2, \dots$ and W are graphons defined on the same space Ω, then $\|T_{W_n} - T_W\|_{S_p} \to 0$ if and only if $\|W_n - W\|_\square \to 0$.

(ii) *For $p = 2$, if $|W| \le 1$, then*

$$\|W\|_{L^1} \le \|T_W\|_{S_2} = \|W\|_{L^2} \le \|W\|_{L^1}^{1/2}.$$

Consequently, if $W_1, W_2, \dots$ and W are graphons defined on the same space Ω, then $\|T_{W_n} - T_W\|_{S_2} \to 0$ if and only if $\|W_n - W\|_{L^1} \to 0$.

Proof. (i): The first inequality is in Lemma E.6 and the second is trivial, since the operator norm $\|T_W\|_{2,2} = \sup_n |\lambda_n|$. Further, by this and (E.21),

$$\|T_W\|_{S_p}^p = \sum_n |\lambda_n|^p \le \sum_n |\lambda_n|^2 \sup_n |\lambda_n|^{p-2} = \|T_W\|_{S_2}^2 \|T_W\|_{2,2}^{p-2}. \tag{E.24}$$

We have $\|T_W\|_{S_2} = \|W\|_{L^2(\Omega^2)} \le 1$ by (E.22), and $\|T_W\|_{2,2} \le \sqrt{2}\|W\|_{\square,2}^{1/2}$ by Lemma E.6, and the result follows.

(ii): Immediate by (E.22) and standard inequalities (e.g. Hölder). $\square$

In particular, by (i) with $p = 4$ and (E.23), if $|W| \le 1$, then $\|W\|_{\square,2} \le t(C_4, W)^{1/4} \le \sqrt{2}\|W\|_{\square,2}^{1/4}$, or

$$\tfrac{1}{4}t(C_4, W) \le \|W\|_{\square,2} \le t(C_4, W)^{1/4}.$$

This was proved in [14, Lemma 7.1] (by a slightly diferent argument, using a version of (C.3)), where also an application is given.

Remark E.8. There is no corresponding result for $p < 2$. In fact, $\|T_W\|_{S_p}$ may be infinite for a graphon W. To see this, let first W be constant $1/2$ on $[0, 1]$ and let (G_n) be a quasirandom sequence of graphs with $G_n \to W$. Let $W_n := W_{G_n}$, so $\delta_\square(W_n, W) \to 0$. By [14, Lemma 5.3], we may label the graphs G_n such that $\|W_n - W\|_\square \to 0$.

By (E.22), $\|T_{W_n-W}\|_{S_2} = \|W_n - W\|_{L^2} = 1/2$. On the other hand, arguing as in (E.24),

$$\|T_{W_n-W}\|_{S_2}^2 \le \|T_{W_n-W}\|_{S_p}^p \|T_{W_n-W}\|_{2,2}^{2-p} \le \sqrt{2}\|T_{W_n-W}\|_{S_p}^p \|W_n - W\|_{\square,2}^{2-p}.$$

Since the left-hand side is constant and the last factor tends to 0, it follows that $\|T_{W_n-W}\|_{S_p} \to \infty$. Further, $\|W_n - W\|_\infty \le 1$. It is now an easy consequence of the closed graph theorem that there exist bounded functions W on $[0, 1]^2$ such that $\|T_W\|_{S_p} = \infty$, and by linearity there must exist such a graphon. (An explicit W is given by a well-known analytic construction [30, §III.10.3, p. 118]: let $W(x, y) = f(x - y)$ on $[0, 1]^2$, where f is a continuous even function with period 1 on $\mathbb{R}$ such that $\sum |\widehat{f}(n)|^p = \infty$ for all $p < 2$; such a function was constructed by Carleman [16], see also [61, V.4.9].)

Appendix F. The weak topology on $\mathcal{W}(\Omega)$

Consider the space $\mathcal{W} = \mathcal{W}(\Omega)$ of graphons on a fixed probability space Ω. We have discussed two different metrics on this space, given by the norms $\| \|_{L^1}$ and $\| \|_\square$; these give two different topologies on $\mathcal{W}(\Omega)$.

Another topology on $\mathcal{W}(\Omega)$ is the *weak topology* σ, regarding $\mathcal{W}(\Omega)$ as a subset of $L^1(\Omega^2)$. This topology is generated by the functionals $\chi_h : W \mapsto \int_{\Omega^2} hW$ for $h \in L^\infty(\Omega^2)$, in the standard sense that it is the weakest topology that makes all these maps continuous. Actually, since the functions in $\mathcal{W}(\Omega)$ are uniformly bounded, we obtain the same topology from many different families of such functionals.

We state this also for subsets of $L^1(\Omega)$ and writing $\chi_h(f) := \int_\Omega hf$ for $h \in L^\infty(\Omega)$ and $f \in L^1(\Omega)$. Thus the weak topology on $L^1(\Omega)$ (or a subset of it) is the topology generated by χ_h, $h \in L^\infty(\Omega)$. Recall further that a subset $\mathcal{H}$ of a topological vector space is *total* if the set of linear combinations of elements of $\mathcal{H}$ is dense in the space.

Lemma F.1. (i) *Let $\mathcal{H}$ be a total set in $L^1(\Omega)$, and let $\mathcal{X}$ be a subset of $L^1(\Omega, \mu)$ consisting of uniformly bounded functions: $\sup_{f\in\mathcal{X}} \|f\|_\infty < \infty$. Then the functionals $\{\chi_h : h \in \mathcal{H}\}$ generate the weak topology on $\mathcal{X}$.*

(ii) *Let $\mathcal{H}$ be a total set in $L^1(\Omega^2)$. Then the functionals $\{\chi_h : h \in \mathcal{H}\}$ generate the weak topology on $\mathcal{W}(\Omega)$.*

Proof. (i): Let $\tau_{\mathcal{H}}$ be the topology on $\mathcal{X}$ generated by $\{\chi_h : h \in \mathcal{H}\}$, and let $\mathcal{H}'$ be the set of all $g \in L^1(\Omega)$ such that χ_g is continuous $(\mathcal{X}, \tau_{\mathcal{H}}) \to \mathbb{R}$. By the definition of $\tau_{\mathcal{H}}$, $\mathcal{H} \subseteq \mathcal{H}'$; further, $\mathcal{H}'$ and $\mathcal{H}$ generate the same topology, i.e., $\tau_{\mathcal{H}} = \tau_{\mathcal{H}'}$.

$\mathcal{H}'$ is clearly a linear subspace of $L^1(\Omega)$, and since we have assumed that $\mathcal{H}$ is total, $\mathcal{H}'$ is dense in $L^1(\Omega)$. If $g \in L^1(\Omega)$, there thus exists a sequence $g_n \in \mathcal{H}'$ with $\|g_n - g\|_{L^1} \to 0$. Since the functions in $\mathcal{X}$ are uniformly bounded, this means that $\chi_{g_n} \to \chi_g$ uniformly on $\mathcal{X}$, and thus χ_g too is $\tau_{\mathcal{H}}$-continuous; hence $g \in \mathcal{H}'$. Consequently, $\mathcal{H}' = L^1(\Omega)$, and thus $\tau_{\mathcal{H}} = \tau_{\mathcal{H}'} = \tau_{L^1(\Omega)}$. Thus every total $H \subseteq L^1(\Omega)$ generates the same topology. One such $\mathcal{H}$ is $L^\infty(\Omega)$ which defines the weak topology (by definition).

(ii): This is a special case, since Ω^2 is another probability space. □

In particular, the weak topology on $\mathcal{W}(\Omega)$ is also the topology generated by the functionals $W \mapsto \int_{\Omega^2} hW$, $h \in L^1(\Omega^2)$, i.e., it equals the *weak* topology* on $\mathcal{W}(\Omega)$, regarded as a subset of $L^\infty(\Omega^2)$.

Remark F.2. Another example of a total set in $L^1(\Omega^2)$ is the set of rectangle indicators $\mathbf{1}_S(x)\mathbf{1}_T(y)$ for $S, T \subseteq \Omega$. Thus the weak topology is also generated by the functionals $W \mapsto \int_{S\times T} W$. Note that the metric given by $\| \ \|_\square$ uses the same functionals, but with an important difference: $\|W_n - W\|_\square \to 0$ if and only if $\int_{S\times T} W_n \to \int_{S\times T} W$ *uniformly* for all $S, T \subseteq \Omega$, while $W_n \to W$ in the weak topology if and only if each $\int_{S\times T} W_n \to \int_{S\times T} W$, without any uniformity requirement. (Similarly, $\|W_n - W\|_{L^1} \to 0$ if and only if $\int hW_n \to \int hW$ uniformly for all h with $\|h\|_\infty \le 1$.)

Lemma F.3. *The weak topology is weaker than the cut norm topology. I.e., the identity maps $(\mathcal{W}, \| \ \|_{L^1}) \to (\mathcal{W}, \| \ \|_\square) \to (\mathcal{W}, \sigma)$ are continuous.*

Proof. Immediate by Remark F.2. □

Theorem F.4. *The topological space $(\mathcal{W}(\Omega), \sigma)$ is compact.*

Proof. $\mathcal{W}$ is a weak* closed subset of the unit ball of $L^\infty(\Omega^2) = L^1(\Omega^2)^*$, so this follows from the Banach–Alaoglu theorem. □

One advantage with the weak topology is thus that it is compact, in contrast to the topologies defined by the norms $\| \ \|_\square$ and $\| \ \|_{L^1}$ which are not compact (in general, e.g. if $\Omega = [0, 1]$), see Example F.6 below. (Recall that, nevertheless, the quotient space $(\widehat{\mathcal{W}}, \delta_\square)$ is compact, and that this is a very important property.)

However, a serious drawback with the weak topology is that the quotient map $\mathcal{W}(\Omega) \to \widehat{\mathcal{W}}$ is *not* continuous in the weak topology. Equivalently, the homomorphism densities $t(F, W)$ defined in Appendix C are *not* continuous in the weak topology (for every fixed F). More precisely, for example

$W \mapsto t(K_3, W)$ is not continuous in the weak topology on $\mathcal{W}([0,1])$, see Example F.6.

Remark F.5. There are graphs F such that $W \mapsto t(F, W)$ is weakly continuous (i.e., continuous for σ), for example K_2 since $t(K_2, W) = \int_{\Omega^2} W$. We show in Lemma F.7 below that K_2 is essentially the only such exceptional case.

Example F.6. Take $\Omega = [0,1]$. Let $g_n(x) = \operatorname{sgn}(\sin(2\pi n x))$ and $W_n(x,y) = \frac{1}{2} - \frac{1}{2} g_n(x) g_n(y)$. Then $g_n(x) \in \{\pm 1\}$ and W_n is $\{0,1\}$-valued; in fact, W_n equals $W^{\mathsf{V}}_{K_{n,n}}$ for a complete bipartite graph $K_{n,n}$. (A less combinatorial alternative is to take $g_n(x) = \sin(2\pi n x)$.)

We have $g_n = g_1^{\varphi_n}$ and $W_n = W_1^{\varphi_n}$, where $\varphi_n(x) = nx \mod 1$ as in Example 8.2. Consequently, $W_n \cong W_1$, and thus $W_n = W_1$ in the quotient space $\widehat{\mathcal{W}}$, i.e. $\delta_\square(W_n, W_1) = 0$; in particular, $W_n \to W_1$ in $(\widehat{\mathcal{W}}, \delta_\square)$.

On the other hand, for any $h \in L^1([0,1]^2)$, $\int_{[0,1]^2} h(x,y) g_n(x) g_n(y) \to 0$, and thus $W_n \to \frac{1}{2}$ in $(\mathcal{W}([0,1]), \sigma)$.

If the quotient map $\mathcal{W}([0,1]) \to \widehat{\mathcal{W}}$ were continuous for σ, then $W_n \to \frac{1}{2}$ in $\widehat{\mathcal{W}}$, and since we already know $W_n \to W_1$ in $\widehat{\mathcal{W}}$, we would have $W_1 = \frac{1}{2}$ in $\widehat{\mathcal{W}}$, i.e., $W_1 \cong \frac{1}{2}$, which contradicts e.g. Corollary 8.12. Consequently, the quotient map is *not* continuous $(\mathcal{W}([0,1]), \sigma) \to (\widehat{\mathcal{W}}, \delta_\square)$.

This also shows that $(\mathcal{W}([0,1]), \| \ \|_\square)$ and, a fortiori, $(\mathcal{W}([0,1]), \| \ \|_{L^1})$ are not compact. Indeed, if one of these spaces were compact, then W_n would have a convergent subsequence in it, and thus in $(\mathcal{W}, \| \ \|_\square)$, with a limit W say. Since both maps $(\mathcal{W}, \| \ \|_\square) \to (\mathcal{W}, \sigma)$ and $(\mathcal{W}, \| \ \|_\square) \to (\widehat{\mathcal{W}}, \delta_\square)$ are continuous, the subsequence would converge to W in both $(\mathcal{W}, \sigma)$ and $(\widehat{\mathcal{W}}, \delta_\square)$ too; hence both $W = \frac{1}{2}$ a.e. and $W \cong W_1$, so again $W_1 \cong \frac{1}{2}$, a contradiction.

Furthermore, with $W = \frac{1}{2}$, so $W_n \to W$ weakly, $t(K_3, W_n) = 0$, while $t(K_3, W) = \frac{1}{8} > 0$; hence, $t(K_3, W)$ is not weakly continuous.

Lemma F.7. *The map $W \mapsto t(F, W)$ is weakly continuous (for $\Omega = [0,1]$, say) if and only if F is a disjoint union of isolated vertices and edges.*

Proof. Let F have m vertices and e edges. If every component of F is a vertex or an edge, then $t(F, W) = \left(\int_{\Omega^2} W\right)^e$, which is weakly continuous.

Conversely, suppose that F is a graph such that $W \mapsto t(F, W)$ is weakly continuous. Let $\alpha \in (0, 1/2)$ be rational and let $W_n := W^{\mathsf{V}}_{G_n}$, where G_n is the complete bipartite graph $K_{\alpha n, n - \alpha n}$ (for n such that αn is an integer). Taking the vertices of G_n in suitable (e.g. random) order, we have $W_n \to W$ weakly, where $W = 2\alpha(1-\alpha)$ is a constant graphon. Thus, by assumption, $t(F, W_n) \to t(F, W)$.

If F is not bipartite, then $t(F, W_n) = t(F, G_n) = 0$, while $t(F, W) > 0$, a contradiction.

If F is bipartite, suppose first that F is connected, so $e \geq m - 1$ edges. Then F has a bipartition where the smallest part has $k \leq m/2$ vertices, and

thus

$$t(F,W_n)=t(F,G_n)\geq \alpha^k(1-\alpha)^{m-k}\geq 2^{-m}\alpha^{m/2}, \tag{F.1}$$

while

$$t(F,W)=\bigl(2\alpha(1-\alpha)\bigr)^e\leq 2^e\alpha^{m-1}. \tag{F.2}$$

If $m\geq 3$, then $m/2<m-1$, and thus we can choose α so small that $t(F,W_n)>2t(F,W)$ for all n, a contradiction. Hence $m\leq 2$.

If F is bipartite and disconnected, we use the same argument for every component of F, noting that $t(F,W_n)=t(F,W)$ if F has at most two vertices. It follows that no component of F can have more than two vertices. □

See Chatterjee and Varadhan [17] for a recent application of the weak topology on $\mathcal{W}$.

Appendix G. Separability in Lebesgue spaces

In many cases, the Banach space $L^1(\Omega,\mathcal{F},\mu)$ is separable. For example, this is the case if $\Omega=[0,1]$ with any Borel measure μ. (One example of a countable dense set is the set of polynomials with rational coefficients; this is dense e.g. by the monotone class theorem [37, Theorem A.1].) Hence, by Theorem A.4, $L^1(\Omega,\mathcal{F},\mu)$ is separable for every Borel probability space $(\Omega,\mathcal{F},\mu)$. This includes almost all examples used in graph limit theory.

However, there are cases when $L^1(\Omega,\mathcal{F},\mu)$ is non-separable. For example, this is the case when (Ω,μ) is an uncountable product $([0,1],\nu)^{\mathbb{R}}$ or $(\{0,1\},\nu)^{\mathbb{R}}$, with ν the uniform distribution, say. (Any uncountable product of non-trivial spaces will do.) In this case there are some technical difficulties and we sometimes have to be more careful.

Recall that the elements f of $L^1(\Omega,\mathcal{F},\mu)$ formally are equivalence classes of functions, so to define pointwise values $f(x)$ we have to make a choice of representative of f. This is usually harmless, but it may be a serious problem if we want to define $f(x)$ for many f simultaneously, in particular if we want to define a measurable evaluation map $(f,x)\mapsto f(x)$ on $L^1(\Omega,\mu)\times\Omega\to\mathbb{R}$.

The following lemma shows that this is possible when $L^1(\Omega,\mu)$ is separable, and more generally on $A\times\Omega$ when $A\subseteq L^1(\Omega,\mu)$ is a separable subspace. Note, however, that there is *no* such measurable evaluation map in general, without separability assumption, see Example G.2 below. This justifies stating and proving the lemma carefully, although it may look obvious.

Lemma G.1. *If A is a closed separable subspace of $L^1(\Omega,\mathcal{F},\mu)$, then there is a measurable function $\Phi:A\times\Omega\to\mathbb{R}$ such that for every $f\in A$, $\Phi(f,x)=f(x)$ for a.e. $x\in\Omega$.*

Proof. There exists a countable dense set $D\subset A$. Each element of D is an element of $L^1(\Omega,\mathcal{F},\mu)$, i.e., an equivalence class of measurable functions on Ω; we fix one representative for each element of D and regard the elements of D as these fixed functions. Write $D=\{d_1,d_2,\dots\}$ with some arbitrary ordering of the elements.

Since D is dense in A, we may recursively define maps $H_i : A \to D$ such that

$$\Big\| f - \sum_{i=1}^{k} H_i(f) \Big\|_{L^1} \le 2^{-k}, \qquad k \ge 1, \tag{G.1}$$

by defining $H_k(f)$ as the first element of D that satisfies (G.1). Then each $H_i : A \to D$ is measurable. Further, (G.1) implies $\|H_i(f)\|_{L^1} \le 3 \cdot 2^{-i}$ for $i \ge 2$, so $\int_\Omega \sum_{i=1}^\infty |H_i(f)| \, \mathrm{d}\mu = \sum_{i=1}^\infty \|H_i(f)\|_{L^1} < \infty$ for every $f \in A$, which implies that $\sum_{i=1}^\infty H_i(f)(x)$ converges absolutely a.e. Moreover, (G.1) implies by dominated convergence $\|f - \sum_{i=1}^\infty H_i(f)(x)\|_{L^1} = 0$, so $\sum_{i=1}^\infty H_i(f)(x) = f$ a.e. We now define

$$\Phi(f,x) := \begin{cases} \sum_{i=1}^\infty H_i(f)(x), & \text{if the sum converges;} \\ 0 & \text{otherwise.} \end{cases}$$

Each map $(f,x) \mapsto H_i(f)(x)$ is measurable, and thus Φ is measurable. $\square$

Example G.2. Let Ω_0 be the two-point set $\{0,1\}$, with uniform measure $\mu_0\{0\} = \mu_0\{1\} = 1/2$, and let (Ω,μ) be the uncountable product $(\Omega_0,\mu_0)^{\mathbb{R}}$. Any measurable function $\Phi : L^1(\Omega,\mu) \times \Omega \to \mathbb{R}$ depends only on countably many coordinates in $L^1(\Omega,\mu) \times \Omega = L^1(\Omega,\mu) \times \Omega_0^{\mathbb{R}}$, i.e., there is a countable set $C \subset \mathbb{R}$ such that if $x = (x_r)_{r\in\mathbb{R}}$ and $y = (y_r)_{r\in\mathbb{R}}$ are elements of $\Omega = \Omega_0^{\mathbb{R}}$ with $x_r = y_r$ for $r \notin C$, then

$$\Phi(f,x) = \Phi(f,y) \qquad \text{for all } f \in L^1(\Omega,\mu). \tag{G.2}$$

Fix $s \notin C$ and define $\sigma : \Omega \to \Omega$ by $\sigma : (x_r)_r \mapsto (x'_r)_r$ with $x'_r = x_r$ for $r \ne s$ and $x'_s = 1 - x_s$; note that σ is measure-preserving. By (G.2), $\Phi(f,\sigma(x)) = \Phi(f,x)$ for every f and $x \in \Omega$. If $\Phi(f,x) = f(x)$ for a.e. x, then thus $f(x) = f(\sigma(x))$ for a.e. x, which obviously is incorrect for the coordinate function $f(x) = x_s$.

Consequently, there exists no measurable evaluation map $\Phi : L^1(\Omega,\mu) \times \Omega \to \mathbb{R}$ such that $\Phi(f,x) = f(x)$ for every f and a.e. x.

In fact, it can be shown (again using the monotone class theorem) that if A is any measurable space and $\Phi : A \times \Omega \to \mathbb{R}$ is measurable and such that $x \mapsto \Phi(\alpha,x) \in L^1(\Omega,\mathcal{F},\mu)$ for every $\alpha \in A$, then these function all lie in some separable subspace of $L^1(\Omega,\mathcal{F},\mu)$. This shows that the condition in Lemma G.1 that A be separable is both necessary and sufficient for the conclusion of the lemma.

References

[1] D. J. Aldous, Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** (1981), no. 4, 581–598.

[2] D. J. Aldous, Exchangeability and related topics. *École d'été de probabilités de Saint-Flour, XIII – 1983*, 1–198, Lecture Notes in Math. 1117, Springer, Berlin, 1985.

[3] N. Alon & A. Naor, Approximating the cut-norm via Grothendieck's inequality. *SIAM J. Comput.* **35** (2006), no. 4, 787–803.
[4] T. Austin, On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys* **5** (2008), 80–145.
[5] J. Bergh & J. Löfström, *Interpolation Spaces. An Introduction.* Springer, Berlin, 1976.
[6] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.
[7] R. Blei, *Analysis in Integer and Fractional Dimensions.* Cambridge University Press, Cambridge, 2001.
[8] B. Bollobás, *Random Graphs.* 2nd ed., Cambridge Univ. Press, Cambridge, 2001.
[9] B. Bollobás, S. Janson & O. Riordan, The phase transition in inhomogeneous random graphs. *Random Struct. Alg.* **31** (2007), 3–122.
[10] B. Bollobás, S. Janson & O. Riordan, The cut metric, random graphs, and branching processes. *J. Statist. Phys.* **140**:2 (2010), 289–335.
[11] B. Bollobás, S. Janson & O. Riordan, Monotone graph limits and quasimonotone graphs. Preprint, 2011. `arXiv:1101.4296.`
[12] B. Bollobás & O. Riordan, Metrics for sparse graphs. *Surveys in Combinatorics 2009*, LMS Lecture Notes Series 365, Cambridge Univ. Press, 2009, pp. 211–287.
[13] C. Borgs, J. T. Chayes & L. Lovász, Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.* **19** (2010), no. 6, 1597–1619.
[14] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós & K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.* **219** (2008), 1801–1851.
[15] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós & K. Vesztergombi, Convergent sequences of dense graphs II: Multiway cuts and statistical physics. Preprint, 2007. `http://research.microsoft.com/~borgs/`
[16] T. Carleman, Über die Fourierkoeffizienten einer stetigen Funktion. *Acta Math.* **41** (1916), no. 1, 377–384.
[17] S. Chatterjee & S. R. S. Varadhan, The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, to appear.
[18] J. A. Clarkson & C. R. Adams, On definitions of bounded variation for functions of two variables. *Trans. Amer. Math. Soc.* **35** (1933), no. 4, 824–854.
[19] D. L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.
[20] C. Dellacherie & P.-A. Meyer, *Probabilités et potentiel.* Édition entièrement refondue, Hermann, Paris, 1975; English transl.: *Probabilities and Potential.* North-Holland, Amsterdam, 1978.
[21] P. Diaconis & D. Freedman, On the statistics of vision: the Julesz conjecture. *J. Math. Psych.* **24** (1981), no. 2, 112–138.
[22] P. Diaconis, S. Holmes & S. Janson, Threshold graph limits and random threshold graphs. *Internet Mathematics* **5** (2009), no. 3, 267–318.

[23] P. Diaconis, S. Holmes & S. Janson, Interval graph limits. Preprint, 2011. `arXiv:1102.2841.`

[24] P. Diaconis & S. Janson, Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** (2008), 33–61.

[25] N. Dunford & J.T. Schwartz, *Linear Operators. I. General Theory.* Interscience Publishers, New York, 1958.

[26] G. Elek & B. Szegedy, A measure-theoretic approach to the theory of dense hypergraphs. Preprint, 2008. `arXiv:0810.4062.`

[27] R. Engelking, *General Topology.* Revised and completed ed., Heldermann Verlag, Berlin, 1989.

[28] M. Fréchet, Sur les fonctionnelles bilinéaires. *Trans. Amer. Math. Soc.* **16** (1915), no. 3, 215–234.

[29] A. Frieze & R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

[30] I. C. Gohberg & M. G. Kreĭn, *Introduction to the Theory of Linear Nonselfadjoint Operators.* (Russian). Nauka, Moscow, 1965; English transl.: American Mathematical Society, Providence, R.I., 1969.

[31] A. Grothendieck, Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo* **8** (1953), 1–79.

[32] U. Haagerup, The best constants in the Khintchine inequality, *Studia Math.* **70** (1981), 231–283.

[33] U. Haagerup, A new upper bound for the complex Grothendieck constant. *Israel J. Math.* **60** (1987), no. 2, 199–224.

[34] J. Haezendonck, Abstract Lebesgue-Rohlin spaces. *Bull. Soc. Math. Belg.* **25** (1973), 243–258.

[35] D. Hoover, Relations on Probability Spaces and Arrays of Random Variables. Preprint, Institute for Advanced Study, Princeton, NJ, 1979.

[36] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath & R. M. Sampaio, Limits of permutation sequences. Preprint, 2011. `arXiv:1103.5844.`

[37] S. Janson, *Gaussian Hilbert Spaces.* Cambridge Univ. Press, Cambridge, 1997.

[38] S. Janson, Standard representation of multivariate functions on a general probability space. *Electronic Comm. Probab.* **14** (2009), paper 34, 343–346.

[39] S. Janson, Quasi-random graphs and graph limits. *European J. Combin.*, to appear.

[40] S. Janson, Poset limits and exchangeable random posets. *Combinatorica*, to appear.

[41] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs.* Wiley, New York, 2000.

[42] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.

[43] O. Kallenberg, *Probabilistic Symmetries and Invariance Principles.* Springer, New York, 2005.

[44] I. Kolossváry & B. Ráth, Multigraph limits and exchangeability. Preprint, 2009. `arXiv:0910.0547.`

[45] J.-L. Krivine, Constantes de Grothendieck et fonctions de type positif sur les sphéres. *Adv. in Math.* **31** (1979), no. 1, 16–30.

[46] J. E. Littlewood, On bounded bilinear forms in an infinite number of variables, *Quart. J. Math. Oxford* **1** (1930), 164–174.

[47] L. Lovász & V. T. Sós, Generalized quasirandom graphs. *J. Combin. Theory Ser. B* **98** (2008), no. 1, 146–163.

[48] L. Lovász & B. Szegedy, Limits of dense graph sequences. *J. Comb. Theory Ser. B* **96** (2006), 933–957.

[49] L. Lovász & B. Szegedy, Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17** (2007), no. 1, 252–270.

[50] L. Lovász & B. Szegedy, Random graphons and a weak positivstellensatz for graphs. Preprint, 2009. `arXiv:0902.1327.`

[51] L. Lovász & B. Szegedy, Regularity partitions and the topology of graphons. Preprint, 2010. `arXiv:1002.4377.`

[52] L. Lovász & B. Szegedy, Limits of compact decorated graphs. Preprint, 2010. `arXiv:1010.5155.`

[53] M. Morse, Bimeasures and their integral extensions. *Ann. Mat. Pura Appl. (4)* **39** (1955), 345–356.

[54] K. R. Parthasarathy, *Probability Measures on Metric Spaces.* Academic Press, New York, 1967.

[55] O. Pikhurko, An analytic approach to stability. *Discrete Math.* **310** (2010), no. 21, 2951–2964.

[56] J. A. Reeds, A new lower bound on the real Grothendieck constant. Preprint, AT&T Bell Laboratories, Murray Hill, NJ, 1992. `http://www.dtc.umn.edu/~reedsj/bound2.dvi`

[57] V. A. Rohlin, On the fundamental ideas of measure theory. (Russian) *Mat. Sbornik N.S.* **25(67)** (1949), 107–150. English transl.: *Amer. Math. Soc. Translation 1952* (1952), no. 71, 55 pp.

[58] S. J. Szarek, On the best constants in the Khinchin inequality, *Studia Math.* **58** (1976), 197–208.

[59] F. Treves, *Topological Vector Spaces, Distributions and Kernels.* Academic Press, New York, 1967.

[60] N. Th. Varopoulos, Tensor algebras and harmonic analysis. *Acta Math.* **119** (1967), 51–112.

[61] A. Zygmund, *Trigonometric Series.* 2nd ed., Cambridge Univ. Press, Cambridge, 1959

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`