MONOTONE GRAPH LIMITS AND QUASIMONOTONE GRAPHS

BÉLA BOLLOBÁS, SVANTE JANSON, AND OLIVER RIORDAN

Abstract. The recent theory of graph limits gives a powerful framework for understanding the properties of suitable (convergent) sequences (G_n) of graphs in terms of a limiting object which may be represented by a symmetric function W on $[0,1]$, i.e., a *kernel* or *graphon*. In this context it is natural to wish to relate specific properties of the sequence to specific properties of the kernel. Here we show that the kernel is monotone (i.e., increasing in both variables) if and only if the sequence satisfies a 'quasi-monotonicity' property defined by a certain functional tending to zero. As a tool we prove an inequality relating the cut and L^1 norms of kernels of the form $W_1 - W_2$ with W_1 and W_2 monotone that may be of interest in its own right; no such inequality holds for general kernels.

1. Introduction

Recently, Lovász and Szegedy [20] and Borgs, Chayes, Lovász, Sós and Vesztergombi (see, e.g., [5]) developed a rich theory of *graph limits*, associating limit objects to suitable sequences (G_ν) of (dense) graphs with $|G_\nu| \to \infty$, where $|G_\nu|$ denotes the number of vertices of G_ν. The basics of this theory are outlined in Section 2 below; see also Diaconis and Janson [8]. These graph limits (which are not themselves graphs) can be represented in several different ways; perhaps the most important is that every graph limit can be represented by a *kernel* (or *graphon*) on [0,1], i.e., a symmetric measurable function $W : [0,1]^2 \to [0,1]$. However, this representation is in general not unique, see e.g. [20, 4, 8, 3]. More generally, kernels can be defined on any probability space, see Section 2.

We use Γ to denote an arbitrary graph limit, and write Γ_W for the graph limit defined by a kernel W. We say that two kernels W and W' are *equivalent* if they define the same graph limit, i.e., if $\Gamma_W = \Gamma_{W'}$. We write $G_\nu \to \Gamma$ when the sequence (G_ν) converges to Γ (see [20], [5] and Section 2 below for definitions); if Γ is represented by a kernel W, i.e., if $\Gamma = \Gamma_W$, we also write $G_\nu \to W$.

Date: 21 January, 2011.

2000 *Mathematics Subject Classification.* 05C99.

The first author's research was supported in part by NSF grants CNS-0721983, CCF-0728928 and DMS-0906634, and ARO grant W911NF-06-1-0076.

Part of this research was carried when SJ visited the Isaac Newton Institute, Cambridge, during the programme Stochastic Processes in Communication Sciences, 2010.

Following [8], we denote the set of all graph limits by $\mathcal{U}_\infty$, and note that $\mathcal{U}_\infty$ is a compact metric space. Another version of the important compactness property for graph limits is that every sequence (G_ν) of graphs with $|G_\nu| \to \infty$ has a convergent subsequence, i.e., a subsequence converging to some $\Gamma \in \mathcal{U}_\infty$.

Given a suitable class $\mathcal{F}$ of graphs, it seems interesting to study the *graph limits of* $\mathcal{F}$, i.e., the set of graph limits arising as limits of sequences of graphs in $\mathcal{F}$. One interesting example is the class of *threshold graphs*, which has several different characterizations, see e.g. [23]. One of them is the monotonicity property of the neighbourhoods $N(v)$ of the vertices:

$$\begin{aligned}&\text{There exists a (linear) ordering } \prec \text{ of the vertices such that}\\ &\text{if } v \prec w, \text{ then } N(v) \setminus \{v,w\} \subseteq N(w) \setminus \{v,w\}.\end{aligned} \tag{1.1}$$

The graph limits of threshold graphs were studied by Diaconis, Holmes and Janson [7] (see also [21]), who showed that they are exactly the graph limits that can be represented by kernels W that take values in $\{0,1\}$ only and are *increasing*, in that

$$W(x_1,y_1) \le W(x_2,y_2) \qquad \text{if } 0 \le x_1 \le x_2 \le 1,\ 0 \le y_1 \le y_2 \le 1. \tag{1.2}$$

In other words, W is the indicator function of a symmetric increasing subset of $[0,1]^2$. (In this paper, 'increasing' should always be interpreted in the weak sense, i.e., as 'non-decreasing'.) Moreover, the representation by such a W is unique, if, as is usual, we identify functions that are equal a.e.

Note that the monotonicity properties in (1.1) and (1.2) are obviously related; this is perhaps best seen if (1.1) is rewritten as a monotonicity property of the adjacency matrix of the graph (with some exceptions at the diagonal), so even without the detailed technical study in [7], the condition (1.2) should not be surprising.

Increasing and decreasing kernels define the same set of graph limits, by the change of variables $x \mapsto 1-x$. Hence we shall talk about *monotone* kernels rather than increasing kernels, but for simplicity (and without loss of generality) we consider only increasing ones, so in this paper 'monotone' is regarded as synonymous with 'increasing'.

The main purpose of the present paper is to study the larger class of graph limits represented by arbitrary monotone kernels (taking any values in [0,1], rather than just the values 0 and 1), and the corresponding sequences of graphs. We shall also study analytic properties of monotone kernels themselves.

Definition. Let $\mathcal{W}_\uparrow$ be the set of monotone kernels on $[0,1]$, i.e., the set of all symmetric measurable functions $W : [0,1]^2 \to [0,1]$ that satisfy (1.2).

Let $\mathcal{U}_\uparrow$ be the corresponding class of graph limits, i.e., the class of graph limits that can be represented as Γ_W for some $W \in \mathcal{W}_\uparrow$. We call these graph limits *monotone*.

By definition, every monotone graph limit can be represented by a monotone kernel W on $[0,1]$, but note that a monotone graph limit may also have many representations by non-monotone kernels. For example, a monotone kernel can be rearranged by an arbitrary measure-preserving bijection from $[0,1]$ to itself, which will in general destroy monotonicity.

The classes $\mathcal{W}_\uparrow$ of monotone kernels and $\mathcal{U}_\uparrow$ of monotone graph limits are studied in Section 4. We show there that $\mathcal{W}_\uparrow$ is a compact subset of $L^1([0,1]^2)$, and that $\mathcal{U}_\uparrow$ is a compact subset of $\mathcal{U}_\infty$. In addition, we consider monotone kernels defined on other (ordered) probability spaces, showing that each such kernel is equivalent to a monotone kernel on $[0,1]$, so the class $\mathcal{U}_\uparrow$ is not enlarged by allowing arbitrary probability spaces.

Definition. A sequence (G_ν) of graphs with $|G_\nu| \to \infty$ is *quasimonotone* if it converges to the set $\mathcal{U}_\uparrow$, in the sense that each convergent subsequence has as its limit a graph limit in $\mathcal{U}_\uparrow$. In this case we will also say that (G_ν) is a sequence of quasimonotone graphs.

In particular, a sequence (G_ν) converging to a graph limit in $\mathcal{U}_\uparrow$ is quasimonotone. Note that it makes no formal sense to ask whether an individual graph is quasimonotone; just as for quasirandomness, quasimonotonicity is a property of sequences of graphs.

Example 1.1 (Threshold graphs are quasimonotone)**.** As noted above, each convergent sequence of threshold graphs converges to a limit represented by a 0/1-valued kernel $W \in \mathcal{W}_\uparrow$. Hence every sequence of threshold graphs (with orders tending to ∞) is quasimonotone.

Example 1.2 (Quasirandom graphs are quasimonotone)**.** Quasirandom graphs were introduced by Thomason [25, 26] as sequences (G_ν) of graphs that have certain properties typical of random graphs. A number of different such properties turn out to be equivalent, and there are thus many equivalent characterizations, see Chung, Graham and Wilson [6]. Another characterization, found by Lovász and Szegedy [20], is that a sequence (G_ν) is quasirandom if and only if it converges to a graph limit represented by a constant kernel $W(x,y) = p$, for some $p \in [0,1]$. (See also [19] and [13].) Since a constant function is monotone, $W \in \mathcal{W}_\uparrow$, and thus every quasirandom sequence of graphs is quasimonotone.

Example 1.3 (Random graphs are quasimonotone)**.** The sequence of random graphs $G(\nu, p)$ with some fixed $p \in [0,1]$ and $\nu = 1, 2, \dots$ (coupled in the natural way for different ν) is a.s. quasirandom, and thus a.s. quasimonotone.

Our main result (Theorem 1.5 below) is that quasimonotone graphs can be characterized by a weakening of (1.1). As is typical for conditions concerning convergence to graph limits, this weakening involves taking averages over subsets of the vertex set V, rather than imposing a condition for all vertices, and allows for a small 'error', making the condition asymptotic.

Given a graph G with vertex set $V = V(G)$, a vertex v of G and a subset A of V, let

$$e(v, A) := |N(v) \cap A| = |\{w \in A : w \sim v\}|$$

denote the number of edges from v to A.

Let x_+ denote the *positive part* of x, i.e., $\max\{x, 0\}$. Writing $n := |G| = |V|$, given a (linear) order $\prec$ on V and a subset $A \subseteq V$, define

$$\Omega_0(G, \prec, A) := \frac{1}{n^3} \sum_{v \prec w} \bigl(e(v, A \setminus \{w\}) - e(w, A \setminus \{v\})\bigr)_+ \tag{1.3}$$

$$= \frac{1}{n^3} \sum_{v \prec w} \bigl(e(v, A \setminus \{v, w\}) - e(w, A \setminus \{v, w\})\bigr)_+, \tag{1.4}$$

$$\Omega_0(G, \prec) := \max_{A \subseteq V} \Omega_0(G, \prec, A), \quad \text{and} \tag{1.5}$$

$$\Omega_0(G) := \min_{\prec} \Omega_0(G, \prec). \tag{1.6}$$

In the last line the minimum is taken over all $n!$ orders on V. The normalization by n^3 ensures that $0 \le \Omega_0 < 1$. In fact, $\Omega_0 < 1/2$, and this bound can be improved further, but this is not important for our purposes since we are interested in small values of Ω_0.

Note that $\Omega_0(G) = 0$ if and only if there exists an order $\prec$ such that $\Omega_0(G, \prec, A) = 0$ for every A, i.e., $e(v, A \setminus \{v, w\}) \le e(w, A \setminus \{v, w\})$ for all A and $v \prec w$, which easily is seen to be equivalent to (1.1), giving the following result.

Proposition 1.4. *A graph G is a threshold graph if and only if $\Omega_0(G) = 0$.* □

Note that Ω_0 is not intended as a measure of how far a graph is from being a threshold graph (for such a measure, see Section 8). Rather, we may think (informally!) of a typical quasimonotone graph as being similar to a random graph in which edges are independent, and the probability p_{ij} of an edge ij is increasing in i and in j. In such a graph, one cannot expect the neighbourhoods of different vertices to be even approximately nested. But one can expect that for all 'large' sets A of vertices, for most $i < j$, $e(i, A)$ will be smaller than (or at least not much larger than) $e(j, A)$. The idea is that a small value of $\Omega_0(G)$ detects this phenomenon, without relying on any given labelling of the vertices.

Some variations of the functional Ω_0 will be defined in Section 3, where we shall show that they are asymptotically equivalent for our purposes.

Our main result is the following, proved in Section 7. (All unspecified limits in this paper are taken as $\nu \to \infty$.)

Theorem 1.5. *Let (G_ν) be a sequence of graphs with $|G_\nu| \to \infty$. Then (G_ν) is quasimonotone if and only if $\Omega_0(G_\nu) \to 0$.*

We state a special case separately.

Theorem 1.6. *Let (G_ν) be a sequence of graphs with $|G_\nu| \to \infty$, and suppose that (G_ν) is convergent, i.e., $G_\nu \to \Gamma$ for some graph limit $\Gamma \in \mathcal{U}_\infty$. Then $\Gamma \in \mathcal{U}_\uparrow$ if and only if $\Omega_0(G_\nu) \to 0$.*

We give several results on monotone graph limits in Sections 4–6. These include a characterization in terms of a functional $\Omega(W)$ for kernels, analogous to Ω_0 for graphs. Along the way we prove some results about monotone kernels that may be of interest in their own right. For example, on functions that may be written as the difference between two monotone kernels, the L^1 norm and the cut norm may be bounded in terms of each other; see Theorem 5.5.

Remark 1.7. Lovász and Szegedy [22] have studied the class of graph limits represented by 0/1-valued kernels (and the corresponding graph properties); with a slight variation of their terminology we call such graph limits *random-free*. In contrast to the monotone case, it can be shown that *every* representing kernel of a random-free limit is a.e. 0/1-valued; see [14]. It follows that the graph limits that are both monotone and random-free are exactly the threshold graph limits.

In Section 8, we consider the functional obtained by taking the supremum over A inside the sum in (1.3) instead of outside as in (1.5). We shall show that this stronger functional characterizes convergence to threshold graph limits instead of monotone graph limits; we call the corresponding sequences of graphs *quasithreshold*.

1.1. **A problem.** The convergence $G_\nu \to \Gamma$ of a sequence (G_ν) of graphs to a graph limit Γ can be expressed using the *homomorphism numbers* $t(F,\cdot)$: $G_\nu \to \Gamma$ if and only if $t(F,G_\nu) \to t(F,\Gamma)$ for every fixed graph F; see e.g. [20], [5], [8] for definitions and further results. In particular, the graph limit Γ is characterized by the family $(t(F,\Gamma))_F$. The families $(t(F,\Gamma))_F$ that appear are characterized algebraically by Lovász and Szegedy [20].

Problem 1.8. Characterize the families $(t(F,\Gamma))_F$ that appear for $\Gamma \in \mathcal{U}_\uparrow$.

The rest of this paper is organized as follows. In the next section we review some basic properties of the cut metric that we shall rely on throughout the paper. In Section 3 we introduce some variants of the functional Ω_0 for graphs. In Section 4 we define analogous functionals for kernels and state several key properties; these are proved in the next two sections, and then our main results are deduced in Section 7. Finally, in Section 8 we discuss related functionals characterizing quasithreshold graphs.

2. Kernels and graph limits

We state here some standard definitions and results that we shall use later in the paper. For proofs and further details, see e.g. Borgs, Chayes, Lovász, Sós and Vesztergombi [5], Bollobás and Riordan [3], or Janson [12, 14].

Let $(\mathcal{S},\mathcal{F},\mu)$ be a probability space; for simplicity, we will usually abbreviate the notation to $\mathcal{S}$ or $(\mathcal{S},\mu)$.

A *kernel* (or *graphon*) on $\mathcal{S}$ is a symmetric measurable function $\mathcal{S}^2 \to [0,1]$. We let $\mathcal{W}(\mathcal{S})$ denote the set of all kernels on $\mathcal{S}$.

If W is an integrable function on $\mathcal{S}^2$, we define its *cut norm* by

$$\|W\|_\square := \sup_{\|f\|_\infty,\|g\|_\infty\le 1}\Bigl|\int_{\mathcal{S}^2} W(x,y)f(x)g(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\Bigr|, \tag{2.1}$$

where $\|\cdot\|_\infty$ denotes the norm in L^∞. In other words, the supremum in (2.1) is taken over all (real-valued) functions f and g with values in $[-1,1]$. (Several other versions exist, which are equivalent within constants.) By considering the supremum over f with g fixed, and vice versa, it is easy to see that the supremum is unchanged if we restrict f and g to take values in $\{\pm1\}$, so we have

$$\|W\|_\square = \sup_{f,g:\mathcal{S}\to\{\pm1\}}\Bigl|\int_{\mathcal{S}^2} W(x,y)f(x)g(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\Bigr|. \tag{2.2}$$

This norm defines a metric $\|W_1-W_2\|_\square$ for kernels on the same probability space $\mathcal{S}$; as usual, we identify kernels that are equal a.e.

The cut norm may be used to define another (semi)metric $\delta_\square$, the *cut metric*, as follows. If $\varphi:\mathcal{S}_1\to\mathcal{S}_2$ is a measure-preserving map between two probability spaces and W is a kernel on $\mathcal{S}_2$, we let W^φ be the kernel on $\mathcal{S}_1$ defined by $W^\varphi(x,y):=W\bigl(\varphi(x),\varphi(y)\bigr)$. Let W_1 be a kernel on a probability space $\mathcal{S}_1$ and W_2 a kernel on a possibly different probability space $\mathcal{S}_2$. Then

$$\delta_\square(W_1,W_2) := \inf_{\varphi_1,\varphi_2}\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_\square, \tag{2.3}$$

where the infimum is taken over all *couplings* (φ_1,φ_2) of $\mathcal{S}_1$ and $\mathcal{S}_2$, i.e., over all pairs of measure-preserving maps $\varphi_1:\mathcal{S}_3\to\mathcal{S}_1$ and $\varphi_2:\mathcal{S}_3\to\mathcal{S}_2$ from a third probability space $\mathcal{S}_3$. It is not difficult to verify that $\delta_\square$ satisfies the triangle inequality (see e.g. [14]), but note that $\delta_\square(W_1,W_2)$ may be 0 even if $W_1\neq W_2$, for example if $W_1=W_2^\varphi$ for some measure-preserving $\varphi:\mathcal{S}_1\to\mathcal{S}_2$. Hence, $\delta_\square$ is really a semimetric (but is usually called a metric for simplicity).

Note that $\delta_\square(W_1,W_2)$ is defined for kernels on different spaces. Moreover, it is invariant under measure-preserving maps: $\delta_\square(W_1^{\varphi_1},W_2^{\varphi_2})=\delta_\square(W_1,W_2)$ for any measure-preserving maps $\varphi_k:\mathcal{S}_k'\to\mathcal{S}_k$, $k=1,2$.

Although we allow couplings (φ_1,φ_2) defined on an arbitrary third space $\mathcal{S}_3$, in (2.3) it suffices to consider the case when $\mathcal{S}_3=\mathcal{S}_1\times\mathcal{S}_2$, with a measure μ having marginals μ_1 and μ_2, taking for φ_1 and φ_2 the projections $\pi_k:\mathcal{S}_1\times\mathcal{S}_2\to\mathcal{S}_k$, $k=1,2$. In fact, for an arbitrary coupling (φ_1,φ_2) defined on a space $(\mathcal{S}_3,\mu_3)$, the mapping $(\varphi_1,\varphi_2):\mathcal{S}_3\to\mathcal{S}_1\times\mathcal{S}_2$ maps μ_3 to a measure μ on $\mathcal{S}_1\times\mathcal{S}_2$ with the right marginals, and it is easily seen that $\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_\square=\|W_1^{\pi_1}-W_2^{\pi_2}\|_\square$.

Although this will be of much lesser importance, we also define the corresponding rearrangement-invariant version of the L^1 distance:

$$\delta_1(W_1,W_2) := \inf_{\varphi_1,\varphi_2}\|W_1^{\varphi_1}-W_2^{\varphi_2}\|_{L^1(\mathcal{S}_3^2)}. \tag{2.4}$$

The coupling definition (2.3) of the cut metric is valid for all $\mathcal{S}_1$ and $\mathcal{S}_2$, but in common special cases it is possible, and often convenient, to use other, equivalent, definitions. For example, if $\mathcal{S}_1 = \mathcal{S}_2 = [0,1]$ (equipped with the Lebesgue measure, as always), then as shown by Borgs, Chayes, Lovász, Sós and Vesztergombi [5, Lemma 3.5],

$$\delta_\square(W_1, W_2) := \inf_\varphi \|W_1 - W_2^\varphi\|_\square, \tag{2.5}$$

taking the infimum over all measure-preserving bijections $[0,1] \to [0,1]$.

We say that two kernels W_1 and W_2 are *equivalent* if $\delta_\square(W_1, W_2) = 0$. The set of equivalence classes is thus a metric space with the metric $\delta_\square$. A central result [20, 5] is that these equivalence classes are in one-to-one correspondence with the graph limits. In other words, each kernel W defines a graph limit Γ_W, every graph limit can be represented by a kernel in this way, and two kernels define the same graph limit if and only if they are equivalent. Thus, the cut metric defines the same notion of equivalence as the one mentioned in the introduction. Furthermore, W_1 and W_2 are equivalent if and only if $\delta_1(W_1, W_2) = 0$, see e.g. [14].

Every kernel is equivalent to a kernel on $[0,1]$, so it suffices to consider such kernels. (We shall not use this restiction in the present paper, however.)

One manifestation of the connection between graph limits and kernels is the following: If G is a graph with vertices labelled 1,2,...,n, let $A_G(i,j) := \mathbf{1}\{i \sim j\}$ define its adjacency matrix, and let

$$W_G(x,y) := A_G\bigl(\lceil nx \rceil, \lceil ny \rceil\bigr).$$

This defines a kernel W_G on $[0,1]$ (or rather on $(0,1]$, which is equivalent). A sequence of graphs with $|G_\nu| \to \infty$ converges to the graph limit $\Gamma = \Gamma_W$ if and only if $\delta_\square(W_{G_\nu}, W) \to 0$.

Note that W_G depends on the labelling of the vertices of G, but only in a rather trivial way, and different labellings yield equivalent kernels. Here, in the study of monotone kernels, the ordering is relevant. If G is a graph with a given order $\prec$ on V, we therefore define $W_G = W_{G,\prec}$ as above, but using the labelling of the vertices with $1 \prec 2 \prec \cdots$, ignoring the original labelling, if any.

3. Further measures of quasimonotonicity

In Section 1 we defined a functional Ω_0 that measures, in an averaged sense, how far the adjacency matrix of a graph is from being monotone. There are several natural variations of the definition; we shall concentrate on two.

Firstly, in (1.3) and (1.4), we were careful to exclude v and w from the set A; this had the advantage of making $\Omega_0(G)$ exactly zero when G is a threshold graph. But most of the time it is more convenient not to do this. Instead, we consider

$$\Omega_1(G, \prec, A) := \frac{1}{n^3} \sum_{v \prec w} \bigl(e(v,A) - e(w,A)\bigr)_+, \tag{3.1}$$

which differs from (1.4) in that we count all edges into A, and not just the edges into $A \setminus \{v, w\}$. This changes each edge count by at most 1, so

$$|\Omega_0(G, \prec, A) - \Omega_1(G, \prec, A)| < 1/n. \tag{3.2}$$

As in (1.5) and (1.6), we set

$$\Omega_1(G, \prec) := \max_{A \subseteq V} \Omega_1(G, \prec, A), \quad \text{and} \tag{3.3}$$

$$\Omega_1(G) := \min_{\prec} \Omega_1(G, \prec). \tag{3.4}$$

Before turning to our second variant, let us note a basic property of Ω_0. Let $\overline{e}(v, A)$ denote the number of edges from v to A in the complement G^{c} of G. If $v \notin A$, then $\overline{e}(v, A) = |A| - e(v, A)$. Hence, for any v, w and A,

$$\overline{e}(w, A \setminus \{v, w\}) - \overline{e}(v, A \setminus \{v, w\}) = e(v, A \setminus \{v, w\}) - e(w, A \setminus \{v, w\}).$$

From (1.4) it follows that $\Omega_0(G^{\mathrm{c}}, \succ, A) = \Omega_0(G, \prec, A)$, where, naturally, $\succ$ denotes the reverse of the order $\prec$. Thus $\Omega_0(G^{\mathrm{c}}, \succ) = \Omega_0(G, \prec)$ and $\Omega_0(G^{\mathrm{c}}) = \Omega_0(G)$.

For Ω_1 one can show similarly, or deduce using (3.2), that $|\Omega_1(G^{\mathrm{c}}) - \Omega_1(G)| \le 2/n$, say.

Despite the above symmetry property of Ω_0, the following 'locally symmetrized' version of the definition turns out to have technical advantages. Given a graph G, an order $\prec$ on $V(G)$, and $A \subseteq V(G)$, set

$$\Omega_2(G, \prec, A) := \Omega_1(G, \prec, A) + \Omega_1(G, \prec, V \setminus A), \tag{3.5}$$

$$\Omega_2(G, \prec) := \max_{A \subseteq V} \Omega_2(G, \prec, A) \tag{3.6}$$

and

$$\Omega_2(G) := \min_{\prec} \Omega_2(G, \prec). \tag{3.7}$$

Of course, we could define a corresponding symmetrization of Ω_0, but we shall not bother.

It is easily seen that all our functionals Ω_j take values in $[0, 1]$ (in fact, in $[0, \frac{1}{2})$). We have the following relations.

Lemma 3.1. *If G is a graph with $|G| = n$, then*

$$|\Omega_0(G) - \Omega_1(G)| < 1/n, \tag{3.8}$$

and

$$\Omega_1(G) \le \Omega_2(G) \le 2\Omega_1(G). \tag{3.9}$$

Consequently, if (G_ν) is a sequence of graphs with $|G_\nu| \to \infty$, then $\Omega_j(G_\nu) \to 0$ for some j if and only if this holds for all $j = 0, 1, 2$.

Proof. The inequality (3.8) is immediate from (3.2).

The definition (3.5) implies that

$$\Omega_1(G, \prec) \le \Omega_2(G, \prec) \le 2\Omega_1(G, \prec), \tag{3.10}$$

which in turn implies (3.9). ∎

Remark 3.2. Instead of summing in (1.4) or (3.1), in analogy with the standard definition of ε-regular partitions (see e.g. [2, Section IV.5]), we may count the number of 'bad' pairs (v,w) of vertices $v \prec w$ where the difference $e(v,A)-e(w,A)$ is larger than εn, for some small ε. This suggests the following definition: with $\prec$ an order on the vertex set V, $n := |V|$, and A a subset of V, set

$$\Omega_1'(G,\prec,A) := \inf\Bigl\{\varepsilon > 0 : \bigl|\bigl\{v \prec w : e(v,A) > e(w,A) + \varepsilon n\bigr\}\bigr| \le \varepsilon n^2\Bigr\},$$

and define $\Omega_1'(G)$ by taking the maximum over A with $\prec$ fixed, and then minimizing over $\prec$. It is a standard observation that if $x_1,\dots,x_a$ take values in $[0,b]$, then $\sum_i x_i \ge \varepsilon ab$ implies that there are at least $\varepsilon a/2$ of the x_i that are at least $\varepsilon b/2$, and that if at least εa of the x_i are at least εb, then the sum is at least $\varepsilon^2 ab$. Using this it is easy to check that Ω_1 and Ω_1' are bounded by suitable functions of each other. In fact, it turns out that

$$\tfrac12\Omega_1(G) \le \Omega_1'(G) \le \Omega_1(G)^{1/2}.$$

We can also define corresponding modifications of the other Ω_j.

Remark 3.3. Proposition 1.4 says that a graph G is a threshold graph if and only if $\Omega_0(G) = 0$. This does not hold for Ω_1; in fact, if G contains an edge vw, with $v \prec w$, then $\Omega_1(G,\prec,\{w\}) \ge n^{-3}e(v,\{w\}) = n^{-3}$ by (3.1); hence $\Omega_1(G) \ge n^{-3}$ unless G is empty. Consequently, $\Omega_1(G) > 0$ for every non-empty graph G. On the other hand, Proposition 1.4 and Lemma 3.1 show that $\Omega_1(G) \le 1/n$ for every threshold graph.

We defined each $\Omega_j(G)$ by taking the minimum of $\Omega_j(G,\prec)$ over all possible orderings $\prec$ of the vertices. As the next lemma shows, for Ω_2, ordering the vertices by their degrees $d(v) := e(v,V)$ (resolving ties arbitrarily) is optimal. This is the main reason for considering Ω_2.

Lemma 3.4. *Let $<$ be an order on V such that $v < w \implies d(v) \le d(w)$. Then $\Omega_2(G) = \Omega_2(G,<)$.*

Proof. The inequality $\Omega_2(G) \le \Omega_2(G,<)$ is immediate from the definition (3.7), so it suffices to prove the reverse inequality.

Let $\prec$ be any order on V. If $v < w$, then $e(v,V) = d(v) \le d(w) = e(w,V)$ and thus, for $A \subseteq V$,

$$\begin{aligned} e(v,A) - e(w,A) &= e(v,V) - e(w,V) + e(w,V\setminus A) - e(v,V\setminus A) \\ &\le e(w,V\setminus A) - e(v,V\setminus A). \end{aligned} \tag{3.11}$$

Let $f(v,w,A) := \bigl(e(v,A)-e(w,A)\bigr)_+$ and $g(v,w,A) := f(v,w,A)+f(v,w,V\setminus A)$. By (3.11), if $v < w$, then $f(v,w,A) \le f(w,v,V\setminus A)$ and thus

$$g(v,w,A) \le f(w,v,V\setminus A) + f(w,v,A) = g(w,v,A). \tag{3.12}$$

Using (3.12) for $v < w$ with $v \succ w$, we obtain

$$\begin{aligned}\Omega_2(G,<,A) &:= \frac{1}{n^3}\sum_{v<w} g(v,w,A)\\ &= \frac{1}{n^3}\sum_{\substack{v<w\\ v\prec w}} g(v,w,A) + \frac{1}{n^3}\sum_{\substack{v<w\\ v\succ w}} g(v,w,A)\\ &\le \frac{1}{n^3}\sum_{\substack{v<w\\ v\prec w}} g(v,w,A) + \frac{1}{n^3}\sum_{\substack{w>v\\ w\prec v}} g(w,v,A)\\ &= \frac{1}{n^3}\sum_{v\prec w} g(v,w,A) = \Omega_2(G,\prec,A).\end{aligned}$$

Hence, by (3.6), $\Omega_2(G,<) \le \Omega_2(G,\prec)$. Since $\prec$ is arbitrary, this yields $\Omega_2(G,<) = \Omega_2(G)$. □

As an immediate consequence of Lemmas 3.4 and 3.1, we have the following result for Ω_1.

Corollary 3.5. *Let $<$ be an order on V such that $v < w \implies d(v) \le d(w)$. Then $\Omega_1(G) \le \Omega_1(G,<) \le 2\Omega_1(G)$.*

Proof. By (3.10), Lemma 3.4 and (3.9),

$$\Omega_1(G) \le \Omega_1(G,<) \le \Omega_2(G,<) = \Omega_2(G) \le 2\Omega_1(G).$$

(Alternatively, one can use a simplified version of the proof of Lemma 3.4.) □

Using a symmetrized version of Ω_0, or otherwise, it is easy to prove the corresponding result for Ω_0.

Remark 3.6. If G is regular, then any order $<$ satisfies the condition of Lemma 3.4 and Corollary 3.5, so these results show that $\Omega_2(G,<)$ is the same for all orders, and $\Omega_1(G,<)$ is the same for all orders within a factor of 2; the latter holds also for Ω_0.

The factor 2 in Corollary 3.5 is annoying but not really harmful for our purposes. It is best possible, as shown by the following example.

Example 3.7. Consider a balanced complete bipartite graph $G = K_{m,m}$ (so $n = 2m$), with bipartition (V_1, V_2). Given an order $\prec$ on the vertex set $V_1 \cup V_2$, let $N_{ij} := \big|\{(x,y) \in V_i \times V_j : x \prec y\}\big|$. Note that

$$N_{12} + N_{21} = \big|V_1 \times V_2\big| = m^2. \tag{3.13}$$

Let $A \subseteq V = V_1 \cup V_2$ and let $a_i = |A \cap V_i|$, $i = 1, 2$. Then $e(v, A) = a_2$ if $v \in V_1$ and $e(v, A) = a_1$ if $v \in V_2$. Hence,

$$\begin{aligned}n^3\Omega_1(G,\prec,A) &= \sum_{v\prec w}\big(e(v,A) - e(w,A)\big)_+\\ &= N_{12}(a_2 - a_1)_+ + N_{21}(a_1 - a_2)_+.\end{aligned} \tag{3.14}$$

Since a_1 and a_2 can be freely chosen in $\{0,\dots,m\}$, we have $a_1-a_2\in\{-m,\dots,m\}$, and maximizing over A yields

$$n^3\Omega_1(G,\prec)=m\max\{N_{12},N_{21}\}. \tag{3.15}$$

If $\prec_1$ is an order with all elements of V_1 coming first, then $N_{12}=m^2$ and $N_{21}=0$, and thus

$$\Omega_1(G,\prec_1)=m^3/n^3=1/8.$$

On the other hand, if m is even and $\prec_2$ is an order which starts with $m/2$ elements of V_1, continues with all of V_2, and finishes with the remaining half of V_1, then $N_{12}=N_{21}=m^2/2$, and thus

$$\Omega_1(G,\prec_2)=\tfrac12 m^3/n^3=1/16. \tag{3.16}$$

Thus $\Omega_1(G,\prec_1)=2\Omega_1(G,\prec_2)$ although G is regular and Corollary 3.5 applies to every order.

For Ω_0, the ratio between $\Omega_0(G,\prec_1)$ and $\Omega_0(G,\prec_2)$ is $2-O(1/n)$ by (3.2).

Note that for any order $\prec$, (3.13) implies $\max\{N_{12},N_{21}\}\geq m^2/2$, and thus (3.15) yields

$$\Omega_1(G)\geq n^{-3}m^3/2=1/16. \tag{3.17}$$

Consequently, if m is even, then (3.16) shows that

$$\Omega_1(G)=\Omega_1(G,\prec_2)=1/16 \qquad (m\text{ even}). \tag{3.18}$$

On the other hand, if m is odd, then since $N_{12}+N_{21}=m^2$ is odd, for any order $\prec$ we have $\max\{N_{12},N_{21}\}\geq(m^2+1)/2$, and this is attained for some $\prec$. Thus (3.15) now yields

$$\Omega_1(G)=n^{-3}m(m^2+1)/2>1/16 \qquad (m\text{ odd}). \tag{3.19}$$

We thus have

$$\begin{cases}\Omega_1(K_{m,m})=1/16, & m\text{ even},\\ \Omega_1(K_{m,m})=(1+m^{-2})/16>1/16, & m\text{ odd}.\end{cases} \tag{3.20}$$

For Ω_2, the situation is simpler. It follows from (3.14) that $n^3\Omega_1(G,\prec,V\setminus A)=N_{12}(a_2-a_1)_-+N_{21}(a_1-a_2)_-$, and thus, using (3.13),

$$n^3\Omega_2(G,\prec,A)=N_{12}|a_2-a_1|+N_{21}|a_1-a_2|=m^2|a_1-a_2|. \tag{3.21}$$

Maximizing over A we find $\Omega_2(G,\prec)=m^3/n^3=1/8$ for every order $\prec$, cf. Remark 3.6, and thus $\Omega_2(G)=1/8$.

If we modify G by adding a perfect matching inside V_2 (assuming m is even) then every order $<$ satisfying the condition of Corollary 3.5 is of the type $\prec_1$. The added edges change each $e(v,A)$ by at most 1, and thus each $\Omega_j(G,\prec,A)$ is changed by at most $1/n$. Hence this yields an example where $\Omega_j(G,<)=(2-O(1/n))\Omega_j(G)$ for $j=0,1$, for every order $<$ considered in Corollary 3.5.

4. Monotone kernels and graph limits

We begin by extending the definition of monotone kernels to other probability spaces.

Definition. An *ordered probability space* $(\mathcal{S}, \prec) = (\mathcal{S}, \mathcal{F}, \mu, \prec)$ is a probability space $(\mathcal{S}, \mathcal{F}, \mu)$ with a (linear) order $\prec$ that is measurable, i.e., $\{(x, y) : x \prec y\}$ is a measurable subset of $\mathcal{S} \times \mathcal{S}$.

Note that it follows that $\{(x, y) : x \succ y\}$ and $\{(x, y) : x = y\}$ are measurable.

All orders considered in this paper are assumed to be measurable, even if we only sometimes say so explicitly. Similarly, we only consider subsets and functions that are measurable.

The standard example of an ordered probability space is $[0, 1]$ with Lebesgue measure and the standard order. $[0, 1]$ is always equipped with these unless we say otherwise.

Definition. Let $(\mathcal{S}, \prec)$ be an ordered probability space. A *monotone kernel* on $(\mathcal{S}, \prec)$ is a kernel $W : \mathcal{S}^2 \to [0, 1]$ such that

$$W(x_1, y_1) \le W(x_2, y_2) \qquad \text{if } x_1 \preceq x_2,\ y_1 \preceq y_2. \tag{4.1}$$

Let $\mathcal{W}_\uparrow(\mathcal{S}, \prec)$ be the set of monotone kernels on $(\mathcal{S}, \prec)$, noting that $\mathcal{W}_\uparrow = \mathcal{W}_\uparrow([0, 1])$. We shall prove the following properties of $\mathcal{W}_\uparrow(\mathcal{S}, \prec)$ in Sections 5 and 6.

Theorem 4.1. *Let $(\mathcal{S}, \prec)$ be an ordered probability space.*
(i) *$\mathcal{W}_\uparrow(\mathcal{S}, \prec)$ is a compact subset of $L^1(\mathcal{S}^2)$.*
(ii) *Two kernels in $\mathcal{W}_\uparrow(\mathcal{S}, \prec)$ are equivalent if and only if they are a.e. equal.*
(iii) *The metrics $\|W_1 - W_2\|_{L^1}$, $\delta_1(W_1, W_2)$, $\|W_1 - W_2\|_\square$, and $\delta_\square(W_1, W_2)$ are equivalent on $\mathcal{W}_\uparrow(\mathcal{S}, \prec)$, i.e., induce the same topology.*

Recall that $\mathcal{U}_\uparrow$ denotes the set of monotone graph limits, i.e., the class of graph limits that can be represented as Γ_W for some $W \in \mathcal{W}_\uparrow = \mathcal{W}_\uparrow([0, 1])$.

Corollary 4.2. *Each monotone graph limit has a representation as Γ_W for some $W \in \mathcal{W}_\uparrow = \mathcal{W}_\uparrow([0, 1])$ with W unique up to equality a.e. Furthermore, there is a homeomorphism between $\mathcal{U}_\uparrow$ and $\mathcal{W}_\uparrow([0, 1])$, regarded as a subset of $L^1([0, 1]^2)$.*

Proof. Immediate from Theorem 4.1 and the fact that the metric on the set of graph limits is equivalent to $\delta_\square$ on the corresponding kernels. $\square$

In Section 1 we defined $\mathcal{U}_\uparrow$ as the set of graph limits that can be represented by some $W \in \mathcal{W}_\uparrow([0, 1])$. The following theorem shows that we may allow monotone kernels on arbitrary ordered probability spaces without changing $\mathcal{U}_\uparrow$, i.e.,

$$\mathcal{U}_\uparrow = \{\Gamma : \exists\, (\mathcal{S}, \prec) \text{ and } W \in \mathcal{W}_\uparrow(\mathcal{S}, \prec) \text{ such that } \Gamma = \Gamma_W\}.$$

This version of the definition is perhaps more natural than considering $[0,1]$ only; on the other hand, it is often convenient to use $[0,1]$.

Theorem 4.3. *Let $(\mathcal{S},\prec)$ be an ordered probability space, and let $W\in\mathcal{W}_{\uparrow}(\mathcal{S},\prec)$. Then there is a monotone kernel $W'\in\mathcal{W}_{\uparrow}([0,1])$ that is equivalent to W. Equivalently, $\Gamma_W\in\mathcal{U}_{\uparrow}$.*

We shall next define two quantitative measures of how far a kernel is from being monotone, in analogy with (1.3)–(1.6) (or, more closely, (3.1), (3.3) and (3.4)), and (3.5)–(3.7).

Given $W\in L^1(\mathcal{S}^2)$, a (measurable) order $\prec$ on $\mathcal{S}$, and a (measurable) subset A of $\mathcal{S}$, set

$$\Omega_1(W,\prec,A):=$$
$$\iint_{x\prec y}\left(\int_A W(x,z)\,\mathrm{d}\mu(z)-\int_A W(y,z)\,\mathrm{d}\mu(z)\right)_+\mathrm{d}\mu(x)\,\mathrm{d}\mu(y),\tag{4.2}$$

$$\Omega_2(W,\prec,A):=\Omega_1(W,\prec,A)+\Omega_1(W,\prec,\mathcal{S}\setminus A),\tag{4.3}$$

and, for $j=1,2$,

$$\Omega_j(W,\prec):=\sup_{A\subseteq\mathcal{S}}\Omega_j(W,\prec,A),\tag{4.4}$$

$$\Omega_j(W):=\inf_{\prec}\Omega_j(W,\prec),\tag{4.5}$$

where the infimum is over all measurable orders on $\mathcal{S}$. Note that

$$\Omega_1(W)\le\Omega_2(W)\le2\Omega_1(W).\tag{4.6}$$

For $A\subseteq\mathcal{S}$, let $W_A(x):=\int_A W(x,z)\,\mathrm{d}\mu(z)$. Then (4.2) can be written as

$$\Omega_1(W,\prec,A)=\iint_{x\prec y}\Bigl(W_A(x)-W_A(y)\Bigr)_+\mathrm{d}\mu(x)\,\mathrm{d}\mu(y).\tag{4.7}$$

Remark 4.4. It is easily seen that

$$\Omega_1(W,\prec)=\sup_{f,g}\iiint_{x\prec y}\bigl(W(x,z)-W(y,z)\bigr)f(x,y)g(z)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z),\tag{4.8}$$

where the supremum is taken over all $f:\mathcal{S}^2\to\{0,1\}$ and $g:\mathcal{S}\to\{0,1\}$, and that allowing all $f:\mathcal{S}^2\to[0,1]$ and $g:\mathcal{S}\to[0,1]$ yields the same result. Thus $\Omega_1(W,\prec)$ can be seen as a one-sided version of the cut norm of the function $\bigl(W(x,z)-W(y,z)\bigr)\mathbf{1}_{\{x\prec y\}}$ on $\mathcal{S}^2\times\mathcal{S}$.

Similarly, $\Omega_2(W,\prec)$ equals

$$\sup_{f_1,f_2,g}\iiint_{x\prec y}\bigl(W(x,z)-W(y,z)\bigr)\bigl(f_1(x,y)g(z)+f_2(x,y)(1-g(z))\bigr)\cdot\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z),\tag{4.9}$$

where the supremum is taken either over all $f_1,f_2:\mathcal{S}^2\to\{0,1\}$ and $g:\mathcal{S}\to\{0,1\}$, or over all $f_1,f_2:\mathcal{S}^2\to[0,1]$ and $g:\mathcal{S}\to[0,1]$.

In the light of (4.6), Ω_1 and Ω_2 are essentially equivalent. In particular $\Omega_1(W) = 0 \iff \Omega_2(W) = 0$. When the difference is not important, we simply write Ω; formally, this may be read as Ω_1. Occasionally, there are advantages to considering one or the other variant.

Theorem 4.5. *Let $(\mathcal{S}, \prec)$ be an ordered probability space and let W be a kernel on $\mathcal{S}$. Then $\Omega(W, \prec) = 0$ if and only if W is a.e. equal to a monotone kernel.*

As noted above, Ω_j, $j = 1, 2$, is an analogue of Ω_j defined earlier for graphs. Indeed, there is a simple relation.

Lemma 4.6. *If G is a graph with an order $\prec$ on the vertex set V, and $<$ denotes the standard order on $[0, 1]$, then $\Omega_j(W_G, <) = \Omega_j(G, \prec)$ for $j = 1, 2$.*

For Ω_2, we shall show that Lemma 4.6 implies a corresponding result after minimizing over the relevant orderings.

Lemma 4.7. *If G is a graph, then $\Omega_2(W_G) = \Omega_2(G)$.*

Note that W_G depends on the labelling of the vertices in G, but this is harmless since the different versions differ by measure-preserving bijections of $[0, 1]$ (in fact, permutations of subintervals) and obviously have the same $\Omega_j(W_G)$.

Remark 4.8. Let $G = K_{m,m}$ as in Example 3.7. Then W_G does not depend on m, and one can check that $\Omega_1(W_G) = 1/16$. For m odd, we have $\Omega_1(G) > 1/16$ by (3.19). Thus we can have $\Omega_1(W_G) < \Omega_1(G)$. It seems likely that the difference is bounded by some function tending to 0 as $n \to \infty$, but we have not proved anything stronger than $\Omega_1(W_G) \le \Omega_1(G) \le 2\Omega_1(W_G)$, which follows from Lemma 4.7 and the relationship between Ω_1 and Ω_2.

Remark 4.9. Given a graph G, define W_G^V as the adjacency matrix of G, regarded as a kernel on $V = V(G)$, which we regard as a probability space with the uniform probability measure (each point has mass $1/|G|$). It is easily verified that $\Omega_1(W_G^V, \prec, A) = \Omega_1(G, \prec, A)$ for every order $\prec$ on V and every set $A \subseteq V$. Hence $\Omega_1(W_G^V, \prec) = \Omega_1(G, \prec)$ for every order $\prec$ and $\Omega_1(W_G^V) = \Omega_1(G)$, and the same holds for Ω_2.

Note that W_G^V and W_G are equivalent kernels. It follows from Lemma 4.7 that $\Omega_2(W_G^V) = \Omega_2(W_G)$, but Remark 4.8 shows that $\Omega_1(W_G^V) > \Omega_1(W_G)$ if $G = K_{m,m}$ with m odd. (See also Corollary 6.7 and Remark 6.8 below.)

Remark 4.10. In (4.5), we take the infimum over all measurable orders on $\mathcal{S}$. In general, this may be problematic, since there are probability spaces with no measurable orders, see Example 4.12 below. In such cases, we interpret (4.5) as $\Omega_j(W) = \infty$ (or perhaps 1), but this has the unhappy consequence that two equivalent kernels W_1 and W_2 may have $\Omega_2(W_1) \neq \Omega_2(W_2)$. For example, let W_1 and W_2 both be constant $1/2$, with W_1 defined on $[0,1]$ and W_2 on a space $\mathcal{S}$ with no measurable order; then $\Omega_2(W_1) = 0$ and $\Omega_2(W_2) = \infty$. In the sequel we therefore consider only $\mathcal{S}$ that have at least one measurable order. Even in this case, equivalent kernels may have different Ω_1; see Remark 4.9. We will show in Corollary 6.7 that there is no such problem for Ω_2. The case $\Omega(W) = 0$ is covered by the following theorem.

Theorem 4.11. *Let W be a kernel on a probability space $\mathcal{S}$ with at least one measurable order. Then the following are equivalent.*

(i) $\Omega(W) = 0$.
(ii) *There exists a measurable order $\prec$ on $\mathcal{S}$ such that W is a.e. equal to a monotone kernel on $(\mathcal{S}, \prec)$.*
(iii) *W is equivalent to a monotone kernel on some ordered probability space.*
(iv) *W is equivalent to a monotone kernel on $[0,1]$.*
(v) *Γ_W is a monotone graph limit.*

Example 4.12. Let $\mathcal{S} = [0,1]$, but equipped with the σ-field $\mathcal{F}_0$ consisting of the subsets of $\mathcal{S}$ that are either countable or have a countable complement. For the measure μ we take the restriction of the Lebesgue measure to $\mathcal{F}$. (Thus, $\mu(A) = 0$ if A is countable, and $\mu(A) = 1$ otherwise.)

Let $\mathcal{C}$ be the family of countable subsets of $\mathcal{S}$. The σ-field $\mathcal{F} \times \mathcal{F}$ is contained in the σ-field

$$\bigl\{A \subseteq \mathcal{S}^2 : \exists B_1, B_2 \in \mathcal{C} \text{ such that } A \text{ or } \mathcal{S} \setminus A \subseteq (B_1 \times \mathcal{S}) \cup (\mathcal{S} \times B_2)\bigr\}.$$

Thus, if $\prec$ is a measurable order, then there exist $B_1, B_2 \in \mathcal{C}$ such that either

$$\{(x,y) : x \prec y\} \subseteq (B_1 \times \mathcal{S}) \cup (\mathcal{S} \times B_2)$$

or

$$\{(x,y) : x \succeq y\} \subseteq (B_1 \times \mathcal{S}) \cup (\mathcal{S} \times B_2);$$

in the latter case we have

$$\{(x,y) : x \prec y\} \subset \{(x,y) : x \preceq y\} \subseteq (B_2 \times \mathcal{S}) \cup (\mathcal{S} \times B_1).$$

However, in both cases we find that if we choose two distinct $x, y \notin (B_1 \cup B_2)$, then neither $x \prec y$ nor $y \prec x$ holds, which is a contradiction. Thus $(\mathcal{S}, \mathcal{F}, \mu)$ is a probability space supporting no measurable orders.

5. Proofs of Theorems 4.1–4.3

A *downset* in an ordered set $(\mathcal{S}, \prec)$ is a subset A such that if $x \prec y$ and $y \in A$, then $x \in A$. We begin with two lemmas concerning simple (and certainly well-known) properties of downsets; for completeness we give full proofs.

Lemma 5.1. (i) *If A and B are downsets in a linearly ordered set $(\mathcal{S}, \prec)$, then $A \subseteq B$ or $B \subseteq A$.*

(ii) *If A and B are downsets in an ordered probability space $(\mathcal{S}, \prec)$ with $\mu(A) < \mu(B)$, then $A \subset B$.*

Proof. (i): Otherwise there would exist $x \in A \setminus B$ and $y \in B \setminus A$, but then neither $x \prec y$, $y \prec x$ nor $x = y$ is possible.

(ii): Now $B \subseteq A$ is impossible, and the result follows by (i). □

Lemma 5.2. *If $(\mathcal{S}, \prec)$ is an ordered probability space without atoms, then for every $t \in [0, 1]$ there exists a downset $D(t)$ with $\mu(D(t)) = t$. Furthermore, $D(t) \subset D(u)$ when $t < u$.*

Proof. It suffices to prove the first statement; the second then follows by Lemma 5.1(ii).

For $x \in \mathcal{S}$, let D_x be the downset $\{y \in \mathcal{S} : y \preceq x\}$. Let $X = X_0, X_1, X_2, \ldots$ be an i.i.d. sequence of random points in $\mathcal{S}$ (with the distribution μ). Since there are no atoms, $\mathbb{P}(X_i = X_j) = 0$ for all $i \neq j$. Thus, for every n, $X_0, \ldots, X_n$ are a.s. distinct, and by symmetry, all $(n+1)!$ orderings of them have the same probability $1/(n + 1)!$. Hence,

$$\mathbb{E}\bigl(\mu(D_X)^n\bigr) = \mathbb{P}(X_1, \ldots, X_n \prec X_0) = \frac{n!}{(n + 1)!} = \frac{1}{n + 1}, \qquad n \geq 1.$$

Consequently, $\mu(D_X)$ has the same moments as the uniform distribution $U(0, 1)$, and thus $\mu(D_X) \sim U(0, 1)$.

It follows that the set $\{\mu(D_x) : x \in \mathcal{S}\}$ is a dense subset of [0,1]. Hence, for every $t \in (0, 1]$, there exists a sequence $(x_i)_i$ in $\mathcal{S}$ such that $\mu(D_{x_i}) \nearrow t$ as $i \to \infty$. Then $D_{x_i} \subset D_{x_{i+1}}$ for $i \geq 1$ by Lemma 5.1(ii), and we can take $D(t) := \bigcup_{i=1}^{\infty} D_{x_i}$, which is a downset with $\mu(D(t)) := \lim_{i\to\infty} \mu(D_{x_i}) = t$. For $t = 0$ we take $D(0) := \emptyset$. □

Given an integrable function W on $\mathcal{S}^2$ and $A, B \subseteq \mathcal{S}$ with $\mu(A), \mu(B) > 0$, let

$$\overline{W}(A, B) := \frac{1}{\mu(A)\mu(B)} \iint_{A\times B} W(x, y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \tag{5.1}$$

denote the average of W over $A \times B$. If $\mathcal{P} = \{A_i\}$ is a finite partition of $\mathcal{S}$, we say that a function on $\mathcal{S}^2$ is a $\mathcal{P}$-*step function* if it is constant on each set $A_i \times A_j$. (A *step function* on $\mathcal{S}^2$ is a $\mathcal{P}$-step function for some finite partition $\mathcal{P}$.) If $W \in L^1(\mathcal{S}^2)$, we let $W_{\mathcal{P}}$ be the $\mathcal{P}$-step function defined by

$$W_{\mathcal{P}}(x, y) = \overline{W}(A_i, A_j) \quad \text{for} \quad x \in A_i,\, y \in A_j. \tag{5.2}$$

If some A_i has measure 0, then $W_{\mathcal{P}}$ is not defined everywhere, but it is always defined a.e., which suffices for us. Note that $W_{\mathcal{P}}$ is the conditional expectation of W given the σ-field $\mathcal{F}_{\mathcal{P}} \times \mathcal{F}_{\mathcal{P}}$, where $\mathcal{F}_{\mathcal{P}}$ is the finite σ-field on $\mathcal{S}$ generated by $\mathcal{P}$. It follows that $\|W_{\mathcal{P}}\|_\square \leq \|W\|_\square$ and $\|W_{\mathcal{P}}\|_{L^1} \leq \|W\|_{L^1}$. If W is a kernel, then $W_{\mathcal{P}}$ is also a kernel. A kernel that is also a step function, such as $W_{\mathcal{P}}$, is called a *step kernel*.

Suppose now that $(\mathcal{S},\mu,\prec)$ is an atomless ordered probability space, and let $D(t)$, $0\le t\le 1$, be an increasing family of downsets in $\mathcal{S}$ with $\mu(D(t))=t$ as in Lemma 5.2, with $D(0)=\emptyset$ and $D(1)=\mathcal{S}$.

For $n\ge 1$ and $i=1,\dots,n$, define

$$A_i = A_{ni} := D(i/n)\setminus D((i-1)/n). \tag{5.3}$$

Then $\mathcal{P}_n := \{A_{ni}\}_i$ is a partition of $\mathcal{S}$ into n sets of the same measure $1/n$. Furthermore, if $i<j$, then $A_{ni}\prec A_{nj}$, meaning that if $x\in A_{ni}$ and $y\in A_{nj}$, then $x\prec y$.

Given a kernel W on $\mathcal{S}$, let $w_{ij}^{(n)} := \overline{W}(A_{ni},A_{nj})$ and let W_n be the step kernel $W_{\mathcal{P}_n}$; thus $W_n = w_{ij}^{(n)}$ on $A_{ni}\times A_{nj}$. Define the step kernels $W_n^{\pm}$ by $W_n^+(x,y) := w_{i+1,j+1}^{(n)}$ and $W_n^-(x,y) := w_{i-1,j-1}^{(n)}$ on $A_{ni}\times A_{nj}$, where $w_{ij}^{(n)}=0$ if i or $j=0$ and $w_{ij}^{(n)}=1$ if i or $j=n+1$.

If W is monotone, then the matrix $(w_{ij}^{(n)})_{ij}$ is increasing along each row and column, and thus W_n is a monotone step kernel.

Lemma 5.3. *Let W be a monotone kernel on an atomless ordered probability space $(\mathcal{S},\prec)$. Then $W_n^-\le W\le W_n^+$, $W_n^-\le W_n\le W_n^+$ and*

$$\|W_n-W\|_{L^1(\mathcal{S}^2)} \le \|W_n^+-W_n^-\|_{L^1(\mathcal{S}^2)} \le 4/n.$$

Proof. If $(x,y)\in A_{ni}\times A_{nj}$ and $(x',y')\in A_{n,i+1}\times A_{n,j+1}$ (with $i,j\le n-1$), then $W(x,y)\le W(x',y')$, and averaging over (x',y') it follows that $W(x,y)\le w_{i+1,j+1}^{(n)} = W_n^+(x,y)$. This inequality evidently holds also if i or $j=n$. Hence $W\le W_n^+$. Similarly, $W\ge W_n^-$.

Averaging over each $A_{ni}\times A_{nj}$, it follows that $W_n^-\le W_n\le W_n^+$. (This also follows directly from the monotonicity of $w_{ij}^{(n)}$.) Consequently, $|W_n-W|\le W_n^+-W_n^-$, and thus

$$\begin{aligned}\|W_n-W\|_{L^1(\mathcal{S}^2)} &\le \iint_{\mathcal{S}^2}\bigl(W_n^+-W_n^-\bigr) = \iint_{\mathcal{S}^2} W_n^+ - \iint_{\mathcal{S}^2} W_n^- \\ &= n^{-2}\sum_{i,j=2}^{n+1} w_{ij}^{(n)} - n^{-2}\sum_{i,j=0}^{n-1} w_{ij}^{(n)} \le 2n^{-2}\sum_{i=n}^{n+1}\sum_{j=2}^{n+1} w_{ij}^{(n)} \\ &\le 4/n. \end{aligned}$$

$\square$

Trivially, for any kernel W we have $\|W\|_\square \le \|W\|_{L^1(\mathcal{S}^2)}$. In general there is no reverse inequality. However, if $\mathcal{P}$ is a partition of $\mathcal{S}$ into n sets and W is a $\mathcal{P}$-step function, then it is trivial to bound $\|W\|_{L^1(\mathcal{S}^2)}$ from above by a polynomial times $\|W\|_\square$. Indeed, one can write $\|W\|_{L^1(\mathcal{S}^2)}$ as a sum of n integrals of the form in (2.1), in each taking g to be 1 on one part of $\mathcal{P}$ and zero elsewhere, and choosing the sign of f on each part appropriately. In fact, the correct polynomial order is $\sqrt{n}$, as shown in [14].

Lemma 5.4. *Let $\mathcal{S}$ be a probability space and $\mathcal{P}$ a partition of $\mathcal{S}$ into n sets. If W is a $\mathcal{P}$-step function, then $\|W\|_{L^1(\mathcal{S}^2)} \le \sqrt{2n}\|W\|_\square$. Furthermore, for any $W \in L^1(\mathcal{S}^2)$ we have*

$$\|W_{\mathcal{P}}\|_{L^1(\mathcal{S}^2)} \le \sqrt{2n}\|W\|_\square. \tag{5.4}$$

Proof. It suffices to prove the first statement; the second follows immediately, since $W_{\mathcal{P}}$ is a $\mathcal{P}$-step function, and $\|W_{\mathcal{P}}\|_\square \le \|W\|_\square$.

The statement and proof are (essentially) present in Remark 9.8 of [14]. Nevertheless, let us write out the proof.

In 1930, Littlewood [18] proved that there is a constant $c \le \sqrt{3}$ such that for any n-by-n array of real numbers a_{ij} we have

$$\begin{aligned}\sum_{i=1}^n\Big(\sum_{j=1}^n |a_{ij}|^2\Big)^{1/2} \quad &\le \quad c\max_{\varepsilon_i,\varepsilon'_j=\pm1}\sum_{i=1}^n\sum_{j=1}^n \varepsilon_i\varepsilon'_j a_{ij}\\ &= \quad c\max_{\varepsilon_i=\pm1}\sum_{j=1}^n\Big|\sum_{i=1}^n\varepsilon_i a_{ij}\Big| = c\max_{\varepsilon_j=\pm1}\sum_{i=1}^n\Big|\sum_{j=1}^n\varepsilon_j a_{ij}\Big|.\end{aligned}$$

Later it was noticed (see [27], Ch. 5 and [1]) that this inequality of Littlewood's could be deduced from a special case of an inequality that had been proved some years earlier by Khintchine [15]. In 1976, Szarek [24] proved that the best constant in Littlewood's inequality (in fact, in the corresponding inequality of Khintchine) is $\sqrt{2}$. For some related results, see, e.g., [9], [10], [11], [16] and [17].

As noted in [14], using the Cauchy–Schwartz inequality and Littlewood's inequality, with the constant $c = \sqrt{2}$ proved by Szarek, it follows that

$$\sum_{i=1}^n\sum_{j=1}^n |a_{ij}| \le \sum_{i=1}^n n^{1/2}\Big(\sum_{j=1}^n |a_{ij}|^2\Big)^{1/2} \le \sqrt{2n}\max_{\varepsilon_i,\varepsilon'_j=\pm1}\sum_{i=1}^n\sum_{j=1}^n \varepsilon_i\varepsilon'_j a_{ij}. \tag{5.5}$$

Returning to the proof of Lemma 5.4, let the parts of $\mathcal{P}$ be $A_1,\dots,A_n$, and set $a_{ij} = \mu(A_i)\mu(A_j)W_{ij}$, where W_{ij} is the value of W on $A_i \times A_j$. Then $\|W\|_{L^1(\mathcal{S}^2)} = \sum_{ij}|a_{ij}|$. In the definition (2.1) of the cut norm, restricting our attention to functions $f, g : \mathcal{S} \to \{\pm1\}$ that are constant on each A_i, we find that

$$\|W\|_\square \ge \max_{\varepsilon_i,\varepsilon'_j=\pm1}\sum_{i=1}^n\sum_{j=1}^n \varepsilon_i\varepsilon'_j a_{ij}$$

(in fact, equality holds), so the result follows from (5.5). $\square$

As noted in [14], it is easy to check that the factor $\sqrt{2n}$ is best possible apart from the constant, for example by considering 0/1-valued kernels associated to random graphs. For arbitrary *monotone* kernels, the lemmas above allow us to bound the L^1-norm in terms of the cut norm.

Theorem 5.5. *If W_1 and W_2 are monotone kernels on an ordered probability space $(\mathcal{S}, \prec)$, then*

$$\|W_1 - W_2\|_{L^1(\mathcal{S}^2)} \le 10\|W_1 - W_2\|_\square^{2/3}. \tag{5.6}$$

Proof. Suppose first that $\mathcal{S}$ is atomless. Let $n \ge 1$ and consider the partition $\mathcal{P}_n = \{A_{ni}\}_i$ defined in (5.3) and the step kernels $W_{k,n} = (W_k)_{\mathcal{P}_n}$, $k = 1, 2$. Lemma 5.4 yields

$$\|W_{1,n} - W_{2,n}\|_{L^1(\mathcal{S}^2)} = \|(W_1 - W_2)_{\mathcal{P}_n}\|_{L^1(\mathcal{S}^2)} \le \sqrt{2n}\|W_1 - W_2\|_\square. \tag{5.7}$$

By Lemma 5.3, we have $\|W_k - W_{k,n}\|_{L^1(\mathcal{S}^2)} \le 4/n$, so by the triangle inequality

$$\|W_1 - W_2\|_{L^1(\mathcal{S}^2)} \le \|W_{1,n} - W_{2,n}\|_{L^1(\mathcal{S}^2)} + 8/n \le \sqrt{2n}\|W_1 - W_2\|_\square + 8/n.$$

The result for atomless $\mathcal{S}$ now follows by choosing $n := \lceil \|W_1 - W_2\|_\square^{-2/3} \rceil \le 2\|W_1 - W_2\|_\square^{-2/3}$. (In the case $\|W_1 - W_2\|_\square = 0$, we let $n \to \infty$.)

If $\mathcal{S}$ has atoms, we consider the atomless probability space $\widehat{\mathcal{S}} := \mathcal{S} \times [0, 1]$ with the lexicographic order. Let $\pi : \widehat{\mathcal{S}} \to \mathcal{S}$ be the projection onto the first coordinate and let $\widehat{W_k} := W_k^\pi$ be the extension of W_k to $\widehat{\mathcal{S}}$. The proof just given applies to $\widehat{\mathcal{S}}$, and thus

$$\|W_1 - W_2\|_{L^1(\mathcal{S}^2)} = \|\widehat{W}_1 - \widehat{W}_2\|_{L^1(\widehat{\mathcal{S}}^2)} \le 10\|\widehat{W}_1 - \widehat{W}_2\|_\square^{2/3} = 10\|W_1 - W_2\|_\square^{2/3}.$$

$\square$

Example 5.6. It is easy to see that (5.6) is tight apart from the constant. Indeed, let $\mathcal{S}$ be the discrete probability space with n equiprobable elements $\{0, 1, \dots, n-1\}$, and choose two 0/1-valued kernels on $\mathcal{S}$ with $\|W_1 - W_2\|_{L^1(\mathcal{S}^2)} = \Theta(1)$ and $\|W_1 - W_2\|_\square = \Theta(n^{-1/2})$. For example, we may take kernels corresponding to two independent instances of the random graph $G(n, 1/2)$. Let W be the function defined by $W(i, j) = i + j$. Then it is easy to see that $W_i' = (W_i + W)/(2n)$ is a monotone kernel for each i. Since $\|W_1' - W_2'\|_{L^1(\mathcal{S}^2)} = \|W_1 - W_2\|_{L^1(\mathcal{S}^2)}/(2n) = \Theta(n^{-1})$ and $\|W_1' - W_2'\|_\square = \|W_1 - W_2\|_\square/(2n) = \Theta(n^{-3/2})$, this gives *monotone* kernels W_1' and W_2' with $\|W_1' - W_2'\|_{L^1(\mathcal{S}^2)} = \Theta(\|W_1' - W_2'\|_\square^{2/3})$.

Our next aim is to prove the rather unsurprising fact that if we start from two monotone kernels, then 'rearranging' one or both does not bring them any closer in the L^1 distance. First we need a preparatory lemma; this can be viewed as a continuous, coupling version of the trivial observation that if we wish to minimize $\sum_{i=1}^n |a_i - b_i|$ (or, equivalently, $\sum (a_i - b_i)_+$) where the values in each sequence are given but we are allowed to permute them, then we should sort both sequences into ascending order.

Lemma 5.7. *If $h_1, h_2 : \mathcal{S} \to \mathbb{R}$ are increasing integrable functions on an ordered probability space $(\mathcal{S}, \mu, \prec)$, and $\varphi_1, \varphi_2 : \mathcal{S}' \to \mathcal{S}$ are measure-preserving* *maps from a probability space* $(\mathcal{S}', \mu')$ *to* $(\mathcal{S}, \mu)$*, then*

$$\int_{\mathcal{S}'} (h_1^{\varphi_1} - h_2^{\varphi_2})_+ \, \mathrm{d}\mu' \geq \int_{\mathcal{S}} (h_1 - h_2)_+ \, \mathrm{d}\mu \tag{5.8}$$

and $\|h_1^{\varphi_1} - h_2^{\varphi_2}\|_{L^1(\mathcal{S}')} \geq \|h_1 - h_2\|_{L^1(\mathcal{S})}$*.*

Proof. For any integrable function on any measure space we have $\|h\|_{L^1} = \int (h)_+ + \int (-h)_+$, so it suffices to prove the first statement.

For any function f and real number t, let $B_f(t) := \{x : f(x) \leq t\}$. Fubini's theorem yields

$$\begin{aligned} \int_{\mathcal{S}} (h_1 - h_2)_+ \, \mathrm{d}\mu &= \int_{\mathcal{S}} \int_{-\infty}^{\infty} \mathbf{1}\{h_1(x) > t \geq h_2(x)\} \, \mathrm{d}t \, \mathrm{d}\mu(x) \\ &= \int_{-\infty}^{\infty} \int_{\mathcal{S}} \mathbf{1}\{x \in B_{h_2}(t) \setminus B_{h_1}(t)\} \, \mathrm{d}\mu(x) \, \mathrm{d}t \\ &= \int_{-\infty}^{\infty} \mu\bigl(B_{h_2}(t) \setminus B_{h_1}(t)\bigr) \, \mathrm{d}t. \end{aligned}$$

Similarly,

$$\int_{\mathcal{S}'} (h_1^{\varphi_1} - h_2^{\varphi_2})_+ \, \mathrm{d}\mu' = \int_{-\infty}^{\infty} \mu'\bigl(B_{h_2^{\varphi_2}}(t) \setminus B_{h_1^{\varphi_1}}(t)\bigr) \, \mathrm{d}t.$$

Since the φ_i are measure preserving, we have $\mu'\bigl(B_{h_i^{\varphi_i}}(t)\bigr) = \mu'\bigl(\varphi_i^{-1}(B_{h_i}(t))\bigr) = \mu(B_{h_i}(t))$. Since h_1 and h_2 are increasing, $B_{h_1}(t)$ and $B_{h_2}(t)$ are downsets, so by Lemma 5.1 they are nested. The result follows by noting that $\mu(X \setminus Y) \geq (\mu(X) - \mu(Y))_+$, with equality if X and Y are nested. □

Lemma 5.8. *If* W_1 *and* W_2 *are monotone kernels on an ordered probability space* $(\mathcal{S}, \prec)$*, then* $\delta_1(W_1, W_2) = \|W_1 - W_2\|_{L^1(\mathcal{S}^2)}$*.*

Proof. Suppose that φ_1, φ_2 are measure-preserving maps $\mathcal{S}' \to \mathcal{S}$ for some probability space $(\mathcal{S}', \mu')$. Then, using Lemma 5.7 on each coordinate separately,

$$\begin{aligned} &\|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{L^1((\mathcal{S}')^2)} \\ &= \int_{\mathcal{S}'} \int_{\mathcal{S}'} \bigl|W_1(\varphi_1(x), \varphi_1(y)) - W_2(\varphi_2(x), \varphi_2(y))\bigr| \, \mathrm{d}\mu'(x) \, \mathrm{d}\mu'(y) \\ &\geq \int_{\mathcal{S}'} \int_{\mathcal{S}} \bigl|W_1(t, \varphi_1(y)) - W_2(t, \varphi_2(y))\bigr| \, \mathrm{d}\mu(t) \, \mathrm{d}\mu'(y) \\ &\geq \int_{\mathcal{S}} \int_{\mathcal{S}} \bigl|W_1(t, u) - W_2(t, u)\bigr| \, \mathrm{d}\mu(t) \, \mathrm{d}\mu(u) = \|W_1 - W_2\|_{L^1(\mathcal{S}^2)}, \end{aligned}$$

where for the last step we first apply Fubini's Theorem to change the order of integration. The result follows by the definition (2.4). □

With a little more work, we obtain a corresponding result for the cut norm and cut metric. Unfortunately, we need to consider a variant of the definition.

If W is an integrable function on $\mathcal{S}^2$, let

$$\|W\|_{\square,1} := \sup_{f,g:\mathcal{S}\to\{0,1\}} \left| \int_{\mathcal{S}^2} W(x,y) f(x) g(y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \right|, \tag{5.9}$$

where the supremum is over all pairs of measurable 0/1-valued functions on $\mathcal{S}$. (We could equally well consider functions taking values in $[0,1]$; the value of the supremum does not change.) Expressing each of the functions f, g in (2.2) as the difference of two 0/1-valued functions, we see that

$$\|W\|_{\square,1} \le \|W\|_{\square} \le 4\|W\|_{\square,1}, \tag{5.10}$$

so for all questions concerning convergence, the norms are equivalent.

In analogy with (2.3), given $W_i \in L^1(\mathcal{S}_i^2)$, $i = 1,2$, let

$$\delta_{\square,1}(W_1, W_2) := \inf_{\varphi_1,\varphi_2} \|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{\square,1}, \tag{5.11}$$

where, as in (2.3), the infimum is taken over all couplings (φ_1, φ_2) of $\mathcal{S}_1$ and $\mathcal{S}_2$.

Lemma 5.9. *If W_1 and W_2 are monotone kernels on an ordered probability space $(\mathcal{S}, \prec)$, then $\delta_{\square,1}(W_1,W_2) = \|W_1 - W_2\|_{\square,1}$.*

Proof. Suppose that φ_1, φ_2 are measure-preserving maps $\mathcal{S}' \to \mathcal{S}$ for some probability space $\mathcal{S}'$. It suffices to show that $\|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{\square,1} \ge \|W_1 - W_2\|_{\square,1}$.

Given a probability space $(\mathcal{S}, \mu)$, an integrable function W on $\mathcal{S}^2$, and two functions $f, g : \mathcal{S} \to \{0,1\}$, set

$$I_{f,g}(W) := \int_{\mathcal{S}^2} W(x,y) f(x) g(y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y),$$

so $\|W\|_{\square,1} = \sup_{f,g} |I_{f,g}(W)|$. Swapping W_1 and W_2 if necessary, we may assume that $\|W_1 - W_2\|_{\square,1} = \sup_{f,g} I_{f,g}(W_1 - W_2)$. Hence, fixing (arbitrary) functions $f, g : \mathcal{S} \to \{0,1\}$, it suffices to prove that

$$\sup_{f',g'} I_{f',g'}(W_1^{\varphi_1} - W_2^{\varphi_2}) \ge I_{f,g}(W_1 - W_2), \tag{5.12}$$

since $\|W_1^{\varphi_1} - W_2^{\varphi_2}\|_{\square}$ is at least the left-hand side.

The first statement (5.8) of Lemma 5.7 says exactly that if h_1 and h_2 are increasing, integrable functions on $(\mathcal{S}, \mu, \prec)$ and φ_1, $\varphi_2 : (\mathcal{S}', \mu') \to (\mathcal{S}, \mu)$ are measure-preserving, then

$$\begin{aligned} \max_{f':\mathcal{S}'\to\{0,1\}} &\int_{\mathcal{S}'} \big(h_1(\varphi_1(x)) - h_2(\varphi_2(x))\big) f'(x) \,\mathrm{d}\mu'(x) \\ &\ge \max_{f:\mathcal{S}\to\{0,1\}} \int_{\mathcal{S}} \big(h_1(t) - h_2(t)\big) f(t) \,\mathrm{d}\mu(t), \end{aligned} \tag{5.13}$$

where the maximization is over all $\{0,1\}$-valued functions on the relevant space; the corresponding supremum is clearly attained. We shall use this inequality twice; in particular, we shall twice use the observation that a specific f on the right is 'beaten' by some f' on the left.

Let $h_i(t) = \int_{\mathcal{S}} W_i(t,u)g(u)\,\mathrm{d}\mu(u)$. Then (since $g(u)$ is non-negative), h_i is monotone. Applying (the observation following) (5.13) to these functions and our function f, we find that there is some $f' : \mathcal{S}' \to \{0,1\}$ such that

$$\int_{\mathcal{S}'} \left(\int_{\mathcal{S}} (W_1(\varphi_1(x),u) - W_2(\varphi_2(x),u))\, g(u)\,\mathrm{d}\mu(u) \right) f'(x)\,\mathrm{d}\mu'(x)$$
$$\geq \int_{\mathcal{S}} \left(\int_{\mathcal{S}} (W_1(t,u) - W_2(t,u))\, g(u)\,\mathrm{d}\mu(u) \right) f(t)\,\mathrm{d}\mu(t) = I_{f,g}(W_1 - W_2).$$

Using Fubini's Theorem, we may rewrite the left-hand side as

$$I := \int_{\mathcal{S}} \left(\int_{\mathcal{S}'} \big(W_1(\varphi_1(x),u) - W_2(\varphi_2(x),u)\big) f'(x)\,\mathrm{d}\mu'(x) \right) g(u)\,\mathrm{d}\mu(u).$$

Let $h_i'(u) = \int_{\mathcal{S}'} W_i(\varphi_i(x),u) f'(x)\,\mathrm{d}\mu'(x)$. Then the h_i' are again monotone, so applying (5.13) to these functions and g gives a $g' : \mathcal{S}' \to \{0,1\}$ such that

$$\int_{\mathcal{S}'} \left(\int_{\mathcal{S}'} \big(W_1(\varphi_1(x),\varphi_1(y)) - W_2(\varphi_2(x),\varphi_2(y))\big) f'(x)\,\mathrm{d}\mu'(x) \right) g'(y)\,\mathrm{d}\mu'(y) \geq I.$$

But now the left-hand side is simply $I_{f',g'}(W_1^{\varphi_1} - W_2^{\varphi_2})$, so we have $I_{f',g'}(W_1^{\varphi_1} - W_2^{\varphi_2}) \geq I \geq I_{f,g}(W_1 - W_2)$, establishing (5.12). $\square$

In the light of (5.10), Lemma 5.9 has the following immediate corollary.

Lemma 5.10. *If W_1 and W_2 are monotone kernels on an ordered probability space $(\mathcal{S}, \prec)$, then $\delta_\square(W_1,W_2) \geq \|W_1 - W_2\|_\square/4$.* $\square$

It seems plausible that $\delta_\square(W_1,W_2) = \|W_1 - W_2\|_\square$ for monotone kernels, but we do not have a proof (or indeed a strong feeling that this is actually true).

We are now ready to bound the L^1 distance with 'rearrangement' in terms of the cut metric, when the kernels in question are monotone.

Lemma 5.11. *If W_1 and W_2 are monotone kernels on an ordered probability space $(\mathcal{S}, \mu, \prec)$, then*

$$\delta_1(W_1,W_2) \leq 26\, \delta_\square(W_1,W_2)^{2/3}. \tag{5.14}$$

Proof. Combining Lemma 5.8, Theorem 5.5 and Lemma 5.10, we have

$$\delta_1(W_1,W_2) = \|W_1 - W_2\|_{L^1(\mathcal{S}^2)} \leq 10\|W_1 - W_2\|_\square^{2/3} \leq 10(4\delta_\square(W_1,W_2))^{2/3}, \tag{5.15}$$

giving the result. $\square$

Remark 5.12. Using Theorem 4.3 (which is proved below), Lemma 5.11 immediately extends to monotone kernels defined on possibly different ordered probability spaces.

Remark 5.13. The exponent 2/3 in (5.14) is best possible, as shown by the kernels W_1', W_2' in Example 5.6. Indeed, for these kernels, the first inequality in (5.15) is tight up to the constant. The second inequality is always tight up to the constant $4^{2/3}$ since, by definition, $\delta_\square(W_1,W_2) \leq \|W_1 - W_2\|_\square$.

We are now ready to prove the first few results in Section 4.

Proof of Theorem 4.1. The equivalence of the different metrics in (iii) follows from Theorem 5.5, Lemmas 5.8 and 5.10 (see (5.15)) and the inequality $\delta_\square(W_1, W_2) \le \delta_1(W_1, W_2)$.

As a special case, for two kernels $W_1, W_2 \in \mathcal{W}_\uparrow(\mathcal{S})$,

$$\delta_\square(W_1, W_2) = 0 \iff \|W_1 - W_2\|_{L^1(\mathcal{S}^2)} = 0 \iff W_1 = W_2 \text{ a.e.},$$

which establishes (ii).

For (i), we show that $\mathcal{W}_\uparrow(\mathcal{S})$ is closed and totally bounded as a subset of $L^1(\mathcal{S}^2)$. First, if $W_\nu \in \mathcal{W}_\uparrow(\mathcal{S})$ and $W_\nu \to W$ in $L^1(\mathcal{S}^2)$ as $\nu \to \infty$, then there is a subsequence that converges a.e. to W, and replacing W by the $\limsup$ of that subsequence, we see that $W \in \mathcal{W}_\uparrow(\mathcal{S})$. Hence, $\mathcal{W}_\uparrow(\mathcal{S})$ is closed.

Next, first assume that $\mathcal{S}$ is atomless. By Lemma 5.3, for every n there is a partition $\mathcal{P}_n$ such that for every kernel $W \in \mathcal{W}_\uparrow(\mathcal{S})$, there is a $\mathcal{P}_n$-step kernel W_n with $\|W - W_n\|_{L^1(\mathcal{S}^2)} \le 4/n$. If F_n is the finite set of $\mathcal{P}_n$-step kernels taking values in $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$, then there always exists a $W_n' \in F_n$ with $\|W_n - W_n'\|_{L^1(\mathcal{S}^2)} \le 1/n$, and thus $\|W - W_n'\|_{L^1(\mathcal{S}^2)} \le 5/n$. Since n is arbitrary, this shows that $\mathcal{W}_\uparrow(\mathcal{S})$ is totally bounded.

If $\mathcal{S}$ has atoms, we consider as above $\widehat{\mathcal{S}} = \mathcal{S} \times [0,1]$ and $\pi : \widehat{\mathcal{S}} \to \mathcal{S}$; then $W \mapsto W^\pi$ is an isometric embedding of $L^1(\mathcal{S}^2)$ into $L^1(\widehat{\mathcal{S}}^2)$. This embeds $\mathcal{W}_\uparrow(\mathcal{S})$ into $\mathcal{W}_\uparrow(\widehat{\mathcal{S}})$, and since the latter is totally bounded, $\mathcal{W}_\uparrow(\mathcal{S})$ is too. □

Proof of Theorem 4.3. If $\mathcal{S}$ has atoms, we replace it, as above, by $\widehat{\mathcal{S}} = \mathcal{S} \times [0,1]$; thus we may assume that $\mathcal{S}$ is atomless. By Lemma 5.3, there is a sequence of step kernels W_n that converges to W in $L^1(\mathcal{S}^2)$. Each W_n is obviously equivalent to the monotone step kernel W_n' on $[0,1]$ defined by $W_n' = w_{ij}^{(n)}$ on $I_i \times I_j$, where $I_i := ((i-1)/n, i/n]$. We have $\|W_n' - W_m'\|_{L^1([0,1]^2)} = \|W_n - W_m\|_{L^1(\mathcal{S}^2)}$, and thus (W_n') is a Cauchy sequence in $L^1([0,1]^2)$. Hence there is some W' such that $W_n' \to W'$ in $L^1([0,1]^2)$, and Theorem 4.1(i) implies that $W' \in \mathcal{W}_\uparrow([0,1])$. For every n,

$$\begin{aligned}\delta_\square(W, W') &\le \delta_\square(W, W_n) + \delta_\square(W_n, W_n') + \delta_\square(W_n', W') \\ &\le \frac{4}{n} + 0 + \|W_n' - W'\|_{L^1([0,1]^2)}.\end{aligned}$$

Since $W_n' \to W'$ in $L^1([0,1]^2)$, it follows that $\delta_\square(W, W') = 0$, so W' and W are equivalent. □

6. Proofs of Theorems 4.5–4.11

In this section we prove the remaining results in Section 4, namely, Theorem 4.5, Lemmas 4.6 and 4.7, and Theorem 4.11.

We start with a technical lemma, which is fairly obvious but nevertheless deserves to be stated precisely.

Lemma 6.1. *Suppose that* $(\mathcal{S}_1, \mu_1, \prec_1)$ *and* $(\mathcal{S}_2, \mu_2, \prec_2)$ *are ordered probability spaces, and that* $\mathcal{S}_1 \times \mathcal{S}_2$ *is equipped with a probability measure* μ *such that the projection* π_1 *onto* $\mathcal{S}_1$ *is measure-preserving. Let* $\prec_1^*$ *be the lexicographic order on* $\mathcal{S}_1 \times \mathcal{S}_2$*. If* W *is a kernel on* $\mathcal{S}_1$*, then for* $j = 1, 2$*,*

$$\Omega_j(W, \prec_1) = \Omega_j(W^{\pi_1}, \prec_1^*).$$

In most applications, we take $\mu = \mu_1 \times \mu_2$.

Proof. Writing $x \in \mathcal{S} := \mathcal{S}_1 \times \mathcal{S}_2$ as $x = (x_1, x_2)$, by (4.8), $\Omega_1(W^{\pi_1}, \prec_1^*)$ is equal to

$$\sup_{f,g} \iiint_{x \prec_1^* y} \bigl(W(x_1, z_1) - W(y_1, z_1)\bigr) f(x, y) g(z) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \,\mathrm{d}\mu(z), \tag{6.1}$$

where the supremum is over all $f : \mathcal{S}^2 \to [0, 1]$ and $g : \mathcal{S} \to [0, 1]$.

Let $\mathcal{F}_1$ be the σ-field on $\mathcal{S}$ obtained by pulling back that on $\mathcal{S}_1$. Thus the $\mathcal{F}_1$-measurable functions are all functions of the form $h(x_1, x_2) = h_1(x_1)$ for measurable h_1 on $\mathcal{S}_1$. In (6.1) we may replace f and g by their conditional expectations given $\mathcal{F}_1 \times \mathcal{F}_1$ and $\mathcal{F}_1$, respectively. Recalling that $\prec_1^*$ is lexicographic, and noting that the integrand vanishes when $x_1 = y_1$, (6.1) reduces to

$$\sup_{f_1, g_1} \iiint_{x_1 \prec_1 y_1} \bigl(W(x_1, z_1) - W(y_1, z_1)\bigr) f_1(x_1, y_1) g_1(z_1) \,\mathrm{d}\mu_1(x_1) \,\mathrm{d}\mu_1(y_1) \,\mathrm{d}\mu_1(z_1),$$

with the supremum over $f_1 : \mathcal{S}_1^2 \to [0, 1]$ and $g_1 : \mathcal{S}_1 \to [0, 1]$. By (4.8), this is simply $\Omega_1(W, \prec_1)$.

(In the special case when $\mu = \mu_1 \times \mu_2$, the argument above is equivalent to simply integrating over x_2, y_2, z_2 in (6.1).)

For Ω_2, the argument is similar, using (4.9) instead of (4.8). □

Proof of Theorem 4.5. Here it makes no difference whether we consider Ω_1 or Ω_2, so we simply write Ω.

If $W = W'$ a.e. where W' is monotone, then we have $\Omega(W, \prec, A) = \Omega(W', \prec, A) = 0$ for all $A \subseteq \mathcal{S}$, and hence $\Omega(W, \prec) = 0$.

Conversely, suppose that $\Omega(W, \prec) = 0$. Let $A, B, C, D \subseteq \mathcal{S}$ have positive measures, and suppose that $A \prec B$. Since $\Omega(W, \prec) = 0$, we have $\Omega(W, \prec, C) = 0$ and thus by (4.7) $W_C(x) \le W_C(y)$ for a.e. (x, y) with $x \prec y$, and in particular for a.e. $(x, y) \in A \times B$. Averaging over all such (x, y) yields $\overline{W}(A, C) \le \overline{W}(B, C)$. Similarly, by symmetry, if $C \prec D$, then $\overline{W}(B, C) \le \overline{W}(B, D)$. Consequently, letting $A \preceq B$ mean $A \prec B$ or $A = B$,

$$\overline{W}(A, C) \le \overline{W}(B, D) \qquad \text{if } A \preceq B,\ C \preceq D. \tag{6.2}$$

Assuming still that $A, B, C, D \subseteq \mathcal{S}$ have positive measures, suppose that $A \prec B$ and $C \prec D$. If $A_1 \subseteq A$ and $C_1 \subseteq C$, then (6.2), applied to A_1, B, C_1, D, yields

$$\iint_{A_1 \times C_1} W \le (\mu \times \mu)(A_1 \times C_1) \overline{W}(B, D).$$

Since every measurable subset of $A \times C$ can be approximated (in measure) by a finite disjoint union of rectangle sets $A_i \times C_i$, and W is bounded, it follows that

$$\iint_E W \le (\mu \times \mu)(E)\overline{W}(B, D) \quad \text{for every } E \subseteq A \times C.$$

Taking $E := \{(x, y) \in A \times C : W(x, y) > \overline{W}(B, D)\}$, we obtain $\mu \times \mu(E) = 0$, and thus

$$W(x, y) \le \overline{W}(B, D) \quad \text{a.e. on } A \times C \text{ when } A \prec B \text{ and } C \prec D. \tag{6.3}$$

Similarly, by reversing the inequalities,

$$W(x, y) \ge \overline{W}(B, D) \quad \text{a.e. on } A \times C \text{ when } A \succ B \text{ and } C \succ D. \tag{6.4}$$

Suppose now that $\mathcal{S}$ is atomless, and consider, for a given n, the partition $\mathcal{P} = (A_i)_1^n$ defined in (5.3). By (6.2), $W_n := W_{\mathcal{P}}$ is a monotone kernel. By (6.3) and (6.4), $W_n^-(x, y) \le W(x, y) \le W_n^+(x, y)$ a.e. on each $A_i \times A_j$, and thus a.e. on $\mathcal{S}^2$. Further, by averaging this or directly from (6.2), also $W_n^- \le W_n \le W_n^+$. It follows as in the proof of Lemma 5.3 that

$$\|W_n - W\|_{L^1(\mathcal{S}^2)} \le 4/n. \tag{6.5}$$

Now consider the sequence W_{2^k}, $k \ge 1$. By (6.5) and the Borel–Cantelli lemma, or by the martingale convergence theorem, $W_{2^k} \to W$ a.e. as $k \to \infty$. Hence, if we define $W' := \limsup_{k\to\infty} W_{2^k}$, then $W = W'$ a.e. and W' is a monotone kernel. This completes the proof when $\mathcal{S}$ is atomless.

If $\mathcal{S}$ has atoms, we may either modify the argument above, or use our standard trick of replacing $\mathcal{S}$ by $\mathcal{S} \times [0, 1]$, using Lemma 6.1; this gives a monotone kernel W' on $\mathcal{S} \times [0, 1]$ with $W'((x, a), (y, b)) = W(x, y)$ for a.e. $(x, a, y, b) \in (\mathcal{S} \times [0, 1])^2$, and thus W is a.e. equal to the monotone kernel W'' on $\mathcal{S}$ defined by $W''(x, y) = \int_0^1 \int_0^1 W'((x, a), (y, b))\,\mathrm{d}a\,\mathrm{d}b$. □

Proof of Lemma 4.6. Let $I_i := ((i - 1)/n, i/n]$ and for $A \subseteq [0, 1]$, set $A_i := A \cap I_i$. For $j = 1, 2$, by (4.2) and (4.3), $\Omega_j(W_G, <, A)$ depends only on the numbers $a_i := \mu(A_i) \in [0, 1/n]$; moreover, since the function $u \mapsto u_+$ is convex, $\Omega_j(W_G, <, A)$ is a convex function of $(a_1, \dots, a_n)$; hence it attains its maximum when each a_i is either 0 or $1/n$. In other words, it suffices to consider $A = \bigcup_{i\in B} I_i$ for some $B \subseteq V$. In this case, it is easily seen that $\Omega_j(W_G, <, A) = \Omega_j(G, \prec, B)$, noting that $\int_A W_G(x, z)\,\mathrm{d}z = \int_A W_G(y, z)\,\mathrm{d}z$ if $x, y \in I_i$ for some i. The result follows by taking the maximum over $B \subseteq V$. □

Lemma 6.2. *Let $(\mathcal{S}, \prec)$ be an ordered probability space, and let $j \in \{1, 2\}$.*

(i) *If $W_1, W_2 \in L^1(\mathcal{S}^2)$, then*

$$\Omega_j(W_1 + W_2, \prec, A) \le \Omega_j(W_1, \prec, A) + \Omega_j(W_2, \prec, A),$$
$$\Omega_j(W_1 + W_2, \prec) \le \Omega_j(W_1, \prec) + \Omega_j(W_2, \prec).$$

(ii) *If $W \in L^1(\mathcal{S}^2)$, then $\Omega_j(W, \prec) \le j\|W\|_\square$.*

(iii) *If $W_1, W_2 \in L^1(\mathcal{S}^2)$, then $\bigl|\Omega_j(W_1, \prec) - \Omega_j(W_2, \prec)\bigr| \le j\|W_1 - W_2\|_\square$.*

Proof. (i): An immediate consequence of the inequality $(a+b)_+ \le a_+ + b_+$ for real a and b, and the definitions (4.2)–(4.4).

(ii): By (4.7) and Fubini's theorem,

$$\begin{aligned}
\Omega_1(W, \prec, A) &\le \iint_{x \prec y} \Big(\big|W_A(x)\big| + \big|W_A(y)\big| \Big) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \\
&= \int_{\mathcal{S}} \mu\{y : y \succ x\} \big|W_A(x)\big| \, \mathrm{d}\mu(x) + \int_{\mathcal{S}} \mu\{x : x \prec y\} \big|W_A(y)\big| \, \mathrm{d}\mu(y) \\
&= \int_{\mathcal{S}} \mu\{z : z \neq x\} \big|W_A(x)\big| \, \mathrm{d}\mu(x) \le \int_{\mathcal{S}} \big|W_A(x)\big| \, \mathrm{d}\mu(x) \\
&= \iint_{\mathcal{S}^2} W(x,y) f(x) g(y) \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) \le \|W\|_\square,
\end{aligned}$$

where $f(x) := \operatorname{sign}(W_A(x))$ and $g(y) := \mathbf{1}_A(y)$; the final inequality follows from the definition (2.1) of the cut norm. Now apply (4.3), if $j = 2$, and take the supremum over A.

(iii): A simple consequence of (i), applied to the sums $W_1 + (W_2 - W_1)$ and $W_2 + (W_1 - W_2)$, and (ii). $\square$

The function $W_{\mathcal{S}} = W_{\mathcal{S}}(x) := \int_{\mathcal{S}} W(x,y) \, \mathrm{d}\mu(y)$ is known as the *marginal* of W. (There is also a second marginal, obtained by integrating over the first variable. Here we consider only symmetric functions, so the two marginals coincide.) It is well known that the marginal of a kernel is the natural analogue of the degree sequence of a graph, see e.g. [7]. We have the following analogue of Lemma 3.4.

Lemma 6.3. *Let $<$ be a (measurable) order on $\mathcal{S}$ and assume that $x < y \implies W_{\mathcal{S}}(x) \le W_{\mathcal{S}}(y)$. Then $\Omega_2(W, <) = \Omega_2(W)$.*

Proof. Follow the proof of Lemma 3.4, replacing sums by integrals and degrees by the values of $W_{\mathcal{S}}$. $\square$

Remark 6.4. For Ω_1, it follows by (4.6) that $\Omega_1(W, <) \le 2\Omega_1(W)$. The factor 2 here is best possible, just as in Corollary 3.5. This can be seen by taking $W = W_G$ where G is the complete bipartite graph $K_{m,m}$ considered in Example 3.7.

Corollary 6.5. *Let $\mathcal{S}$ be a probability space and W a kernel on $\mathcal{S}$. Then $\Omega_2(W) = 0$ if and only if there exists an order $\prec$ on $\mathcal{S}$ such that $\Omega_2(W, \prec) = 0$.*

Proof. The 'if' direction is clear. Thus, assume $\Omega_2(W) = 0$. Then there exists a measurable order $\prec_0$ on $\mathcal{S}$. Define an order $\prec$ on $\mathcal{S}$ by

$$x \prec y \quad \text{if} \quad W_{\mathcal{S}}(x) < W_{\mathcal{S}}(y) \text{ or } (W_{\mathcal{S}}(x) = W_{\mathcal{S}}(y) \text{ and } x \prec_0 y). \tag{6.6}$$

This is a measurable order to which Lemma 6.3 applies, so $\Omega_2(W, \prec) = \Omega_2(W) = 0$. $\square$

Of course, the same result for Ω_1 follows by (4.6).

Proof of Lemma 4.7. Recall that $W_G = W_{G,\prec}$ depends on a labelling of the vertices of G, via the associated order $\prec$ on $V(G)$. However, $\Omega_2(W_{G,\prec})$ is independent of the order $\prec$.

For any order $\prec$ on $V = V(G)$, Lemma 4.6 shows that, using $\prec$ to define W_G, and writing $<$ for the standard order on $[0,1]$, we have $\Omega_2(W_G) \le \Omega_2(W_G, <) = \Omega_2(G, \prec)$. Thus $\Omega_2(W_G) \le \Omega_2(G)$.

Conversely, let $\prec$ be an order on V such that $v \prec w \implies d(v) \le d(w)$, and use this order to define W_G. Then W_G satisfies the assumption of Lemma 6.3 with the standard order $<$ on $[0,1]$, and thus $\Omega_2(W_G, <) = \Omega_2(W_G)$. Hence, by Lemma 4.6 again,

$$\Omega_2(G) \le \Omega_2(G, \prec) = \Omega_2(W_G, <) = \Omega_2(W_G). \qquad \square$$

Our next lemma shows that Ω_2 is continuous with respect to the cut metric.

Lemma 6.6. *If W_1 and W_2 are kernels on probability spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, and there exists a measurable order on $\mathcal{S}_1$, then $\Omega_2(W_1) \le \Omega_2(W_2) + 2\delta_\square(W_1, W_2)$.*

Proof. Recall that the set of step functions is dense in $L^1(\mathcal{S}_1^2)$. Hence, for any $\varepsilon > 0$, there exists a step kernel W_1' on $\mathcal{S}_1$ with $\|W_1 - W_1'\|_\square \le \|W_1 - W_1'\|_{L^1(\mathcal{S}_1^2)} < \varepsilon$. By Lemma 6.2(iii), replacing W_1 by W_1' changes $\Omega_2(W_1)$ by less than 2ε, and the same holds for $\delta_\square(W_1, W_2)$. Hence, it suffices to prove the result when W_1 is a step kernel.

Consequently, assume that W_1 is a $\mathcal{P}$-step kernel, for a finite partition $\mathcal{P} = (A_i)_i$ of $\mathcal{S}_1$. Then its marginal $W_{1,\mathcal{S}}$ is constant on each A_i, and we may assume that $A_1, A_2, \dots$ are labelled such that $W_{1,\mathcal{S}}(x) \le W_{1,\mathcal{S}}(y)$ if $x \in A_i$, $y \in A_j$ with $i < j$. Let $\prec_0$ be a measurable order on $\mathcal{S}_1$, and define $\prec_1$ by

$$x \prec_1 y' \quad \text{if} \quad x \in A_i \text{ and } y \in A_j \text{ with } (i < j \text{ or } (i = j \text{ and } x \prec_0 y)).$$

Let $\prec_2$ be any measurable order on $\mathcal{S}_2$. Consider a coupling (π_1, π_2) defined on $(\mathcal{S}_1 \times \mathcal{S}_2, \mu)$ for some μ. Let $\prec_1^*$ be the lexicographic order on $\mathcal{S}_1 \times \mathcal{S}_2$, and let $\prec_2^*$ be the lexicographic order with the factors in opposite order. By Lemma 6.1,

$$\Omega_2(W_k, \prec_k) = \Omega_2(W_k^{\pi_k}, \prec_k^*), \qquad k = 1, 2. \tag{6.7}$$

Moreover, Lemma 6.3 applies to $\prec_1^*$ and $W_1^{\pi_1}$ and shows that

$$\Omega_2(W_1^{\pi_1}, \prec_1^*) = \Omega_2(W_1^{\pi_1}) \le \Omega_2(W_1^{\pi_1}, \prec_2^*), \tag{6.8}$$

and by Lemma 6.2(iii),

$$\Omega_2(W_1^{\pi_1}, \prec_2^*) \le \Omega_2(W_2^{\pi_2}, \prec_2^*) + 2\|W_1^{\pi_1} - W_2^{\pi_2}\|_\square. \tag{6.9}$$

Combining (6.7)–(6.9), we find

$$\Omega_2(W_1, \prec_1) \le \Omega_2(W_2, \prec_2) + 2\|W_1^{\pi_1} - W_2^{\pi_2}\|_\square,$$

and the result follows by taking the infimum over all couplings such (π_1, π_2), i.e., over all probability measures μ with the right marginals, and then over all orders $\prec_2$. $\square$

Corollary 6.7. *If W_1 and W_2 are equivalent kernels on probability spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ that have measurable orders, then $\Omega_2(W_1) = \Omega_2(W_2)$, and $\frac{1}{2}\Omega_1(W_2) \le \Omega_1(W_1) \le 2\Omega_1(W_2)$.*

Proof. We have $\delta_\square(W_1, W_2) = 0$; the first statement follows by Lemma 6.6. To deduce the second, use (4.6). ∎

Remark 6.8. The equivalent of Lemma 6.6 for Ω_1 does not hold, and the inequalities $\frac{1}{2}\Omega_1(W_2) \le \Omega_1(W_1) \le 2\Omega_1(W_2)$ in Corollary 6.7 are best possible. In fact, if $W_m := W^V_{K_{m,m}}$ is the kernel defined in Remark 4.9 for the bipartite graph $K_{m,m}$, then W_m is equivalent to $W_{K_{m,m}}$ (defined on $[0,1]$), but $W_{K_{m,m}}$ is the same for all m. Hence, all W_m are equivalent. Nevertheless, Remark 4.9 and (3.20) show that $\Omega_1(W_m) = \Omega_1(K_{m,m}) = (1+m^{-2})/16$ if m is odd, while $\Omega_1(W_m) = \Omega_1(K_{m,m}) = 1/16$ if m is even. In particular, $\Omega_1(W_1) = 1/8 = 2\Omega_1(W_2)$.

On the other hand, for kernels W_1, W_2 on the standard space $\mathcal{S} = [0,1]$ (and thus for kernels on any atomless Borel spaces), it follows from (2.5) and Lemma 6.2(iii) that $\left|\Omega_1(W_1) - \Omega_1(W_2)\right| \le \delta_\square(W_1, W_2)$, since clearly $\Omega_1(W_2^\varphi) = \Omega_1(W_2)$ for a measure-preserving bijection φ. In particular, $\Omega_1(W_1) = \Omega_1(W_2)$ for any two equivalent kernels on $[0,1]$. Hence the unruly behaviour of Ω_1 is caused by the atoms.

Proof of Theorem 4.11. (i) $\implies$ (ii). We use Ω_2. If $\Omega_2(W) = 0$, then by Corollary 6.5 there exists an order $\prec$ on $\mathcal{S}$ such that $\Omega_2(W, \prec) = 0$, and Theorem 4.5 shows that W is a.e. equal to a monotone kernel on $(\mathcal{S}, \prec)$.

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

(iii) $\implies$ (i). If W is equivalent to a monotone kernel W' on some probability space $\mathcal{S}'$, then $\delta_\square(W, W') = 0$ and $\Omega_2(W') = 0$, and thus $\Omega_2(W) = 0$ by Lemma 6.6.

(iii) $\iff$ (iv) $\iff$ (v). By Theorem 4.3. ∎

7. Proof of Theorems 1.5–1.6

After the preparation above, the proofs are simple.

Proof of Theorem 1.6. Let W be a kernel on $[0,1]$ representing Γ, i.e., $\Gamma = \Gamma_W$ and $G_\nu \to W$. Since $G_\nu \to W$, we have $\delta_\square(W_{G_\nu}, W) \to 0$.

Suppose first that $\Gamma \in \mathcal{U}_\uparrow$; we then may choose $W \in \mathcal{W}_\uparrow$, and thus $\Omega_2(W, <) = 0$ so $\Omega_2(W) = 0$. Then, by Lemmas 4.7 and 6.6,

$$\Omega_2(G_\nu) = \Omega_2(W_{G_\nu}) \le \Omega_2(W) + 2\delta_\square(W_{G_\nu}, W) = 2\delta_\square(W_{G_\nu}, W) \to 0.$$

Hence $\Omega_2(G_\nu) \to 0$, and by Lemma 3.1, $\Omega_0(G_\nu) \to 0$ as well.

Conversely, suppose that $\Omega_0(G_\nu) \to 0$, and thus by Lemma 3.1 $\Omega_2(G_\nu) \to 0$. Then, by Lemmas 6.6 and 4.7 again,

$$\Omega_2(W) \le \Omega_2(W_{G_\nu}) + 2\delta_\square(W_{G_\nu}, W) = \Omega_2(G_\nu) + 2\delta_\square(W_{G_\nu}, W) \to 0,$$

and thus $\Omega_2(W) = 0$. Hence, $\Gamma = \Gamma_W \in \mathcal{U}_\uparrow$ by Theorem 4.11. ∎

Proof of Theorem 1.5. If $\Omega_0(G_\nu) \to 0$, then the same holds for every subsequence. Hence Theorem 1.6 shows that every convergent subsequence has a limit that is in $\mathcal{U}_\uparrow$, which by definition says that (G_ν) is quasimonotone.

Conversely, suppose that (G_ν) is quasimonotone but $\Omega_0(G_\nu) \not\to 0$. We can then find $\varepsilon > 0$ and a subsequence along which $\Omega_0(G_\nu) > \varepsilon$. By restricting to a suitable subsubsequence, we may further assume that (G_ν) converges to some limit Γ. By the assumption that (G_ν) is quasimonotone, $\Gamma \in \mathcal{U}_\uparrow$ and thus by Theorem 1.6, $\Omega_0(G_\nu) \to 0$ along the subsubsequence, a contradiction. $\square$

8. Quasithreshold graphs

In the definition (1.5) of $\Omega_0(G, \prec)$, we take the maximum over A of the sum in (1.3). If instead we take the maximum inside the sum, then we obtain the functional

$$\Omega_0^*(G, \prec) := \frac{1}{n^3} \sum_{v \prec w} \bigl|N(v) \setminus (N(w) \cup \{w\})\bigr|, \tag{8.1}$$

since $\max_A\bigl(e(v, A \setminus \{w\}) - e(w, A \setminus \{v\})\bigr)_+$ is obtained by taking (for example) $A = N(v) \setminus N(w)$. From Ω_1, we similarly obtain the slightly simpler functional

$$\Omega_1^*(G, \prec) := \frac{1}{n^3} \sum_{v \prec w} \bigl|N(v) \setminus N(w)\bigr| = \Omega_0^*(G, \prec) + O(1/n). \tag{8.2}$$

For a kernel W on an ordered probability space $(\mathcal{S}, \mu, \prec)$, taking the supremum over A inside the double integral in (4.2), we define

$$\Omega^*(W, \prec) := \iiint_{x \prec y} \bigl(W(x, z) - W(y, z)\bigr)_+ \,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z). \tag{8.3}$$

(Cf. (4.8).) For any graph G with an ordering $\prec$ of the vertices, corresponding to Lemma 4.6 we have

$$\Omega^*(W_G, <) = \Omega_1^*(G, \prec). \tag{8.4}$$

Obviously, $\Omega_0^*(G, \prec) \geq \Omega_0(G, \prec)$, and similarly for Ω_1^* and Ω^*.

Let

$$\Omega_j^*(G) := \min_{\prec} \Omega_j^*(G, \prec) \quad (j = 0, 1), \qquad \Omega^*(W) := \inf_{\prec} \Omega^*(W, \prec). \tag{8.5}$$

For kernels, we can use Ω^* instead of Ω to characterize monotonicity, cf. Theorems 4.5 and 4.11.

Theorem 8.1. *Let $(\mathcal{S}, \mu, \prec)$ be an ordered probability space and W a kernel on $(\mathcal{S}, \mu)$. Then $\Omega^*(W, \prec) = 0$ if and only if W is a.e. equal to a monotone kernel.*

Proof. If W is a.e. equal to a monotone kernel, then $W(x, z) \leq W(y, z)$ for a.e. (x, y, z) with $x \prec y$, and thus $\Omega^*(W, \prec) = 0$. The converse follows by Theorem 4.5, since $\Omega_1(W, \prec) \leq \Omega^*(W, \prec)$. $\square$

Theorem 8.2. *Let W be a kernel on a probability space $\mathcal{S}$ with at least one measurable order. Then $\Omega^*(W) = 0$ if and only if W is a.e. equal to a monotone kernel on $(\mathcal{S}, \prec)$ for some order $\prec$ on $\mathcal{S}$.*

Theorem 4.11 gives further equivalent conditions, for example that Γ_W is a monotone graph limit.

Proof. If $\Omega^*(W) = 0$, then $\Omega_1(W) = 0$, since $\Omega_1(W) \le \Omega^*(W)$. Hence the conclusion follows by Theorem 4.11.

Conversely, if W is a.e. equal to a monotone kernel om $(\mathcal{S}, \prec)$, then $\Omega^*(W) \le \Omega^*(W, \prec) = 0$ by Theorem 8.1. □

For a sequence of graphs, we cannot replace Ω_0 by Ω_0^* in Theorem 1.5. In fact, we have the following result, which shows that $\Omega_0^*(G_\nu) \to 0$ characterizes *threshold* graph limits rather than monotone graph limits. (Recall that threshold graph limits are the monotone graph limits that correspond to 0/1-valued kernels; see Remark 1.7.)

As usual, we define the *edit distance* $d_{\mathrm{e}}(G, G')$ of two graphs on the same vertex set $V(G) = V(G')$ by $d_{\mathrm{e}}(G, G') = |E(G) \triangle E(G')|$. If $\mathcal{A}$ is a class of graphs, then

$$d_{\mathrm{e}}(G, \mathcal{A}) := \inf\bigl\{d_{\mathrm{e}}(G, G') : G' \in \mathcal{A} \text{ and } V(G') = V(G)\bigr\}. \tag{8.6}$$

Theorem 8.3. *Let (G_ν) be a sequence of graphs with $|G_\nu| \to \infty$. Then the following are equivalent.*

(i) $\Omega_0^*(G_\nu) \to 0$.
(ii) *Every convergent subsequence of (G_ν) has a limit that is a threshold graph limit.*
(iii) $d_{\mathrm{e}}(G_\nu, \mathcal{T}) = o\bigl(|G_\nu|^2\bigr)$, *where $\mathcal{T}$ is the class of threshold graphs.*
(iv) *There exists a sequence of threshold graphs G'_ν with $V(G'_\nu) = V(G_\nu)$ and $\bigl|E(G_\nu) \triangle E(G'_\nu)\bigr| = o\bigl(|G_\nu|^2\bigr)$.*
(v) *There exists a sequence of threshold graphs G'_ν with $V(G'_\nu) = V(G_\nu)$ and $\|W_{G_\nu} - W_{G'_\nu}\|_{L^1(\mathcal{S}^2)} = o(1)$.*
(vi) *There exists a sequence of threshold graphs G'_ν with $V(G'_\nu) = V(G_\nu)$ and $\|W_{G_\nu} - W_{G'_\nu}\|_\square = o(1)$.*

We say that a sequence (G_ν) of graphs with $|G_\nu| \to \infty$ is *quasithreshold* if it satisfies one, and thus all, of the conditions in Theorem 8.3.

As a special case of the equivalence (i) $\iff$ (ii), we see that if $G_\nu \to \Gamma$, then Γ is a threshold graph limit if and only if $\Omega_0^*(G_\nu) \to 0$; cf. Theorem 1.6.

The proof of Theorem 8.3 is simpler than the proof of Theorem 1.5, but we will nevertheless need some other results first. One complication is that there is no analogue of Lemma 6.2(iii); as is shown by the following example, $\Omega^*(W, \prec)$ is not continuous for the cut norm.

Example 8.4. Let $W = 1/2$ be constant on $[0, 1]^2$, and let (G_n) be a sequence of graphs with $|G_n| = n$ and $G_n \to W$, i.e., (G_n) is a sequence of quasirandom graphs. (E.g., let G_n be random graphs $G(n, 1/2)$.) Then, for every $\varepsilon > 0$, $\big||N(v) \setminus N(w)| - n/4\big| \le \varepsilon n$ for all but $o(n^2)$ pairs $(v, w) \in V_{G_n}^2$, and thus for any order $\prec$, $\big|n^3\Omega_1^*(G_n, \prec) - n^3/8\big| \le \varepsilon n^3 + o(n^3)$, so $\big|\Omega_1^*(G_n, \prec) - 1/8\big| \le \varepsilon + o(1)$. Since ε is arbitrary, it follows that

$$\Omega^*(W_{G_n}) = \Omega_1^*(G_n) \to \tfrac{1}{8} \neq 0 = \Omega^*(W),$$

although $\|W_{G_n} - W\|_\square \to 0$.

Ω^* is obviously continuous in the stronger L^1 norm. It is possible to prove Theorem 8.3 using this fact and Lemma 8.13 below, but it is simpler to use another extension of Ω_1^* to kernels.

Definition. If $(\mathcal{S}, \mu)$ is an atomless probability space and $\prec$ an order on $\mathcal{S}$, let

$$\widetilde{\Omega}^*(W, \prec) := \iiint_{x \prec y} W(x, z)\big(1 - W(y, z)\big) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) \,\mathrm{d}\mu(z). \tag{8.7}$$

If $\mathcal{S}$ has atoms, we add half the integral over $x = y$ (and any z), i.e., we add $\frac{1}{2} \iint W(x, z)\big(1 - W(x, z)\big)\mu\{x\} \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(z)$.

The definition in the case that $\mathcal{S}$ has atoms is such that $\widetilde{\Omega}^*(W, \prec) = \widetilde{\Omega}^*(\widehat{W}, \widehat{\prec})$, where $\widehat{W}$ is the extension of W to the atomless probability space $\widehat{\mathcal{S}} := \mathcal{S} \times [0, 1]$ and $\widehat{\prec}$ is the lexicographic order on $\widehat{\mathcal{S}}$.

Note that if W is 0/1-valued, then $\widetilde{\Omega}^*(W, \prec) = \Omega^*(W, \prec)$. In particular, for any graph with an order $\prec$ on $V = V(G)$, by (8.4),

$$\Omega_1^*(G, \prec) = \Omega^*(W_G, <) = \widetilde{\Omega}^*(W_G, <). \tag{8.8}$$

For our purposes $\widetilde{\Omega}^*$ is better than Ω^* in two different ways. The first is that, unlike Ω^*, $\widetilde{\Omega}^*$ is continuous with respect to the cut norm. Before proving this, we recall a basic property of the cut norm. (See e.g. [14] for a proof.)

Lemma 8.5. *If $W \in L^1(\mathcal{S}^2)$, then $\|W_{\mathcal{S}}\|_{L^1(\mathcal{S})} \le \|W\|_\square$.* $\square$

Recall that, by definition, a kernel W takes values in $[0, 1]$.

Lemma 8.6. *Let $(\mathcal{S}, \prec)$ be an ordered probability space. If W_1 and W_2 are kernels on $\mathcal{S}$, then $\big|\widetilde{\Omega}^*(W_1, \prec) - \widetilde{\Omega}^*(W_2, \prec)\big| \le 2\|W_1 - W_2\|_\square$.*

Proof. We may assume that $\mathcal{S}$ is atomless. (Otherwise we consider $\mathcal{S}\times[0,1]$.) In this case, writing U_x for $\{y : y \succ x\}$, we have the alternative formula

$$\begin{aligned}
\widetilde{\Omega}^*(W,\prec) &= \iint W(x,z)\mu(U_x)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(z) \\
&\qquad - \iiint_{x\prec y} W(x,z)W(y,z)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z) \\
&= \iint \mu(U_x)W(x,z)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(z) \\
&\qquad - \frac{1}{2}\iiint W(x,z)W(y,z)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(z) \\
&= \iint \mu(U_x)W(x,z)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(z) - \frac{1}{2}\int W_{\mathcal{S}}(z)^2\,\mathrm{d}\mu(z). \qquad (8.9)
\end{aligned}$$

By the definition (2.1) of the cut norm,

$$\left|\iint \mu(U_x)\big(W_1(x,z)-W_2(x,z)\big)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(z)\right| \le \|W_1-W_2\|_\square.$$

Recalling that $|W_j|\le 1$ and using Lemma 8.5 on W_1-W_2,

$$\begin{aligned}
&\left|\int_{\mathcal{S}}\big(W_{1,\mathcal{S}}(z)^2 - W_{2,\mathcal{S}}(z)^2\big)\,\mathrm{d}\mu(z)\right| \\
&\qquad = \left|\int_{\mathcal{S}}\big(W_{1,\mathcal{S}}(z)-W_{2,\mathcal{S}}(z)\big)\big(W_{1,\mathcal{S}}(z)+W_{2,\mathcal{S}}(z)\big)\,\mathrm{d}\mu(z)\right| \\
&\qquad\qquad \le 2\|W_{1,\mathcal{S}}(z)-W_{2,\mathcal{S}}(z)\|_{L^1(\mathcal{S})} \le 2\|W_1-W_2\|_\square.
\end{aligned}$$

Applying (8.9) to W_1 and W_2, the result follows. □

Theorem 8.7. *Let $(\mathcal{S},\prec)$ be an ordered probability space and W a kernel on $(\mathcal{S},\prec)$. Then $\widetilde{\Omega}^*(W,\prec)=0$ if and only if W is a.e. equal to a $0/1$-valued monotone kernel.*

Proof. As usual, we may assume for simplicity that $\mathcal{S}$ is atomless. Suppose first $\widetilde{\Omega}^*(W,\prec)=0$. For $a>0$, let $E_a := \{(x,y)\in\mathcal{S}^2 : a\le W(x,y)\le 1-a\}$, and, for $z\in\mathcal{S}$, let $E_a(z) := \{x\in\mathcal{S} : (x,z)\in E_a\}$ be the corresponding section.

If $x,y\in E_a(z)$, then $W(x,z)(1-W(y,z))\ge a^2$, and thus, for each z,

$$\begin{aligned}
&\iint_{x\prec y} W(x,z)\big(1-W(y,z)\big)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&\qquad\qquad \ge a^2\mu\times\mu\big\{(x,y)\in E_a(z)^2 : x\prec y\big\} = \frac{1}{2}a^2\mu(E_a(z))^2.
\end{aligned}$$

Hence,

$$0=\widetilde{\Omega}^*(W,\prec) \ge \int_{\mathcal{S}} \frac{1}{2}a^2\mu(E_a(z))^2\,\mathrm{d}\mu(z),$$

and thus $\mu(E_a(z))=0$ for a.e. z, so $\mu\times\mu(E_a)=\int_{\mathcal{S}}\mu(E_a(z))\,\mathrm{d}\mu(z)=0$. Consequently, E_a is a null set for every $a>0$. Hence, $W(x,y)\in\{0,1\}$

a.e. Thus W is a.e. 0/1-valued, which implies that $\Omega^*(W,\prec) = \widetilde{\Omega}^*(W,\prec) = 0$; hence Theorem 8.1 shows that W is a.e. equal to a monotone kernel W'. Finally, W' is a.e. 0/1-valued, and thus a.e. equal to the 0/1-valued monotone kernel $\mathbf{1}\{W'>0\}$.

The converse is obvious. □

We also have an analogue of Lemma 6.3. To prove this, we shall need the following 'rearrangement' inequality.

Lemma 8.8. *Let $\prec$ and $<$ be two orders on an atomless probability space $\mathcal{S}$, and let f be a bounded function on $\mathcal{S}$. If $x<y \implies f(x)\le f(y)$, then $\iint_{x\prec y} f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \ge \iint_{x<y} f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$.*

Proof. Consider first one arbitrary order $\prec$. Let $D_y := \{x : x \prec y\}$ and set $\varphi(y) := \mu(D_y)$, and let $D(t)$ be as in Lemma 5.2. Then D_y and $D(\varphi(y))$ are two downsets with the same measure, and thus they differ only by a null set, cf. Lemma 5.1.

Let $F(y) := \int_{x\prec y} f(x)\,\mathrm{d}\mu(x)$ and define $\alpha(t) := \int_{D(t)} f(x)\,\mathrm{d}\mu(x)$. Then

$$F(y) = \int_{D_y} f = \int_{D(\varphi(y))} f = \alpha(\varphi(y)).$$

It was noted in the proof of Lemma 5.2 that if X has distribution μ, then $\varphi(X)$ has distribution $U(0,1)$. Equivalently, the function $\varphi : \mathcal{S} \to [0,1]$ maps μ to the uniform measure on [0,1]. Hence,

$$\iint_{x\prec y} f(x)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) = \int_{\mathcal{S}} F(y)\,\mathrm{d}\mu(y) = \int_{\mathcal{S}} \alpha(\varphi(y))\,\mathrm{d}\mu(y) = \int_0^1 \alpha(t)\,\mathrm{d}t.$$

Now write $\alpha = \alpha_\prec$ and compare $\alpha_\prec(t)$ and $\alpha_<(t)$. Both are integrals of f over sets of measure t, and for $\alpha_<$ the set is such that if x is in the set and y is not, then $x<y$ and thus $f(x)\le f(y)$. It follows easily that $\alpha_<(t)$ is the minimum of $\int_E f\,\mathrm{d}\mu$ over all set E of measure t, and thus in particular $\alpha_<(t)\le\alpha_\prec(t)$ for any other order $\prec$. Consequently, $\int_0^1 \alpha_<(t)\,\mathrm{d}t \le \int_0^1 \alpha_\prec(t)\,\mathrm{d}t$, and the result follows. □

Lemma 8.9. *Let $<$ be a (measurable) order on $\mathcal{S}$ and assume that $x<y \implies W_{\mathcal{S}}(x)\le W_{\mathcal{S}}(y)$. Then, $\widetilde{\Omega}^*(W,<) = \widetilde{\Omega}^*(W)$.*

Proof. We may again assume for simplicity that $\mathcal{S}$ is atomless. Let $\prec$ be any order on $\mathcal{S}$. We again use (8.9), which we write as

$$\widetilde{\Omega}^*(W,\prec) = \int_{\mathcal{S}} \mu(U_x) W_{\mathcal{S}}(x)\,\mathrm{d}\mu(x) - \frac{1}{2}\int_{\mathcal{S}} W_{\mathcal{S}}(x)^2\,\mathrm{d}\mu(x).$$

The second integral does not depend on $\prec$. Moreover, the first integral equals $\iint_{x\prec y} W_{\mathcal{S}}(x)$, which by Lemma 8.8 is minimized by taking $\prec$ equal to $<$. Hence $\widetilde{\Omega}^*(W,\prec) \ge \widetilde{\Omega}^*(W,<)$, and the result follows. □

Remark 8.10. It follows by (8.8) that the corresponding result holds for graphs and Ω_1^*: i.e., ordering the vertices by their degrees achieves the minimum $\min_{\prec} \Omega_1^*(G, \prec)$.

Our next result shows that $\widetilde{\Omega}^*$ characterizes kernels that yield threshold graph limits. Note the parallel and contrast to Theorems 4.11 and 8.2.

Theorem 8.11. *Let W be a kernel on a probability space $\mathcal{S}$ with at least one measurable order. Then the following are equivalent.*

(i) $\widetilde{\Omega}^*(W) = 0$.
(ii) *There exists an order $\prec$ on $\mathcal{S}$ such that W is a.e. equal to a* $0/1$*-valued monotone kernel on* $(\mathcal{S}, \prec)$.
(iii) *W is equivalent to a* $0/1$*-valued monotone kernel on some ordered probability space.*
(iv) *W is equivalent to a* $0/1$*-valued monotone kernel on* $[0,1]$.
(v) *Γ_W is a threshold graph limit.*

Proof. (i) $\Longrightarrow$ (ii). There exists a measurable order $\prec_0$ on $\mathcal{S}$. As in the proof of Corollary 6.5, we define an order $\prec$ on $\mathcal{S}$ by (6.6). Lemma 8.9 applies and yields $\widetilde{\Omega}^*(W, \prec) = \widetilde{\Omega}^*(W) = 0$, and the result follows by Theorem 8.7.

(ii) $\Longrightarrow$ (i). Theorem 8.7 yields $\widetilde{\Omega}^*(W, \prec) = 0$ and thus $\widetilde{\Omega}^*(W) \leq \widetilde{\Omega}^*(W, \prec) = 0$.

(ii) $\Longleftrightarrow$ (iii) $\Longleftrightarrow$ (iv). Every kernel equivalent to an a.e. $0/1$-valued kernel is itself a.e. $0/1$-valued, see Remark 1.7 and [14]. Furthermore, arguing as in the proof of Theorem 8.7, a monotone kernel W that is a.e. $0/1$-valued is a.e. equal to the $0/1$-valued monotone kernel $\mathbf{1}\{W > 0\}$. Hence, (ii) $\Longleftrightarrow$ (iii) $\Longleftrightarrow$ (iv) follows from the corresponding equivalences in Theorem 4.11.

(iv) $\Longleftrightarrow$ (v). As noted in the introduction, this was proved by Diaconis, Holmes and Janson [7]. ☐

We need some more preparation before the proof of Theorem 8.3.

Lemma 8.12. *Let W_1 and W_2 be kernels on a probability space $\mathcal{S}$ with W_1* $0/1$*-valued, and let W_1' be a* $0/1$*-valued step kernel with n steps. Then*

$$\|W_1 - W_2\|_{L^1(\mathcal{S}^2)} \leq n^2 \|W_1 - W_2\|_\square + 2\|W_1 - W_1'\|_{L^1(\mathcal{S}^2)}.$$

Proof. Let $\{A_i\}_1^n$ be a partition of $\mathcal{S}$ 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

$$\iint_{A_i \times A_j} |W_1' - W_2| = \iint_{A_i \times A_j} W_2 \leq \|W_1 - W_2\|_\square + \iint_{A_i \times A_j} W_1$$
$$= \|W_1 - W_2\|_\square + \iint_{A_i \times A_j} |W_1 - W_1'|.$$

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

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

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

Lemma 8.13. *Let W and $W_1, W_2, \dots$ be kernels on a probability space $\mathcal{S}$, and assume that W is 0/1-valued. Then $\|W_n - W\|_\square \to 0$ as $n \to \infty$ if and only if $\|W_n - W\|_{L^1(\mathcal{S}^2)} \to 0$.*

Proof. Assume $\|W_n - W\|_\square \to 0$. W is the indicator function $\mathbf{1}_A$ of a measurable set $A \subseteq \mathcal{S}^2$. 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 kernel W' such that $\|W - W'\|_{L^1} < \varepsilon$. Let the corresponding partition have $N = N(\varepsilon)$ parts. Lemma 8.12 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. □

Proof of Theorem 8.3. Note first that (i) is equivalent to $\Omega_1^*(G_\nu) \to 0$ by (8.2), and that $\Omega_1^*(G_\nu) = \widetilde{\Omega}^*(W_{G_\nu})$ by (8.8).

(i) $\Longrightarrow$ (ii). Assume (i) and consider a subsequence that converges. We thus assume that there exists a graph limit Γ with $G_\nu \to \Gamma$. Let W be a kernel on [0,1] representing Γ.

We have $G_\nu \to W$, and thus $\delta_\square(W_{G_\nu}, W) \to 0$. Moreover, by [5, Lemma 5.3] we may choose the labelling of the vertices in G_ν such that

$$\|W_{G_\nu} - W\|_\square \to 0. \tag{8.10}$$

This labelling yields an order $<$ on $V(G_\nu)$. Let $\prec$ be an order on $V(G_\nu)$ achieving the minimum in (8.5) for $\Omega_1^*(G_\nu)$, i.e., such that

$$\Omega_1^*(G_\nu, \prec) = \Omega_1^*(G_\nu) = o(1). \tag{8.11}$$

In general $\prec$ differs from $<$, but it clearly corresponds to some order $\prec_\nu$ on $[0, 1]$ and, by (8.8) again,

$$\Omega_1^*(G_\nu, \prec) = \Omega^*(W_{G_\nu}, \prec_\nu) = \widetilde{\Omega}^*(W_{G_\nu}, \prec_\nu). \tag{8.12}$$

By Lemma 8.6 and (8.10)–(8.12), we then have

$$\widetilde{\Omega}^*(W, \prec_\nu) \le \widetilde{\Omega}^*(W_{G_\nu}, \prec_\nu) + 2\|W - W_{G_\nu}\|_\square \to 0,$$

as $\nu \to \infty$; hence $\widetilde{\Omega}^*(W) = 0$ and $\Gamma = \Gamma_W$ is a threshold graph limit by Theorem 8.11.

(ii) $\Longrightarrow$ (iii) Suppose that (iii) fails; then there exists $\varepsilon > 0$ and a subsequence for which $d_{\mathrm{e}}(G_\nu, \mathcal{T}) > \varepsilon |G_\nu|^2$. We may select a subsubsequence such that G_ν converges; we shall show that (ii) implies (iii) in this case, which yields a contradiction.

Suppose then that $G_\nu \to \Gamma$ for some graph limit Γ, and that (ii) holds. By assumption, Γ is a threshold graph limit. Let W be a kernel on $[0,1]$ representing Γ. By the result of Diaconis, Holmes and Janson [7] discussed in the introduction, we may choose W to be monotone and 0/1-valued.

We have $G_\nu \to W$, and thus $\delta_\square(W_{G_\nu}, W) \to 0$. As above, by [5, Lemma 5.3] we may choose the labelling of the vertices in G_ν such that $\|W_{G_\nu} - W\|_\square \to 0$. By Lemma 8.13, this implies $\|W_{G_\nu} - W\|_{L^1} \to 0$.

Since, by assumption, Γ is a threshold graph limit, there exists a sequence of threshold graphs G'_ν such that $G'_\nu \to \Gamma$, and we may further assume that $|G'_\nu| = |G_\nu|$. (For example, we may a.s. take G'_ν as the random graph $G(n_\nu, W)$ with $n_\nu = |G_\nu|$.) Then also $\delta_\square(W_{G'_\nu}, W) \to 0$, and by [5, Lemma 5.3] again we may choose the labelling of the vertices in G'_ν such that $\|W_{G'_\nu} - W\|_\square \to 0$, and thus by Lemma 8.13 $\|W_{G'_\nu} - W\|_{L^1} \to 0$. Consequently,

$$\|W_{G_\nu} - W_{G'_\nu}\|_{L^1} \le \|W_{G_\nu} - W\|_{L^1} + \|W - W_{G'_\nu}\|_{L^1} \to 0.$$

We may identify the vertex sets of G_ν and G'_ν. Then

$$d_{\mathrm{e}}(G_\nu, \mathcal{T}) \le \bigl|E(G_\nu) \triangle E(G'_\nu)\bigr| = \tfrac{1}{2}|G_\nu|^2 \|W_{G_\nu} - W_{G'_\nu}\|_{L^1} = o(|G_\nu|^2).$$

(iii) $\Longleftrightarrow$ (iv) by the definition (8.6).

(iv) $\Longleftrightarrow$ (v) by

$$\|W_{G_\nu} - W_{G'_\nu}\|_{L^1(\mathcal{S}^2)} = 2|G_\nu|^{-2} \bigl|E(G_\nu) \triangle E(G'_\nu)\bigr|.$$

(v) $\Longrightarrow$ (vi) since $\|\cdot\|_\square \le \|\cdot\|_{L^1(\mathcal{S}^2)}$.

(vi) $\Longrightarrow$ (i). Let $<$ be the order on $V(G_\nu) = V(G'_\nu)$ defined by the degrees of the vertices in G'_ν. Then, since G'_ν is a threshold graph, $N_{G'_\nu}(v) \subseteq N_{G'_\nu}(w) \cup \{w\}$ whenever $v < w$, and thus $\Omega^*_0(G'_\nu, <) = 0$ by (8.1).

By (8.2), (8.8) and Lemma 8.6,

$$\begin{aligned} \Omega^*_0(G_\nu, <) &= \Omega^*_0(G_\nu, <) - \Omega^*_0(G'_\nu, <) = \Omega^*_1(G_\nu, <) - \Omega^*_1(G'_\nu, <) + o(1) \\ &= \widetilde{\Omega}^*(W_{G_\nu}, <) - \widetilde{\Omega}^*(W_{G'_\nu}, <) + o(1) \\ &\le 2\|W_{G_\nu} - W_{G'_\nu}\|_\square + o(1) = o(1). \end{aligned}$$

Hence $\Omega^*_0(G_\nu) \to 0$. □

References

[1] R. Blei, *Analysis in Integer and Fractional Dimensions.* Cambridge University Press, Cambridge, 2001.

[2] B. Bollobás, *Modern Graph Theory*, Springer, New York, 1998.

[3] B. Bollobás and O. Riordan, Metrics for sparse graphs. *Surveys in Combinatorics 2009*, LMS Lecture Notes Series 365, Cambidge Univ. Press, 2009, pp. 211–287. `arXiv:0708.1919.`

[4] 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.

[5] 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.

[6] F. R. K. Chung, R. L. Graham & R. M. Wilson, Quasi-random graphs. *Combinatorica* **9** (1989), no. 4, 345–362.

[7] P. Diaconis, S. Holmes and S. Janson, Threshold graph limits and random threshold graphs. *Internet Mathematics* **5** (2009), no. 3, 267–318.

[8] P. Diaconis & S. Janson, Graph limits and exchangeable random graphs. *Rendiconti di Matematica* **28** (2008), 33–61.

[9] T. Figiel, P. Hitczenko, W.B. Johnson, G. Schechtman, and J. Zinn, Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities, *Trans. Amer. Math. Soc.* **349** (1997), 997–1027.

[10] U. Haagerup, Les meilleures constantes de l'inégalité de Khintchine, *C. R. Acad. Sci. Paris Sér. A-B* **286** (1978), A259–A262.

[11] U. Haagerup, The best constants in the Khintchine inequality, *Studia Math.* **70** (1981), 231–283.

[12] S. Janson, Standard representation of multivariate functions on a general probability space. *Electronic Comm. Probab.* **14** (2009), paper 34, 343–346.

[13] S. Janson, Quasi-random graphs and graph limits. *Eur. J. Comb.*, to appear. `arXiv:0905.3241v1.`

[14] S. Janson, Graphons, cut norm and distance, couplings and rearrangements. Preprint, 2010. `arXiv:1009.2376v2.`

[15] A. Khintchine, Über dyadische Brüche, *Math. Z.* **18** (1923), 109–116.

[16] H. König, S. Kwapień, Best Khintchine type inequalities for sums of independent, rotationally invariant random vectors, *Positivity* **5** (2001), 115–152.

[17] R. Latała, Estimation of moments of sums of independent real random variables, *Ann. Probab.* **25** (1997), 1502–1513.

[18] J. E. Littlewood, On bounded bilinear forms in an infinite number of variables, *Quart. J. Math. Oxford* **1** (1930), 164–174.

[19] L. Lovász and V. T. Sós, Generalized quasirandom graphs. *J. Comb. Theory B* **98** (2008), no. 1, 146–163.

[20] L. Lovász and B. Szegedy, Limits of dense graph sequences. *J. Comb. Theory B* **96** (2006), no. 6, 933–957.

[21] L. Lovász and B. Szegedy, Finitely forcible graphons. Preprint, 2009. `arXiv:0901.0929.`

[22] L. Lovász and B. Szegedy, Regularity partitions and the topology of graphons. Preprint, 2010. arXiv:1002.4377.

[23] N. V. R. Mahadev and U. N. Peled, *Threshold Graphs and Related Topics*. North-Holland, Amsterdam, 1995.

[24] S.J. Szarek, On the best constants in the Khinchin inequality, *Studia Math.* **58** (1976), 197–208.

[25] A. Thomason, Pseudorandom graphs. *Random Graphs '85 (Poznań, 1985)*, 307–331, North-Holland, Amsterdam, 1987.

[26] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs. *Surveys in Combinatorics 1987 (New Cross, 1987)*, 173–195, London Math. Soc. Lecture Note Ser. **123**, Cambridge Univ. Press, Cambridge, 1987.

[27] A. Zygmund, *Trigonometric series, Vol. I*, 3rd ed., Cambridge Math. Library, Cambridge University Press, 2002, xiv + 383 pp.

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK and Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA

E-mail address: b.bollobas@dpmms.cam.ac.uk

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/

Mathematical Institute, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK

E-mail address: riordan@maths.ox.ac.uk