Regularity partitions and the topology of graphons

László Lovász*
Institute of Mathematics
Eötvos Loránd University
and
Balázs Szegedy
Department of Mathematics
University of Toronto

October 31, 2018

Contents

Abstract

We highlight a topological aspect of the graph limit theory. Graphons are limit objects for convergent sequences of dense graphs. We introduce the representation of a graphon

*Research supported by OTKA grant No. 67867 and ERC Advanced Research Grant No. 227701

on a unique metric space and we relate the dimension of this metric space to the size of regularity partitions. We prove that if a graphon has an excluded induced sub-bigraph then the underlying metric space is compact and has finite packing dimension. It implies in particular that such graphons have regularity partitions of polynomial size.

1 Introduction

One can define convergence of a growing graph sequence [4, 3, 5], and construct a limit object to such a sequence [11] in the form of a symmetric measurable function $W: J \times J \to [0,1]$, where J is any probability space (one may assume here that $J = [0,1]$ with the Lebesgue measure, but this is not always convenient). We call the pair (J, W) a *graphon*.

The goal of this paper is to show that one can introduce also a topology on J (in fact, a metric), and that topological properties of this space are related to combinatorial properties of the graphon (or of the graphs whose limit it represents). A related metric was introduced in [12], and the topology on J was used in [13].

The theory of graph limits is tied to the Regularity Lemma of Szemerédi [14, 15] in several ways. In [12] it was shown that the Regularity Lemma is equivalent to the compactness of the space of graphons in an appropriate metric, and also to a "dimensionality" of particular graphons. This paper relates to the latter result.

The metric in question is simply the L_1 metric on functions $W(x, .)$, $x \in J$. This metric itself can be weird (it may not even be defined on all points of J). We show in Section 3 that that every graphon is "equivalent" (technically: weakly isomorphic, see the end of Section 2) to a graphon (J, W) with special properties: J is a complete separable metric space, and the probability measure on J has full support. We call such graphons *pure*. We also prove that the pure version of a graphon is uniquely determined up to changing the function W on a 0-set in each row. We define another metric in which J is compact, and characterize the cases when the two define the same topology. We prove that several important functions defined on J are continuous in this topology, which shows that it is indeed the "right" topology to define on J.

In Section 4 we show that topological properties of pure graphons are related to their graph-theoretic properties. Our main result states tha *if we exclude any bipartite graph from the graphon, then J must be compact and finite dimensional.*

In [12] it was shown that weak regularity partitions of a graphon (J, W) (which generalize weak regularity partitions of graphs in a natural way) correspond to covering J with sets of small diameter. In Section 5 we give a stronger and cleaner version of this result. Combined with the results in Section 4, we obtain the following fact: If a graph does not contain a fixed bipartite graph F as an induced sub-bigraph, then it has polynomial size strong regularity partitions (in the error bound ε).

A motivation for our paper comes from extremal combinatorics. In [13] we study the structure of graphons that arise as unique solutions of extremal problems involving the densities of finitely many subgraphs (we call such graphons *finitely forcible*). Such graphons come up naturally in extremal graph theory. Quite interestingly, all the examples of finitely forcible graphons produced in [13] have a compact and finite dimensional underlying metric space. The question arises wether every extremal problem (involving a finite number of subgraph densities) has a solution of this type.

Finally we mention that graph limit theory has a close connection to the theory of dynamical systems. Probability spaces with measure preserving actions can often be endowed by a natural topology in which the action is continuous. The corresponding theory is called topological dynamics. Informally speaking, we can say that the relationship between graphons and topological graphons is similar to the relationship between dynamics and topological dynamics.

2 Preliminaries

We make a technical but useful distinction between bipartite graphs and bigraphs. A *bipartite graph* is a graph (V, E) whose node set has a partition into two classes such that all edges connect nodes in different classes. A *bigraph* is a triple (U_1, U_2, E) where U_1 and U_2 are finite sets and $E \subseteq U_1 \times U_2$. So a bipartite graph becomes a bigraph if we fix a bipartition and specify which bipartition class is first and second. On the other hand, if $F = (V, E)$ is a graph, then (V, V, E') is an associated bigraph, where $E' = \{(x, y) :\ xy \in E\}$. This bigraph is obtained from F by a standard construction of doubling the nodes.

If $G = (V, E)$ is a graph, then an *induced sub-bigraph* of G is determined by two subsets $S, T \subseteq V$, and its edge set consists of those pairs $(x, y) \in S \times T$ for which $xy \in E$ (so this is an induced subgraph of the bigraph associated with G).

Let $J_i = (\Omega_i, \mathcal{A}_i, \pi_i)$ $(i = 1, 2)$ be (standard) probability spaces. A measurable function $W :\ J_1 \times J_2 \to [0, 1]$ is called a *bigraphon*. A *graphon* is a special bigraphon where $J_1 = J_2 = J$ and W is symmetric: $W(x, y) = W(y, x)$ for all $x, y \in J$.

For a fixed probability space J, graphons can be considered as elements of the space $L_\infty(J \times J)$. The norm that it most important in their is study is, however, not the L_∞ norm, but the *cut-norm*, defined by

$$\|W\|_\square = \sup_{S,T\subseteq J} \Big| \int\limits_{S\times T} W(x, y)\, dx\, dy \Big|.$$

We will also use the L_1 norm

$$\|W\|_1 = \int\limits_{J\times J} |W(x, y)|\, dx\, dy.$$

A graphon (J, W) is called a *stepfunction*, if there is a partition of J into a finite number of measurable sets $S_1, \dots, S_n$ so that W is constant on every $S_i \times S_j$. The partition classes will be called the *steps* of the stepfunction.

Every graph $F = (V, E)$ can be considered as a graphon, if we consider V as a finite probability space with the uniform measure, and E, as the indicator function of adjacency. We can resolve the atoms into intervals of length $1/|V|$, to get a graphon $([0, 1], W_F)$ (which is a stepfunction). More explicitly, we split $[0, 1]$ in $|V|$ equal intervals L_i, and define $W_F(x, y) = E(i, j)$ for $ix \in L_i$ and $y \in L_j$. This graphon is weakly isomorphic to (V, E) (see below).

In a similar way, every bigraph can be considered as a finite bigraphon, and defines a bigraphon $([0, 1], [0, 1], W_F)$.

Remark 2.1 We could consider the version of this notion where $J_1 = J_2$ but W is not necessarily symmetric. Such a structure arises as the limit object of a convergent sequence of directed graphs with no parallel edges, and therefore can be called a *digraphon*. We do not need them in this paper.

Every bigraphon (J_1, J_2, W) can be considered as a linear kernel operator $L_1(J_1) \to L_\infty(J_2)$, defined by

$$f \mapsto \int\limits_J W(.,y)f(y)\,dy.$$

Of course, this operator reamin well-defined if we increase the subscript in L_1 in the domain and lower the subscript in L_∞ in the range. In the case of a graphon (J, W), it is useful to consider it as an operator $L_2(J) \to L_2(J)$, since it is then a Hilbert-Schmidt operator, and a rich theory is applicable. In particular, we know that it has a discrete spectrum.

If (J_1, J_2, U) and (J_2, J_3, W) are two bigraphons, we can define their *operator product* $(J_1, J_3, U \circ W)$ by

$$(U \circ W)(x,y) = \int\limits_{J_2} U(x,z)W(z,y)\,dz.$$

(We will write dz instead of $d\pi_2(z)$, where π_2 is the measure on J_2: integrating over J_2 means that we integrate with respect to the probability measure of J_2.)

The notion of the density of a graph in a graphon has been introduced in [7]. Here we need several versions, which unfortunately leads to some messy notation. For a graphon (J, W) and graph $F = (V, E)$, we associate a variable $x_v \in J$ with every node $v \in V$, and define

$$t(F,W;x) = \prod_{uv \in E(F)} W(x_u, x_v), \qquad t(F,W) = \int\limits_{J^V} t(F,W;x)\,dx.$$

We can think of $t(F, W)$ as "counting subgraphs isomorphic to F". We also need the induced version:

$$t_{\text{ind}}(F,W;x) = \prod_{uv \in E(F)} W(x_u, x_v) \prod_{\substack{u,v \in V \\ uv \notin E(F)}} (1 - W(x_u, x_v))$$

$$t_{\text{ind}}(F,W) = \int\limits_{J^V} t_{\text{ind}}(F,W;x)\,dx.$$

For any subset $S \subseteq V$, we define $t_S(F,W;.) : \ J^S \to \mathbb{R}$ by integrating only over variables corresponding to $V \setminus S$: If x' and x'' denote the restrictions of $x \in J^V$ to S and $V \setminus S$, respectively, then

$$t_S(F,W;x') = \int\limits_{J^{V \setminus S}} t(F,W;x)\,dx''.$$

Note that $t_\emptyset(F,W) = t(F,W)$ and $t_V(F,W;.) = t(F,W;.)$.

These quantities have obvious analogues for bigraphs and bigraphons. For a bigraphon (J_1, J_2, W) and bipartite graph (U_1, U_2, E), we introduce variables $x_u \in J_1$ $(u \in U_1)$ and $y_v \in J_2$

$(v \in U_2)$, and define

$$t^{\mathsf{b}}(F,W;x,y) = \prod_{uv\in E(F)} W(x_u,y_v), \qquad t^{\mathsf{b}}(F,W) = \int\limits_{J_1^{U_1}} \int\limits_{J_2^{U_2}} t^{\mathsf{b}}(F,W;x,y)\,dy\,dx.$$

Again, we define an induced version:

$$t^{\mathsf{b}}_{\mathrm{ind}}(F,W;x,y) = \prod_{ij\in E(F)} W(x_i,y_j) \prod_{\substack{i\in U_1, j\in U_2\\ ij\notin E(F)}} (1-W(x_i,y_j))$$

$$t^{\mathsf{b}}_{\mathrm{ind}}(F,W) = \int\limits_{J_1^{U_1}} \int\limits_{J_2^{U_2}} t^{\mathsf{b}}_{\mathrm{ind}}(F,W;x,y)\,dy\,dx.$$

Assume that subsets $S_i \subseteq U_i$ are specified. We define the function $t^{\mathsf{b}}(F,W;.) : J_1^{S_1} \times J_2^{S_2} \to \mathbb{R}$ by

$$t^{\mathsf{b}}_{S_1,S_2}(F,W;x',y') = \int\limits_{J_1^{U_1\setminus S_1}} \int\limits_{J_2^{U_2\setminus S_2}} t^{\mathsf{b}}(F,W;x,y)\,dy''\,dx'',$$

where, similarly as above, x' and x'' denote the restrictions of $x \in J_1^{U_1}$ to S_1 and $U_1 \setminus S_1$, respectively, and similarly for y. We can define $t^{\mathsf{b}}_{\mathrm{ind};S_1,S_2}(F,W)(x',y')$ analogously.

Two graphons (J,W) and (J',W') are *weakly isomorphic* if for every graph F, $t(F,W) = t(F,W')$. Various characterizations of weak isomorphism were given in [2]. Every graphon is weakly isomorphic to a graphon on $[0,1]$ (with the Lebesgue measure), and also to a (possibly different) graphon which is twin-free in the sense that $W(x,.)$ and $W(x',.)$ differ on a set of positive measure for all $x \neq x'$.

3 The topology of graphons

3.1 The neighborhood distance

Let (J,W) be a graphon. We can endow the space J with a distance function by

$$r_W(x,y) = \|W(x,.) - W(y,.)\|_1.$$

This function is defined for almost all pairs x,y; we can delete those points from J where $W(x,.) \notin L_1(W)$ (a set of measure 0), to have r_W defined on all pairs. It is clear that r_W is a pre-metric (it is symmetric and satisfies the triangle inequality). We call r_W the *neighborhood distance* on W.

We also define metrics on bigraphons, endowing the spaces J_1 and J_2 with distance functions by

$$r_1(x,y) = \|W(x,.) - W(y,.)\|_1 \qquad (x,y \in J_1),$$
$$r_2(x,y) = \|W(.,x) - W(.,y)\|_1 \qquad (x,y \in J_2).$$

These functions are defined for almost all pairs x,y.

Example 1 Let S^k denote the unit sphere in $\mathbb{R}^{k+1}$, consider the uniform probability measure on it, and let $W(x, y) = 1$ if $x \cdot y \geq 0$ and $W(x, y) = 0$ otherwise. Then (S^k, W) is a graphon, in which the neighborhood distance of two points $a, b \in S^k$ is just their spherical distance (normalized by dividing by π). Furthermore, $1 - 2(W \circ W)(x, y)$ is just the spherical distance of x and y, and from here is is easy to see that the similarity distance is within constant factors of the neighborhood distance.

Example 2 Let (M, d) be a metric space, and let π be a Borel probability measure on M. Assume that the diameter of M is at most 1. Then d can be viewed as a graphon on (M, d). For $x, y \in M$, we have

$$r_d(x, y) = \int\limits_M |d(x, z) - d(y, z)| \, d\pi(z) \leq \int\limits_M d(x, y) \, d\pi(z) = d(x, y),$$

so the identity map $(M, d) \to (M, r_d)$ is contractive. This implies that if (M, d) is compact, and/or finite dimensional (in many senses of dimension), then so is (M, r_d). For most "everyday" metric spaces like (like segments, spheres, or balls) $r_d(x, y)$ can be bounded from below by $\Omega(d(x, y))$, in which case (M, d) and (M, r_d) are homeomorphic.

More generally, if $F : [0, 1] \to [0, 1]$ is a continuous function, then $W(x, y) = F(d(x, y))$ defines a graphon, and the identity map $(M, d) \to (M, r_W)$ is continuous.

Example 3 Finitely forcible graphons, mentioned in the introduction, give interesting examples, for whose details we refer to [13]. One class is stepfuctions (equivalent to finite weighted graphs), which were proved to be finitely forcible by Lovász and Sós [10]; for these, the underlying metric space is finite. Other examples introduced in [13] provide as underlying topologies an interval, the Cantor set, and the one-point compactification of $\mathbb{N}$.

3.2 Pure [bi]graphons

A bigraphon (J_1, J_2, W) is *pure* if (J_i, r_i) is a complete separable metric space and the probability measure has full support (i.e., every open set has positive measure). This definition includes that $r_i(x, y)$ is defined for all $x, y \in J_i$ and $r_i(x, y) > 0$ if $x \neq y$, i.e., the bigraphon has no "twin points". We say that a graphon is *pure*, if the underlying metric probability space is complete, separable and the probability measure has full support.

Theorem 3.1 *Every [bi]graphon is weakly isomorphic to a pure [bi]graphon.*

Remark 3.2 It was shown in [2] that every graphon is weakly isomorphic to a graphon on a standard probability space with no parallel points, which means that for any two points $x, x' \in J$, $W(x, .)$ and $W(x', .)$ differ on a set of positive measure. Lemma 3.4 can be considered as a strengthening of this result.

Proof. We give the proof for bigraphons; the case of graphons is similar. We assume that J_1 and J_2 are standard probability spaces; this can be achieved similarly as for graphons. Let

T_1 be the set of functions $f \in L_1[J_2]$ such that for every L_1-neighborhood U of f, the set $\{x \in J_1 :\ W(x,.) \in U\}$ has positive measure.

Claim 3.3 *For almost every point $x \in J_1$, $W(x,.) \in T_1$.*

Indeed, it is clear that for almost all $x \in J_1$, $W(x,.) \in L_1[J_2]$. Every function $g \in L_1[J_2] \setminus T_1$ has an open neighborhood U_g in $L_1[J_2]$ such that $\pi_1\{x \in J_1 :\ W(x,.) \in U_g\} = 0$. Let $U = \bigcup_{g \notin T_1} U_g$. Since $L_1[J_2]$ is separable, U equals the union of some countable subfamily $\{U_{g_i} :\ i \in \mathbb{N}\}$ and thus $\pi_1\{x \in J_1 :\ W(x,.) \in U\} = 0$. Since if $W(x,.) \notin T_1$ then $W(x,.) \in U$, this proves the Claim.

Clearly T_1 inherits a metric from $L_1[J_2]$, and it is complete and separable in this metric. The functions $W(x,.)$ are everywhere dense in $T_1(W)$ and have measure 1. It also inherits a probability measure π_1' from J_1 through

$$\pi_1'(X) = \pi_1\{x \in \Omega_1 :\ W(x,.) \in X\}.$$

So T_1 is a complete separable metric space with a probability measure on its Borel sets. It also follows from the definition of T_1 that every open set has positive measure.

Define $\widetilde{W} :\ T_1 \times J_2 \to [0,1]$ by $\widetilde{W}(f,y) = f(y)$ for $f \in T_1$ and $y \in J_2$. Then we can replace J_1 by T_1 and W by $\widetilde{W}$, to get a weakly isomorphic graphon. Similarly, we can replace J_2 by T_2. $\square$

We say that two graphons (J,W) and (J',W') are *isometric* if there is an isometric bijection $\phi :\ J \to J'$ that is measure preserving, and $W'(\phi(x),\phi(y)) = W(x,y)$ for almost all $x,y \in J$. The definition for bigraphons is slightly more complicated: two bigraphons (J_1,J_2,W) and (J_1',J_2',W') are *isometric* if there are isometric, measure preserving bijections $\phi_1 :\ J_1 \to J_1'$ and $\phi_2 :\ J_2 \to J_2'$ such that $W'(\phi_1(x),\phi_2(y)) = W(x,y)$ for almost all $(x,y) \in J_1 \times J_2$.

Theorem 3.4 *If two pure [bi]graphons are weakly isomorphic, then they are isometric.*

Proof. We describe the proof for graphons. Theorem 2.1 (a) in [2] says that if two graphons (J,W) and (J',W') are weakly isomorphic, and they have no twins, then one can delete delete 0-sets $S \subseteq J$ and $S' \subseteq J'$ such that there is a bijective measure preserving map $\phi :\ J \setminus S \to J' \setminus S'$ such that $W'(\phi(x),\phi(y)) = W(x,y)$ for almost all $(x,y) \in J \times J$. We may even assume that for every $x \in J \setminus S$, $W'(\phi(x),\phi(y)) = W(x,y)$ holds for almost all y (and vice versa), since this can be achieved by deleting further 0-sets. Clearly ϕ preserves the metric.

We also know that $J \setminus S$ is dense in J (since (J,W) is pure and so its probability measure has full support), and so J is the completion of $J \setminus S$ (and similarly for J'). Hence ϕ extends to an isometry between J and J', which shows that (J,W) and (J',W') are isometric graphons. $\square$

Remark 3.5 Is purity the ultimate normalization of a graphon? There is still some freedom left: we can change the value of W on a symmetric subset of $J \times J$ that intersects every fiber $J \times \{v\}$ in a set of measure. We can take the integral of W (which is a measure ω on J), and then the derivative of ω wherever this exists. This way we get back W almost everywhere, and a well defined value for some further points. What is left undefined is the set of "essential discontinuity" of W (of measure 0). It would be interesting to relate this set to combinatorial properties of W.

3.3 Density functions on pure [bi]graphons

The following technical Lemma will be very useful in the study of r_W and related distance functions.

Lemma 3.6 (a) *Let (J, W) be a graphon, F, a graph, and $S \subseteq V$, an independent set of nodes. Then the function $t = t_S(F, W; .): \ J^S \to \mathbb{R}$ satisfies*

$$|t(x) - t(x')| \le |E| \max_{i \in S} r_W(x_i, x'_i).$$

(b) *Let (J_1, J_2, W) be a bigraphon, let $F = (U_1, U_2, E)$ be a bigraph, and let $S_i \subseteq U_i$ be such that no edge connects S_1 to S_2. Then the function $t = t^{\mathsf{b}}_{S_1,S_2}(F, W, .): J_1^{S_1} \times J_2^{S_2} \to \mathbb{R}$ satisfies*

$$|t(x, y) - t(x', y')| \le |E| \max\{\max_{i \in S_1} r_1(x_i, x'_i), \max_{j \in S_2} r_2(y_j, y'_j)\}.$$

Remark 3.7 (i) It follows that the functions t in (a) and (b) are Lipschitz (and hence continuous).

(ii) In both parts (a) and (b) of the Lemma, the graph F could have multiple edges.

Proof. We describe the proof of (a); the proof of (b) is similar. For each $i \in U \setminus S$, let $x_i = x'_i$ be a variable. Let $E = \{u_1v_1, \dots u_mv_m\}$, where we may assume that $v_i \in U \setminus S$. Then

$$t(x) - t(x') = \int\limits_{J^{U \setminus S}} \prod_{i=1}^m W(x_{u_i}, x_{v_i}) \, dy - \int\limits_{J^{U \setminus S}} \prod_{i=1}^m W(x'_{u_i}, x'_{v_i}) \, dy$$

$$= \sum_{j=1}^m \int\limits_{J^{U \setminus S}} \prod_{i<j} W(x_{u_i}, x_{v_i})(W(x_{u_j}, x_{v_j}) - W(x'_{u_j}, x'_{v_j})) \prod_{j>i} W(x'_{u_i}, x'_{v_i}), dy$$

and hence

$$|t(x) - t(x')| \le \sum_{j=1}^m \int\limits_{J^{U \setminus S}} |W(x_{u_j}, x_{v_j}) - W(x'_{u_j}, x'_{v_j})| \, dy.$$

By the assumption that $v_i \in U \setminus S$, we have $x_{v_j} = x'_{v_j}$ for every j, and so

$$|t(x) - t(x')| \le \sum_{j=1}^m r_W(x_{u_j}, x'_{u_j}) \le |E| \max_{1 \le i \le k} r_W(x_i, x'_i),$$

which proves the assertion. □

Lemma 3.6 has an important corollaries for pure graphons, which are closely related to Lemma 2.8 in [13]. We do not formulate all versions, just a few that we need.

Corollary 3.8 *Let (J, W) be a pure graphon, and let F be a graph and let $S \subseteq V$, where S is independent. Then $t_S(F, W; x)$ is a continuous function of $x \in J^S$.*

Applying this when F is a path of length 2, we get:

Corollary 3.9 *For every pure graphon (J,W), $W \circ W$ is a continuous function on J.*

Another application of Corollary 3.8 gives:

Corollary 3.10 *Let (J,W) be a pure graphon, and let $F_1,\dots,F_m$ be graphs whose node set contains a common set S, which is independent in each. Let $T \subseteq S$, and let $a_1,\dots,a_m$ be real numbers. Let $x \in J^T$, and assume that the equation*

$$\sum_{i=1}^{m} a_i t_S(F_i,W;x,y) = 0 \tag{1}$$

holds for almost all $y \in J^{S\setminus T}$. Then it holds for all $y \in J^{S\setminus T}$.

Proof. By Corollary 3.8, the left hand side of (1) is a continuous function of (x,y), and so it remains a continuous function of y if we fix x. Hence the set where it is not 0 is an open subset of $J^{S\setminus T}$. Since the graphon is pure, it follows that this set is either empty of has positive measure. □

We formulate one similar corollary for bigraphons.

Corollary 3.11 *Let (J_1,J_2,W) be a pure bigraphon, and let $F_1,\dots,F_m$ be bigraphs with the same bipartition classes U_1 and U_2. Let $a_1,\dots,a_m$ be real numbers. Assume that the equation*

$$\sum_{i=1}^{m} a_i t^{\mathsf{b}}_{U_1}(F_i,W;x) = 0 \tag{2}$$

holds for almost all $x \in J_1^{U_1}$. Then it holds for all $x \in J_1^{U_1}$.

3.4 The similarity distance

It turns out (it was already noted in [12]) that the distance function $r_{W\circ W}$ defined by the operator square of W is also closely related to combinatorial properties of a graphon. We call this the *similarity distance* (for reasons that will become clear later). In explicit terms, we have

$$\begin{aligned} r_{W\circ W}(a,b) &= \int_J \Big| \int_J W(a,y)W(y,x)\,dy - \int_J W(b,y)W(y,x)\,dy \Big|\,dx \\ &= \int_J \Big| \int_J W(x,y)\big(W(y,a)-W(y,b)\big)\,dy \Big|\,dx. \end{aligned} \tag{3}$$

Remark 3.12 Let $\mathbf{X},\mathbf{Y},\mathbf{Z}$ be independent uniform random points from J, then we can rewrite the definitions of these distances as

$$r_W(a,b) = \mathsf{E}_{\mathbf{X}}|W(\mathbf{X},a)-W(\mathbf{X},b)|, \tag{4}$$

$$r_{W\circ W}(a,b) = \mathsf{E}_{\mathbf{X}}\big|\mathsf{E}_{\mathbf{Y}}(W(\mathbf{X},\mathbf{Y})(W(\mathbf{Y},a)-W(\mathbf{Y},b)))\big|. \tag{5}$$

This formulation shows that this distance can be computed with arbitrary precision from a bounded size sample. We do not go into the details of this.

Lemma 3.13 *If (J, W) is a pure graphon, then the similarity distance $r_{W \circ W}$ is a metric.*

So $(J, r_{W\circ W})$ is a metric space, and hence Huasdorff. We will show later that it is always compact.

Proof. The only nontrivial part of this lemma is that $r_{W\circ W}(x, y) = 0$ implies that $x = y$. The condition $r_{W\circ W}(x, y) = 0$ implies that for almost all $u \in J$ we have $(W \circ W)(x, u) = (W \circ W)(y, u)$, or more explicitly

$$\int\limits_J (W(x, z) - W(y, z))W(z, u)\, dz = 0.$$

Using that (J, W) is pure, Corollary 3.11 implies that this holds for every $u \in J$. in particular, it holds for $u = x$ and $u = y$. Taking the difference, we get that

$$\int\limits_J (W(x, z) - W(y, z))(W(z, x) - W(z, y))\, dz = 0,$$

and hence $W(x, z) = W(y, z)$ almost everywhere. Using again that (J, W) is pure, we get that $x = y$. □

For every $x \in J$, the function $W(x, .)$ is in $L_\infty(J)$, and hence the weak topology of $L_1(J)$ gives a topology on J. It is well known that when restricted to $L_\infty(J)$, this topology is the weak-$*$ topology on $L_\infty(J)$, and hence it is metrizable, and the unit ball of $L_\infty(J)$ is compact in it (Alaoglu's Theorem). A sequence of points (x_n) is convergent in this topology if and only if

$$\int_A W(x_n, y)\, dy \to \int_A W(x, y)\, dy$$

for every measurable set $A \subseteq J$. We call this the *weak topology* on J. We need this name only temporarily, since we are going to show that $r_{W\circ W}$ gives a metrization of the weak topology.

Theorem 3.14 *For any pure graphon, the metric $r_{W\circ W}$ defines exactly the weak topology.*

Proof. First we show that the weak topology is finer than the topology of $(J, r_{W\circ W})$. Suppose that $x_n \to x$ in the weak topology, and consider

$$r_{W\circ W}(x_n, x) = \int\limits_J \Big| \int\limits_J \big(W(x_n, y) - W(x, y)\big)W(y, z)\, dy \Big|\, dz.$$

Here the inner integral tends to 0 for every z, by the weak convergence $x_n \to x$. Since it also remains bounded, it follows that the outer integral tends to 0. This implies that $x_n \to x$ in $(J, r_{W\circ W})$.

From here, the equality of the two topologies follows by general arguments: the weak topology is compact, and the coarser topology of $r_{W\circ W}$ is Hausdorff, which implies that they are the same. □

Corollary 3.15 *For every pure graphon (J, W), the space $(J, r_{W\circ W})$ is compact.*

To compare the topology of (J, r_W) with these, note that for any two points $x, y \in J$, we have

$$r_{W\circ W}(x,y) \le r_W(x,y), \tag{6}$$

which implies that the topology of (J, r_W) is finer than the topology of $(J, r_{W\circ W})$.

3.5 Compact Graphons

Graphons for which the finer space (J, r_W) is also compact seem to have a special importance in combinatorics. Let us call such a graphon a **compact graphon**.

Proposition 3.16 *A pure graphon (J, W) is compact if and only if (J, r_W) and $(J, r_{W\circ W})$ define the same topologies.*

Proof. If the topologies (J, r_W) and $(J, r_{W\circ W})$ are the same, then (J, r_W) is compact by Corollary 3.15. Conversely, if (J, r_W) is compact then, by the argument used before in the proof of Theorem 3.14, the coarser Hausdorff topology of $(J, r_{W\circ W})$ must be the same. □

Example 4 Let $J = [0, 1]$, $f(y) = \lfloor \log(1/y) \rfloor$, and define

$$W(x,y) = \begin{cases} x_{f(y)}, & \text{if } x > 1/2 \text{ and } y \le 1/2, \\ y_{f(x)}, & \text{if } x \le 1/2 \text{ and } y > 1/2, \\ 0, & \text{otherwise,} \end{cases}$$

where $x = 0.x_1x_2\ldots$ and $y = 0.y_1y_2\ldots$ are the binary expansions of x and y, respectively. Then selecting one point from each interval $[2^{-k+1}, 2^{-(k)}]$, we get an infinite number of points in $([0, 1], r_2)$ mutually at distance $1/4$, so (J, W_r) is not compact, but by Corollary 3.15, $(J, r_{W\circ W})$ is compact. So the two topologies are different.

We conclude this section with an observation relating the topology of J to spectral theory.

Lemma 3.17 *Let (J, W) be a pure graphon. Then every eigenfunction $f \in L_2(J)$ of W as a kernel operator belonging to a nonzero eigenvalue is continuous in the metric $r_{W\circ W}$ (and therefore also in r_W).*

Proof. It suffices to prove that f is continuous in (J, r_W), since we can apply the argument to the graphon $(J, W \circ W)$, which also has f as an eigenvector.

First, we have

$$|f(x)| = \frac{1}{|\lambda|}\left|\int_J W(x,y)f(y)\,dy\right| \le \frac{1}{|\lambda|}\|f\|_1 \le \frac{1}{|\lambda|}\|f\|_2,$$

and so f is bounded. We know by Corollary 3.9 that $W \circ W$ is continuous in (J, r_W), and hence so is

$$f = \frac{1}{\lambda^2}\int_J (W\circ W)(x,y)f(y)\,dy.$$

□

4 Thin graphons

4.1 The main theorem

We say that a bigraphon W is *thin* if there is a bigraph F such that $t^{\mathsf{b}}_{\rm ind}(F,W)=0$. Trivially, if W is thin, then so is its complementary bigraphon $1-W$.

We call a graphon *thin* if it is thin as a bigraphon. (Note: for this, it is not enough to require $t_{\rm ind}(F,W)=0$ for some bipartite graph F. For example, consider the graphon $U:\ [0,1]^2\to[0,1]$ defined by $U(x,y)=U(y,x)=1/2$ if $x\in[0,1/2]$ and $y\in(1/2,1]$, and $U(x,y)=1$ otherwise. As a bigraphon, this is not thin, but satisfies $t_{\rm ind}(F,W)=0$ for every bigraph with at least 3 nodes in one of the classes.

The *(upper) packing dimension* of a metric space (M,d) is defined as

$$\limsup_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log(1/\varepsilon)},$$

where $N(\varepsilon)$ is the maximum number of points in M mutually at distance at least ε. So this dimension is finite if and only if there is a $d\ge 0$ such that every set of points mutually at distance at least ε has at most ε^{-d} elements. It is easy to see that we could use instead of $N(\varepsilon)$ the minimum number of sets of diameter at most ε covering the space.

Our main goal is to prove:

Theorem 4.1 *If a pure bigraphon* (J_1,J_2,W) *is thin, then* (a) $W(x,y)\in\{0,1\}$ *almost everywhere,* (b) J_1,J_2 *are compact, and* (c) J_1,J_2 *have finite packing dimension.*

Remark 4.2 The proof will show that if $t_{\rm ind}(F,W)=0$ for a bigraph F with k nodes, then the packing dimension of J_i is bounded by $10|F|$.

Before giving the proof, we describe a class of examples, and then recall some facts about the Vapnik-Červonenkis dimension.

Example 5 Let V be a finite or countable set, π, a probability measure on V, and define $J_1=[0,1]^V$, $J_2=[0,1]\times V$, with the power measure μ_1 on J_1 and the product measure μ_2 on J_2. We define a bigraphon on $J_1\times J_2$ by

$$W(x,y)=\mathbf{1}_{t\le x_i}$$

for $x=(x_i:\ i\in S)$ and $y=(t,i)$. We can metrize this bigraphon by

$$r_1(x,x')=\sum_{i\in V}\pi(i)|x_i-x'_i|$$

for $x=(x_i:\ i\in S),\ x'=(x'_i:\ i\in S)\in J_1$, and

$$r_2(y,y')=\begin{cases}|t-t'| & \text{if } i=1',\\ t+t'-2tt' & \text{otherwise.}\end{cases}$$

for $y=(t,i),\ y'=(t',i')\in J_2$.

If V is finite, then (J_1, r_1) has dimension $|V|$, while (J_2, r_2) has dimension 1, and both are compact. These facts also follow if we observe that W is thin. Indeed, if F denotes the matching with $|V|+1$ edges, then $t^{\mathsf{b}}_{\text{ind}}(F, W) = 0$, since among any $|V|+1$ points in J_2, there are two points of the form $y = (t, i)$ and $y' = (t', i)$ with $t < t'$, and then $W(., (t, i)) \geq W(., (t', i))$.

If V is infinite, then (J_1, r_1) is infinite dimensional but compact, while (J_2, r_2) is not compact.

Example 6 Let $J_1 = J_2 = [0, 1]$, and let $W(x, y) = x_{f(y)}$, where $x = 0.x_1x_2\dots$ is the binary expansion of x, and $f(y) = \lceil \log(1/y) \rceil$. Then for $x = 0.x_1x_2\dots$ and $x' = 0.x'_1x'_2\dots$ we have $r_1(x, x') = \sum_{k=1}^{\infty} 2^{-k}|x_k - x'_k|$, and from here is is easy to see that $([0, 1], r_1)$ is compact. Furthermore, if $S \subseteq [0, 1]$ is a set of points mutually more than 2^{-n} apart, then any two elements of S must differ in one of their first n digits, and so their number is at most 2^n. Hence the packing dimension of $([0, 1], r_1)$ is 1.

On the other hand, selecting a point $y_k \in [2^{-k}, 2^{-(k-1)}]$, we get an infinite number of points in $([0, 1], r_2)$ mutually at distance 1/2, so this space is not compact and infinite dimensional.

4.2 Vapnik-Červonenkis dimension

For any set V and family of subsets $\mathcal{H} \subseteq 2^V$, a set $S \subseteq V$ is called *shattered*, if for every $X \subseteq S$ there is a $Y \in \mathcal{H}$ such that $X = Y \cap S$. The *Vapnik-Červonenkis dimension* or *VC-dimension* $\dim_{\text{VC}}(\mathcal{H})$ of a family of sets is the supremum of cardinalities of shattered sets [16]. For us, k will be always finite.

Let V be a probability space and $\mathcal{H}$, a family of measurable subsets of V. A finite subfamily $\mathcal{H}'$ is *qualitatively independent* if all the $2^{|\mathcal{H}'|}$ atoms of the set algebra they generate have positive measure. The *dual essential Vapnik-Červonenkis dimension*, or briefly *DE-dimension*, of $\mathcal{H}$ is a supremum of all cardinalities of qualitatively independent subfamilies of $\mathcal{H}$.

We recall two basic facts about VC-dimension:

Lemma 4.3 (Sauer-Shelah Lemma) *If a family $\mathcal{H}$ of subsets of an m-element set has VC-dimension k, then*

$$|\mathcal{H}| \leq 1 + m + \cdots + \binom{m}{k}.$$

For a family $\mathcal{H}$ of sets, we denote by $\tau(\mathcal{H})$ the minimum cardinality of a set meeting every member of $\mathcal{H}$. The following basic fact about VC-dimension was proved by Komlós, Pach and Woeginger [9], based on the results of Vapnik and Červonenkis [16] (we do not state it in its sharpest form):

Theorem 4.4 *Let J be a probability space and, $\mathcal{H}$ a family of measurable subsets of J such that every $A \in \mathcal{H}$ has measure at least ε. Suppose that $\mathcal{H}$ has finite VC-dimension k. Then*

$$\tau(\mathcal{H}) \leq 8k\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}.$$

We need a couple of further facts. For a family $\mathcal{H}$ of sets, let $\mathcal{H}(\triangle)\mathcal{H} = \{A\triangle B : \ A, B \in \mathcal{H}\}$.

Lemma 4.5 *For every family of sets,* $\dim_{\text{VC}}(\mathcal{H}(\triangle)\mathcal{H}) \leq 10 \dim_{\text{VC}}(\mathcal{H})$.

Proof. Set $k = \dim_{\mathrm{VC}}(\mathcal{H})$. Let S be a subset of $V = \cup\mathcal{H}$ with m elements that is shattered by $\mathcal{H}(\triangle)\mathcal{H}$. Then every $X \subseteq S$ arises as $X = (A\triangle B) \cap S$, where $A, B \in \mathcal{H}$. Since $(A\triangle B) \cap S = (A \cap S)\triangle(B \cap S)$, the number of different sets of the form $A \cap S$ is at least $2^{m/2}$. By the Sauer-Shelah Lemma, this implies that

$$2^{m/2} \le 1 + m + \cdots + \binom{m}{k},$$

whence $m \le 10k$ follows by standard calculation. □

Lemma 4.6 *Let $\mathcal{H}$ be a family of measurable sets in a probability space with VC-dimension k such that $\pi(A\triangle B) \ge \varepsilon$ for all $A, B \in \mathcal{H}$. Then $|\mathcal{H}| \le (80k)^k\varepsilon^{-20k}$.*

Proof. Consider the family $\mathcal{H}' = \mathcal{H}(\triangle)\mathcal{H}$. Every $A \in \mathcal{H}'$ has $\pi(A) \ge 1/\varepsilon$, and $\dim_{\mathrm{VC}}(\mathcal{H}') \le 10k$ by Lemma 4.5. Hence by Theorem 4.4, we have

$$\tau(\mathcal{H}') \le 80k\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}.$$

Let $S \subseteq \cup\mathcal{H}$ be a set of size $\tau(\mathcal{H}')$ meeting every symmetric difference $A\triangle B$ $(A, B \in \mathcal{H})$. Then the sets $S \cap A$, $A \in \mathcal{H}$ are all different. By the Sauer-Shelah Lemma, this implies that

$$|\mathcal{H}| \le 1 + |S| + \cdots + \binom{|S|}{10k} < |S|^{10k} \le \left(80k\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}\right)^{10k} < (80k)^{10k}\varepsilon^{-20k}.$$

□

4.3 VC-dimension and graphons

Lemma 4.7 *Let (J_1, J_2, W) be a pure 0-1 valued bigraphon. Then W is thin if and only if the DE-dimension of the family $\mathcal{R}_W = \{\mathrm{supp}(W(x, .)) :\ x \in T_1\}$ is finite.*

Proof. Suppose that this dimension is infinite. We claim that $t^{\mathsf{b}}_{\mathrm{ind}}(F, W) > 0$ for every bipartite graph $F = (U, U', E)$. Let $S \subseteq J_1$ be a set such that the subfamily $\{\mathrm{supp}(W(x, .)) :\ x \in T_1\}$ is qualitatively independent. To each $i \in U$, assign a value $x_i \in S$ bijectively. By Corollary 3.11, the set of points $y \in J_2$ such that $\mathrm{supp}(W(., y)) \cap S = \{x_i :\ i \in N(j)\}$ has positive measure for each $j \in U'$. Hence $t^{\mathsf{b}}_{\mathrm{ind}}(F, W) > 0$.

Conversely, suppose that $k = \dim(\mathcal{R}_W)$ is finite. Let F denote the bipartite graph with $k+1$ nodes in one class U and 2^{k+1} nodes in the other class U', in which the nodes in U' have all different neighborhoods. Then $t^{\mathsf{b}}_{\mathrm{ind}}(F, W) = 0$. □

Remark 4.8 The proof above in fact gives the following quantitative result: $t^{\mathsf{b}}_{\mathrm{ind}}(F, W) = 0$ for some bigraph F with k nodes in its smaller bipartition class if and only if $\dim_{\mathrm{DE}}(\mathcal{R}_W) < k$.

Proof of Theorem 4.1. We may assume that W is pure.

(a) Suppose that the bigraph $F=(U_1,U_2,E)$ satisfies $t^{\mathsf{b}}_{\text{ind}}(F,W)=0$. Then for almost all $x\in J_1^{U_1}$, we have $t^{\mathsf{b}}_{U_1,\text{ind}}(F,W;x)=0$. By Corollary 3.11, it follows that $t^{\mathsf{b}}_{U_1,\text{ind}}(F,W;x)=0$ for every x. In particular, $t^{\mathsf{b}}_{U_1,\text{ind}}(F,W;x_0,\dots,x_0)=0$ for all $x_0\in J_1$. But for this substitution,

$$t^{\mathsf{b}}_{U_1,\text{ind}}(F,W;x_0,\dots,x_0)=\int\limits_{J_2^{V_2}}\prod_{j\in J_2}W(x_0,y_j)^{d_F(j)}(1-W(x_0,y_j))^{|U_1|-d_F(j)},$$

and so for every x_0 we must have $W(x_0,y_0)\in\{0,1\}$ for almost all y_0.

(b) By Theorem 3.16 it suffices to prove that if $W(x_n,.)$, $n=1,2,\dots$ weakly converges to f, i.e.,

$$\lim_{n\to\infty}\int\limits_S W(x_n,y)\,dy\to\int\limits_S f(y)\,dy$$

for every measurable set $S\subseteq J_2$, then it is also convergent in L_1.

Claim 4.9 *The weak limit function f is almost everywhere* 0-1 *valued.*

Suppose not, then there is an $\varepsilon>0$ and a set $Y\subseteq J_2$ with positive measure such that $\varepsilon\le f(x)\le 1-\varepsilon$ for $x\in Y$. Let $S_n=\operatorname{supp}(W(x_n,.))\cap Y$. We select, for every $k\ge 1$, k indices $n_1,\dots n_k$ so that the Boolean algebra generated by $S_{n_1},\dots S_{n_k}$ has 2^k atoms of positive measure. If we have this for some k, then for every atom A of the boolean algebra

$$\lambda(A\cap S_n)=\int\limits_A W(x,y_n)\,dx\longrightarrow\int\limits_A f(x)\,dx\qquad(n\to\infty),$$

and so if n is large enough then

$$\frac{\varepsilon}{2}\lambda(A)\le\lambda(A\cap S_n)\le\Bigl(1-\frac{\varepsilon}{2}\Bigr)\lambda(A).$$

If n is large enough, then this holds for all atoms A, and so S_n cuts every previous atom into two sets with positive measure, and we can choose $n_{k+1}=n$.

But this means that the DE-dimension of the supports of the $W(x,.)$ is infinite, contradicting Lemma 4.7. This proves Claim 4.9.

So we know that $f(x)\in\{0,1\}$ for almost all x, and hence

$$\|f-W(.,y_n)\|_1=\int\limits_{\{f=1\}}(1-W(x,y_n))\,dx+\int\limits_{\{f=0\}}W(x,y_n)\,dx\longrightarrow 0.$$

Thus $W(.,y_n)\to f$ in L_1, which we wanted to prove.

(c) Let $F=(U_1,U_2,E)$ be a bigraph such that $t^{\mathsf{b}}_{\text{ind}}(F,W)=0$, and let $U_i=[k_i]$. We show that the packing dimension of J_1 is at most $10k_2$. To this end, we show that if any two elements of a finite set $Z\subseteq J_1$ are at a distance at least ε, then $|Z|\le c(k)\varepsilon^{-2k_2}$. Let $\mathcal{H}=\{\operatorname{supp}(W(x,.)):\ x\in Z\}$, then

$$\pi_2(X\triangle Y)\ge\varepsilon \tag{7}$$

for any two distinct sets $X,Y\in\mathcal{H}$.

Let A be the union of all atoms of the set algebra generated by $\mathcal{H}$ that have measure 0. Clearly A itself has measure 0, and hence the family $\mathcal{H}' = \{X \setminus A : \ X \in \mathcal{H}\}$ still has property (7).

We claim that $\mathcal{H}'$ has VC-dimension less than k_2. Indeed, suppose that $J_2 \setminus A$ contains a shattered set S with $|S| = k_2$. To each $j \in U_2$, assign a point $q_j \in S$ bijectively. To each $i \in U_1$, assign a point $p_i \in Z$ such that $q_j \in \text{supp}(W(p_i, .))$ if and only if $ij \in E$. This is possible since S is shattered. Now fixing the p_i, for each j there is a subset of J_2 of positive measure whose points are contained in exactly the same members of $\mathcal{H}'$ as q_j, since $q_j \notin A$. This means that the function $t = t^{\mathsf{b}}_{J_1,\text{ind}}(F, W; .) : \ V_1^{J_1} \to \mathbb{R}$ satisfies $t(p) > 0$. Corollary 3.11 implies that $t(x) > 0$ for a positive fraction of the choices of $x \in J_1^{V_1}$, and hence $t^{\mathsf{b}}_{\text{ind}}(F, W) > 0$, a contradiction.

Applying Lemma 4.6 we conclude that $|Z| = |\mathcal{H}| \leq (80k_2)^{10k_2}\varepsilon^{-20k_2}$. □

4.4 Hereditary properties and thin bigraphons

A graph property $\mathcal{P}$ is a class of finite graphs closed under isomorphism. The property is called *hereditary*, if whenever $G \in \mathcal{P}$, then every induced subgraph is also in $\mathcal{P}$.

Let $\mathcal{P}$ be any graph property. We denote by $\overline{\mathcal{P}}$ its *closure*, i.e., the class of graphons (J, W) that arise as limits of graph sequences in $\mathcal{P}$. For every graphon W, let $\mathcal{I}(W)$ denote the set of those graphs F for which $t_{\text{ind}}(F, W) > 0$. Clearly, $\mathcal{I}(W)$ is a hereditary graph property.

Let $\mathcal{P}$ be a hereditary property of graphs. Then

$$\cup_{W \in \overline{\mathcal{P}}} \mathcal{I}(W) \subseteq \mathcal{P}. \tag{8}$$

Indeed, if $F \notin \mathcal{P}$, then $t_{\text{ind}}(F, G) = 0$ for every $G \in \mathcal{P}$, since $\mathcal{P}$ is hereditary. This implies that $t_{\text{ind}}(F, W) = 0$ for all $W \in \overline{\mathcal{P}}$, and so $F \notin \mathcal{I}(W)$.

Equality does not always hold in (8). For example, we can always add a bigraph G and all its induced subgraphs to $\mathcal{P}$ without changing $\overline{\mathcal{P}}$. As a less trivial example, consider all bigraphs with degrees bounded by 10. This property is hereditary, and $\overline{\mathcal{P}}$ consists of a single bigraphon (the identically 0 function).

Proposition 4.10 *For a hereditary property $\mathcal{P}$ of graphs equality holds in* (8) *if and only if for every graph $G \in \mathcal{P}$ and $v \in V(G)$, if we add a new node v' and connect it to all neighbors of v, then at least one of the two graphs obtained by joining or not joining v and v' has property $\mathcal{P}$.*

Proof. Suppose that this condition holds. Let $F \in \mathcal{P}$ have n nodes, and let $F(k)$ denote a graph in $\mathcal{P}$ obtained from F by a repetition of this operation so that each original node has k copies. Then $t_{\text{ind}}(F, F(k)) \geq 1/n^n$. Let W be the limit graphon of some subsequence of the $F(k)$ $(k \to \infty)$, then $W \in \overline{\mathcal{P}}$. Furthermore, clearly $t_{\text{ind}}(F, W) > 0$, and so $F \in \mathcal{I}(W)$.

Conversely, assume that $F = (V, E) \in \mathcal{I}(W)$ for some $W \in \overline{\mathcal{P}}$, so that $t_{\text{ind}}(F, W) > 0$. Let F' and F'' be the two graphs obtained from F by doubling a node v ($vv' \notin E(F')$, but $vv' \in E(F'')$), then

$$\int\limits_{J^V} t_{\text{ind}}(F, W; x)\, dx > 0$$

implies that there is a positive measure of choices for the values of x_u $(u \in V(F) \setminus v)$, for which the set X of the choices of x_v with $t_{\rm ind}(F, W; x) > 0$ has positive measure. Clearly either $W(x, y) < 1$ for a positive measure of choices of $(x, y) \in Y$ or this holds for $W(x, y) > 0$. One or the alternative, say the first one, holds for a positive measure of choices for the values of x_u $(u \in V(F) \setminus v)$. But then $t(F', W) > 0$. □

All of the above notions and simple facts extend to bigraphs and bigraphons trivially.

Let us turn to thin graphons and bigraphons. The significance of thin bigraphons is supported by the following observation:

Proposition 4.11 *Let $\mathcal{P}$ be a hereditary bigraph property that does not contain all bigraphs. Then every bigraphon in its closure is thin.*

Proposition 4.11 and Theorem 4.1 imply:

Corollary 4.12 *Let $\mathcal{P}$ be a hereditary bigraph property that does not contain all bigraphs. Then for every pure bigraphon $(J_1, J_2, , W)$ in its closure, W is 0-1 valued almost everywhere, and J_1 and J_2 are compact and their dimension is bounded by a finite number depending on $\mathcal{P}$ only.*

By this corollary, we can define, for every nontrivial hereditary property of bigraphs, a finite dimension. It would be interesting to find further combinatorial properties of this dimension.

The natural analogue of this corollary for graph properties fails to hold.

Example 7 Let $\mathcal{P}$ be the property of a graph that it is triangle-free. Then every bipartite graphon is in its closure, but such graphons need not be 0-1 valued, and their topology need not be finite dimensional or compact.

However, if we include the (seemingly) simplest of the conclusions of Corollary 4.12 as a hypothesis, then we can extend it to all graphs. A graph property $\mathcal{P}$ is *random-free*, if every $W \in \overline{\mathcal{P}}$ is 0-1 valued almost everywhere.

Theorem 4.13 *Let $\mathcal{P}$ be a hereditary random-free graph property. Then for every pure graphon (J, W) in its closure, J is compact and finite dimensional.*

Before proving this theorem, we need some preparation.

Lemma 4.14 *For a hereditary graph property $\mathcal{P}$, the following are equivalent:*

(i) *$\mathcal{P}$ is random-free;*

(ii) *there is a bigraph F such that $t^{\mathsf{b}}(F, W) = 0$ for all $W \in \overline{\mathcal{P}}$;*

(iii) *there is a bipartite graph F with bipartition (U_1, U_2) such that no graph obtained from F by adding edges within U_1 and U_2 has property $\mathcal{P}$.*

Proof. (i)⇒(iii): Assume that (iii) does not hold, then for every bigraph F there is a graph $\hat{F} \in \mathcal{P}$ and a partition $V(\hat{F}) = \{U_1(\hat{F}), U_2(\hat{F})\}$ such that the bigraph between $U_1(\hat{F})$ and $U_2(\hat{F})$ is isomorphic to F. We want to show that $\mathcal{P}$ is not random-free.

Let $(F_1, F_2, \dots)$ be a quasirandom sequence of bigraphs with edge density $1/2$, with the same number of nodes in each bipartition class. Consider the graphs $\hat{F}_n$, and let F'_n and F''_n denote the subgraphs of $\hat{F}_n$ induced by $U_1(\hat{F}_n)$ and $U_2(\hat{F}_n)$, respectively. By selecting a subsequence we may assume that the graph sequences $(F'_1, F'_2, \dots)$ $(F''_1, F''_2, \dots)$ are convergent. By Lemma 4.16 in [5], we can order the nodes of F'_n so that $W_{F'_n}$ converges to a graphon $([0,1], W')$ in the cut norm $\|.\|_\square$, and similarly, $W_{F''_n}$ converges to a graphon $([0,1], W'')$ in the cut norm. We order the nodes of $\hat{F}_n$ so that the nodes in F'_n preceed the nodes of F''_n, and keep the above ordering otherwise. Then trivially $W_{\hat{F}_n}$ converges to the graphon

$$U(x,y) = \begin{cases} W'(2x, 2y) & \text{if } x, y < 1/2, \\ W''(2x-1, 2y-1) & \text{if } x, y > 1/2, \\ 1/2 & \text{otherwise.} \end{cases}$$

So $U \in \overline{\mathcal{P}}$ is not 0-1 valued, and so $\mathcal{P}$ is not random-free.

(ii)$\Rightarrow$(i): Suppose that $\mathcal{P}$ is not random-free, and let $(J, W) \in \overline{\mathcal{P}}$ be a graphon that is not 0-1 valued almost everywhere. Then by Theorem 4.1, it is not thin as a bigraphon, which means that for every bigraph $F = (U_1, U_2, E)$, $t^{\mathsf{b}}_{\text{ind}}(F, W) > 0$, so (ii) is not satisfied.

(iii)$\Rightarrow$(ii): Consider a bigraph $F = (U_1, U_2, E)$ as in (iii), and consider it as a bipartite graph on $V = U_1 \cup U_2$ (we assume that $U_1 \cap U_2 = \emptyset$). Suppose that it does not satisfy (ii), then there is a graphon $W \in \overline{\mathcal{P}}$ such that $t(F, W; x) > 0$ for a positive measure of choices of the $x \in J^V$. For every such choice, we define a graph F' by connecting those pairs $\{i, j\}$ of nodes of F for which $W(x_i, x_j) > 0$ and either $i, j \in U_1$ or $i, j \in U_2$. The same supergraph F' will occur for a positive measure of choices of the x_i, and for this F' we have $t_{\text{ind}}(F', W) > 0$, so using (8), we get $F' \in \mathcal{I}(W) \subseteq \mathcal{P}$, a contradiction. $\square$

Proof of Theorem 4.13. By Lemma 4.14, there is a bigraph F such that $t^{\mathsf{b}}(F, W) = 0$ for all $W \in \overline{\mathcal{P}}$. Thus Theorem 4.1 implies the assertion. $\square$

5 Regularity partitions

5.1 Weak and strong regularity partitions

The Regularity Lemma of Szemerédi [14, 15], and various weaker and stronger versions of it are basic tools in the study of large graphs and graphons [12]. Our goal is to show that it is also closely related to the topology of graphons.

Let (J, W) be a graphon and $\mathcal{P}$, a partition of J into measurable sets with positive measure. For $x \in J$, let $S(x)$ denote the partition class containing x. Define

$$f_{\mathcal{P}}(x) = \frac{1}{\pi(S(x))} \int\limits_{S(x)} f(x)\, dx$$

for a function $f \in L_1(J)$, and

$$W_{\mathcal{P}}(x,y) = \frac{1}{\pi(S(x))\pi(S(y))} \int\limits_{S(x)\times S(y)} W(x,y)\, dx.$$

We say that $\mathcal{P}$ is a *weak regularity partition* with error ε, if $\|W - W_{\mathcal{P}}\|_\square \le \varepsilon$.

We define a *Szemerédi partition* of a graphon with error ε as a partition $\mathcal{P} = \{S_1 \cup \dots \cup S_k\}$ of J into measurable sets such that

$$|\langle W - W_{\mathcal{P}}, H\rangle| \le \varepsilon \tag{9}$$

for every function $H: \ J \times J \to [0,1]$ that is 0-1 valued and whose support is the union of product sets $R_{ij} = R'_{ij} \times R'_{ij} \subseteq S_i \times S_j$ $(i, j \in [k])$. To relate this to the weak partitions, we note that $\|W - W_{\mathcal{P}}\|_\square \le \varepsilon$ can be expressed as (9) for all functions h of the form $\mathbf{1}_{S\times T}$. (The formulation above is not a direct generalization of Szemerédi's definition, but it is closest in our setting; cf. [12].)

A *strong regularity partition* of a graph was introduced by Alon, Fischer, Krivelevich and M. Szegedy [1]. Here the error is specified by an infinite sequence $\mathcal{E} = (\varepsilon_0, \varepsilon_1, \dots)$ of positive numbers. Again recasting it in our setting, $\mathcal{P}$ is a strong regularity partition with error $\mathcal{E}$ of a graphon (J, W) if there is a graphon (J, U) such that

$$\|W - U\|_1 \le \varepsilon_0 \qquad \text{and} \qquad \|U - W_{\mathcal{P}}\|_\square \le \varepsilon_{|\mathcal{P}|}.$$

Even stronger would be, of course, to require that $\|W - W_{\mathcal{P}}\|_1 \le \varepsilon$ (equivalently, (9) holds for all measurable functions $H: \ J \times J \to [-1,1]$). In this case we call $\mathcal{P}$ an *ultra-strong regularity partition* with error ε.

The following result is a graphon version of the original Szemerédi's Regularity Lemma [14, 15], its "weak" form due to Frieze and Kannan [8], and its strong form due to Alon, Fischer, Krivelevich and M. Szegedy [1]. It was proved for graphons in [12].

Theorem 5.1 *Let (J, W) be a graphon on an atomfree probability space. Then*

(a) *for every $\varepsilon > 0$ (J, W) has a Szemerédi partition with error ε into no more than $T(\varepsilon)$ classes, where $T(\varepsilon)$ depends only on ε;*

(b) *for every $\varepsilon > 0$ (J, W) has a weak regularity partition with error ε into no more than $2^{2/\varepsilon^2}$ classes.*

(c) *for every sequence $\mathcal{E} = (\varepsilon_0, \varepsilon_1, \dots)$ of positive numbers, (J, W) has a strong regularity partition of (J, W) with error $\mathcal{E}$ into no more than $T(\mathcal{E})$ classes, where $T(\mathcal{E})$ depends only on $\mathcal{E}$.*

Remark 5.2 (i) We note that every graphon has an ultra-strong partition with error ε by standard results in analysis, but the number of classes cannot be bounded uniformly by any function of ε.

(ii) In the usual formulation, partitions in the Regularity Lemma are equitable, i.e., the partition classes are as equal as possible. For graphons on atomless probability spaces, the classes can be required to have the same measure. In fact, it is easy to see that the partitions constructed e.g. in Corollary 5.4 and Theorem 5.8 below can be repartitioned so that the classes will be as equal as possible, the error is at most doubled, and the number of classes is increased by a factor of at most $\lceil 1/\varepsilon \rceil$.

Several other analytic aspects and versions of the Regularity Lemma were proved in [12]. One of these results made a connection between regularity partitions and partitions of J into sets with small diameter in the $r_{W\circ W}$ metric. Here we prove a stronger, cleaner version of that result, and then show how to combine it with our results on thin graphons to get better bounds on the number of partition classes in weak regularity partitions of this graphons.

5.2 Voronoi cells and regularity partitions

We show that Voronoi cells in the metric spaces (J, R_W) and $(J, R_{W\circ W})$ are intimately related to different versions of the Regularity Lemma.

Let (J, d) be a metric space and let π be a probability measure on its Borel sets. We say that a set $S \subseteq J$ is an *average ε-net*, if $\int_J d(x, S)\, d\pi(x) \le \varepsilon$.

Let $S \subseteq J$ be a finite set and $s \in S$. The *Voronoi cell* of S with center s is the set of all points $x \in J$ for which $d(x, s) \le d(x, y)$ for all $y \in S$. Clearly, the Voronoi cells of S cover J. (We can break ties arbitrarily to get a partition.)

Theorem 5.3 *Let (J, W) be a graphon, and let $\varepsilon > 0$.*

(a) *Let S be an average ε-net in the metric space $(S, r_{W\circ W})$. Then the Voronoi cells of S form a weak regularity partition with error at most $8\sqrt{\varepsilon}$.*

(b) *Let $\mathcal{P} = \{J_1, \dots, J_k\}$ be a weak regularity partition with error ε. Then there are points $v_i \in J_i$ such that the set $S = \{v_1, \dots, v_k\}$ is an average (4ε)-net in the metric space $(S, r_{W\circ W})$.*

Proof. (a) Let $\mathcal{P}$ be the partition into the Voronoi cells of S. Let us write $R = W - W_{\mathcal{P}}$. We want to show that $\|R\|_\square \le 8\sqrt{\varepsilon}$. It suffices to show that for any 0-1 valued function f,

$$\langle f, Rf\rangle \le 2\sqrt{\varepsilon}. \tag{10}$$

Let us write $g = f - f_{\mathcal{P}}$, where $f_{\mathcal{P}}(x)$ is obtained by replacing $f(x)$ by the average of f over the class of $\mathcal{P}$ containing x. Clearly $\langle f_{\mathcal{P}}, Rf_{\mathcal{P}}\rangle = 0$, and so

$$\langle f, Rf\rangle = \langle g, Rf\rangle + \langle f_{\mathcal{P}}, Rf\rangle = \langle f, Rg\rangle + \langle f_{\mathcal{P}}, Rg\rangle \le 2\|Rg\|_1 \le 2\|Rg\|_2. \tag{11}$$

For each $x \in J$, let $\varphi(x) \in S$ be the center of the Voronoi cell containing x, and define $W'(x, y) = W(x, \phi(y))$ and similarly $R'(x, y) = R(x, \phi(y))$. Then using that $(W - R)g = W_{\mathcal{P}}g = 0$, $W - W' = R - R'$ and $R'g = 0$, we get

$$\begin{aligned}\|Rg\|_2^2 &= \langle Rg, Rg\rangle = \langle Wg, (R - R')g\rangle = \langle Wg, (W - W')g\rangle = \langle g, W(W - W')g\rangle\\ &\le \|W(W - W')\|_1 = \int\limits_{J^2}\Big|\int\limits_J W(x, y)(W(y, z) - W(y, \varphi(z))\, dy\Big|\, dx\, dz\\ &= \int\limits_J r_W(z, \varphi(z)) = \mathsf{E}_{\mathbf{X}}(r_W(\mathbf{X}, S)) \le \varepsilon.\end{aligned}$$

This proves (10).

(b) Suppose that $\mathcal{P}$ is a weak Szemerédi partition with error ε. Let $R = W - W_{\mathcal{P}}$, then we know that $\|R\|_\square \le \varepsilon$.

For every $x \in [0,1]$, define

$$F(x) = \int\limits_J \Big| \int\limits_J R(x,y)W(y,z)\,dy \Big|\,dz.$$

Then we have

$$\int\limits_J F(x)\,dx = \int\limits_{J^3} s(x,z)R(x,y)W(y,z)\,dx\,dy\,dz,$$

where $s(x,z)$ is the sign of $\int R(x,y)W(y,z)$. For every $z \in J$,

$$\int\limits_{J^2} s(x,z)R(x,y)W(y,z)\,dx\,dy \le 2\|R\|_\square \le 2\varepsilon,$$

and hence

$$\int\limits_J F(x)\,dx \le 2\varepsilon. \tag{12}$$

Let $x, y \in J$ be two points in the same partition class of $\mathcal{P}$. Then $W_{\mathcal{P}}(x,s) = W_{\mathcal{P}}(y,s)$ for every $s \in J$, and hence

$$\begin{aligned} r_W(x,y) &= \int\limits_J \Big| \int\limits_J (W(x,s) - W(y,s))W(s,z)\,ds \Big|\,dz \\ &= \int\limits_J \Big| \int\limits_J (R(x,s) - R(y,s))W(s,z)\,ds \Big|\,dz \\ &\le \int\limits_J \Big| \int\limits_J R(x,s)W(s,z)\,ds \Big|\,dz + \int\limits_J \Big| \int\limits_J R(y,s)W(s,z)\,ds \Big|\,dz \\ &= F(x) + F(y). \end{aligned}$$

For every set $T \in \mathcal{P}$, let $v_T \in T$ be a point "below average" in the sense that

$$F(v_T) \le \frac{1}{\pi(T)} \int\limits_T F(x)\,dx,$$

and let $S = \{v_T :\ T \in \mathcal{P}\}$. Then using (12),

$$\begin{aligned} \mathsf{E}_{\mathbf{X}} d(\mathbf{X}, S) &\le \sum_{T\in\mathcal{P}} \int\limits_T d(x, v_T)\,dx \le \sum_{T\in\mathcal{P}} \int\limits_T (F(x) + F(v_T))\,dx \\ &\le \int\limits_J F(x)\,dx + \sum_{T\in\mathcal{P}} \lambda(T)F(v_T) \le 2\int\limits_J F(x)\,dx \le 4\varepsilon. \end{aligned}$$

This proves the Theorem. $\square$

Theorems 5.3 and 4.1 imply the following Corollary (we prove a stronger result in the next section).

Corollary 5.4 *For every bigraph $F = (V, E)$ there is a constant $c_F > 0$ such that if G is a graph not containing F as an induced sub-bigraph, then for every $\varepsilon > 0$, G has a weak regularity partition with error ε with at most $c_F \varepsilon^{-10|V|}$ classes.*

Remark 5.5 The conclusion does not remain true if the subgraph we exclude is nonbipartite. Any bipartite graph will then satisfy the condition, and some bipartite graphs are known to need an exponential (in $1/\varepsilon$) number of classes in their weak regularity partitions.

5.3 Edit distance

We conclude with deriving bounds on the size of the Szemerédi partitions and approximations in L_1, using the packing dimension of (J, r_W). In the graph theoretic case, this corresponds to approximation in edit distance.

Lemma 5.6 *Let W be a graphon such that (J, r_W) can be covered by m balls of radius ε. Then there is a stepfunction U with $m(1/\varepsilon)^m$ steps such that $\|W - U\|_1 \le 2\varepsilon$.*

Remark 5.7 If W is 0-1 valued, then the bound on the number of classes can be improved to $m2^m$.

Proof. Let $\mathcal{P} = \{J_1, J_2, \dots, J_m\}$ be a partition of J into measurable sets such that for every i there is $x_i \in J$ with $\|W(x_i, .) - W(x, .)\|_1 \le \varepsilon$ for every $x \in J_i$. Let $W'(x, y) = W(x_i, y)$ for $x \in J_i$, then trivially $\|W - W'\|_1 \le \varepsilon$. Let $\mathcal{Q}_i$ be a partition of J into $1/\varepsilon$ measurable classes so that $W(x_i, .)$ varies at most ε on each class of $\mathcal{Q}_i$. For $x \in J_i$ and $y \in S \in \mathcal{Q}_i$, define

$$U(x, y) = \frac{1}{\pi(S)} \int_S W'(x, z)\, dz.$$

Then clearly $|U(x, y) - W'(x, y)| \le \varepsilon$ for all $x, y \in J$, and hence $\|U - W\|_1 \le \|U - W'\|_1 + \|W - W'\|_1 \le 2\varepsilon$. It is obvious that U is a stepfunction in the partition generated by $\mathcal{P}$ and $\mathcal{Q}_1, \dots, \mathcal{Q}_m$, which has at most $m(1/\varepsilon)^m$ classes. □

We obtain from this lemma:

Theorem 5.8 *Let W be a graphon such that (J, r_W) has packing dimension d, then for every $0 < \varepsilon < 1$ it has an ultra-strong partition with error ε and with at most $\varepsilon^{-O(\varepsilon^{-d})}$ classes.*

Proof. Consider a maximal packing in (J, r_W) of balls with radius $\varepsilon/8$; this consists of $m = O(\varepsilon^{-d})$ balls. The balls with the same centers and with radius $\varepsilon/4$ cover J, so Lemma 5.6 there is a stepfunction U with $m(4/\varepsilon)^m \le \varepsilon^{-c\varepsilon^{-d}}$ steps such that $\|W - U\|_1 \le \varepsilon/2$. For the partition $\mathcal{P}$ into the steps of U, we have

$$\|W - W_{\mathcal{P}}\|_1 \le 2\|W - U\|_1 \le \varepsilon$$

(the first inequality follows by easy computation). □

For thin graphons, we get a stronger bound.

Theorem 5.9 *Let W be a thin graphon in which a bigraph $F = (V, E)$ is excluded as an induced sub-bigraph. Then for every $0 < \varepsilon < 1$, it has an ultra-strong partition with error ε and with $O(\varepsilon^{-10|V|^2})$ classes.*

Proof. Theorem 4.1 implies that W is 0-1 valued and it has finite packing dimension at most $10|V|$. Similarly to the proof of lemma 5.6, let $\mathcal{P} = \{J_1, J_2, \dots, J_m\}$ be a partition of J with $m = O(\varepsilon^{-|V|})$ into measurable sets such that for every i there is an $x_i \in J$ with $\|W(x_i, .) - W(x, .)\|_1 \leq \varepsilon$ for every $x \in J_i$. Let $W'(x, y) = W(x_i, y)$ for $x \in J_i$, then $\|W' - W\|_1 \leq \varepsilon$. Let S_i be the support of the function $W(x_i, .)$, and let A be the set of atoms of the Boolean algebra generated by $\{S_1, S_2, \dots, S_m\}$ with positive measure. For every atom $a \in A$, let $F_a \subseteq [m]$ denote the index set $\{i | a \subseteq S_i\}$ and let $\mathcal{F}$ denote the set system $\{F_a | a \in A\}$. Since F is not an induced sub-bigraph, $\mathcal{F}$ has VC-dimension less than $|V|$, and so by lemma 4.3 we obtain that $|A| \leq O(m^{|V|-1})$. The joint refinement $\mathcal{P}_2$ of A and $\mathcal{P}$ is of size at most $O(\varepsilon^{-10|F|^2})$. This completes the proof since W' is a stepfunction with steps in $\mathcal{P}_2$. □

It is easy to see that in the definition of ultra-strong regularity partitions of 0-1 valued graphons, we can replace $W_{\mathcal{P}}$ by a 0-1 valued stepfunction with the same steps, at the cost of doubling the error. Together with Remark 5.2, we can apply this to a (large) finite graph G. To state the result, we need a definition. Let H be a simple graph, and let us replace each node v of H by a set S_v of "twin" nodes, where two nodes $x \in S_u$ and $y \in S_v$ are connected if and only if $uv \in E(H)$. For each $u \in V(H)$, either connect all pairs of nodes in S_u, or none of them. We call every graph obtained this way a *blow-up* of H.

Corollary 5.10 *For every bigraph F there is a constant $c_F > 0$ such that if G is a graph not containing F as an induced sub-bigraph, then for every $\varepsilon > 0$, we can change $\varepsilon|G|^2$ edges of G so that the resulting graph is a blow-up of a graph with at most $c_F\varepsilon^{-10|F|^2}$ nodes.*

Let us say that a graphon W has *polynomial L_1-complexity* if there is a $d > 0$ such that for every $\varepsilon > 0$ there is a stepfunction W' with $O(\varepsilon^{-d})$ steps satisfying $\|W - W'\|_1 \leq \varepsilon$. We can define *polynomial $\square$-complexity* analogously. As we have pointed out, polynomial $\square$-complexity corresponds to the finite dimensionality of the metric space of $W \circ W$. Theorem 5.9 implies that every thin graphon has polynomial L_1-complexity.

If W has polynomial complexity, then the structure of W can be described by a polynomial number (in $1/\varepsilon$) of real parameters with an error ε in the appropriate norm. The set of graphons with polynomial complexity is closed under many natural operations such as operator product, tensor product, etc.

It could be interesting to study other aspects of this complexity notion. We offer a conjecture relating our complexity notion to extremal combinatorics. It is supported by examples in [13].

Conjecture 5.11 Let $F_1, F_2, \dots, F_n$ be a set of finite graphs, $t_1, t_2, \dots, t_m$ be real numbers in $[0, 1]$ and $\mathcal{S}$ be the set of graphons W with $t(F_i, W) = t_i$ for $1 \leq i \leq n$. Then $\mathcal{S}$ is either empty or it contains a graphon of polynomial L_1-complexity.

References

[1] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy: Efficient testing of large graphs, *Combinatorica* **20** (2000), 451–476.

[2] C. Borgs, J. Chayes, L. Lovász: Moments of Two-Variable Functions and the Uniqueness of Graph Limits, *Geometric and Functional Analysis* (to appear)

`http://www.cs.elte.hu/~lovasz/limitunique.pdf`

[3] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Counting graph homomorphisms, in: *Topics in Discrete Mathematics* (ed. M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas, P. Valtr), Springer (2006), 315–371.

[4] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy and K. Vesztergombi: Graph Limits and Parameter Testing, Proc. 38th Annual ACM Symp. on Theory of Computing 2006, 261–270.

[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent Graph Sequences I: Subgraph frequencies, metric properties, and testing, *Advances in Math.* (2008), 10.1016/j.aim.2008.07.008.

[6] P. Erdös, L. Lovász, J. Spencer: Strong independence of graphcopy functions, in: *Graph Theory and Related Topics*, Academic Press, 165-172.

[7] M. Freedman, L. Lovász, A. Schrijver: Reflection positivity, rank connectivity, and homomorphisms of graphs, *J. Amer. Math. Soc.* **20** (2007), 37–51.

[8] A. Frieze and R. Kannan: Quick approximation to matrices and applications, *Combinatorica* **19**, 175–220.

[9] J. Komlós, J. Pach and G. Woeginger: Almost Tight Bounds for epsilon-Nets, *Discr. Comput. Geometry* **7** (1992), 163–173.

[10] L. Lovász, V.T. Sós: Generalized quasirandom graphs, *J. Comb. Th. B* **98** (2008), 146–163.

[11] L. Lovász, B. Szegedy: Limits of dense graph sequences, *J. Comb. Theory B* **96** (2006), 933–957.

[12] L. Lovász and B. Szegedy: Szemerédi's Lemma for the analyst, *Geom. Func. Anal.* **17** (2007), 252–270.

[13] L. Lovász and B. Szegedy: Finitely forcible graphons (submitted)

`http://arxiv.org/abs/0901.0929`

[14] E. Szemerédi: On sets of integers containing no k elements in arithmetic progression", *Acta Arithmetica* **27** (1975) 199-245.

[15] E. Szemerédi: Regular partitions of graphs, *Colloque Inter. CNRS* (J.-C. Bermond, J.-C. Fournier, M. Las Vergnas and D. Sotteau, eds.) (1978) 399–401.

[16] V. Vapnik, A. Chervonenkis: On the uniform convergence of relative frequencies of events to their probabilities, *Theor. Prob. Appl.* **16** (1971), 264–280.