Ordered Graph Limits and Their Applications

Omri Ben-Eliezer

Center of Mathematical Sciences and Applications, Harvard University, Cambridge, MA, USA
omribene@cmsa.fas.harvard.edu

Eldar Fischer

Faculty of Computer Science, Technion - Israel Institute of Technology, Haifa, Israel
eldar@cs.technion.ac.il

Amit Levi

Cheriton School of Computer Science, University of Waterloo, Canada
amit.levi@uwaterloo.ca

Yuichi Yoshida

Principles of Informatics Research Division, National Institute of Informatics (NII), Tokyo, Japan
yyoshida@nii.ac.jp

Abstract

The emerging theory of graph limits exhibits an analytic perspective on graphs, showing that many important concepts and tools in graph theory and its applications can be described more naturally (and sometimes proved more easily) in analytic language. We extend the theory of graph limits to the ordered setting, presenting a limit object for dense vertex-ordered graphs, which we call an *orderon*. As a special case, this yields limit objects for matrices whose rows and columns are ordered, and for dynamic graphs that expand (via vertex insertions) over time. Along the way, we devise an ordered locality-preserving variant of the cut distance between ordered graphs, showing that two graphs are close with respect to this distance if and only if they are similar in terms of their ordered subgraph frequencies. We show that the space of orderons is compact with respect to this distance notion, which is key to a successful analysis of combinatorial objects through their limits. For the proof we combine techniques used in the unordered setting with several new techniques specifically designed to overcome the challenges arising in the ordered setting.

We derive several applications of the ordered limit theory in extremal combinatorics, sampling, and property testing in ordered graphs. In particular, we prove a new ordered analogue of the well-known result by Alon and Stav [RS&A'08] on the furthest graph from a hereditary property; this is the first known result of this type in the ordered setting. Unlike the unordered regime, here the Erdős–Rényi random graph $\mathbf{G}(n,p)$ with an ordering over the vertices is *not* always asymptotically the furthest from the property for some p. However, using our ordered limit theory, we show that random graphs generated by a stochastic block model, where the blocks are consecutive in the vertex ordering, are (approximately) the furthest. Additionally, we describe an alternative analytic proof of the ordered graph removal lemma [Alon et al., FOCS'17].

2012 ACM Subject Classification Mathematics of computing → Functional analysis; Mathematics of computing → Nonparametric representations; Mathematics of computing → Extremal graph theory; Theory of computation → Streaming, sublinear and near linear time algorithms

Keywords and phrases graph limits, ordered graph, graphon, cut distance, removal lemma

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.42

Related Version A full version of the paper is available at `https://arxiv.org/abs/1811.02023`.

Funding *Omri Ben-Eliezer*: Part of this work was done while the author was at Tel Aviv University.
Amit Levi: Research supported by the David R. Cheriton Graduate Scholarship. Part of this work was done while the author was visiting the Technion.
Yuichi Yoshida: Research supported by JSPS KAKENHI Grant Number JP17H04676.

12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 42; pp. 42:1–42:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Large graphs appear in many applications across all scientific areas. Naturally, it is interesting to try to understand their structure and behavior: When can we say that two graphs are similar (even if they do not have the same size)? How can the convergence of graph sequences be defined? What properties of a large graph can we capture by taking a small sample from it?

The theory of graph limits addresses such questions from an analytic point of view. The investigation of convergent sequences of dense graphs was started to address three seemingly unrelated questions asked in different fields: statistical physics, theory of networks and the Internet, and quasi-randomness. A comprehensive series of papers [12, 13, 28, 21, 29, 14, 11, 30, 15] laid the infrastructure for a rigorous study of the theory of dense graph limits, demonstrating various applications in many areas of mathematics and computer science. The book of Lovász on graph limits [27] presents these results in a unified form.

A sequence $\{G_n\}_{n=1}^{\infty}$ of finite graphs, whose number of vertices tends to infinity as $n \to \infty$, is considered *convergent*[1] if the frequency[2] of any fixed graph F as a subgraph in G_n converges as $n \to \infty$. The limit object of a convergent sequence of (unordered) graphs in the dense setting, called a *graphon*, is a measurable symmetric function $W\colon [0,1]^2 \to [0,1]$, and it was proved in [28] that, indeed, for any convergent sequence $\{G_n\}$ of graphs there exists a graphon serving as the limit of G_n in terms of subgraph frequencies. Apart from their role in the theory of graph limits, graphons are useful in probability theory, as they give rise to exchangeable random graph models; see e.g. [17, 33]. An analytic theory of convergence has been established for many other types of discrete structures. These include sparse graphs, for which many different (and sometimes incomparable) notions of limits exist – see e.g. [16, 10] for two recent papers citing and discussing many of the works in this field; permutations, first developed in [25] and further investigated in several other works; partial orders [26]; and high dimensional functions over finite fields [35]. The limit theory of dense graphs has also been extended to hypergraphs, see [36, 18] and the references within.

In this work we extend the theory of dense graph limits to the ordered setting, establishing a limit theory for vertex-ordered graphs in the dense setting, and presenting several applications of this theory. An *ordered graph* is a symmetric function $G\colon [n]^2 \to \{0,1\}$. G is *simple* if $G(x,x) = 0$ for any x. A *weighted ordered graph* is a symmetric function $F\colon [n]^2 \to [0,1]$. Unlike the unordered setting, where $G, G'\colon [n]^2 \to \Sigma$ are considered isomorphic if there is a permutation π over $[n]$ so that $G(x,y) = G'(\pi(x),\pi(y))$ for any $x \neq y \in [n]$, in the ordered setting, the automorphism group of a graph G is trivial: G is only isomorphic to itself through the identity function.

For simplicity, we consider in the following only graphs (without edge colors). All results here can be generalized in a relatively straightforward manner to edge-colored graph-like ordered structures, where pairs of vertices may have one of $r \geq 2$ colors (the definition above corresponds to the case $r = 2$). This is done by replacing the range $[0,1]$ with the $(r-1)$-dimensional simplex (corresponding to the set of all possible distributions over $[r]$).

Two interesting special cases of two-dimensional ordered structures for which our results naturally yield a limit object are *images*, i.e., ordered matrices, and *dynamic graphs* with vertex insertions. Specifically, (binary) $m \times n$ images can be viewed as ordered bipartite

---

[1] In unordered graphs, this is also called *convergence from the left*; see the discussion on [14].

[2] The frequency of F in G is roughly defined as the ratio of induced subgraphs of G isomorphic to F among all induced subgraphs of G on $|F|$ vertices.

graphs $I\colon [m] \times [n] \to \{0,1\}$, and our results can be adapted to get a bipartite ordered limit object for them as long as $m = \Theta(n)$. Meanwhile, a dynamic graph with vertex insertions can be viewed as a sequence $\{G_i\}_{i=1}^{\infty}$ of ordered graphs, where G_{i+1} is the result of adding a vertex to G_i and connecting it to the previous vertices according to some prescribed rule. It is natural to view such dynamic graphs that evolve with time as ordered ones, as the time parameter induces a natural ordering. Thus, our work gives, for example, a limit object for time-series where there are pairwise relations between events occurring at different times.

As we shall see in Subsection 1.2, the main results proved in this paper are, in a sense, natural extensions of results in the unordered setting. However, proving them requires machinery that is heavier than that used in the unordered setting: the tools used in the unordered setting are not rich enough to overcome the subtleties materializing in the ordered setting. In particular, the limit object we use in the ordered setting – which we call an *orderon* – has a 4-dimensional structure that is more complicated than the analogous 2-dimensional structure of the graphon, the limit object for the unordered setting. The tools required to establish the ordered theory are described next.

1.1 Main ingredients

Let us start by considering a simple yet elusive sequence of ordered graphs, which has the makings of convergence. The *odd-clique* ordered graph H_n on $2n$ vertices is defined by setting $H_n(i,j) = 1$, i.e., having an edge between vertices i and j, if and only if $i \neq j$ and i, j are both odd, and otherwise setting $H_n(i,j) = 0$. In this subsection we closely inspect this sequence to demonstrate the challenges arising while trying to establish a theory for ordered graphs, and the solutions we propose for them. First, let us define the notions of subgraph frequency and convergence.

The (induced) frequency $t(F,G)$ of a simple ordered graph F on k vertices in an ordered graph G with n vertices is the probability that, if one picks k vertices of G uniformly and independently (repetitions are allowed) and reorders them as $x_1 \leq \cdots \leq x_k$, F is isomorphic to the induced ordered subgraph of G over $x_1, \ldots, x_k$. (The latter is defined as the ordered graph H on k vertices satisfying $H(i,j) = G(x_i, x_j)$ for any $i, j \in [k]$.) A sequence $\{G_n\}_{n=1}^{\infty}$ of ordered graphs is *convergent* if $|V(G_n)| \to \infty$ as $n \to \infty$, and the frequency $t(F, G_n)$ of any simple ordered graph F converges as $n \to \infty$. Observe that the odd-clique sequence $\{H_n\}$ is indeed convergent: The frequency of the empty k-vertex graph in H_n tends to $(k+1)2^{-k}$ as $n \to \infty$, the frequency of any non-empty k-vertex ordered graph containing only a clique and a (possibly empty) set of isolated vertices tends to 2^{-k}, and the frequency of any other graph in H_n is 0.[3]

In light of previous works on the unordered theory of convergence, we look for a limit object for ordered graphs that has the following features.

Representation of finite ordered graphs. The limit object should have a natural and consistent representation for finite ordered graphs. As in graphons, we allow graphs G and H to have the same representation when one is a blowup[4] of the other.

Usable distance notion. Working directly with the definition of convergence in terms of subgraph frequencies is difficult. The limit object we seek should be endowed with a metric, like the cut distance for unordered graphs (see discussion below), that should be easier to work with and must have the following property: A sequence of ordered graph is convergent (in terms of frequencies) if and only if it is Cauchy in the metric.

[3] To see why the sum of frequencies is 1, note that for $k \geq \ell \geq 2$, the number of k-vertex ordered graphs consisting of an ℓ-vertex clique and $k - \ell$ isolated vertices is $\binom{k}{\ell}$.

[4] A graph G on nt vertices is an ordered t-blowup of H on n vertices if $G(x,y) = H(\lceil x/t \rceil, \lceil y/t \rceil)$ for any x and y.

Completeness and compactness. The space of limit objects must be complete with respect to the metric: Cauchy sequences should converge in this metric space. Combined with the previous requirements, this will ensure that any convergent sequence of ordered graphs has a limit (in terms of ordered frequencies), as desired. It is even better if the space is compact, as compactness is essentially an "ultimately strong" version of Szemerédi's regularity lemma [34], and will help to develop applications of the theory in other areas.

Additionally, we would like the limit object to be as simple as possible, without unnecessary over-representation. In the unordered setting, the metric used is the *cut distance*, introduced by Frieze and Kannan [22, 23] and defined as follows. First, we define the *cut norm* $\|W\|_\square$ of a function $W\colon [0,1]^2 \to \mathbb{R}$ as the supremum of $|\int_{S\times T} W(x,y)dxdy|$ over all measurable subsets $S, T \subseteq [0,1]$. The *cut distance* between graphons W and W' is the infimum of $\|W^\phi - W'\|_\square$ over all measure-preserving bijections $\phi\colon [0,1] \to [0,1]$, where $W^\phi(x,y) \stackrel{\text{def}}{=} W(\phi(x), \phi(y))$.

For the ordered setting, we look for a similar metric; the cut distance itself does not suit us, as measure-preserving bijections do not preserve ordered subgraph frequencies in general. A first intuition is then to try graphons as the limit object, endowed with the metric $d_\square(W,W') \stackrel{\text{def}}{=} \|W - W'\|_\square$. However, this metric does not satisfy the second requirement: the odd-clique sequence is convergent, yet it is not Cauchy in $d_\square$, since $d_\square(H_n, H_{2n}) = 1/2$ for any n. Seeing that $d_\square$ seems "too strict" as a metric and does not capture the similarities between large odd-clique graphs well, it might make sense to use a slightly more "flexible" metric, which allows for measure-preserving bijections, as long as they do not move any of the points too far from its original location. In view of this, we define the *cut-shift distance* between two graphons W, W' as

$$d_\triangle(W,W') \stackrel{\text{def}}{=} \inf_f \left(\text{Shift}(f) + \|W^f - W'\|_\square\right), \tag{1}$$

where $f\colon [0,1] \to [0,1]$ is a measure-preserving bijection, $\text{Shift}(f) = \sup_{x\in[0,1]} |f(x) - x|$, and $W^f(x,y) = W(f(x), f(y))$ for any $x, y \in [0,1]$. As we show in this paper (Theorem 2 below), the cut-shift distance settles the second requirement: a sequence of ordered graphs is convergent *if and only if* it is Cauchy in the cut-shift distance.

Consider now graphons as a limit object, coupled with the cut-shift distance as a metric. Do graphons satisfy the third requirement? In particular, does there exist a graphon whose ordered subgraph frequencies are equal to the limit frequencies for the odd-clique sequence? The answers to both of these questions are negative: it can be shown that such a graphon cannot exist in view of Lebesgue's density theorem, which states that there is no measurable subset of $[0,1]$ whose density in every interval (a,b) is $(b-a)/2$ (see e.g. Theorem 2.5.1 in the book of Franks on Lebesgue measure [20]). Thus, we need a somewhat richer ordered limit object that will allow us to "bypass" the consequences of Lebesgue's density theorem. Consider for a moment the graphon representations of the odd clique graphs. In these graphons, the domain $[0,1]$ can be partitioned into increasingly narrow intervals that alternately represent odd and even vertices. Intuitively, it seems that our limit object needs to be able to contain infinitesimal odd and even intervals at any given location, leading us to the following limit object candidate, which we call an *orderon*.

An orderon is a symmetric measurable function $W\colon ([0,1]^2)^2 \to [0,1]$ viewed, intuitively and loosely speaking, as follows. In each point $(x,a) \in [0,1]^2$, corresponding to an infinitesimal "vertex" of the orderon, the first coordinate, x, represents a location in the linear order of $[0,1]$. Each set $\{x\} \times [0,1]$ can thus be viewed as an infinitesimal probability space of vertices that have the same location in the linear order. The role of the second coordinate is to allow "variability" (in terms of probability) of the infinitesimal "vertex" occupying this point in the order. The definition of the frequency $t(F, W)$ of a simple ordered graph $F = ([k], E)$ in an orderon W is a natural extension of frequency in graphons. First, define the random variable $\mathbf{G}(k, W)$ as follows: Pick k points in $[0, 1]^2$ uniformly and independently, order them according to the first coordinate as $(x_1, a_1), \ldots, (x_k, a_k)$ with $x_1 \leq \cdots \leq x_k$, and then return a k-vertex graph G, in which the edge between each pair of vertices i and j exists with probability $W((x_i, a_i), (x_j, a_j))$, independently of other edges. The frequency $t(F, W)$ is defined as the probability that the graph generated according to $\mathbf{G}(k, W)$ is isomorphic to F.

Consider the orderon W satisfying $W((x, a), (y, b)) = 1$ if and only if $a, b \leq 1/2$, and otherwise $W((x, a), (y, b)) = 0$. W now emerges as a natural limit object for the odd-clique sequence: one can verify that the subgraph frequencies in it are as desired.

The cut-shift distance for orderons is defined similarly to (1), except that f is now a measure-preserving bijection from $[0, 1]^2$ to $[0, 1]^2$ and $\mathrm{Shift}(f) = \sup_{(x,a)\in[0,1]^2} |\pi_1(f(x, a)) - x|$, where $\pi_1(y, b) \stackrel{\text{def}}{=} y$ is the projection to the first coordinate.

1.2 Main results

Let $\mathcal{W}$ denote the space of orderons endowed with the cut-shift distance. In view of Lemma 19 below, $d_\triangle$ is a pseudo-metric for $\mathcal{W}$. By identifying $W, U \in \mathcal{W}$ whenever $d_\triangle(W, U) = 0$, we get a metric space $\widetilde{\mathcal{W}}$. The following result is the main component for the viability of our limit object, settling the third requirement above.

▶ **Theorem 1.** *The space $\widetilde{\mathcal{W}}$ is compact with respect to $d_\triangle$.*

The proof of Theorem 1 is significantly more involved than the proof of its unordered analogue. While at a very high level, the roadmap of the proof is similar to that of the unordered one, our setting induces several new challenges, and to handle them we develop new *shape approximation* techniques. These are presented along the proof of the theorem in Section 4.

The next result shows that convergence in terms of frequencies is equivalent to being Cauchy in $d_\triangle$. This settles the second requirement.

▶ **Theorem 2.** *Let $\{W_n\}_{n=1}^\infty$ be a sequence of orderons. Then $\{W_n\}$ is Cauchy in $d_\triangle$ if and only if $t(F, W_n)$ converges for any fixed simple ordered graph F.*

As a corollary of the last two results, we get the following.

▶ **Corollary 3.** *For every convergent sequence of ordered graphs $\{G_n\}_{n\in\mathbb{N}}$, there exists an orderon $W \in \mathcal{W}$ such that $t(F, G_n) \to t(F, W)$ for every ordered graph F.*

The next main result is a sampling theorem, stating that a large enough sample from an orderon is almost always close to it in cut-shift distance. For this, we define the orderon representation W_G of an n-vertex ordered graph G by setting $W_G((x, a), (y, b)) = G(Q_n(x), Q_n(y))$ for any x, a, y, b, where we define $Q_n(x) = \lceil nx \rceil$ for $x > 0$ and $Q_n(0) = 1$. This addresses the first requirement.

▶ **Theorem 4.** *Let k be a positive integer and let $W \in \mathcal{W}$ be an orderon. Let $G \sim \mathbf{G}(k, W)$. Then,*

$$d_\triangle(W, W_G) \leq C \left(\frac{\log \log k}{\log k} \right)^{1/3}$$

holds with probability at least $1 - C \exp(-\sqrt{k}/C)$ for some constant $C > 0$.

Theorem 4 implies, in particular, that ordered graphs are a dense subset in $\mathcal{W}$.

▶ **Corollary 5.** *For every orderon W and every $\varepsilon > 0$, there exists a simple ordered graph G on at most $2^{\varepsilon^{-3+o(1)}}$ vertices such that $d_\triangle(W, W_G) \le \varepsilon$.*

Our next result asserts that any orderon $W \in \mathcal{W}$ can be approximated in L_1-distance by an orderon U with a finite block structure, with the added property that any ordered finite structure that appears with positive density in U also has positive density in W.[5] The orderon U is described as follows. the point set $[0,1]^2$ is divided into b "blocks", which are subsets of the form $[(i-1)/b, i/b] \times [0,1]$ for some $i \in [b]$. Each block is decomposed into l "layers", of the form $[(i-1)/b, i/b] \times [(j-1)/l, j/l]$ where $j \in [l]$. The value of $U((x,a),(y,b))$ is now only dependent on which blocks x, y belong to, which layers a, b belong to, and possibly whether $x < y$. For example, the orderon U representing the limit of the odd-clique sequence (defined by $U((x,a),(y,b)) = 1$ if $a, b \le 1/2$, and $U((x,a),(y,b)) = 0$ elsewhere) has one block and two layers in it. Roughly speaking, one can think of such U as the orderon representation of a "pixelized" ordered graph, where each vertex (block) consists of multiple "pixels" (here a pixel corresponds to a block-layer pair), and there is a weighted edge[6] between each pair of pixels. Therefore we call an orderon U with such structure a *pixelized* orderon and term our result the *pixelization lemma.*

▶ **Theorem 6** (Pixelization lemma; informal)**.** *For any orderon W and $\varepsilon > 0$, there exists a pixelized orderon U so that $d_1(U, W) \le \varepsilon$, satisfying the following: for all ordered graphs F with $t(F, U) > 0$, we have $t(F, W) > 0$.*

We note that the pixelized structure of U is necessary for this statement to be correct; it is no longer correct in general if we insist that U must be the orderon representation of a standard edge-weighted ordered graph.

The pixelization lemma is especially useful for applications where the L_1-distance comes into play. Two such applications, reproving the ordered graph removal lemma [2] and proving a new result in extremal combinatorics, are described next.

1.3 The furthest ordered graph from a hereditary property

Here and in the next subsection we describe three applications of our ordered limit theory. We start with an extensive discussion on the first application: A new result on the maximum edit[7] distance $d_1(G, \mathcal{H})$ of an ordered graph G from a hereditary[8] property $\mathcal{H}$.

For a hereditary property $\mathcal{H}$ of simple ordered graphs, define $\overline{d_\mathcal{H}} = \sup_G d_1(G, \mathcal{H})$ where G ranges over all simple graphs (of any size). The parameter $d_\mathcal{H}$ has been widely investigated for unordered graphs. A well-known surprising result of Alon and Stav [6] states, roughly speaking, that $d_\mathcal{H}$ is always "achieved" by the Erdős–Rényi random graph $\mathbf{G}(n,p)$ for an appropriate choice of p and large enough n.

[5] A weaker result, in which the L_1-distance is replaced by the cut-shift distance, is not hard to prove using our previous main results; we note that it is indeed strictly weaker since the L_1-distance between any two orderons U and W is always at least as large as (and sometimes much larger than) $d_\triangle(U, W)$.

[6] In fact, a weighted bi-directed edge, with possibly different weights in the the different directions.

[7] For our purposes, define the edit (or Hamming) distance between two ordered graphs G and G' on n vertices as the smallest number of entries that one needs to change in the adjacency matrix A_G of G to make it equal to $A_{G'}$, divided by n^2. For this matter, the adjacency matrix A_G of a graph G over vertices $v_1 < \ldots < v_n$ is a binary $n \times n$ matrix where $A_G(i,j) = 1$ if and only if there is an edge between v_i and v_j in G. The distance between G and a property $\mathcal{P}$ of ordered graphs is $\min_{G'} d_1(G, G')$ where G' ranges over all graphs G' of the same size as G. The definition for unordered graphs is similar; the only difference is in the notion of isomorphism.

[8] A property of (ordered or unordered) graphs is *hereditary* if it is closed under taking induced subgraphs.

▶ **Theorem 7** ([6]). *For any hereditary property $\mathcal{H}$ of unordered graphs there exists $p_{\mathcal{H}} \in [0,1]$ satisfying the following. A graph $G \sim \mathbf{G}(n, p_{\mathcal{H}})$ satisfies $d_1(G, \mathcal{H}) \geq \overline{d_{\mathcal{H}}} - o(1)$ with high probability.*

In other words, a random graph $\mathbf{G}(n, p_{\mathcal{H}})$ is with high probability asymptotically (that is, up to relative edit distance of $o(1)$) the furthest from the property $\mathcal{H}$. From the analytic perspective, Lovász and Szegedy [30] were able to reprove (and extend) this result using graph limits.

The surprising result of Alon and Stav has led naturally to a very interesting and highly non-trivial question, now known as the (extremal) *graph edit distance problem* [31], which asks the following: Given a hereditary property of interest $\mathcal{H}$, what is the value (or values) $p_{\mathcal{H}}$ that maximizes the distance of $\mathbf{G}(n,p)$ from $\mathcal{H}$? The general question of determining $p_{\mathcal{H}}$ given any $\mathcal{H}$ is currently wide open, although there have been many interesting developments for various classes of hereditary properties; see [31] for an extensive survey of previous works and useful techniques.

While the situation in unordered graphs, and even in (unordered) directed graphs [7] and matrices [32] has been thoroughly investigated, for ordered graphs no result in the spirit of Theorem 7 is known. The first question that comes to mind is whether the behavior in the ordered setting is similar to that in the unordered case: Is it true that for any hereditary property $\mathcal{H}$ of *ordered* graphs there exists $p = p_{\mathcal{H}}$ for which $G \sim \mathbf{G}(n,p)$ satisfies $d_1(G, \mathcal{H}) \geq \overline{d_{\mathcal{H}}} - o(1)$ with high probability?

As we show, the answer is in fact *negative*. Consider the ordered graph property $\mathcal{H}$ defined as follows: $G \in \mathcal{H}$ if and only if there do not exist vertices $u_1 < u_2 \leq u_3 < u_4$ in G where u_1u_2 is a non-edge and u_3u_4 is an edge. $\mathcal{H}$ is clearly a hereditary property, defined by a finite family of forbidden ordered subgraphs. In the full version [9], we prove that the typical distance of $G \sim \mathbf{G}(n,p)$ from $\mathcal{H}$ is no more than $1/4 + o(1)$ (the maximum is asymptotically attained for $p = 1/2$). In contrast, we show there exists a graph G satisfying $d_1(G, \mathcal{H}) = 1/2 - o(1)$, which is clearly the furthest possible up to the $o(1)$ term (every graph G is $1/2$-close to either the complete or the empty graph, which are in $\mathcal{H}$), and is substantially further than the typical distance of $\mathbf{G}(n,p)$ for any choice of p. This shows that Theorem 7 *cannot be true* for the ordered setting.

However, the news are not all negative: We present a positive result in the ordered setting, which generalizes the unordered statement in some sense, and whose proof makes use of our ordered limit theory. While it is no longer true that $\mathbf{G}(n,p)$ generates graphs that are asymptotically the furthest from $\mathcal{H}$, we show that a random graph generated according to a *consecutive stochastic block model* is approximately the furthest. A *stochastic block model* [1] with M blocks is a well-studied generalization of $\mathbf{G}(n,p)$, widely used in the study of community detection, clustering, and various other problems in mathematics and computer science. A stochastic block model is defined according to the following three parameters: n, the total number of vertices; $(q_1, \ldots, q_M)$, a vector of probabilities that sum up to one; and a symmetric $M \times M$ matrix of probabilities p_{ij}. A graph on n vertices is generated according to this model as follows. First, we assign each of the vertices independently[9] to one of M parts $A_1, \ldots, A_M$, where the probability of any given vertex to fall in A_i is q_i. Then, for any $(i,j) \in [M]^2$, and any pair of disjoint vertices $u \in A_i$ and $v \in A_j$, we add an edge between u and v with probability p_{ij}. By *consecutive*, we mean that all vertices assigned to A_i precede (in the vertex ordering) all vertices assigned to A_{i+1}, for any $i \in [M-1]$. Our main result now is as follows.

[9] In some contexts, the stochastic block model is defined by determining the *exact* number of vertices in each A_i in advance, rather than assigning the vertices independently; all results here are also true for this alternative definition.

▶ **Theorem 8.** *Let $\mathcal{H}$ be a hereditary property of simple ordered graphs and let $\varepsilon > 0$. There exists a consecutive stochastic block model with at most $M = M_{\mathcal{H}}(\varepsilon)$ blocks with equal containment probabilities (i.e., $q_i = 1/M$ for any $i \in [M]$), satisfying the following. A graph G on n vertices generated by this model satisfies $d_1(G, \mathcal{H}) \geq \overline{d_{\mathcal{H}}} - \varepsilon$ with probability that tends to one as $n \to \infty$.*

The proof, given in the full version of this paper [9], is a good example of the power of the analytic perspective, combining our ordered limit theory with standard measure-theoretic tools and a few simple lemmas proved in [30].

1.4 Sampling and property testing

We finish by showing two additional applications of the ordered limit theory. These applications are somewhat more algorithmically oriented – concerning sampling and property testing – and illustrate the use of our theory for algorithmic purposes. The first of them is concerned with naturally estimable ordered graph parameters, defined as follows.

▶ **Definition 9** (naturally estimable parameter). *An ordered graph parameter f is* naturally estimable *if for every $\varepsilon > 0$ and $\delta > 0$ there is a positive integer $k = k(\varepsilon, \delta) > 0$ satisfying the following. If G is an ordered graph with at least k nodes and $G|_{\boldsymbol{k}}$ is the subgraph induced by a uniformly random ordered set of exactly k nodes of G, then*

$$\Pr_{G|_{\boldsymbol{k}}}[|f(G) - f(G|_{\boldsymbol{k}})| > \varepsilon] < \delta.$$

The following result provides an analytic characterization of ordered natural estimability, providing a method to study estimation problems on ordered graphs from the analytic perspective.

▶ **Theorem 10.** *Let f be a bounded simple ordered graph parameter. Then, the following are equivalent:*

1. *f is naturally estimable.*
2. *For every convergent sequence $\{G_n\}_{n\in\mathbb{N}}$ of ordered simple graphs with $|V(G_n)| \to \infty$, the sequence of numbers $\{f(G_n)\}_{n\in\mathbb{N}}$ is convergent.*
3. *There exists a functional $\widehat{f}(W)$ over $\mathcal{W}$ that satisfies the following:*
   a. *$\widehat{f}(W)$ is continuous with respect to $d_{\triangle}$.*
   b. *For every $\varepsilon > 0$, there is $k = k(\varepsilon)$ such that for every ordered graph G with $|V(G)| \geq k$, it holds that $\left|\widehat{f}(W_G) - f(G)\right| \leq \varepsilon$.*

Our third application is a new analytic proof of the ordered graph removal lemma of [2], implying that every hereditary property of ordered graphs (and images over a fixed alphabet) is testable, with one-sided error, using a constant number of queries. (For the relevant definitions, see [2] and Definition 9 here.)

▶ **Theorem 11** ([2]). *Let $\mathcal{H}$ be a hereditary property of simple ordered graphs, and fix $\varepsilon, c > 0$. Then there exists $k = k(\mathcal{H}, \varepsilon, c)$ satisfying the following: For every ordered graph G on $n \geq k$ vertices that is ε-far from $\mathcal{H}$, the probability that $G|_{\boldsymbol{k}}$ does not satisfy $\mathcal{H}$ is at least $1 - c$.*

Our proof of Theorem 11 utilizes the analytic tools developed in this work, and bypasses the need for many of the sophisticated combinatorial techniques from [2], resulting in an arguably cleaner proof.

1.5 Related work

The theory of graph limits has strong ties to the area of property testing, especially in the dense setting. Regularity lemmas for graphs, starting with the well-known regularity lemma of Szemerédi [34], later to be joined by the weaker (but more efficient) versions of Frieze and Kannan [22, 23] and the stronger variants of Alon et al. [3], among others, have been very influential in the development of property testing. For example, regularity was used to establish the testability of all hereditary properties in graphs [5], the relationship between the testability and estimability of graph parameters [19], and combinatorial characterizations of testability [4].

The analytic theory of convergence, built using the cut distance and its relation to the weak regularity lemma, has proved to be an interesting alternative perspective on these results. Indeed, the aforementioned results have equivalent analytic formulations, in which both the statement and the proof seem cleaner and more natural. A recent line of work has shown that many of the classical results in property testing of dense graphs can be extended to dense ordered graph-like structures, including vertex-ordered graphs and images. In [2], it was shown that the testability of hereditary properties extends to the ordered setting (see Theorem 11 above). Shortly after, in [8] it was proved that characterizations of testability in unordered graphs can be partially extended to similar characterizations in ordered graph-like structures, provided that the property at stake is sufficiently "well-behaved" in terms of order.

Graphons and their sparse analogues have various applications in different areas of mathematics, computer science, and even social sciences. The connections between graph limits and real-world large networks have been very actively investigated; see the survey of Borgs and Chayes [16]. Graph limits have applications in probability and data analysis [33]. Graphons were used to provide new analytic proofs of results in extremal graph theory; see Chapter 16 in [27]. Through the notion of free energy, graphons were also shown to be closely connected to the field of statistical physics [15]. We refer the reader to [27] for more details.

We remark that an independent work, by Frederik Garbe, Robert Hancock, Jan Hladky, and Maryam Sharifzadeh, investigates an alternative limit object for the ordered setting in the context of latin squares. See [24] for their findings, as well as connections between orderons and their limit object, called a latinon.

1.6 Organization

Due to space limitations, much of the technical content is missing from this version of the paper. Specifically, we only include here the following components. In Section 2, we present basic definitions for our ordered limit theory. Section 3 contains the main ingredients for the the proof that ordered graphs are dense in the space of orderons (some technical details are relegated to the full version). Section 4 presents the proof that the space of orderons is compact (Theorem 1). The reader is referred to [9] for a full version of this paper, including proofs of all results stated in this manuscript.

2 Preliminaries

In this section we formally describe some of the basic ingredients of our theory, including the limit object – the *orderon*, and several distance notions including the cut-norm for orderons (both unordered and ordered variants are presented), and the cut-shift distance. We then show that the latter is a pseudo-metric for the space of orderons. This will later allow us to view the space of orderons as a metric space, by identifying orderons of cut-shift distance 0.

The measure used here is the Lebesgue measure, denoted by λ. We start with the formal definition of an orderon.

▶ **Definition 12** (orderon). *An* orderon *is a measurable function $W : \left([0,1]^2\right)^2 \to [0,1]$ that is symmetric in the sense that $W((x,a),(y,b)) = W((y,b),(x,a))$ for all $(x,a),(y,b) \in [0,1]^2$. For the sake of brevity, we also denote $W((x,a),(y,b))$ by $W(v_1,v_2)$ for $v_1,v_2 \in [0,1]^2$.*

We denote the set of all orderons by $\mathcal{W}$.

▶ **Definition 13** (measure-preserving bijection). *A map $g\colon [0,1]^2 \to [0,1]^2$ is* measure preserving *if the pre-image $g^{-1}(X)$ is measurable for every measurable set X and $\lambda(g^{-1}(X)) = \lambda(X)$. A* measure preserving bijection *is a measure preserving map whose inverse map exists (and is also measure preserving).*

Let $\mathcal{F}$ denote the collection of all measure preserving bijections from $[0,1]^2$ to itself. Given an orderon $W \in \mathcal{W}$ and $f \in \mathcal{F}$, we define W^f as the unique orderon satisfying $W^f((x,a),(y,b)) = W(f(x,a),f(y,b))$ for any $x,a,y,b \in [0,1]$. Additionally, denote by $\pi_1\colon [0,1]^2 \to [0,1]$ the projection to the first coordinate, that is, $\pi_1(x,a) = x$ for any $(x,a) \in [0,1]^2$.

2.1 Cut-norm and ordered cut-norm

The definition of the (unordered) cut-norm for orderons is analogous to the corresponding definition for graphons.

▶ **Definition 14** (cut-norm). *Given a symmetric measurable function $W\colon ([0,1]^2)^2 \to \mathbb{R}$, we define the* cut-norm *of W as*

$$\|W\|_\square \stackrel{\text{def}}{=} \sup_{S,T\subseteq[0,1]^2} \left| \int_{(x,a)\in S\ (y,b)\in T} W((x,a),(y,b))dxdadydb \right| .$$

As we are working with ordered objects, the following definition of *ordered cut-norm* will sometimes be of use (see the full version [9] for more details). Given $v_1,v_2 \in [0,1]^2$, we write $v_1 \le v_2$ to denote that $\pi_1(v_1) \le \pi_1(v_2)$. Let $\mathbf{1}_E$ be the indicator function for the event E.

▶ **Definition 15** (ordered cut-norm). *Let $W\colon ([0,1]^2)^2 \to \mathbb{R}$ be a symmetric measurable function. The* ordered cut norm *of W is defined as*

$$\|W\|_{\square'} = \sup_{S,T\subseteq[0,1]^2} \left| \int_{(v_1,v_2)\in S\times T} W(v_1,v_2)\mathbf{1}_{v_1\le v_2} dv_1 dv_2 \right| .$$

We mention two important properties of the ordered-cut norm. The first is a standard smoothing lemma, and the second is a relation between the ordered cut-norm and the unordered cut-norm. The proof of both lemmas can be found in the full version [9].

▶ **Lemma 16.** *Let $W \in \mathcal{W}$ and $\mu,\nu : [0,1]^2 \to [0,1]$. Then,*

$$\left| \int_{v_1,v_2} \mu(v_1)\nu(v_2)W(v_1,v_2)\mathbf{1}_{v_1\le v_2} dv_1 dv_2 \right| \le \|W\|_{\square'}.$$

▶ **Lemma 17.** *Let $W\colon ([0,1]^2)^2 \to [-1,1]$ be a symmetric measurable function. Then,*

$$\frac{\|W\|_{\square'}^2}{4} \le \|W\|_\square \le 2\|W\|_{\square'} .$$

2.2 The cut and shift distance

The next notion of distance is a central building block in this work. It can be viewed as a locality preserving variant of the unordered cut distance, which accounts for order changes resulting from applying a measure preserving function.

▶ **Definition 18.** *Given two orderons $W, U \in \mathcal{W}$ we define the* CS-distance *(cut-norm+shift distance) as:*

$$d_{\triangle}(W,U) \stackrel{\text{def}}{=} \inf_{f\in\mathcal{F}} \left(\mathrm{Shift}(f) + \|W - U^f\|_{\square}\right),$$

where $\mathrm{Shift}(f) \stackrel{\text{def}}{=} \sup_{x,a\in[0,1]} |x - \pi_1(f(x,a))|$.

▶ **Lemma 19.** *$d_{\triangle}$ is a pseudo-metric on the space of orderons.*

For the proof, see the full version [9].

3 Block orderons and their density in $\mathcal{W}$

Here we show that weighted ordered graphs are dense in the space of orderons coupled with the cut-shift distance. To start, we have to define the orderon representation of a weighted ordered graph, called a *naive block orderon*. A naive n-block orderon is defined as follows.

▶ **Definition 20** (naive block orderon). *Let $m \in \mathbb{N}$. For $z \in (0,1]$, we denote $Q_n(z) = \lceil nz \rceil$; we also set $Q_n(0) = 1$. An m-*block naive orderon *is a function $W\colon \left([0,1]^2\right)^2 \to [0,1]$ that can be written, for some weighted ordered graph G on n vertices, as*

$$W((x,a),(y,b)) = G(Q_n(x), Q_n(y)), \qquad \forall x,a,y,b \in [0,1].$$

Following the above definition, we denote by W_G the naive block orderon defined using G, and view W_G as the orderon "representing" G in $\mathcal{W}$. Similarly to the unordered setting, this representation is slightly ambiguous (but this will not affect us). Indeed, it is not hard to verify that two weighted ordered graphs F and G satisfy $W_F = W_G$ if and only if both F and G are blowups of some weighted ordered graph H. Here, a weighted ordered graph G on nt vertices is a *t-blowup* of a weighted ordered graph H on n vertices if $G(x,y) = H(\lceil x/t \rceil, \lceil y/t \rceil)$ for any $x, y \in [nt]$.

We call an orderon $U \in \mathcal{W}$ a *step function with at most k steps* if there is a partition $\mathcal{R} = \{S_1, \ldots, S_k\}$ of $[0,1]^2$ such that U is constant on every $S_i \times S_j$.

▶ Remark 21 (The name choices). The definition of a step function in the space of orderons is the natural extension of a step function in graphons. Note that a naive block orderon is a special case of a step function, where the steps S_i are rectangular (this is why we call these "block orderons"). The "naive" prefix refers to the fact that we do not make use of the second coordinate in the partition.

For every $W \in \mathcal{W}$ and every partition $\mathcal{P} = \{S_1, \ldots, S_k\}$ of $[0,1]^2$ into measurable sets, let $W_{\mathcal{P}}\colon ([0,1]^2)^2 \to [0,1]$ denote the step function obtained from W by replacing its value at $((x,a),(y,b)) \in S_i \times S_j$ by the average of W on $S_i \times S_j$. That is,

$$W_{\mathcal{P}}((x,a),(y,b)) = \frac{1}{\lambda(S_i)\lambda(S_j)} \int_{S_i\times S_j} W((x',a'),(y',b'))dx'da'dy'db',$$

Where i and j are the unique indices such that $(x,a) \in S_i$ and $(y,b) \in S_j$, respectively.

The next lemma is an extension of the regularity lemma to the setting of Hilbert spaces.

▶ **Lemma 22** ([29], Lemma 4.1). *Let $\{\mathcal{K}_i\}_i$ be arbitrary non-empty subsets of a Hilbert space $\mathcal{H}$. Then, for every $\varepsilon > 0$ and $f \in \mathcal{H}$ there is an $m \le \lceil 1/\varepsilon^2 \rceil$ and there are $f_i \in \mathcal{K}_i$ ($1 \le i \le k$) and $\gamma_1 \dots, \gamma_k \in \mathbb{R}$ such that for every $g \in \mathcal{K}_{k+1}$*

$$|\langle g, f - (\gamma_1 f_1 + \cdots + \gamma_k f_k)\rangle| \le \varepsilon \|f\| \|g\|$$

The following is a direct consequence of Lemma 22.

▶ **Lemma 23.** *For every $W \in \mathcal{W}$ and $\varepsilon > 0$ there is a step function $U \in \mathcal{W}$ with at most $\lceil 2^{8/\varepsilon^2} \rceil$ steps such that*

$$\|W - U\|_\square \le \varepsilon \, .$$

Similarly to the graphon case, the step function U might not be a stepping of W. However, it can be shown that these steppings are almost optimal.

▷ Claim 24. Let $W \in \mathcal{W}$, let U be a step function, and let $\mathcal{P}$ denote the partition of $[0,1]^2$ into the steps of U. Then $\|W - W_\mathcal{P}\|_\square \le 2\|W - U\|_\square$.

Using Lemma 23 and Claim 24 we can obtain the following lemma.

▶ **Lemma 25.** *For every function $W \in \mathcal{W}$ and every $\varepsilon > 0$, there is a partition $\mathcal{P}$ of $[0,1]^2$ into at most $2^{\lceil 32/\varepsilon^2 \rceil}$ sets with positive measure such that $\|W - W_\mathcal{P}\|_\square \le \varepsilon$.*

Using the above lemma, we can impose stronger requirements on our partition. In particular, we can show that there exists a partition of $[0,1]^2$ to sets of the same measure. Such a partition is referred to as an *equipartition*. Also, we say that a partition $\mathcal{P}$ *refines* $\mathcal{P}'$, if $\mathcal{P}$ can be obtained from $\mathcal{P}'$ by splitting each $P_j \in \mathcal{P}'$ into a finite number of sets (up to sets of measure 0).

▶ **Lemma 26.** *Fix some $\varepsilon > 0$. Let $\mathcal{P}$ be an equipartition of $[0,1]^2$ into k sets, and fix $q \ge 2k^2 \cdot 2^{162/\varepsilon^2}$ such that k divides q. Then, for any $W \in \mathcal{W}$, there exists an equipartition $\mathcal{Q}$ that refines $\mathcal{P}$ with q sets, such that $\|W - W_\mathcal{Q}\|_\square \le \frac{8\varepsilon}{9} + \frac{2}{k}$.*

The next lemma is an (easier) variant of Lemma 26, in the sense that we refine two given partitions. However, the resulting partition will not be an equipartition.

▶ **Lemma 27.** *Fix some $\varepsilon > 0$ and $d \in \mathbb{N}$. Let $\mathcal{I}_d$ be an equipartition of $[0,1]^2$ into 2^d sets, $\mathcal{P}$ be a partition of $[0,1]^2$ into k sets, and fix $q \ge 2(k \cdot 2^d)^2 \cdot 2^{162/\varepsilon^2}$ such that both k and 2^d divide q. Then, for any $W \in \mathcal{W}$, there exists a partition $\mathcal{Q}$ that refines both $\mathcal{P}$ and $\mathcal{I}_d$ with q sets, such that $\|W - W_\mathcal{Q}\|_\square \le \frac{8\varepsilon}{9} + \frac{2}{k \cdot 2^d}$.*

Proof. Let $\mathcal{P}' = \{P'_1, \dots, P'_{p'}\}$ be a partition of $[0,1]^2$ into $p' \le 2^{162/\varepsilon^2}$ sets such that $\|W - W_{\mathcal{P}'}\|_\square \le \frac{4\varepsilon}{9}$, and let $\mathcal{Q} = \{Q_1, \dots, Q_q\}$ be a common refinement of the three partitions $\mathcal{P}$, $\mathcal{P}'$ and $\mathcal{I}_d$. Note that we do not repartition further to get an equipartition. The rest of the proof is similar to the proof of Lemma 26. ◀

The following theorem shows that naive block orderons are a dense subset in $\mathcal{W}$.

▶ **Theorem 28.** *For every orderon $W \in \mathcal{W}$ and every $\varepsilon > 0$, there exist a naive $\frac{c}{\varepsilon^4} 2^{162/\varepsilon^2}$-block orderon W' (for some constant $c > 0$) such that*

$$d_\triangle(W, W') \le \varepsilon \, .$$

Proof. Fix $\varepsilon > 0$ and $\gamma = \gamma(\varepsilon) > 0$. We consider an interval equipartition $J = \{J_1, \dots, J_{1/\gamma}\}$ of $[0,1]$ (namely, for each $j \in [\frac{1}{\gamma} - 1]$, $J_j = [(j-1)\cdot\gamma, j\cdot\gamma)$, and for $j = 1/\gamma$, $J_j = [(j-1)\cdot\gamma, j\cdot\gamma])$. In addition, let $\mathcal{P} = (J_i \times J_j \mid i,j \in [1/\gamma])$ be an equipartition of $[0,1]^2$. By Lemma 26, there exists an equipartition $\mathcal{Q}$ of $[0,1]^2$ of size $q = \frac{2}{\gamma^4}2^{162/\varepsilon^2}$ that refines $\mathcal{P}$, such that

$$\|W - W_{\mathcal{Q}}\|_\square \le \frac{8\varepsilon}{9} + 2\gamma^2 .$$

Next we construct a small shift measure preserving function f as follows. For every $i \in [1/\gamma]$, consider the collection of sets $\{Q^i_k \mid k \in [\gamma q]\}$ in $\mathcal{Q}$ such that

$$(J_i \times [0,1]) \cap \mathcal{Q} = \{Q^i_k \mid k \in [\gamma q]\} .$$

For each $k \in [\gamma q]$, the function f maps Q^i_k to a rectangular set

$$\left[(i-1)\gamma + \frac{(k-1)}{q}, (i-1)\gamma + \frac{k}{q}\right) \times [0,1] .$$

Finally, for every $i,j \in [q]$ and every $(x,a),(y,b) \in Q_i \times Q_j$, we define

$$W'(f(x,a), f(y,b)) = W_{\mathcal{Q}}((x,a),(y,b))$$

Note that the resulting function W' obeys the definition of a naive q-block orderon and $\mathrm{Shift}(f) \le \gamma$. Therefore, setting $\gamma = \varepsilon/100$, we get that

$$d_\triangle(W, W') \le \gamma + \frac{8\varepsilon}{9} + 2\gamma^2 \le \varepsilon/100 + 8\varepsilon/9 + 2\varepsilon^2/100^2 \le \varepsilon ,$$

as desired. ◀

4 Compactness of the space of orderons

In this section we prove Theorem 1. We construct a metric space $\widetilde{\mathcal{W}}$ from $\mathcal{W}$ with respect to $d_\triangle$, by identifying $W, U \in \mathcal{W}$ with $d_\triangle(W,U) = 0$. Let $\widetilde{\mathcal{W}}$ be the image of $\mathcal{W}$ under this identification. On $\widetilde{\mathcal{W}}$ the function $d_\triangle$ is a distance function.

We start with some definitions and notations. Let $(\Omega, \mathcal{M}, \lambda)$ be some probability space, $\mathcal{P}_\ell = \left\{P^{(\ell)}_i\right\}_i$ a partition of Ω, and let $\beta(\mathcal{P}_\ell : \cdot) : \mathcal{P}_\ell \to [0,1]$ be a function. For $v \in \Omega$, we slightly abuse notation and write $\beta(\mathcal{P}_\ell : v)$ to denote $\beta(\mathcal{P}_\ell : i)$ for $v \in P^{(\ell)}_i$. With this notation, observe that for every ℓ

$$\int_{v\in\Omega} \beta(\mathcal{P}_\ell : v) dv = \sum_{i \in [|\mathcal{P}_\ell|]} \lambda\left(P^{(\ell)}_i\right) \beta(\mathcal{P}_\ell : i) . \tag{2}$$

The following two results serve as useful tools to prove convergence. The first result is known as the *martingale convergence theorem*, see e.g. Theorem A.12 in [27]. The second result is an application of the martingale convergence theorem, useful for our purposes.

▶ **Theorem 29** (see [27], Theorem A.12)**.** *Let $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ be a martingale satisfying $\sup_n \mathbf{E}[|\mathbf{X}_n|] < \infty$. Then $\{\mathbf{X}_i\}_{i\in\mathbb{N}}$ is convergent with probability 1.*

▶ **Lemma 30.** *Let $\{\mathcal{P}_\ell\}_\ell$ be a sequence of partitions of Ω such that for every ℓ, $\mathcal{P}_{\ell+1}$ refines $\mathcal{P}_\ell$. Assume that for every ℓ and $j \in [|\mathcal{P}_\ell|]$, the functions $\beta(\mathcal{P}_\ell : \cdot)$ satisfy*

$$\lambda\left(P^{(\ell)}_j\right) \beta(\mathcal{P}_\ell : j) = \sum_{i \in [|\mathcal{P}_{\ell+1}|]} \lambda\left(P^{(\ell)}_j \cap P^{(\ell+1)}_i\right) \beta(\mathcal{P}_{\ell+1} : i). \tag{3}$$

Then, there is a measurable function $\beta\colon \Omega \to [0,1]$ such that $\beta(v) = \lim_{\ell\to\infty} \beta(\mathcal{P}_\ell : v)$ for almost all $v \in \Omega$.

Proof. Fix some $\ell \in \mathbb{N}$. Let $\mathbf{X}$ be a uniformly distributed random variable in Ω. Let $\psi_\ell \colon \Omega \to [|\mathcal{P}_\ell|]$ be the function mapping each $v \in \Omega$ to its corresponding part in $\mathcal{P}_\ell$ and let $\mathbf{Z}_\ell = \beta(\mathcal{P}_\ell : \mathbf{X})$. We now show that the sequence $(\mathbf{Z}_1, \mathbf{Z}_2 \dots)$ is a martingale. That is, $\mathbf{E}_{\mathbf{X}\sim\Omega}[\mathbf{Z}_{\ell+1} \mid \mathbf{Z}_1, \dots, \mathbf{Z}_\ell] = \mathbf{Z}_\ell$, for every $\ell \in \mathbb{N}$. Note that by the fact that $\mathcal{P}_{\ell+1}$ refines $\mathcal{P}_\ell$, $\psi_\ell(\mathbf{X})$ determines $\psi_i(\mathbf{X})$ for every $i < \ell$. By definition, the value $\beta(\mathcal{P}_\ell : \mathbf{X})$ is completely determined by $\psi_\ell(\mathbf{X})$, and so it suffices to prove that $\mathbf{Z}_\ell = \mathbf{E}_{\mathbf{X}\sim\Omega}[\mathbf{Z}_{\ell+1} \mid \psi_\ell(\mathbf{X})]$. By the fact that for every $j \in [|\mathcal{P}_\ell|]$ Equation (3) holds (and in particular holds for $\psi_\ell(\mathbf{X})$), we can conclude that the sequence $(\mathbf{Z}_1, \mathbf{Z}_2, \dots)$ is a martingale.

Since $\mathbf{Z}_\ell$ is bounded, we can invoke the martingale convergence theorem (Theorem 29) and conclude that $\lim_{\ell\to\infty} \mathbf{Z}_\ell$ exists with probability 1. That is, $\beta(v) = \lim_{\ell\to\infty} \beta(\mathcal{P}_\ell : v)$ exists for almost all $v \in \Omega$. ◀

▶ **Definition 31.** *Fix some $d \in \mathbb{N}$ and define $\mathcal{I}_d = \left\{I_1^{(d)}, \dots, I_{2^d}^{(d)}\right\}$ so that for every $t \in \left[2^d\right]$, $I_t^{(d)} = \left[\frac{t-1}{2^d}, \frac{t}{2^d}\right) \times [0,1]$. We refer to this partition as the* strip partition of order d.

The next lemma states that for any orderon W we can get a sequence of partitions $\{\mathcal{P}_\ell\}_\ell$, with several properties that will be useful later on.

▶ **Lemma 32.** *For any orderon $W \in \mathcal{W}$ and $\ell \in \mathbb{N}$, there is a sequence of partitions $\{\mathcal{P}_\ell\}_\ell$ of $[0,1]^2$ with the following properties.*

1. *$\mathcal{P}_\ell$ has $g(\ell)$ many sets (for some monotone increasing $g\colon \mathbb{N} \to \mathbb{N}$).*
2. *For every ℓ, $\Gamma_\ell \stackrel{\text{def}}{=} \frac{g(\ell)}{g(\ell-1)} \in \mathbb{N}$.*
3. *For every $\ell' \geq \ell$, the partition $\mathcal{P}_{\ell'}$ refines both $\mathcal{P}_\ell$ and the strip partition $\mathcal{I}_{\ell'}$. In particular, for every $j \in [g(\ell-1)]$,*

$$P_j^{(\ell-1)} = \bigcup_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} P_{j'}^{(\ell)} .$$

4. *$W_\ell = (W)_{\mathcal{P}_\ell}$ satisfies $\|W - W_\ell\|_\square \leq \frac{4}{g(\ell-1)2^\ell}$.*

Proof. We invoke Lemma 27 with the trivial partition $\{[0,1]^2\}$ and the strip partition $\mathcal{I}_1$, to get a partition $\mathcal{P}_{n,1}$ with $g(1)$ many sets such that $\mathcal{P}_{n,1}$ refines $\mathcal{I}_1$ and $\|W_n - W_{n,1}\|_\square \leq 1$. For $\ell > 1$, we invoke Lemma 27 with $\mathcal{I}_\ell$ and $\mathcal{P}_{n,\ell-1}$ to get a partition $\mathcal{P}_{n,\ell}$ of size $g(\ell) = (g(\ell-1)\cdot 2^\ell)^2 \cdot 2^{O(g(\ell-1)^2)}$ which refines both $\mathcal{I}_\ell$ and $\mathcal{P}_{n,\ell-1}$ such that $\|W_n - W_{n,\ell}\|_\square \leq \frac{4}{g(\ell-1)2^\ell}$. In order to take care of divisibility, we add empty (zero measure) sets in order to satisfy items (2) and (3). ◀

Consider a sequence of orderons $\{W_n\}_{n\in\mathbb{N}}$. For every $n \in \mathbb{N}$, we use Lemma 32 to construct a sequence of functions $\{W_{n,\ell}\}_\ell$ such that $\|W_n - W_{n,\ell}\|_\square$ is small. For each ℓ, we would like to approximate the shape of the limit partition resulting from taking $n \to \infty$. Inside each strip $I_t^{(\ell)}$, we consider the relative measure of the intersection of each set contained in $I_t^{(\ell)}$, with a finer strip partition $\mathcal{I}_{\ell'}$.

▶ **Definition 33** (shape function). *For fixed $n \in \mathbb{N}$, let $\{\mathcal{P}_{n,\ell}\}_\ell$ be partitions of $[0,1]^2$ with the properties listed in Lemma 32. For every $\ell' > \ell$ and $I_{t'}^{(\ell')} \in \mathcal{I}_{\ell'}$, we define $\alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t') \stackrel{\text{def}}{=} 2^{\ell'} \cdot \lambda\left(P_j^{(n,\ell)} \cap I_{t'}^{(\ell')}\right)$ to be the relative volume of the set $P_j^{(n,\ell)}$ in $I_{t'}^{(\ell')}$.*

For any $\ell' \geq \ell$ and $I_{t'}^{(\ell')} \in \mathcal{I}_{\ell'}$, by the compactness of $[0,1]$, we can select a subsequence of $\{W_n\}_{n\in\mathbb{N}}$ such that $\alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t')$ converges for all $j \in [g(\ell)]$ as $n \to \infty$. Let

$$\alpha_j^{(\ell)}(\mathcal{I}_{\ell'} : t') \stackrel{\text{def}}{=} \lim_{n\to\infty} \alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t') .$$

Next we define the limit density function.

▶ **Definition 34** (density function). *For fixed $n \in \mathbb{N}$, let $\{\mathcal{P}_{n,\ell}\}_\ell$ be partitions of $[0,1]^2$ with the properties listed in Lemma 32. We let $\delta^{(n,\ell)}(\mathcal{P}_{n,\ell} \times \mathcal{P}_{n,\ell} : i, j) \stackrel{\text{def}}{=} W_{n,\ell}((x,a),(y,b))$ for $(x,a) \in P_i^{(n,\ell)}$ and $(y,b) \in P_j^{(n,\ell)}$.*

By the compactness of $[0,1]$, we can select a subsequence of $\{W_n\}_{n\in\mathbb{N}}$ such that $\delta^{(n,\ell)}(\mathcal{P}_{n,\ell} \times \mathcal{P}_{n,\ell} : i, j)$ converge for all $i, j \in [g(\ell)]$ as $n \to \infty$. Let

$$\delta^{(\ell)}(i,j) \stackrel{\text{def}}{=} \lim_{n\to\infty} \delta^{(n,\ell)}(\mathcal{P}_{n,\ell} \times \mathcal{P}_{n,\ell} : i, j) .$$

The following lemma states that by taking increasingly refined strip partitions $\mathcal{I}_{\ell'}$, we obtain a limit shape function for each set contained in any strip of $\mathcal{I}_\ell$.

▶ **Lemma 35.** *For fixed ℓ and $j \in [g(\ell)]$, there is a measurable function $\alpha_j^{(\ell)} : [0,1] \to [0,1]$ such that $\alpha_j^{(\ell)}(x) = \lim_{\ell'\to\infty} \alpha_j^{(\ell)}(\mathcal{I}_{\ell'} : x)$ for almost all $x \in [0,1]$.*

Proof. Fix n, ℓ and $\ell' > \ell$. For every $j \in [g(\ell)]$, by the definition of $\alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t')$ and the strip partition $\mathcal{I}_{\ell'}$

$$\lambda\left(I_{t'}^{(\ell')}\right) \cdot \alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t') = \lambda\left(P_j^{(n,\ell)} \cap I_{t'}^{(\ell')}\right) \qquad \forall t' \in \left[2^\ell\right] .$$

On the other hand, since $\mathcal{I}_{\ell'+1}$ refines $\mathcal{I}_{\ell'}$,

$$\begin{aligned}\lambda\left(P_j^{(n,\ell)} \cap I_{t'}^{(\ell')}\right) &= \lambda\left(P_j^{(n,\ell)} \cap I_{2t'-1}^{(\ell'+1)}\right) + \lambda\left(P_j^{(n,\ell)} \cap I_{2t'}^{(\ell'+1)}\right) \\ &= \lambda\left(I_{2t'-1}^{(\ell'+1)}\right) \cdot \alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'+1} : 2t'-1) + \lambda\left(I_{2t'}^{(\ell'+1)}\right) \cdot \alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'+1} : 2t') .\end{aligned}$$

Therefore, when $n \to \infty$ we get that,

$$\lambda\left(I_{t'}^{(\ell')}\right) \cdot \alpha_j^{(\ell)}(\mathcal{I}_{\ell'} : t') = \lambda\left(I_{2t'-1}^{(\ell'+1)}\right) \cdot \alpha_j^{(\ell)}(\mathcal{I}_{\ell'+1} : 2t'-1) + \lambda\left(I_{2t'}^{(\ell'+1)}\right) \cdot \alpha_j^{(\ell)}(\mathcal{I}_{\ell'+1} : 2t') ,$$

which is exactly the condition in Equation (3). By applying Lemma 30 with the sequence of strip partitions $\{\mathcal{I}_{\ell'}\}_{\ell'}$ on $\alpha_j^{(\ell)}$ the lemma follows. ◀

The next lemma asserts that the limit shape functions behave consistently.

▶ **Lemma 36.** *For every ℓ and $j \in [g(\ell-1)]$,*

$$\alpha_j^{(\ell-1)}(x) = \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \alpha_{j'}^{(\ell)}(x) ,$$

for almost all $x \in [0,1]$.

Proof. Fix some n, ℓ and $\ell' > \ell$. By the additivity of the Lebesgue measure,

$$\alpha_j^{(n,\ell-1)}(\mathcal{I}_{\ell'} : x) = \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \alpha_{j'}^{(n,\ell)}(\mathcal{I}_{\ell'} : x) \qquad \forall x \in [0,1] .$$

By the fact that for every $j \in [g(\ell-1)]$ and $x \in [0,1]$ the sequence $\left\{\alpha_j^{(n,\ell-1)}(\mathcal{I}_{\ell'} : x)\right\}_n$ converges to $\alpha_j^{(\ell-1)}(\mathcal{I}_{\ell'} : x)$ as $n \to \infty$, we get that

$$\alpha_j^{(\ell-1)}(\mathcal{I}_{\ell'} : x) = \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \alpha_{j'}^{(\ell)}(\mathcal{I}_{\ell'} : x) \qquad \forall x \in [0,1] .$$

By applying Lemma 35 on each $j' \in [g(\ell)]$, where $\ell' \to \infty$, we get that

$$\alpha_j^{(\ell-1)}(x) = \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \alpha_{j'}^{(\ell)}(x) \ ,$$

for almost all $x \in [0,1]$. ◀

Using the sequence of $\left\{\alpha_j^{(\ell)}\right\}_j$ we define a limit partition $\mathcal{A}_\ell = \left\{A_1^{(\ell)}, \ldots, A_{g(\ell)}^{(\ell)}\right\}$ of $[0,1]^2$ as follows.

▶ **Definition 37** (limit partition). *For every $\ell \in \mathbb{N}$, let $\mathcal{A}_\ell = \left\{A_1^{(\ell)}, \ldots, A_{g(\ell)}^{(\ell)}\right\}$ be a partition of $[0,1]^2$ such that,*

$$A_j^{(\ell)} = \left\{(x,a) \colon \sum_{i<j} \alpha_i^{(\ell)}(x) \le a < \sum_{i \le j} \alpha_i^{(\ell)}(x)\right\} \qquad \forall j \in [g(\ell)] \ .$$

▶ **Lemma 38.** *For any ℓ, the partition $\mathcal{A}_\ell$ has the following properties*
1. *$\mathcal{A}_\ell$ refines the strip partition $\mathcal{I}_\ell$.*
2. *The partition $\mathcal{A}_\ell$ refines $\mathcal{A}_{\ell-1}$.*
3. *For every $j \in [g(\ell)]$, $\lambda\left(A_j^{(\ell)}\right) = \lim_{n\to\infty} \lambda\left(P_j^{(n,\ell)}\right)$.*

Proof. The first item follows by the fact that each $\alpha_j^{(\ell)}$ is non-zero inside only one strip.

By the definition of the sets $A_j^{(\ell)}$ and Lemma 36 it follows that for each $j \in [g(\ell-1)]$,

$$A_{j'}^{(\ell)} \subset A_j^{(\ell-1)} \qquad \text{for all} \qquad (j-1)\cdot\Gamma_\ell + 1 \le j' \le j \cdot \Gamma_\ell,$$

and therefore,

$$A_j^{(\ell-1)} = \bigcup_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} A_{j'}^{(\ell)},$$

which shows the second item. To prove the third item of the lemma, note that for every n, ℓ and $\ell' > \ell$,

$$\begin{aligned}\lim_{n\to\infty} \lambda\left(P_j^{(n,\ell)}\right) &= \lim_{n\to\infty} \sum_{t'\in[2^{\ell'}]} 2^{-\ell'} \cdot \alpha_j^{(n,\ell)}(\mathcal{I}_{\ell'} : t') \\ &= \sum_{t'\in[2^{\ell'}]} 2^{-\ell'} \cdot \alpha_j^{(\ell)}(\mathcal{I}_{\ell'} : t') = \int_x \alpha_j^{(\ell)}(\mathcal{I}_{\ell'} : x)dx,\end{aligned}$$

where the last equality follows from Equation (2). Finally, by taking $\ell' \to \infty$ and using Lemma 35, we get

$$\lim_{n\to\infty} \lambda\left(P_j^{(n,\ell)}\right) = \int_x \alpha_j^{(\ell)}(x)dx = \lambda\left(A_j^{(\ell)}\right)$$

as desired. ◀

Using the definition of $\delta^{(\ell)}$ and $\mathcal{A}_\ell$, we define a density function on the limit partition. For $(x,a) \in A_i^{(\ell)}$ and $(y,b) \in A_j^{(\ell)}$, let

$$\delta\left(\mathcal{A}_\ell \times \mathcal{A}_\ell : (x,a),(y,b)\right) \stackrel{\text{def}}{=} \delta^{(\ell)}(i,j) \ .$$

▶ **Lemma 39.** *For each $\ell \in \mathbb{N}$ and $i, j \in [g(\ell-1)]$,*

$$\sum_{i'=(i-1)\cdot\Gamma_\ell+1}^{i\cdot\Gamma_\ell} \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \lambda\left(A_{i'}^{(\ell)}\right)\cdot\lambda\left(A_{j'}^{(\ell)}\right)\delta\left(\mathcal{A}_\ell\times\mathcal{A}_\ell : i', j'\right) = \lambda\left(A_i^{(\ell-1)}\right)\cdot\lambda\left(A_j^{(\ell-1)}\right)\delta\left(\mathcal{A}_{\ell-1}\times\mathcal{A}_{\ell-1} : i, j\right) .$$

Proof. Fix n, ℓ and $i, j \in [g(\ell-1)]$. By the definition of the partitions $\mathcal{P}_{n,\ell}$, $\mathcal{P}_{n,\ell-1}$ and the density functions $\delta^{(n,\ell)}$, $\delta^{(n,\ell-1)}$

$$\sum_{i'=(i-1)\cdot\Gamma_\ell+1}^{i\cdot\Gamma_\ell} \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \lambda\left(P_{i'}^{(n,\ell)}\right)\cdot\lambda\left(P_{j'}^{(n,\ell)}\right)\delta^{(n,\ell)}\left(\mathcal{P}_{n,\ell}\times\mathcal{P}_{n,\ell} : i', j'\right) = \lambda\left(P_i^{(n,\ell-1)}\right)\cdot\lambda\left(P_j^{(n,\ell-1)}\right)\delta\left(\mathcal{P}_{n,\ell-1}\times\mathcal{P}_{n,\ell-1} : i, j\right) .$$

By taking the limit as $n \to \infty$ and using the third item of Lemma 38,

$$\sum_{i'=(i-1)\cdot\Gamma_\ell+1}^{i\cdot\Gamma_\ell} \sum_{j'=(j-1)\cdot\Gamma_\ell+1}^{j\cdot\Gamma_\ell} \lambda\left(A_{i'}^{(\ell)}\right)\cdot\lambda\left(A_{j'}^{(\ell)}\right)\delta\left(\mathcal{A}_\ell\times\mathcal{A}_\ell : i', j'\right) = \lambda\left(A_i^{(\ell-1)}\right)\cdot\lambda\left(A_j^{(\ell-1)}\right)\delta\left(\mathcal{A}_{\ell-1}\times\mathcal{A}_{\ell-1} : i, j\right) ,$$

which completes the proof. ◀

The next Lemma asserts that the natural density function of the limit partition is measurable. It follows directly from the combination of Lemma 30 and Lemma 39.

▶ **Lemma 40.** *There exists a measurable function $\delta : ([0,1]^2)^2 \to [0,1]$ such that $\delta((x,a),(y,a)) = \lim_{\ell\to\infty} \delta\left(\mathcal{A}_\ell\times\mathcal{A}_\ell : (x,a),(y,b)\right)$ for almost all $(x,a),(y,b) \in ([0,1]^2)^2$.*

Finally, we are ready to prove Theorem 1.

Proof of Theorem 1. We start by giving a high-level overview of the proof. Let $\{W_n\}_{n\in\mathbb{N}}$ be a sequence of functions in $\mathcal{W}$. We show that there exists a subsequence that has a limit in $\widetilde{\mathcal{W}}$.

For every $n \in \mathbb{N}$, we use Lemma 32 to construct a sequence of functions $\{W_{n,\ell}\}_\ell$ such that $\|W_n - W_{n,\ell}\|_\square \le \frac{4}{g(\ell-1)2^\ell}$. Then, for every fixed $\ell \in \mathbb{N}$, we find a subsequence of $\{W_{n,\ell}\}$ such that their corresponding $\alpha_j^{(n,\ell)}$ and $\delta^{(n,\ell)}(i,j)$ converge for all $i, j \in [g(\ell)]$ (as $n \to \infty$). For every ℓ, we consider the partition $\mathcal{A}_\ell$ (which by Definition 37, is determined by $\{\alpha_j^{(\ell)}\}_j$) and $\delta^{(\ell)}$. Using $\mathcal{A}_\ell$ and $\delta^{(\ell)}$, we can the define the function U_ℓ, such that $W_{n,\ell} \to U_\ell$ almost everywhere as $n \to \infty$.

Given the sequence of functions $\{U_\ell\}_\ell$, we use Lemma 40 to show that $\{U_\ell\}_\ell$ converges to some U almost everywhere as $\ell \to \infty$ (where U is defined according the limit density function δ). Finally we show that for any fixed $\varepsilon > 0$, there is $n_0(\varepsilon)$ such that for any $n > n_0(\varepsilon)$, $d_\triangle(W_n, U) \le \varepsilon$.

Fix some $\varepsilon > 0$ and $\xi(\varepsilon) > 0$ which will be determined later. Consider the sequence $\{U_\ell\}_\ell$ which is defined by the partition $\mathcal{A}_\ell$ and the density function $\delta^{(\ell)}$. By Lemma 40, the sequence $\{U_\ell\}_\ell$ converges (as $\ell \to \infty$) almost everywhere to U, which is defined by the limit density function δ. Therefore, we can find some $\ell > 1/\xi$ such that $\|U_\ell - U\|_1 \le \xi$.

Fixing this ℓ, we show that there is n_0 such that $d_\triangle(W_{n,\ell}, U_\ell) \le 2^{-\ell} + 3\xi$ for all $n > n_0$. We shall do it in two steps by defining an interim function $W'_{n,\ell}$ and using the triangle inequality.

Recall that the function $W_{n,\ell}$ is defined according to the partition $\mathcal{P}_{n,\ell}$ and the density function $\delta^{(n,\ell)}$. Let $W'_{n,\ell}$ be the function defined according to the partition $\mathcal{A}_\ell$ and the density function $\delta^{(n,\ell)}$. That is, for every $(x,a) \in A_i^{(\ell)}$ and $(y,b) \in A_j^{(\ell)}$, $W'_{n,\ell}((x,a),(y,b)) \stackrel{\text{def}}{=} \delta^{(n,\ell)}(\mathcal{P}_{n,\ell} \times \mathcal{P}_{n,\ell} : i,j)$. By the third item of Lemma 38, for every $j \in [g(\ell)]$, $\lambda\left(A_j^{(\ell)}\right) = \lim_{n\to\infty} \lambda\left(P_j^{(n,\ell)}\right)$. Then, we can find $n'_0(\ell)$ such that for all $n > n'_0$,

$$\max\left(\lambda\left(A_j^{(\ell)}\right), \lambda\left(P_j^{(n,\ell)}\right)\right) - \min\left(\lambda\left(A_j^{(\ell)}\right), \lambda\left(P_j^{(n,\ell)}\right)\right) \leq \frac{\xi}{g(\ell)} \qquad \forall j \in [g(\ell)]\,. \tag{4}$$

We define a measure preserving map f from $W_{n,\ell}$ to $W'_{n,\ell}$ as follows. For every strip $I_t^{(\ell)} \in \mathcal{I}_\ell$, we consider all the sets $\{P_{j_1}^{(n,\ell)} \dots, P_{j_t}^{(n,\ell)}\}$ in $\mathcal{P}_{n,\ell}$ such that $\bigcup_{j'=j_1}^{j_t} P_{j'}^{(n,\ell)} = I_t^{(\ell)}$. Similarly, consider all the sets $\{A_{j_1}^{(\ell)} \dots, A_{j_t}^{(\ell)}\}$ in $\mathcal{A}_\ell$ such that $\bigcup_{j'=j_1}^{j_t} A_{j'}^{(\ell)} = I_t^{(\ell)}$. For every $j' \in \{j_1,\dots,j_t\}$, we map an arbitrary subset $S_{j'}^{(n,\ell)} \subseteq P_{j'}^{(n,\ell)}$ of measure $\min\left(\lambda\left(A_{j'}^{(\ell)}\right), \lambda\left(P_{j'}^{(n,\ell)}\right)\right)$ to an arbitrary subset (with the same measure) of $A_{j'}^{(\ell)}$. Next, we map $I_t^{(\ell)} \setminus \bigcup_{j'=j_1}^{j_t} S_{j'}^{(n,\ell)}$ to $I_t^{(\ell)} \setminus \bigcup_{j'=j_1}^{j_t} f(S_{j'}^{(n,\ell)})$. Note that by (4) and the fact that $W_{n,\ell}$ and $W'_{n,\ell}$ have the same density function $\delta^{(n,\ell)}$, the functions $W_{n,\ell}$ and $W'_{n,\ell}$ disagree on a set of measure at most 2ξ. Note that for every $I_t^{(\ell)} \in \mathcal{I}_\ell$, the function f maps sets from $\mathcal{P}_{n,\ell}$ that are contained in $I_t^{(\ell)}$ to sets in $\mathcal{A}_\ell$ that are contained in $I_t^{(\ell)}$, and thus, $\text{Shift}(f) \leq 2^{-\ell}$. Therefore, for $n > n'_0$, we get that $d_\triangle(W_{n,\ell}, W'_{n,\ell}) \leq 2^{-\ell} + 2\xi$, and the first step is complete.

In the second step we bound $d_\triangle(W'_{n,\ell}, U_\ell)$. The two functions $W'_{n,\ell}$ and U_ℓ are defined on the same partition $\mathcal{A}_\ell$, however, their values are determined by the density functions $\delta^{(n,\ell)}$ and $\delta^{(\ell)}$ respectively. By the fact that $\delta^{(n,\ell)}$ converges to $\delta^{(\ell)}$ (as $n \to \infty$), we can find $n''_0(\ell)$ such that for all $n > n''_0$,

$$\left|\delta^{(n,\ell)}(i,j) - \delta^{(\ell)}(i,j)\right| \leq \frac{\xi}{g(\ell)^2} \qquad \forall i,j \in [g(\ell)]\,.$$

Thus, for every $n > n''_0$, it holds that $d_\triangle(W'_{n,\ell}, U_\ell) \leq \|W'_{n,\ell} - U_\ell\|_1 \leq \xi$. By choosing $n_0 = \max(n'_0, n''_0)$ we get that

$$d_\triangle(W_{n,\ell}, U_\ell) \leq d_\triangle(W_{n,\ell}, W'_{n,\ell}) + d_\triangle(W'_{n,\ell}, U_\ell) \leq 2^{-\ell} + 3\xi\,.$$

By putting everything together we get that for every $n > n_0$

$$\begin{aligned} d_\triangle(W_n, U) &\leq d_\triangle(W_n, W_{n,\ell}) + d_\triangle(W_{n,\ell}, U_\ell) + d_\triangle(U_\ell, U) \\ &\leq \|W_n - W_{n,\ell}\|_\square + d_\triangle(W_{n,\ell}, U_\ell) + \|U_\ell - U\|_1 \\ &\leq O\left(\frac{1}{g(\ell-1)2^\ell}\right) + 2^{-\ell} + 3\xi + \xi. \end{aligned}$$

By our choice of $\ell > 1/\xi$ we get that

$$d_\triangle(W_n, U) \leq 6\xi\,.$$

By choosing $\xi = \varepsilon/6$ the theorem follows. ◀

References

1. Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
2. Noga Alon, Omri Ben-Eliezer, and Eldar Fischer. Testing hereditary properties of ordered graphs and matrices. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 848–858, 2017.
3. Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.
4. Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.
5. Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM Journal on Computing*, 37(6):1703–1727, 2008.
6. Noga Alon and Uri Stav. What is the furthest graph from a hereditary property? *Random Structures & Algorithms*, 33(1):87–104, 2008.
7. Maria Axenovich and Ryan R. Martin. Multicolor and directed edit distance. *Journal of Combinatorics*, 2(4), 2011.
8. Omri Ben-Eliezer and Eldar Fischer. Earthmover resilience and testing in ordered structures. In *Proceedings of the 33rd Conference on Computational Complexity (CCC)*, pages 18:1–18:35, 2018.
9. Omri Ben-Eliezer, Eldar Fischer, Amit Levi, and Yuichi Yoshida. Limits of ordered graphs and their applications. Full version of this work. `arXiv:1811.02023`.
10. Christian Borgs, Jennifer Chayes, and David Gamarnik. Convergent sequences of sparse graphs: A large deviations approach. *Random Structures & Algorithms*, 51(1):52–89, 2017.
11. Christian Borgs, Jennifer Chayes, and László Lovász. Moments of two-variable functions and the uniqueness of graph limits. *Geometric and Functional Analysis*, 19(6):1597–1619, 2010.
12. Christian Borgs, Jennifer Chayes, László Lovász, Vera T Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the 38th ACM Symposium on the Theory of Computing (STOC)*, pages 261–270, 2006.
13. Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Counting graph homomorphisms. In *Topics in Discrete Mathematics*, pages 315–371. Springer Berlin Heidelberg, 2006.
14. Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.
15. Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. multiway cuts and statistical physics. *Annals of Mathematics*, 176(1):151–219, 2012.
16. Christian Borgs and Jennifer T. Chayes. Graphons: A nonparametric method to model, estimate, and design algorithms for massive networks. *CoRR*, abs/1706.01143, 2017. `arXiv:1706.01143`.
17. Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII*, 28(1):33–61, 2008.
18. Gábor Elek and Balázs Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Advances in Mathematics*, 231(3):1731–1772, 2012.
19. Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM Journal on Computing*, 37(2):482–501, 2007.
20. John M. Franks. *A (terse) introduction to Lebesgue integration*, volume 48 of *Student Mathematical Library*. American Mathematical Society, 2009.
21. Michael Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *Journal of the American Mathematical Society*, 20:37–51, 2007.

22 Alan Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 12–20, 1996.

23 Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

24 Frederik Garbe, Robert Hancock, Jan Hladký, and Maryam Sharifzadeh. Limits of latin squares. *arXiv preprint*, 2020. Extended abstract appeared in Eurocomb 2019 under the name *Theory of limits of sequences of Latin squares*. arXiv:2010.07854.

25 Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B*, 103(1):93–113, 2013.

26 Svante Janson. Poset limits and exchangeable random posets. *Combinatorica*, 31(5):529–563, 2011.

27 László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Society, 2012.

28 László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

29 László Lovász and Balázs Szegedy. Szemerédi's lemma for the analyst. *Geometric And Functional Analysis*, 17(1):252–270, 2007.

30 László Lovász and Balázs Szegedy. Testing properties of graphs and functions. *Israel Journal of Mathematics*, 178(1):113–156, 2010.

31 Ryan R. Martin. The edit distance in graphs: Methods, results, and generalizations. In Andrew Beveridge, Jerrold R. Griggs, Leslie Hogben, Gregg Musiker, and Prasad Tetali, editors, *Recent Trends in Combinatorics*, pages 31–62. Springer International Publishing, 2016.

32 Ryan R. Martin and Maria Axenovich. Avoiding patterns in matrices via a small number of changes. *SIAM Journal of Discrete Mathematics*, 20(1):49–54, 2006.

33 Peter Orbanz and Daniel M. Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2015.

34 Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes. Colloq. Internat. CNRS*, volume 260, pages 399–401, 1976.

35 Yuichi Yoshida. Gowers norm, function limits, and parameter estimation. In *Proceedings of the 27th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1391–1406, 2016.

36 Yufei Zhao. Hypergraph limits: A regularity approach. *Random Structures & Algorithms*, 47(2):205–226, 2015.