Ordered Graph Limits and Their Applications

Omri Ben-Eliezer∗ Eldar Fischer† Amit Levi‡ Yuichi Yoshida§

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 $G(n, p)$ 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].

∗Tel Aviv University, Israel. Email: omrib@mail.tau.ac.il.
†Technion - Israel Institute of Technology, Israel. Email: eldar@cs.technion.ac.il.
‡University of Waterloo, Canada. Email: amit.levi@uwaterloo.ca. Research supported by the David R. Cheriton Graduate Scholarship. Part of this work was done while the author was visiting the Technion.
§National Institute of Informatics (NII), Japan. Email: yyoshida@nii.ac.jp. Research supported by JSPS KAKENHI Grant Number JP17H04676.
## Contents

1 Introduction .......................... 1
  1.1 Main ingredients .......................... 2
  1.2 Main results .......................... 4
  1.3 The furthest ordered graph from a hereditary property .......................... 5
  1.4 Additional applications .......................... 7
  1.5 Related work .......................... 8

2 Preliminaries .......................... 8
  2.1 Cut-norm and ordered cut-norm .......................... 9
  2.2 The cut and shift distance .......................... 11

3 Block orderons and their density in \( \mathcal{W} \) .......................... 12

4 Compactness of the space of orderons .......................... 15

5 Sampling theorem .......................... 21

6 Subgraph statistics .......................... 25

7 The furthest ordered graph from a hereditary property .......................... 29
  7.1 The random graph \( G(n, p) \) is not always the furthest .......................... 29
  7.2 Proof of Theorem 1.7 .......................... 31

8 Parameter estimation .......................... 36

9 Testability of hereditary ordered graph properties .......................... 38
  9.1 Proof sketch of Lemma 9.1 .......................... 39
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 quasirandomness. A comprehensive series of papers [BCL+06a, BCL+06b, LS06, FLS07, LS07, BCL+08, BCL10, LS10, BCL+12] 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 [Lov12] 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 \( W : [0,1]^2 \to [0,1] \), and it was proved in [LS06] 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. [DJ08, OR15]. 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. [BC17, BCG17] for two recent papers citing and discussing many of the works in this field; permutations, first developed in [HKM+13] and further investigated in several other works; partial orders [Jan11]; and high dimensional functions over finite fields [Yos16]. The limit theory of dense graphs has also been extended to hypergraphs, see [Zha15, ES12] 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 \( G : [n]^2 \to \{0,1\} \). \( G \) is simple if \( G(x,x) = 0 \) for any \( x \). A weighted ordered graph is a symmetric function \( F : [n]^2 \to [0,1] \). Unlike the unordered setting, where \( G, G' : [n]^2 \to \Sigma \) are considered isomorphic if there is a permutation \( \pi \) 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

\(^1\)In unordered graphs, this is also called convergence from the left; see the discussion on [BCL+08].

\(^2\)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.
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 graphs \( I : [m] \times [n] \rightarrow \{0, 1\} \), 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 (and possibly deletions) 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 \( |V(G_n)| \rightarrow \infty \) as \( n \rightarrow \infty \), and the frequency \( t(F,G_n) \) of any simple ordered graph \( F \) converges as \( n \rightarrow \infty \). 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 \( n \rightarrow \infty \), 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.

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 is the situation with graphons, we allow graphs \( H \) and \( G \) to have the same representation when one is a blowup\(^3\) of the other.

\(^3\)A graph \( G \) on \( nt \) vertices is an ordered \( t \)-blowup of \( H \) on \( n \) vertices if \( G(x,y) = H([x/t],[y/t]) \) for any \( x \) and \( y \).
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.

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 [Sze76], 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 [FK96, FK99] and defined as follows. First, we define the cut norm \( \|W\|_\Box \) of a function \( W: [0,1]^2 \to \mathbb{R} \) as the supremum of \( |\int_{S \times T} W(x,y) dx dy| \) over all measurable subsets \( S, T \subseteq [0,1] \). The cut distance between graphons \( W \) and \( W' \) is the infimum of \( \|W^\phi - W'\|_\Box \) over all measure-preserving bijections \( \phi: [0,1] \to [0,1] \), where \( W^\phi(x,y) = W(\phi(x), \phi(y)) \).

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 \( d_\Box(W, W') \) as

\[
d_\Box(W, W') \equiv \inf_f \left( \text{Shift}(f) + \|W^f - W'\|_\Box \right),
\]

where \( f: [0,1] \to [0,1] \) 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 1.2 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 [Fra09]). 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 $W : ([0,1]^2) \rightarrow [0,1]$ 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 $F$ in an orderon $W$ is a natural extension of frequency in graphs. First, define the random variable $G(k,W)$ 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 $G(k,W)$ 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 $\text{Shift}(f) = \sup_{(x,a) \in [0,1]^2} |\pi_1(f(x,a)) - x|$, where $\pi_1(y,b) \overset{\text{def}}{=} y$ 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 2.8 below, $d_{\Delta}$ is a pseudo-metric for $\mathcal{W}$. By identifying $W,U \in \mathcal{W}$ whenever $d_{\Delta}(W,U) = 0$, we get a metric space $\tilde{\mathcal{W}}$. The following result is the main component for the viability of our limit object, settling the third requirement above.

**Theorem 1.1.** The space $\tilde{\mathcal{W}}$ is compact with respect to $d_{\Delta}$.

The proof of Theorem 1.1 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 $d_{\Delta}$. This settles the second requirement.

**Theorem 1.2.** Let $\{W_n\}_{n=1}^{\infty}$ be a sequence of orderons. Then $\{W_n\}$ is Cauchy in $d_{\Delta}$ 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 1.3. 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 \( Q_n(x) = \lfloor nx \rfloor \) for \( x > 0 \) and \( Q_n(0) = 1 \). This addresses the first requirement.

Theorem 1.4. Let \( k \) be a positive integer and let \( W \in \mathcal{W} \) be an orderon. Let \( G \sim G(k, W) \). Then,

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

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

Theorem 1.4 implies, in particular, that ordered graphs are a dense subset in \( \mathcal{W} \).

Corollary 1.5. For every orderon \( W \) and every \( \epsilon > 0 \), there exists a simple ordered graph \( G \) on at most \( 2^{c_3+o(1)} \) vertices such that \( d_{\Delta}(W, W_G) \leq \epsilon \).

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 and perhaps most interesting application: A new result on the maximum edit distance \( d_1(G, \mathcal{H}) \) of an ordered graph \( G \) from a hereditary\(^5\) 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 [AS08b] states, roughly speaking, that \( d_{\mathcal{H}} \) is always “achieved” by the Erdős–Rényi random graph \( G(n, p) \) for an appropriate choice of \( p \) and large enough \( n \).

Theorem 1.6 ([AS08b]). For any hereditary property \( \mathcal{H} \) of unordered graphs there exists \( p_{\mathcal{H}} \in [0, 1] \) satisfying the following. A graph \( G \sim G(n, p_{\mathcal{H}}) \) satisfies \( d_1(G, \mathcal{H}) \geq \overline{d_{\mathcal{H}}} - o(1) \) with high probability.

In other words, a random graph \( G(n, p_{\mathcal{H}}) \) 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 [LS10] 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 [Mar16], which asks the

\(^4\)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 \( P \) 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.

\(^5\)A property of (ordered or unordered) graphs is said to be hereditary if it is closed under taking induced subgraphs.
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 [Mar16] for an extensive survey of previous works and useful techniques.

While the situation in unordered graphs, and even in (unordered) directed graphs [AM11] and matrices [MA06] has been thoroughly investigated, for ordered graphs no result in the spirit of Theorem 1.6 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 \( \mathcal{H} \), Theorem 1.6 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 \( d_1(G, \mathcal{H}) \geq \bar{d}_\mathcal{H} - o(1) \) 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 beginning of Section 7, 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 1.6 cannot be true for the ordered setting. For the exact details, see Subsection 7.1, which is completely elementary and self-contained.

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 [Abb18] 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\(^6\) 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.

**Theorem 1.7.** 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 \( d_1(G, \mathcal{H}) \geq \bar{d}_\mathcal{H} - \varepsilon \) with probability that tends to one as \( n \to \infty \).

The proof, given in Subsection 7.2, is a good example of the power of the analytic perspective.

\(^6\)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.
combining our ordered limit theory with standard measure-theoretic tools and a few simple lemmas proved in [LS10].

1.4 Additional applications

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 1.8 (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 \( G|_k \) is the subgraph induced by a uniformly random ordered set of exactly \( k \) nodes of \( G \), then

\[
\Pr_{G|_k}[|f(G) - f(G|_k)| > \varepsilon] < \delta.
\]

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 1.9. 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 \( \hat{f}(W) \) over \( W \) that satisfies the following:
   (a) \( \hat{f}(W) \) 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 \( \left| \hat{f}(W_G) - f(G) \right| \leq \varepsilon \).

Our third application is a new analytic proof of the ordered graph removal lemma of [ABF17], 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 [ABF17] and Definition 1.8 here.)

Theorem 1.10 ([ABF17]). 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 \( \varepsilon \)-far from \( \mathcal{H} \), the probability that \( G|_k \) does not satisfy \( \mathcal{H} \) is at least \( 1 - c \).

The proof of Theorem 1.10 utilizes the analytic tools developed in this work, and provides an (arguably) cleaner proof compared to the combinatorial proof of [ABF17].
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 Szemérdi [Sze76], later to be joined by the weaker (but more efficient) versions of Frieze and Kannan [FK96, FK99] and the stronger variants of Alon et al. [AFKS00], 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 [AS08a], the relationship between the testability and estimability of graph parameters [FN07], and combinatorial characterizations of testability [AFNS09].

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 [ABF17], it was shown that the testability of hereditary properties extends to the ordered setting (see Theorem 1.10 above). Shortly after, in [BF18] 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 [BC17]. Graph limits have applications in probability and data analysis [OR15]. Graphons were used to provide new analytic proofs of results in extremal graph theory; see Chapter 16 in [Lov12]. Through the notion of free energy, graphons were also shown to be closely connected to the field of statistical physics [BCL+12]. We refer the reader to [Lov12] for more details.

We remark that an ongoing independent work by Frederik Garbe, Robert Hancock, Jan Hladky, and Maryam Sharifzadeh, investigates an alternative limit object for the ordered setting.

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 $\lambda$. We start with the formal definition of an orderon.

**Definition 2.1** (orderon). An orderon is a measurable function $W : ([0,1]^2)^2 \to [0,1]$ 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 \( W \).

**Definition 2.2** (measure-preserving bijection). A map \( g: [0,1]^2 \to [0,1]^2 \) 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 \( F \) denote the collection of all measure preserving bijections from \([0,1]^2\) to itself. Given an orderon \( W \in W \) and \( f \in F \), 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 \( \pi_1: [0,1]^2 \to [0,1] \) 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 2.3** (cut-norm). Given a symmetric measurable function \( W: ([0,1]^2)^2 \to \mathbb{R} \), we define the cut-norm of \( W \) as
\[
\|W\|_{\square} = \sup_{S,T \subseteq [0,1]^2} \left| \int_{(x,a) \in S, (y,b) \in T} W((x,a),(y,b))\, dx\, da\, db \right|.
\]

As we are working with ordered objects, the following definition of ordered cut-norm will sometimes be of use (in particular, see Section 6). Given \( v_1, v_2 \in [0,1]^2 \), we write \( v_1 \leq v_2 \) to denote that \( \pi_1(v_1) \leq \pi_1(v_2) \). Let \( 1_E \) be the indicator function for the event \( E \).

**Definition 2.4** (ordered cut-norm). Let \( W: ([0,1]^2)^2 \to \mathbb{R} \) be a symmetric measurable function. The ordered cut norm of \( W \) is defined as
\[
\|W\|_{\square'} = \sup_{S,T \subseteq [0,1]^2} \left| \int_{(v_1,v_2) \in S \times T} W(v_1,v_2)1_{v_1 \leq v_2} \, dv_1 \, dv_2 \right|.
\]

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.

**Lemma 2.5.** Let \( W \in W \) and \( \mu, \nu : [0,1] \to [0,1] \). Then,
\[
\left| \int_{v_1,v_2} \mu(v_1)\nu(v_2)W(v_1, v_2)1_{v_1 \leq v_2} \, dv_1 \, dv_2 \right| \leq \|W\|_{\square'}.
\]

**Proof:** Fix partitions \( S = \{S_i\} \) and \( T = \{T_j\} \) of \([0,1]^2\). We show below that the claim holds when \( \mu \) and \( \nu \) are step functions on \( S \) and \( T \), respectively. Then, the proof is complete by the fact that all integrable functions are approximable in \( L^1(([0,1]^2)^2) \) by step functions.
Since $\mu$ and $\nu$ are step functions, we can write $\mu = \sum_i a_i 1_{S_i}$ and $\nu = \sum_j b_j 1_{T_j}$ for some vectors $a \in [0, 1]|^{[S]}$ and $b \in [0, 1]|^{[T]}$. We define
\[
f(a, b) \overset{\text{def}}{=} \int_{v_1, v_2} \mu(v_1)\nu(v_2) W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2.
\]
When $a \in \{0, 1\}|^{[S]}$ and $b \in \{0, 1\}|^{[T]}$, we have
\[
|f(a, b)| = \left| \int \sum_i \sum_j a_i b_j 1_{S_i}(v_1) 1_{T_j}(v_2) W(v_1, v_2) dv_1 dv_2 \right|
= \left| \int \bigcup_{i, a_i = 1} s_i \int \bigcup_{j, b_j = 1} t_j W(v_1, v_2) dv_1 dv_2 \right| \leq \|W\|_{\square},
\]
where the last inequality follows from the definition of the ordered cut-norm. As $f(a, b)$ is bilinear in $a$ and $b$, and $|f(a, b)| \leq \|W\|_{\square}$ for any $a \in \{0, 1\}|^{[S]}$ and $b \in \{0, 1\}|^{[T]}$, we have $|f(a, b)| \leq \|W\|_{\square}$ for any $a \in \{0, 1\}|^{[S]}$ and $b \in \{0, 1\}|^{[T]}$.

**Lemma 2.6.** Let $W : ([0, 1]^2)^2 \to [-1, 1]$ be a symmetric measurable function. Then,
\[
\frac{\|W\|_{\square}^2}{4} \leq \|W\|_{\square} \leq 2\|W\|_{\square}.
\]

**Proof:** The inequality $\|W\|_{\square} \leq 2\|W\|_{\square}$ follows immediately from the fact that $W$ is symmetric. For the other inequality, let $\xi = \|W\|_{\square}$, fix $\gamma > 0$, and let $S, T \subseteq [0, 1]^2$ be a pair of sets satisfying
\[
\left| \int_{S \times T} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right| \geq \xi - \gamma.
\]
We partition $[0, 1]^2$ into strips $I = \{I_1, \ldots, I_{2/\xi}\}$, such that for every $j \in [2/\xi]$, $I_j = \left[\frac{j-1}{2}, \frac{j}{2}\right] \times [0, 1]$. For every $j \in [2/\xi]$, let $I(<j) = \bigcup_{i < j} I_i$ (where $I(1) = \emptyset$). Then,
\[
\xi - \gamma \leq \left| \int_{S \times T} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right|
\leq \sum_{i \in [2/\xi]} \left| \int_{(S \cap I_i) \times (T \cap I_i)} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right| + \sum_{j \in [2/\xi]} \left| \int_{(S \cap I(<j)) \times (T \cap I_j)} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right|
\]
Note that by the fact that $|W(v_1, v_2)| \leq 1$ for all $v_1, v_2 \in [0, 1]^2$,
\[
\sum_{i \in [2/\xi]} \left| \int_{(S \cap I_i) \times (T \cap I_i)} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right| \leq \sum_{i \in [2/\xi]} \lambda(I_i \times I_i) \leq \xi/2,
\]
and therefore,
\[
\sum_{j \in [2/\xi]} \left| \int_{(S \cap I(<j)) \times (T \cap I_j)} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right| \geq \xi/2 - \gamma.
\]
On the other hand, the above implies that there exists $j \in [2/\xi]$ such that
\[
\left| \int_{(S \cap I^{(<j)}) \times (T \cap I_j)} W(v_1, v_2) 1_{v_1 \leq v_2} dv_1 dv_2 \right| \geq \xi^2 / 4 - \xi \gamma / 2,
\]
Note that for every $(v_1, v_2) \in (S \cap I^{(<j)}) \times (T \cap I_j)$, we have that $1_{v_1 \leq v_2} = 1$, and thus
\[
\|W\|_{\square} \geq \left| \int_{(S \cap I^{(<j)}) \times (T \cap I_j)} W(v_1, v_2) dv_1 dv_2 \right| \geq \xi^2 / 4 - \frac{\xi \gamma}{2}.
\]
Since the choice of $\gamma$ is arbitrary, the lemma follows.

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 2.7. Given two orderons $W, U \in W$ we define the CS-distance (cut-norm+shift distance) as:
\[
d_{\triangle}(W, U) \overset{\text{def}}{=} \inf_{f \in \mathcal{F}} \left( \text{Shift}(f) + \|W - U^f\|_\square \right),
\]
where $\text{Shift}(f) \overset{\text{def}}{=} \sup_{x, a \in [0, 1]} |x - \pi_1(f(x, a))|$.

Lemma 2.8. $d_\triangle$ is a pseudo-metric on the space of orderons.

Proof: First note that non-negativity follows trivially from the definition. In addition, it is easy to see that $d_\triangle(W, W) = 0$ for any orderon $W$. For symmetry,
\[
d_\triangle(W, U) = \inf_{g \in \mathcal{F}} (\text{Shift}(g) + \|W - U^g\|_\square) = \inf_{g \in \mathcal{F}} \left( \text{Shift}(g^{-1}) + \|W - U^g\|_\square \right)
\]
\[
= \inf_{g^{-1} \in \mathcal{F}} \left( \text{Shift}(g^{-1}) + \|W^{-1} - U\|_\square \right) = \inf_{f \in \mathcal{F}} \left( \text{Shift}(f) + \|U - W^f\|_\square \right)
\]
\[
= d_\triangle(U, W).
\]
Where we used the fact that $g$ is a measure preserving bijection and that $\text{Shift}(g^{-1}) = \text{Shift}(g)$ for any $g \in \mathcal{F}$.

Consider three orderons $W, U, Z$. We now show that $d_\triangle(W, U) \leq d_\triangle(W, Z) + d_\triangle(Z, U)$.
\[
d_\triangle(W, U) = \inf_{f, g \in \mathcal{F}} \left( \text{Shift}(g^{-1} \circ f) + \|W - U^{g^{-1} \circ f}\|_\square \right)
\]
\[
\leq \inf_{f, g \in \mathcal{F}} \left( \text{Shift}(f) + \text{Shift}(g^{-1}) + \|W^g - U^f\|_\square \right)
\]
\[
\leq \inf_{f \in \mathcal{F}} \left( \text{Shift}(f) + \|Z - U^f\|_\square \right) + \inf_{g \in \mathcal{F}} \left( \text{Shift}(g) + \|W^g - Z\|_\square \right)
\]
\[
= d_\triangle(W, Z) + d_\triangle(Z, U),
\]
where the first equality holds since $g^{-1} \circ f$ is a measure preserving bijection, and the last inequality follows from the triangle inequality; note that $\text{Shift}(g^{-1}) = \text{Shift}(g)$ for any $g \in \mathcal{F}$. 


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

In this section 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 3.1** (naive block orderon). Let $m \in \mathbb{N}$ be an integer. For $z \in (0, 1]$, we denote $Q_n(z) = \lfloor nz \rfloor$; we also set $Q_n(0) = 1$. An $m$-block naive orderon is a function $W : ([0, 1]^2)^2 \to [0, 1]$ that can be written as

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

for some weighted ordered graph $G$ on $n$ vertices.

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 $G(x, y) = H([x/t], [y/t])$ 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 $R = \{S_1, \ldots, S_k\}$ of $[0, 1]^2$ such that $U$ is constant on every $S_i \times S_j$.

**Remark** (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 $P = \{S_1, \ldots, S_k\}$ of $[0, 1]^2$ into measurable sets, let $W_P : ([0, 1]^2)^2 \to [0, 1]$ 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,

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

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 3.2** ([LS07] Lemma 4.1). Let $\{K_i\}_i$ be arbitrary non-empty subsets of a Hilbert space $\mathcal{H}$. Then, for every $\varepsilon > 0$ and $f \in \mathcal{H}$ there is an $m \leq \lceil 1/\varepsilon^2 \rceil$ and there are $f_i \in K_i$ ($1 \leq i \leq k$) and $\gamma_1, \ldots, \gamma_k \in \mathbb{R}$ such that for every $g \in K_{k+1}$

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

The next lemma is a direct consequence of Lemma 3.2.

**Lemma 3.3.** 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

$$\|W - U\|_\Box \leq \varepsilon.$$
Proof: We apply Lemma 3.2 to the case where the Hilbert space is $L^2([0,1]^4)$, and each $K_i$ is the set of indicator functions of product sets $S \times S$, where $S \subseteq [0,1]^2$ is a measurable subset. Then for $f \in W$, there is an $f' = \sum_{j=1}^k \gamma_j f_j$, which is a step function with at most $2^k$ steps. Therefore, we get a step function $U \in W$ with at most $2^{1/\varepsilon^2}$ steps such that for every measurable set $S \subseteq [0,1]^2$
\[
\left| \int_{v_1,v_2 \in S \times S} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 \right| \leq \varepsilon .
\]

By the above and the fact that
\[
\left| \int_{v_1,v_2 \in (S \cup T) \times (S \cup T)} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 \right| = \left| \int_{v_1,v_2 \in S \times S} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 \right| + 2 \cdot \left| \int_{v_1,v_2 \in S \times T} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 + \int_{v_1,v_2 \in T \times T} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 \right| \leq \varepsilon ,
\]
we get that for any two measurable sets $S, T \subseteq [0,1]^2$,
\[
\left| \int_{v_1,v_2 \in S \times T} (W(v_1,v_2) - U(v_1,v_2))dv_1dv_2 \right| \leq 2\varepsilon ,
\]
which implies the lemma.

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

Claim 3.4. Let $W \in W$, let $U$ be a step function, and let $P$ denote the partition of $[0,1]^2$ into the steps of $U$. Then $\|W - W_P\|_\square \leq 2\|W - U\|_\square$.

Proof: The proof follows from the fact that $U = U_P$ and the fact that the stepping operator is contractive with respect to the cut norm. More explicitly,
\[
\|W - W_P\|_\square \leq \|W - U\|_\square + \|U - W_P\|_\square = \|W - U\|_\square + \|U_P - W_P\|_\square \leq 2\|W - U\|_\square .
\]

Using Lemma 3.3 and Claim 3.4 we can obtain the following lemma.

Lemma 3.5. For every function $W \in W$ and every $\varepsilon > 0$, there is a partition $P$ of $[0,1]^2$ into at most $2^{32/\varepsilon^2}$ sets with positive measure such that $\|W - W_P\|_\square \leq \varepsilon$.

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 $P$ refines $P'$, if $P$ can be obtained from $P'$ by splitting each $P_j \in P'$ into a finite number of sets (up to sets of measure 0).

Lemma 3.6. Fix some $\varepsilon > 0$. Let $P$ be an equipartition of $[0,1]^2$ into $k$ sets, and fix $q \geq 2k^2 \cdot 2^{162/\varepsilon^2}$ such that $k$ divides $q$. Then, for any $W \in W$, there exists an equipartition $Q$ that refines $P$ with $q$ sets, such that $\|W - W_Q\|_\square \leq \frac{8\varepsilon}{q} + \frac{2}{k}$.
Theorem 3.8. For every orderon $W \in \mathcal{W}$ and every $\varepsilon > 0$, there exist a naive \( \frac{\varepsilon}{\varepsilon} 2^{162/\varepsilon^2} \)-block orderon $W'$ (for some constant $c > 0$) such that

\[ d_\triangle(W, W') \leq \varepsilon. \]
Proof: Fix $\varepsilon > 0$ and $\gamma = \gamma(\varepsilon) > 0$. We consider an interval equipartition $J = \{J_1, \ldots, J_{1/\gamma}\}$ 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 $\mathcal{P} = (J_i \times J_j | i, j \in [1/\gamma])$ be an equipartition of $[0, 1]^2$. By Lemma 3.6, there exists an equipartition $Q$ of $[0, 1]^2$ of size $q = \frac{2}{\sqrt{\gamma}}(2^{162}/\varepsilon^2)$ that refines $\mathcal{P}$, such that

$$\|W - W_Q\|_\square \leq \frac{8\varepsilon}{9} + 2\gamma^2.$$ 

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

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

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

$$W'(f(x, a), f(y, b)) = W_Q((x, a), (y, b))$$

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

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

as desired. 

4 Compactness of the space of orderons

In this section we prove Theorem 1.1. We construct a metric space $\tilde{W}$ from $W$ with respect to $d_\triangle$, by identifying $W, U \in W$ with $d_\triangle(W, U) = 0$. Let $\tilde{W}$ be the image of $W$ under this identification. On $\tilde{W}$ 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 $\Omega$, 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 $\ell$

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

(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 [Lov12]. The second result is an application of the martingale convergence theorem, useful for our purposes.
Theorem 4.1 (see [Lov12], Theorem A.12). Let \( \{X_i\}_{i \in \mathbb{N}} \) be a martingale such that \( \sup_n \mathbb{E}|X_n| < \infty \). Then \( \{X_i\}_{i \in \mathbb{N}} \) is convergent with probability 1.

Lemma 4.2. Let \( \{\mathcal{P}_\ell\}_\ell \) be a sequence of partitions of \( \Omega \) such that for every \( \ell \), \( \mathcal{P}_{\ell+1} \) refines \( \mathcal{P}_\ell \). Assume that for every \( \ell \) and \( j \in [||\mathcal{P}_\ell||] \), the functions \( \beta(\mathcal{P}_\ell : \cdot) \) satisfy
\[
\lambda \left( P_{j^\ell} \right) \beta(\mathcal{P}_\ell : j) = \sum_{i \in [||\mathcal{P}_{\ell+1}||]} \lambda \left( P_{j^\ell} \cap P_{i^{\ell+1}} \right) \beta(\mathcal{P}_{\ell+1} : i). 
\]

Then, there is a measurable function \( \beta : \Omega \rightarrow [0,1] \) 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 \( X \) be a uniformly distributed random variable in \( \Omega \). Let \( \psi : \Omega \rightarrow [||\mathcal{P}_\ell||] \) be the function mapping each \( v \in \Omega \) to its corresponding part in \( \mathcal{P}_\ell \) and let \( Z_\ell = \beta(\mathcal{P}_\ell : X) \). We now show that the sequence \( (Z_1, Z_2, \ldots) \) is a martingale. That is, \( \mathbb{E}_{X \sim \Omega}[Z_{\ell+1} | Z_1, \ldots, Z_\ell] = Z_\ell \), for every \( \ell \in \mathbb{N} \). Note that by the fact that \( \mathcal{P}_{\ell+1} \) refines \( \mathcal{P}_\ell \), \( \psi_\ell(X) \) determines \( \psi_i(X) \) for every \( i < \ell \).

By definition, the value \( \beta(\mathcal{P}_\ell : X) \) is completely determined by \( \psi_\ell(X) \), and so it suffices to prove that \( Z_\ell = \mathbb{E}_{X \sim \Omega}[Z_{\ell+1} | \psi_\ell(X)] \). By the fact that for every \( j \in [||\mathcal{P}_\ell||] \) Equation (3) holds (and in particular holds for \( \psi_\ell(X) \)), we can conclude that the sequence \( (Z_1, Z_2, \ldots) \) is a martingale.

Since \( Z_\ell \) is bounded, we can invoke the martingale convergence theorem (Theorem 4.1) and conclude that \( \lim_{\ell \to \infty} Z_\ell \) exists with probability 1. That is, \( \beta(v) = \lim_{\ell \to \infty} \beta(\mathcal{P}_\ell : v) \) exists for almost all \( v \in \Omega \). \( \blacksquare \)

Definition 4.3. Fix some \( d \in \mathbb{N} \) and define \( \mathcal{I}_d = \{I_1^{(d)}, \ldots, I_{2^d}^{(d)}\} \) so that for every \( t \in [2^d] \),
\[
I_t^{(d)} = \left[ \frac{t-1}{2^d}, \frac{t}{2^d} \right) \times [0,1].
\]
We refer to this partition as the strip partition of order \( d \).

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

Lemma 4.4. For any order \( W \in \mathcal{W} \) 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 \( g : \mathbb{N} \rightarrow \mathbb{N} \)).

2. For every \( \ell \), \( \Gamma_\ell \overset{\text{def}}{=} \frac{\ell}{2^{\ell-1}} \in \mathbb{N} \).

3. For every \( \ell' \geq \ell \), the partition \( \mathcal{P}_{\ell'} \) refines both \( \mathcal{P}_\ell \) and the strip partition \( \mathcal{I}_d \). In particular, for every \( j \in [g(\ell - 1)] \),
\[
P_{j^\ell} = \bigcup_{j' = (j-1) \cdot \Gamma_\ell + 1}^{j \cdot \Gamma_\ell} P_{j'^\ell}.
\]

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

Proof: We invoke Lemma 3.7 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 3.7 with \( I_\ell \) and \( P_{n,\ell-1} \) to get a partition \( P_{n,\ell} \) of size \( g(\ell) = (g(\ell - 1) \cdot 2^{\ell})^2 \cdot 2^{\Omega(g(\ell - 1)^2)} \) which refines both \( I_\ell \) and \( P_{n,\ell-1} \) such that \( \|W_n - W_{n,\ell}\| \leq \frac{4}{g(\ell - 1)^2} \). 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 4.4 to construct a sequence of functions \( \{W_n,\ell\}_\ell \) such that \( \|W_n - W_{n,\ell}\| \) is small. For each \( \ell \), we would like to approximate the shape of the limit partition resulting from taking \( n \to \infty \). Inside each strip \( I_\ell^{(\ell')} \), we consider the relative measure of the intersection of each set contained in \( I_\ell^{(\ell')} \), with a finer strip partition \( I_\ell' \).

**Definition 4.5** (shape function). For fixed \( n \in \mathbb{N} \), let \( \{P_{n,\ell}\}_\ell \) be partitions of \([0,1]^2\) with the properties listed in Lemma 4.4. For every \( \ell' > \ell \) and \( I^{(\ell')}_{\ell'} \subseteq I_\ell' \), we define \( \alpha_j^{(n,\ell)}(I^{(\ell')}_{\ell'}) \) to be the relative volume of the set \( P_j^{(n,\ell)} \) in \( I^{(\ell')}_{\ell'} \).

For any \( \ell' \geq \ell \) and \( I^{(\ell')}_{\ell'} \subseteq I_\ell' \), by the compactness of \([0,1]\), we can select a subsequence of \( \{W_n\}_{n \in \mathbb{N}} \) such that \( \alpha_j^{(n,\ell)}(I^{(\ell')}_{\ell'}) \) converges for all \( j \in [g(\ell)] \) as \( n \to \infty \). Let

\[
\alpha_j^{(\ell)}(I^{(\ell')}_{\ell'}) \overset{\text{def}}{=} \lim_{n \to \infty} \alpha_j^{(n,\ell)}(I^{(\ell')}_{\ell'}) .
\]

Next we define the limit density function.

**Definition 4.6** (density function). For fixed \( n \in \mathbb{N} \), let \( \{P_{n,\ell}\}_\ell \) be partitions of \([0,1]^2\) with the properties listed in Lemma 4.4. We let \( \delta^{(n,\ell)}(P_{n,\ell} \times P_{n,\ell} : i,j) \) be the relative measure of the set \( P_j^{(n,\ell)} \) in \( P_i^{(n,\ell)} \). By the compactness of \([0,1]\), we can select a subsequence of \( \{W_n\}_{n \in \mathbb{N}} \) such that \( \delta^{(n,\ell)}(P_{n,\ell} \times P_{n,\ell} : i,j) \) converge for all \( i,j \in [g(\ell)] \) as \( n \to \infty \). Let

\[
\delta^{(\ell)}(i,j) \overset{\text{def}}{=} \lim_{n \to \infty} \delta^{(n,\ell)}(P_{n,\ell} \times P_{n,\ell} : i,j) .
\]

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

**Lemma 4.7.** For fixed \( \ell \) and \( j \in [g(\ell)] \), there is a measurable function \( \alpha_j^{(\ell)} : [0,1] \to [0,1] \) such that \( \alpha_j^{(\ell)}(x) = \lim_{\ell' \to \infty} \alpha_j^{(\ell)}(I_{\ell'} : x) \) for almost all \( x \in [0,1] \).

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

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

On the other hand, since \( I_{\ell'}^{(\ell'+1)} \) refines \( I_{\ell'} \),

\[
\lambda \left( P_j^{(n,\ell)} \cap I_{\ell'}^{(\ell')} \right) = \lambda \left( P_j^{(n,\ell)} \cap I_{2^{\ell'-1}}^{(\ell'+1)} \right) + \lambda \left( P_j^{(n,\ell)} \cap I_{2^{\ell'-1}}^{(\ell'+1)} \right).
\]

For fixed \( \ell \) and \( n, \ell' > \ell \), we add empty (zero measure) sets in order to satisfy items (2) and (3).
Therefore, when \( n \to \infty \) we get that,
\[
\lambda \left( I_{n'}^{(e)} \right) \cdot \alpha_j^{(e)}(I_{e'} : t') = \lambda \left( I_{2n'-1}^{(e+1)} \right) \cdot \alpha_j^{(e)}(I_{e'+1} : 2t' - 1) + \lambda \left( I_{2n'}^{(e+1)} \right) \cdot \alpha_j^{(e)}(I_{e'+1} : 2t') ,
\]
which is exactly the condition in Equation (3). By applying Lemma 4.2 with the sequence of strip partitions \( \{ I_{e'} \}_{e'} \) on \( \alpha_j^{(e)} \) the lemma follows. 

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

**Lemma 4.8.** For every \( \ell \) 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 \( n, \ell \) and \( \ell' > \ell \). By the additivity of the Lebesgue measure,
\[
\alpha_j^{(n,\ell-1)}(I_{e'} : x) = \sum_{j' = (j-1) \cdot \Gamma_{\ell+1}}^{j \cdot \Gamma_{\ell}} \alpha_{j'}^{(n,\ell)}(I_{e'} : x) \quad \forall x \in [0,1] .
\]

By the fact that for every \( j \in [g(\ell - 1)] \) and \( x \in [0,1] \) the sequence \( \{ \alpha_j^{(n,\ell-1)}(I_{e'} : x) \} \) converges to \( \alpha_j^{(\ell-1)}(I_{e'} : x) \) as \( n \to \infty \), we get that
\[
\alpha_j^{(\ell-1)}(I_{e'} : x) = \sum_{j' = (j-1) \cdot \Gamma_{\ell+1}}^{j \cdot \Gamma_{\ell}} \alpha_{j'}^{(\ell)}(I_{e'} : x) \quad \forall x \in [0,1] .
\]

By applying Lemma 4.7 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 \( \{ \alpha_j^{(\ell)} \} \) we define a limit partition \( A_\ell = \{ A_1^{(\ell)}, \ldots, A_{g(\ell)}^{(\ell)} \} \) of \([0,1]^2\) as follows.

**Definition 4.9** (limit partition). For every \( \ell \in \mathbb{N} \), let \( A_\ell = \{ A_1^{(\ell)}, \ldots, A_{g(\ell)}^{(\ell)} \} \) be a partition of \([0,1]^2\) such that,
\[
A_j^{(\ell)} = \left\{ (x, a) : \sum_{i < j} \alpha_i^{(\ell)}(x) \leq a < \sum_{i \leq j} \alpha_i^{(\ell)}(x) \right\} \quad \forall j \in [g(\ell)] .
\]

**Lemma 4.10.** For any \( \ell \), the partition \( A_\ell \) has the following properties
1. \(A_\ell\) refines the strip partition \(I_\ell\).

2. The partition \(A_\ell\) refines \(A_{\ell-1}\).

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 4.8 it follows that for each \(j \in [g(\ell-1)]\),
\[
A_j^{(\ell)} \subset A_j^{(\ell-1)} \quad \text{for all} \quad (j-1) \cdot \Gamma_\ell + 1 \leq j' \leq j \cdot \Gamma_\ell,
\]
and therefore,
\[
A_j^{(\ell-1)} = \bigcup_{j'=(j-1)\Gamma_\ell+1}^{j\Gamma_\ell} A_j^{(\ell)},
\]
which shows the second item. To prove the third item of the lemma, note that for every \(n, \ell\) and \(\ell' > \ell\),
\[
\lim_{n \to \infty} \lambda \left( P_j^{(n,\ell)} \right) = \lim_{n \to \infty} \sum_{t' \in [2^{\ell'}]} 2^{-\ell'} \cdot \alpha_j^{(n,\ell)} (I_{t'} : t') = \sum_{t' \in [2^{\ell'}]} 2^{-\ell'} \cdot \alpha_j^{(\ell)} (I_{t'} : t') = \int_x \alpha_j^{(\ell)} (I_{t'} : x) dx,
\]
where the last equality follows from Equation (2). Finally, by taking \(\ell' \to \infty\) and using Lemma 4.7, we get
\[
\lim_{n \to \infty} \lambda \left( P_j^{(n,\ell)} \right) = \int_x \alpha_j^{(\ell)} (x) dx = \lambda \left( A_j^{(\ell)} \right).
\]

Using the definition of \(\delta^{(\ell)}\) and \(A_\ell\), we define a density function on the limit partition. For \((x, a) \in A_i^{(\ell)}\) and \((y, b) \in A_j^{(\ell)}\), let
\[
\delta \left( A_\ell \times A_\ell : (x, a), (y, b) \right) \overset{\text{def}}{=} \delta^{(\ell)} (i, j).
\]

**Lemma 4.11.** For each \(\ell \in \mathbb{N}\) and \(i, j \in [g(\ell-1)]\),
\[
\sum_{i'=-(i-1)\Gamma_\ell+1}^{i\Gamma_\ell} \sum_{j'=-(j-1)\Gamma_\ell+1}^{j\Gamma_\ell} \lambda \left( A_{i'}^{(\ell)} \right) \cdot \lambda \left( A_{j'}^{(\ell)} \right) \delta \left( A_\ell \times A_\ell : i', j' \right) = \lambda \left( A_i^{(\ell-1)} \right) \cdot \lambda \left( A_j^{(\ell-1)} \right) \delta \left( A_{\ell-1} \times A_{\ell-1} : i, j \right).
\]

**Proof:** Fix \(n, \ell\) and \(i, j \in [g(\ell-1)]\). By the definition of the partitions \(P_{n,\ell}, P_{n,\ell-1}\) and the density functions \(\delta^{(n,\ell)}, \delta^{(n,\ell-1)}\)
\[
\sum_{i'=-(i-1)\Gamma_\ell+1}^{i\Gamma_\ell} \sum_{j'=-(j-1)\Gamma_\ell+1}^{j\Gamma_\ell} \lambda \left( P_{i'}^{(n,\ell)} \right) \cdot \lambda \left( P_{j'}^{(n,\ell)} \right) \delta^{(n,\ell)} \left( P_{n,\ell} \times P_{n,\ell} : i', j' \right) = \lambda \left( P_i^{(n,\ell-1)} \right) \cdot \lambda \left( P_j^{(n,\ell-1)} \right) \delta \left( P_{n,\ell-1} \times P_{n,\ell-1} : i, j \right).
\]
By taking the limit as \( n \to \infty \) and using the third item of Lemma 4.10,

\[
\sum_{i=0}^{\Gamma_\ell} \sum_{j=0}^{\Gamma_\ell} \lambda(A^{(\ell)}_i) \cdot \lambda(A^{(\ell)}_j) \delta(\mathcal{A}_\ell \times \mathcal{A}_\ell : i, j) = \lambda(A^{(\ell-1)}_i) \cdot \lambda(A^{(\ell-1)}_j) \delta(\mathcal{A}_{\ell-1} \times \mathcal{A}_{\ell-1} : i, j).
\]

The next Lemma asserts that the natural density function of the limit partition is measurable. It follows directly from the combination of Lemma 4.2 and Lemma 4.11.

**Lemma 4.12.** There exists a measurable function \( \delta : ([0,1]^2)^2 \to [0,1] \) such that \( \delta((x,a),(y,a)) = \lim_{\ell \to \infty} \delta(\mathcal{A}_\ell \times \mathcal{A}_\ell : (x,a),(y,a)) \) for almost all \((x,a),(y,b) \in ([0,1]^2)^2\).

Finally, we are ready to prove Theorem 1.1.

**Proof of Theorem 1.1:** 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 \( \mathcal{W} \).

For every \( n \in \mathbb{N} \), we use Lemma 4.4 to construct a sequence of functions \( \{W_{n,\ell}\}_\ell \) such that \( \|W_n - W_{n,\ell}\|_{\square} \leq \frac{4}{g(\ell-1)^2} \). Then, for every fixed \( \ell \in \mathbb{N} \), we find a subsequence of \( \{W_{n,\ell}\} \) such that their corresponding \( a_j^{(n,\ell)} \) and \( \delta(n,\ell)(i,j) \) converge for all \( i,j \in [g(\ell)] \) (as \( n \to \infty \)).

Fixing this \( \ell \), we consider the sequence \( \{U_\ell\}_\ell \) which is defined by the partition \( \mathcal{A}_\ell \) and the density function \( \delta(\ell) \). By Lemma 4.12, the sequence \( \{U_\ell\}_\ell \) converges to some \( U \) almost everywhere to \( \ell \to \infty \), which is defined by the limit density function \( \delta \). Therefore, we can find some \( \ell > 1/\xi \) such that \( ||U_\ell - U||_1 \leq \xi \).

Fixing this \( \ell \), we show that there is \( n_0 \) such that \( d_\Delta(W_{n,\ell},U_\ell) \leq 2^{-\ell} + 3\xi \) 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 \((y,b) \in A_j^{(\ell)} \), \( W'_{n,\ell}((x,a),(y,b)) \triangleq \delta(n,\ell)(\mathcal{P}_{n,\ell} \times \mathcal{P}_{n,\ell} : i,j) \).

By the third item of Lemma 4.10, for every \( j \in [g(\ell)] \), \( \lambda(A_j^{(\ell)}) = \lim_{n \to \infty} \lambda(P_j^{(n,\ell)}) \). Then, we can find \( n_0(\ell) \) such that for all \( n > n_0 \),

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

We define a measure preserving map \( f \) from \( W_{n,\ell} \) to \( W'_{n,\ell} \) as follows. For every strip \( I_i^{(\ell)} \in \mathcal{I}_\ell \), we consider all the sets \( \{P_j^{(n,\ell)}\} \) in \( \mathcal{P}_{n,\ell} \) such that \( \bigcup_{j' = j}^{j_1} P_j^{(n,\ell)} = I_i^{(\ell)} \). Similarly, consider all
the sets \( \{A_{j_1}^{(l)}, \ldots, A_{j_t}^{(l)}\} \) in \( \mathcal{A}_t \) such that \( \bigcup_{j' = j_1}^{j_t} A_{j'}^{(l)} = I_t^{(l)} \). For every \( j' \in \{j_1, \ldots, j_t\} \), we map an arbitrary subset \( S_{j'}^{(n, \ell)} \subseteq P_{j'}^{(n, \ell)} \) of measure \( \min \left( \frac{\lambda(A_{j'}^{(l)})}{\lambda(P_{j'}^{(n, \ell)})} \right) \) to an arbitrary subset (with the same measure) of \( A_{j'}^{(l)} \). Next, we map \( I_t^{(l)} \setminus \bigcup_{j' = j_1}^{j_t} S_{j'}^{(n, \ell)} \) to \( I_t^{(l)} \setminus \bigcup_{j' = j_1}^{j_t} f(S_{j'}^{(n, \ell)}) \). 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 2\( \xi \). Note that for every \( I_t^{(l)} \in \mathcal{I}_t \), the function \( f \) maps sets from \( \mathcal{P}_{n, \ell} \) that are contained in \( I_t^{(l)} \) to sets in \( \mathcal{A}_t \) that are contained in \( I_t^{(l)} \), and thus, \( \text{Shift}(f) \leq 2^{-\ell} \). Therefore, for \( n > n'_0 \), we get that \( d_\Delta(W_{n, \ell}, W'_{n, \ell}) \leq 2^{-\ell} + 2\xi \), and the first step is complete.

In the second step we bound \( d_\Delta(W'_{n, \ell}, U_\ell) \). The two functions \( W'_{n, \ell} \) and \( U_\ell \) are defined on the same partition \( \mathcal{A}_\ell \), however, their values are determined by the density functions \( \delta^{(n, \ell)} \) and \( \delta^{(l)} \) respectively. By the fact that \( \delta^{(n, \ell)} \) converges to \( \delta^{(l)} \) (as \( n \to \infty \)), we can find \( n'_0(\ell) \) such that for all \( n > n'_0 \),
\[
\left| \delta^{(n, \ell)}(i, j) - \delta^{(l)}(i, j) \right| \leq \frac{\xi}{g(\ell)^2} \quad \forall i, j \in [g(\ell)] .
\]
Thus, for every \( n > n''_0 \), it holds that \( d_\Delta(W'_{n, \ell}, U_\ell) \leq \|W'_{n, \ell} - U_\ell\|_1 \leq \xi \). By choosing \( n_0 = \max(n'_0, n''_0) \) we get that
\[
d_\Delta(W_{n, \ell}, U_\ell) \leq d_\Delta(W_{n, \ell}, W'_{n, \ell}) + d_\Delta(W'_{n, \ell}, U_\ell) \leq 2^{-\ell} + 3\xi .
\]
By putting everything together we get that for every \( n > n_0 \)
\[
d_\Delta(W, U) \leq d_\Delta(W, W_{n, \ell}) + d_\Delta(W_{n, \ell}, U_\ell) + d_\Delta(U_\ell, U) \leq \|W_{n, \ell} - U_\ell\|_1 + d_\Delta(W_{n, \ell}, U_\ell) + \|U_\ell - U\|_1 \leq O \left( \frac{1}{g(\ell - 1)^2} \right) + 2^{-\ell} + 3\xi + \xi .
\]
By our choice of \( \ell > 1/\xi \) we get that
\[
d_\Delta(W, U) \leq 6\xi .
\]
By choosing \( \xi = \varepsilon/6 \) the theorem follows.

\[\square\]

\section{Sampling theorem}

In this section we prove Theorem \ref{thm:main}. We start by defining two models of random graphs which are constructed using orderons.

\begin{definition}[ordered \( W \)-random graphs] Given a function \( W \in \mathcal{W} \) and an integer \( n > 0 \), we generate an edge-weighted ordered random graph \( \mathbf{H}(n, W) \) and an ordered random graph \( \mathbf{G}(n, W) \), both on nodes \([n]\), as follows. We generate \( Z_1, \ldots, Z_n \) i.i.d random variables, distributed uniformly in \([0, 1]\), and let \( X_1 \leq \cdots \leq X_n \) be the result of sorting \( Z_1, \ldots, Z_n \). In addition, we generate \( Y_1, \ldots, Y_n \) i.i.d random variables, distributed uniformly in \([0, 1]\). Then, for all \( i, j \in [n] \), we set the edge weight of \((i, j)\) in \( \mathbf{H}(n, W) \) to be \( W((X_i, Y_i), (X_j, Y_j)) \). Also, for all \( i, j \in [n] \), we set \((i, j)\) to be an edge in \( \mathbf{G}(n, W) \) with probability \( W((X_i, Y_i), (X_j, Y_j)) \) independently.
\end{definition}
The proof consists of two main parts. The first (and simpler) part states that large enough samples from a naive block orderon are typically close to it in cut-shift distance. The second and main part shows that samples from orderons that are close with respect to cut-shift distance are typically close as well. We start with the proof of the first part, regarding sampling from naive block orderons.

**Lemma 5.2.** Let $k$ be a positive integer and $W \in \mathcal{W}$ be a naive $m$-block orderon. For any $\varepsilon > 0$, we have

$$
\Pr \left[ \Delta(W, W_{G(k,W)}) > \frac{2m^{3/2}}{k} + \frac{\varepsilon}{\sqrt{k}} \right] \leq \exp \left( -C\varepsilon^2 k \right).
$$

for some constant $C > 0$.

**Proof:** We first show that $\Delta(W, W_{H(k,W)})$ is small with high probability and then discuss how it derives a concentration bound for $\Delta(W, W_{G(k,W)})$.

First, we show that the expectation of $\Delta(W, W_{H(k,W)})$ is at most $2m^{3/2}/k$. Let $\mathcal{P} = \{P_i \mid i \in [m]\}$ be the block partition of $W$. That is, $P_i = [(i-1)/m, i/m]$ for every $i \in [m]$. Note that for any $i \in [m]$, $\lambda(P_i) = 1/m$. Let $Z_1, \ldots, Z_k$ be independent uniformly random variables in $[0,1]$ used to construct $H(k,W)$. For every $i \in [m]$, let $A_i$ be the number of samples $Z_i$ falling into $P_i$. By the fact that the variables are uniform, for every $i \in [m]$,

$$
E_i[Z_i] = \frac{k_i}{m} \quad \text{and} \quad \text{Var}(A_i) = \frac{1}{m} \left( 1 - \frac{1}{m} \right) k < \frac{k}{m}.
$$

We construct a partition $\mathcal{P}' = \{P'_i \mid i \in [m]\}$ of $[0,1]$ using the values $A_i$. For every $i \in [m]$, we define

$$
P'_i = \left[ \frac{\sum_{i' \leq i} A_{i'}}{m}, \frac{\sum_{i' \leq i} A_{i'}}{m} \right].
$$

We construct an orderon $W_{H(k,W)} \in W$ so that the value of $W_{H(k,W)}$ on $(P'_i \times [0,1]) \times (P'_j \times [0,1])$ is the same as the value of $W$ on $(P_i \times [0,1]) \times (P_j \times [0,1])$. Therefore, $W_{H(k,W)}$ agrees with $W$ on a set

$$
Q = \bigcup_{i,j \in [m]} \left( (P_i \cap P'_i) \times [0,1] \right) \times \left( (P_j \cap P'_j) \times [0,1] \right).
$$

We note that $\lambda(P'_i) = A_i/k$ and

$$
\lambda \left( \bigcup_{i \in [m]} (P_i \cap P'_i) \right) \geq 1 - \sum_{i \in [m]} \left| \sum_{i' \leq i} \frac{A_{i'}}{m} - \frac{i}{m} \right| \geq 1 - \sum_{i \in [m]} \sum_{i' \leq i} \left| \frac{A_{i'}}{m} - \frac{1}{m} \right| \geq 1 - m \sum_{i \in [m]} \left| \frac{A_i}{k} - \frac{1}{m} \right|.
$$

Then,

$$
d_{\Delta}(W, W_{H(k,W)}) \leq \|W - W_{H(k,W)}\| \leq 1 - \lambda(Q) = 1 - \left( 1 - m \sum_{i \in [m]} \left| \frac{A_i}{k} - \frac{1}{m} \right| \right)^2
$$

$$
\leq 2m \sum_{i \in [m]} \left| \frac{A_i}{k} - \frac{1}{m} \right| \leq 2m \left( m \sum_{i \in [m]} \left( \frac{1}{m} - \frac{A_i}{k} \right)^2 \right)^{1/2} = \left( \frac{4m^3}{k^2} \sum_{i \in [m]} \left( \frac{k}{m} - \frac{A_i}{k} \right)^2 \right)^{1/2}.
$$
Therefore, by taking expectation

\[ \mathbb{E}\left[d_\Delta(W, W_{H(k,W)})^2\right] \leq \frac{4m^3}{k^2} \sum_{i \in [m]} \text{Var}(A_i) < \frac{4m^3}{k}, \]

and \( \mathbb{E}\left[d_\Delta(W, W_{H(k,W)})\right] < 2m^{3/2}/\sqrt{k} \) holds by Jensen’s inequality.

By applying Azuma’s inequality (see [Lov12, Corollary A.15]), noting that a single change in \( Z_\ell \) changes the value of \( d_\Delta(W, W_{H(k,W)}) \) by at most \( O(1/k) \), we have for any \( \varepsilon > 0 \),

\[ \Pr\left[d_\Delta(W, W_{H(k,W)}) > \frac{2m^{3/2}}{\sqrt{k}} + \frac{\varepsilon}{k}\right] \leq \exp\left(-C'\varepsilon^2/k\right). \tag{5} \]

for some constant \( C' > 0 \).

For an edge-weighted ordered graph \( H \) on nodes \([k]\), we define \( G(H) \) to be the ordered graph obtained by, for all \( i, j \in [n] \), setting \((i, j)\) to be an edge in \( G(H) \) with probability being the weight of \((i, j)\) in \( H \) independently. By [Lov12, Lemma 10.11], we have for any edge-weighted ordered graph \( H \) and \( \varepsilon > 0 \)

\[ \Pr\left[d_\Delta(W_{G(H)}, W_H) > \frac{\varepsilon}{\sqrt{k}}\right] \leq \Pr\left[\|W_{G(H)} - W_H\|_\square > \frac{\varepsilon}{\sqrt{k}}\right] \leq \exp\left(-\varepsilon^2k/100\right). \tag{6} \]

The desired concentration bound is obtained by (5), (6) and a union bound.

Before proceeding to the next lemma, we first recall the notion of a coupling of distributions.

**Definition 5.3** (couplings). Let \( \mathcal{D}_1 \) and \( \mathcal{D}_2 \) be distributions over domains \( \Omega_1 \) and \( \Omega_2 \), respectively. Then, a coupling of \( \mathcal{D}_1 \) and \( \mathcal{D}_2 \) is a distribution \( \mathcal{D} \) over \( \Omega_1 \times \Omega_2 \) such that the marginal distributions of \( \mathcal{D} \) on \( \Omega_1 \) and \( \Omega_2 \) are the same distributions as \( \mathcal{D}_1 \) and \( \mathcal{D}_2 \), respectively.

Let \( W, W' \in W \) be orderons. The following lemma says that the random ordered graphs \( G(k, W) \) and \( G(k, W') \) can be coupled (note that \( G(k, W) \) and \( G(k, W') \) can be regarded as distributions over ordered graphs of \( k \) vertices) so that, when \( d_\Delta(W, W') \) is small, \( d_\Delta(W_{G(k,W)}, W_{G(k,W')}) \) is also small with high probability.

Roughly speaking, the main idea is as follows. The “cut part” follows from results in the unordered setting, specifically Corollary 10.12 in [Lov12]. For the “shift part”, we show that there is a coupling between samples \( G \sim G(k, W) \) and samples \( G' \sim G(k, W') \), so that for most of the pairs \((G, G')\), one can turn \( G \) into \( G' \) by only making local reordering of vertices, where no vertex is moved more than \( O(k \cdot \text{Shift}(f)) \) steps away from its original location. This follows from the fact that if one takes a sample from an orderon, then the location of the \( i \)-th order statistic—the vertex that is the \( i \)-th smallest in the first coordinate—is typically close to its mean.

**Lemma 5.4.** Let \( W, W' \in W \) be orderons and \( k \) be a positive integer. Then, the random ordered graphs \( G(k, W) \) and \( G(k, W') \) can be coupled so that

\[ d_\Delta(W_{G(k,W)}, W_{G(k,W')}) \leq 9d_\Delta(W, W') + \frac{10}{k^{1/4}} \]

holds with probability at least \( 1 - k(e/4)^{4k}d_\Delta(W, W') - 5e^{-\sqrt{k}/10} \).
Proof: For any $\delta > 0$, there exists $f \in \mathcal{F}$ such that $\text{Shift}(f) + \|W^f - W'\|_\square < d_\Delta(W, W') + \delta$. Here, we choose $\delta = d_\Delta(W, W')$ and fix $f$ for this choice. To define the desired coupling between $G \overset{\text{def}}{=} G(k, W)$ and $G' \overset{\text{def}}{=} G(k, W')$, we first define a coupling between $G$ and $G' \overset{\text{def}}{=} G(k, W^f)$ so that $d_\Delta(W_G, W_{G'})$ is small with high probability, and then define a coupling between $G^f$ and $G'$ so that $d_\Delta(W_{G^f}, W_{G'})$ is small with high probability. We obtain the desired bound by chaining the couplings and a union bound.

Recall that, in the construction of $G(k, W)$, we used two sequences of independent random variables $Z = (Z_i)_{i \in [k]}$ and $Y = (Y_i)_{i \in [k]}$. To look at the construction more in detail, it is convenient to introduce another sequence of independent random variables $R = (R_{ij})_{i,j \in [k], i < j}$, where each $R_{ij}$ is uniform over $[0, 1]$. After defining $X = (X_i)_{i \in [k]}$ as in Definition 5.1, we obtain $G(k, W)$ by setting $(i, j)$ to be an edge if $W((X_i, Y_i), (X_j, Y_j)) \geq R_{ij}$ for each $i,j \in [k]$ with $i < j$. To make the dependence on these random variables more explicit, we write $G(Z, Y, R; W)$ to denote the ordered graph obtained from $W$ by using $Z$, $Y$, and $R$.

Let $(Z = (Z_i)_{i \in [k]}, Y = (Y_i)_{i \in [k]}, R = (R_{ij})_{i,j \in [k], i < j})$ be uniform over $[0, 1]^{k^2}$. Then, we define $(Z' = (Z'_i)_{i \in [k]}, Y' = (Y'_i)_{i \in [k]}, R' = (R'_{ij})_{i,j \in [k], i < j})$ so that $(Z'_i, Y'_i) = f^{-1}(Z_i, Y_i)$ for every $i \in [k]$ and $R'_{ij} = R_{ij}$ for every $i,j \in [k]$ with $i < j$. Note that the marginal $(Z', Y', R')$ is uniform over $[0, 1]^{k^2}$, and hence the distribution of $G(Z', Y', R'; W)$ is exactly same as that of $G(k, W^f)$. Now, we couple $G(Z, Y, R; W)$ with $G(Z', Y', R'; W^f)$.

We can naturally define a measure preserving function $g \in \mathcal{F}$ from $W_G$ to $W_{G'}$ with $W^g_G = W_{G'}$ as follows (G and $G'$ are coupled as in the last paragraph). Let $\pi : [k] \to [k]$ be a permutation of $[k]$ such that $Z_{\pi^{-1}(1)} \leq Z_{\pi^{-1}(2)} \leq \cdots \leq Z_{\pi^{-1}(k)}$. Then, $\pi(i)$ is the position of $Z_i$ in this sorted sequence. Similarly we define a permutation $\pi' : [k] \to [k]$ using $Z'$. Then, we arbitrarily choose $g$ so that the part corresponding to $\pi(i)$ is mapped to the part corresponding to $\pi'(i)$, that is, $\{g(v) \mid v \in [(\pi(i) - 1)/k, \pi(i)/k] \times [0, 1]\} = [(\pi'(i) - 1)/k, \pi'(i)/k] \times [0, 1]$ for every $i \in [k]$.

We now show a concentration bound for $\text{Shift}(g)$. For each $i \in [k]$, we consider a segment $A_i = [Z_i - 2d_\Delta(W, W'), Z_i + 2d_\Delta(W, W')]$ in a circular domain $[0, 1]$, where we identify $-x$ with $1 - x$. Letting $M$ be the maximum number of overlaps of the segments at a point $x$ over $x$ in the circular domain, we can upper bound $\text{Shift}(g)$ by $(M - 1)/k$ because $\text{Shift}(f) \leq 2d_\Delta(W, W')$ and the overlap of two segments may cause a shift of $1/k$ in $g$ to map the vertex corresponding to one segment to the vertex corresponding to the other. Let $\mu = 4kd_\Delta(W, W')$ be the average overlap at a point. As the segments $A_1, \ldots, A_k$ are independently and uniformly distributed, by [SV03, Theorem 3.1] (to apply this theorem, we considered the circular domain instead of the interval $[0, 1]$), we have

$$ \Pr[M \geq 2\mu + 1] \leq k \left(\frac{e}{4}\right)^\mu. $$

Hence, we have

$$ \text{Shift}(g) \leq 8d_\Delta(W, W') $$

with probability at least $1 - k(e/4)^{4kd_\Delta(W, W')}$. 

Next, we couple $G^f$ with $G'$ by coupling $G(Z', Y', R'; W^f)$ with $G(Z', Y', R'; W')$. By [Lov12, Corollary 10.12], we have

$$ \|W_{G^f} - W_G\|_\square = \|W_{G(Z', Y', R', W^f)} - W_{G(Z', Y', R', W')}\|_\square \leq \|W^f - W'\|_\square + \frac{10}{k^{1/4}} $$

24
with probability at least $1 - 5e^{-\sqrt{k}/10}$. Note that we can apply the corollary because the sorting process according to $Z'$ during the constructions of $G_f$ and $G'$ does not affect the cut norm.

Now, we combine by chaining the couplings $(G, G')$ and $(G_f, G')$. By a union bound, we have

$$d_\Delta(W_G, W_{G'}) \leq \text{Shift}(g) + \|W_{G'} - W_G\|_\square$$

$$\leq 8d_\Delta(W, W') + \|W_f - W'\|_\square + \frac{10}{k^{1/4}} \leq 9d_\Delta(W, W') + \frac{10}{k^{1/4}}$$

with probability at least $1 - k(\varepsilon/4)^{4k\varepsilon d_\Delta(W, W')} - 5e^{-\sqrt{k}/10}$.

The proof of Theorem 1.4 now follows easily from the above two lemmas and Theorem 3.8. Indeed, from Theorem 3.8 we know that any orderon $W$ has an arbitrarily close naive block orderon $W'$, and from Lemma 5.4 we conclude that the cut-shift distance between samples from $W$ and samples from $W'$ is typically not much larger than $d_\Delta(W, W')$. Finally, Lemma 5.2 implies that $W'$ is typically close in $d_\Delta$ to large samples taken from it.

**Proof of Theorem 1.4:** Let $W'$ be the naive $m$-block orderon obtained by applying Theorem 3.8 on $W$. Let $(G, G')$ be the coupling obtained by applying Lemma 5.4 on $W$ and $W'$. By the triangle inequality, we have

$$d_\Delta(W, W_G) \leq d_\Delta(W, W') + d_\Delta(W', W_{G'}) + d_\Delta(W_{G'}, W_G).$$

By Lemmas 5.2 and 5.4 and the union bound, we have for any $\varepsilon' > 0$

$$d_\Delta(W, W') + \frac{2m^{3/2}}{k} + \frac{\varepsilon'}{\sqrt{k}} + \frac{9d_\Delta(W', W) + \frac{10}{k^{1/4}}}{k^{1/4}} = 10d_\Delta(W, W') + \frac{2m^{3/2} + \frac{\varepsilon'}{\sqrt{k}} + \frac{10}{k^{1/4}}}{k^{1/4}} = 10\varepsilon + \frac{e^{3/2}2^{2062}/e^9}{\sqrt{k}} + \frac{\varepsilon'}{\sqrt{k}} + \frac{10}{k^{1/4}},$$

with probability at least $1 - k(\varepsilon/4)^{4k\varepsilon} - 5e^{-\sqrt{k}/10} - \exp(C(\varepsilon')^2 k)$. By setting $\varepsilon = \Theta(\log \log k / \log k)^{1/3}$ and $\varepsilon' = \Theta(\sqrt{k} / \log k)$, we have the desired bound.

6 Subgraph statistics

For $k \in \mathbb{N}$, let $\Omega = ([0, 1]^2)^k$. We define $\Omega_o \subset \Omega$ such that $v \in \Omega_o$ if and only if $v$ is ordered according to the entries of the first coordinate in the tuple (in such case, we write $v_1 \leq \cdots \leq v_k$). Namely, $v = ((x_1, y_1), \ldots, (x_k, y_k)) \in \Omega_o$ if and only if for every $i \leq j$, $x_i \leq x_j$. In addition, given a set of pairs $E \subseteq \binom{[k]}{2}$ we let $E_< = \{(i, j) \in E \mid i < j\}$. Let us restate the definition of homomorphism density of an ordered graph in an orderon from Subsection 1.1 in a slightly different but equivalent form.

**Definition 6.1** (homomorphism density). Let $F = ([k], E)$ be a simple ordered graph and let $W \in \mathcal{W}$ be an orderon. We define the (induced) homomorphism density of $F$ in $W$ as

$$t(F, W) \overset{\text{def}}{=} k! \cdot \int_{v \in \Omega_o} \prod_{(i, j) \in E_<} W(v_i, v_j) \cdot \prod_{(i, j) \in E_<} (1 - W(v_i, v_j)) dv,$$
or equivalently,

\[ t(F, W) \overset{\text{def}}{=} k! \cdot \int_{v \in \Omega} \left( \prod_{(i,j) \in E_<} W(v_i, v_j) \cdot \prod_{(i,j) \in E_<} (1 - W(v_i, v_j)) \cdot \prod_{i<j} 1_{v_i \leq v_j} \right) dv. \]

Recall the definition of \( t(F, G) \) where \( G \) is an ordered graph, presented in Subsection 1.1. Clearly, \( t(F, G) = t(F, W_G) \) always holds. Our first main result of the section proves the “only if” direction of Theorem 1.2, showing that if a sequence of orderons \( W_n \) is Cauchy in \( d_\triangle \), then it is convergent in terms of subgraph frequencies.

**Lemma 6.2.** Let \( W, U \in \mathcal{W} \). Then, for every simple ordered graph \( F = ([k], E) \)

\[ |t(F, W) - t(F, U)| \leq 6k! \left( \frac{k}{2} \right) \sqrt{d_\triangle(W, U)}. \]

In order to prove the above, we introduce the following two lemmas. Lemma 6.3 considers pairs of orderons that are close in cut-norm, whereas Lemma 6.4 describes the effect of shifts.

**Lemma 6.3.** For any \( W, U \in \mathcal{W} \) and every ordered graph \( F = ([k], E) \)

\[ |t(F, W) - t(F, U)| \leq 2k! \left( \frac{k}{2} \right) \sqrt{\|W - U\|_\square}. \]

**Proof:** Fix some arbitrary ordering on \( \binom{[k]}{2}_< \). For every \((i, j) \in \binom{[k]}{2}_< \) we define the function

\[ \gamma_W(v_i, v_j) = \begin{cases} W(v_i, v_j) \cdot 1_{v_i \leq v_j}, & \text{if } (i, j) \in E(F) \\ (1 - W(v_i, v_j)) \cdot 1_{v_i \leq v_j}, & \text{if } (i, j) \notin E(F) \end{cases}, \]

and define \( \gamma_U \) similarly.

\[ t(F, W) - t(F, U) = k! \int_{v \in \Omega} \left( \prod_{(i,j) \in \binom{[k]}{2}_<} \gamma_W(v_i, v_j) - \prod_{(i,j) \in \binom{[k]}{2}_<} \gamma_U(v_i, v_j) \right) dv. \]

By identifying each \( e_r \in \binom{[k]}{2}_< \) with \((i_r, j_r)\), the integrand can be written as

\[ \sum_{s=1}^{\binom{k}{2}} \left( \prod_{r<s} \gamma_W(v_{i_r}, v_{j_r}) \right) \left( \prod_{r>s} \gamma_U(v_{i_r}, v_{j_r}) \right) \left( \gamma_W(v_{i_s}, v_{j_s}) - \gamma_U(v_{i_s}, v_{j_s}) \right). \]

To estimate the integral of a given term, we fix all variables except \( v_{i_s} \) and \( v_{j_s} \). Then, the integral is of the form

\[ \int_{v_{i_s}, v_{j_s}} g(v_{i_s}) h(v_{j_s}) \left( \gamma_W(v_{i_s}, v_{j_s}) - \gamma_U(v_{i_s}, v_{j_s}) \right) dv_{i_s} dv_{j_s}, \]
where \( g, h : [0, 1]^2 \to [0, 1] \) are some functions. By applying Lemma 2.5, it suffices to provide an upper bound on

\[
\sup_{S,T \subseteq [0,1]^2} \left| \int_{S \times T} \left( \gamma_W(v_i, v_j) - \gamma_U(v_i, v_j) \right) dv_i dv_j \right|.
\]

By using Lemma 2.6, we get that

\[
\sup_{S,T \subseteq [0,1]^2} \left| \int_{S \times T} \left( \gamma_W(v_i, v_j) - \gamma_U(v_i, v_j) \right) dv_i dv_j \right| = \sup_{S,T \subseteq [0,1]^2} \left| \int_{S \times T} \left( W(v_i, v_j) - U(v_i, v_j) \right) 1_{v_i \leq v_j} dv_i dv_j \right| = \|W - U\|_\infty \leq 2\sqrt{\|W - U\|}.
\]

By summing up over all \( \binom{k}{2} \) pairs of vertices, the lemma follows.

**Lemma 6.4.** Let \( U \in \mathcal{W} \) and let \( \phi : [0, 1]^2 \to [0, 1]^2 \) be a measure preserving function. Then, for every ordered graph \( F = ([k], E) \)

\[
|t(F, U) - t(F, U^\phi)| \leq 4k! \binom{k}{2} \cdot \text{Shift}(\phi).
\]

**Proof:** The proof is similar to Lemma 6.3. However, we shall slightly change notation. Let \( \gamma_W \) be defined as follows.

\[
\gamma_W(v_i, v_j) = \begin{cases} W(v_i, v_j), & \text{if } (i, j) \in E(F) \\ 1 - W(v_i, v_j), & \text{if } (i, j) \notin E(F) \end{cases}.
\]

Then, by changing the integration variables in \( t(F, U^\phi) \) from \( v_i, v_j \) to \( \phi^{-1}(v_i), \phi^{-1}(v_j) \), we have

\[
t(F, U) - t(F, U^\phi) = k! \int_{v \in \Omega} \left( \prod_{(i,j) \in \binom{[k]}{2}} \frac{\gamma_U(v_i, v_j)}{1 - \phi^{-1}(v_i, v_j)} \frac{\gamma_U^\phi(\phi^{-1}(v_i), \phi^{-1}(v_j))}{1 - \phi^{-1}(v_i, v_j)} \right) dv
\]

\[
= k! \int_{v \in \Omega} \left( \prod_{(i,j) \in \binom{[k]}{2}} \frac{\gamma_U(v_i, v_j)}{1 - \phi^{-1}(v_i, v_j)} \frac{\gamma_U(v_i, v_j)}{1 - \phi^{-1}(v_i, v_j)} \right) dv
\]

\[
= k! \int_{v \in \Omega} \prod_{(i,j) \in \binom{[k]}{2}} \gamma_U(v_i, v_j) \left( \prod_{(i,j) \in \binom{[k]}{2}} \frac{1}{1 - \phi^{-1}(v_i, v_j)} \right) dv.
\]

Hence,

\[
|t(F, U) - t(F, U^\phi)| \leq k! \int_{v \in \Omega} \left( \prod_{(i,j) \in \binom{[k]}{2}} 1_{v_i \leq v_j} - \prod_{(i,j) \in \binom{[k]}{2}} 1_{\phi^{-1}(v_i) \leq \phi^{-1}(v_j)} \right) dv
\]

\[
= k! \int_{v \in \Omega} \sum_{s=1}^{\binom{k}{2}} \left( \prod_{r<s} 1_{v_r \leq v_j} \prod_{r>s} 1_{\phi^{-1}(v_r) \leq \phi^{-1}(v_j)} \right) \left( 1_{v_{s} \leq v_{j}} - 1_{\phi^{-1}(v_{s}) \leq \phi^{-1}(v_{j})} \right) dv.
\]

27
Similarly to Lemma 6.3, we fix all the variables except \( v_{i_s} \) and \( v_{j_s} \). Then, by using Lemma 2.5, it suffices to estimate
\[
\left| \int_{v_{i_s},v_{j_s}} \left( 1_{v_{i_s} \leq v_{j_s}} - 1_{\phi^{-1}(v_{i_s}) \leq \phi^{-1}(v_{j_s})} \right) dv_{i_s} dv_{j_s} \right| \leq \int_{v_{i_s},v_{j_s}} \left| 1_{v_{i_s} \leq v_{j_s}} - 1_{\phi^{-1}(v_{i_s}) \leq \phi^{-1}(v_{j_s})} \right| dv_{i_s} dv_{j_s}.
\]
Note that whenever the intersection between \([\pi_1(v_{i_s}), \pi_1(\phi^{-1}(v_{i_s}))]\) and \([\pi_1(v_{j_s}), \pi_1(\phi^{-1}(v_{j_s}))]\) is empty, the difference between the indicators is zero. Therefore,
\[
\left| \int_{v_{i_s},v_{j_s}} \left( 1_{v_{i_s} \leq v_{j_s}} - 1_{\phi^{-1}(v_{i_s}) \leq \phi^{-1}(v_{j_s})} \right) dv_{i_s} dv_{j_s} \right| \leq \int_{v_{i_s},v_{j_s}} 1_{[\pi_1(v_{i_s}), \pi_1(\phi^{-1}(v_{i_s}))] \cap [\pi_1(v_{j_s}), \pi_1(\phi^{-1}(v_{j_s}))] \neq \emptyset} dv_{i_s} dv_{j_s} \leq 4\text{Shift}(\phi).
\] (9)

By summing up over all \( \binom{n}{2} \) pairs of vertices, the lemma follows.

**Proof of Lemma 6.2:** For any \( \gamma > 0 \) let \( \phi: [0,1]^2 \to [0,1]^2 \) be the measure preserving function such that \( \text{Shift}(\phi) + \|W - U^\phi\|_\Box \leq d_\Box(W, U) + \gamma \). For this specific \( \phi \) we have that
\[
\|W - U^\phi\|_\Box = \sup_{S,T \subseteq [0,1]^2} \left| \int_{S \times T} W(v_1, v_2) - U(\phi(v_1), \phi(v_2)) dv_1 dv_2 \right| \leq d_\Box(W, U) + \gamma,
\]
and \( \text{Shift}(\phi) \leq d_\Box(W, U) + \gamma \). Then, by assuming \( d_\Box(W, U) > 0 \) (note that the case \( d_\Box(W, U) = 0 \) is covered by considering what happens when \( d_\Box(W, U) \to 0 \)), and using the triangle inequality combined with Lemma 6.3 and Lemma 6.4, we get
\[
|t(F, W) - t(F, U)| \leq |t(F, W) - t(F, U^\phi)| + |t(F, U^\phi) - t(F, U)|
\leq 2k! \left( \frac{k}{2} \right) \sqrt{d_\Box(W, U)} + k! \left( \frac{k}{2} \right) \frac{\gamma}{\sqrt{d_\Box(W, U)}} + 4k! \left( \frac{k}{2} \right) (d_\Box(W, U) + \gamma)
\leq 6k! \left( \frac{k}{2} \right) \sqrt{d_\Box(W, U)} + + k! \left( \frac{k}{2} \right) \frac{\gamma}{\sqrt{d_\Box(W, U)}} + 4k! \left( \frac{k}{2} \right) \sqrt{\gamma}.
\]
As the choice of \( \gamma \) is arbitrary, the lemma follows.

Next we prove a converse statement, showing that if all frequencies of \( k \)-vertex graphs in a pair of orderons \( W \) and \( U \) are very similar, than \( d_\Box(W, U) \) is small. This establishes the “if” component of Theorem 1.2.

**Lemma 6.5.** Let \( k \in \mathbb{N} \) and \( W, U \in \mathcal{W} \). Assume that for every ordered graph \( F \) on \( k \) vertices,
\[
|t(F, W) - t(F, U)| \leq 2^{-k^2}.
\]
Then, \( d_\Box(W, U) \leq 2C \left( \frac{\log \log k}{\log k} \right)^{1/3} \) for some constant \( C > 0 \).

**Proof:** We start by showing that if for some \( k \geq 2 \), the total variation distance between the distribution \( \mathbf{G}(k, W) \) and \( \mathbf{G}(k, U) \) is small then they are close in CS-distance.
Assume that for $U, W \in \mathcal{W}$ and some $k \geq 2$ it holds that
\[
\mathcal{d}_{TV}(G(k, W), G(k, U)) < 1 - \exp\left(\frac{-k}{2 \log k}\right).
\]
This assumption implies that there exists a joint distribution $(G(k, W), G(k, U))$ so that $G(k, W) = G(k, U)$ with probability larger than $\exp\left(\frac{-k}{2 \log k}\right)$. By Theorem 1.4, with probability at least $1 - C \exp(-\sqrt{k/C})$, we have that $d(\overline{W}, G(k, W)) \leq C \left(\frac{\log \log k}{\log k}\right)^{1/3}$ and $d(U, G(k, U)) \leq C \left(\frac{\log \log k}{\log k}\right)^{1/3}$, respectively.

Therefore, by using a union bound, $\Pr[\overline{E}_1 \vee \overline{E}_2 \vee \overline{E}_3] \leq 2C \exp\left(-\sqrt{k/C}\right) + \exp\left(-k/2 \log k\right) < 1$. Hence, there is a positive probability for all the three events to occur, implying that
\[
d(\overline{W}, U) \leq d(\overline{W}, G(k, W)) + d(U, G(k, U)) + d(G(k, U), G(k, W)) \leq 2C \left(\frac{\log \log k}{\log k}\right)^{1/3}.
\]

The lemma follows by noting that,
\[
|t(F, W) - t(F, U)| = \left| \Pr_{G \sim G(k, W)}[G = F] - \Pr_{G \sim G(k, U)}[G = F] \right| \leq 2^{-k^2},
\]
and hence,
\[
\mathcal{d}_{TV}(G(k, W), G(k, U)) = \sum_F \left| \Pr_{G \sim G(k, W)}[G = F] - \Pr_{G \sim G(k, U)}[G = F] \right| \leq 2^{6k} \cdot 2^{-k^2} \leq 2^{-k}
\]
\[
< 1 - \exp\left(\frac{-k}{2 \log k}\right).
\]

7 The furthest ordered graph from a hereditary property

In this section we prove Theorem 1.7. The proof roughly follows along the same lines as the proof of Lovász and Szegedy for the unordered setting [LS10]. However, before proceeding, let us show why an exact analogue of Theorem 1.6 cannot hold for the ordered setting.

7.1 The random graph $G(n, p)$ is not always the furthest

Recall the hereditary property $\mathcal{H}$ defined in Subsection 1.3. For convenience, let us describe $\mathcal{H}$ again. An ordered graph $G$ satisfies $\mathcal{H}$ 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. Here we prove that (a direct analogue of) Theorem 1.6 does not hold for $\mathcal{H}$, that is, for any $p$, a typical graph $G \sim G(n, p)$ is not asymptotically the furthest
The following lemma characterizes the property \( \mathcal{H} \) in a way that will make it fairly straightforward, given an ordered graph \( G \), to estimate the distance \( d_1(G, \mathcal{H}) \).

**Lemma 7.1** (Thresholding lemma). An ordered graph \( G \) on vertices \( v_1 < v_2 < \cdots < v_n \) satisfies \( \mathcal{H} \) if and only if there exists a “threshold” \( i \in [n] \) for which the following two conditions hold.

- For any \( 1 \leq j < j' \leq i \), the vertices \( v_j \) and \( v_{j'} \) are connected in \( G \).
- For any \( i + 1 \leq j < j' \leq n \), the vertices \( v_j \) and \( v_{j'} \) are not connected in \( G \).

**Proof:** Suppose first that \( G \in \mathcal{H} \), and pick the maximal \( i \in [n-1] \) for which all pairs of the form \( v_jv_{j'} \) where \( 1 \leq j < j' \leq i \) are connected in \( G \). The first condition holds trivially, and it only remains to prove the second one. Suppose on the contrary that it does not hold, which means that \( i < n-1 \) and that \( v_j \) and \( v_{j'} \) are connected for some \( i+1 \leq j < j' \leq n \). By the maximality of \( i \), there exists some \( j'' \leq i \) for which \( v_{j''}v_{i+1} \) is not an edge. Since \( i+1 \leq j \), the four vertices \( v_{j''} < v_{i+1} \leq v_j < v_{j'} \) induce the ordered pattern forbidden by \( \mathcal{H} \), a contradiction.

Conversely, if there exists \( i \) for which the two conditions hold, then for any tuple of vertices \( v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4} \) where \( j_1 < j_2 \leq j_3 < j_4 \), at least one of the following holds. Either we have \( j_2 \leq i \), meaning that \( v_{j_1}v_{j_2} \) is an edge, or \( j_3 \geq i+1 \), meaning that \( v_{j_3}v_{j_4} \) is a non-edge. In both cases, the tuple does not violate the condition in the definition of \( \mathcal{H} \).

Using Lemma 7.1, we can estimate the typical distance of a graph \( G \sim G(n,p) \) from \( \mathcal{H} \).

**Lemma 7.2.** For any \( n \in \mathbb{N} \) and \( p \in [0,1] \) (possibly depending on \( n \)), a graph \( G \sim G(n,p) \) satisfies with high probability that \( d_1(G, \mathcal{H}) = p(1-p) + o(1) \leq \frac{1}{4} + o(1) \).

**Proof:** Standard Chernoff-type concentration bounds show that w.h.p. the following holds for all \( i \in [n] \): the number of non-edges among the first \( i \) vertices is \( (1-p)^{\frac{i^2}{2}} + o(n^2) \), and the number of edges among the last \( n-i \) vertices is \( p(n-i)^{\frac{i^2}{2}} + o(n^2) \). Thus, w.h.p. we have

\[
d_1(G, \mathcal{H}) = \frac{1}{\binom{n}{2}} \min_{i \in [n]} \left( (1-p)^{\frac{i^2}{2}} + p\frac{(n-i)^2}{2} + o(n^2) \right) = p(1-p) + o(1)
\]

where the minimum of the sum (up to the \( o(n^2) \) term) is attained for \( i \in \{[np], [np] \} \). The \( p(1-p) \) term is maximized at \( p = 1/2 \), where it equals 1/4.

On the other hand, there exist large graphs whose distance from \( \mathcal{H} \) is much larger than 1/4.

**Lemma 7.3.** Let \( n \in \mathbb{N} \), and consider the graph \( G \) on the vertex set \( [n] \) with the standard ordering, where \( u \) and \( v \) are connected if and only if \( u + v \geq n \). The graph \( G \) satisfies \( d_1(G, \mathcal{H}) \geq 1/2 - o(1) \).

**Proof:** Fix any \( i \in [n] \). Without loss of generality, we may assume that \( i \leq n/2 \) (the case \( i > n/2 \) is symmetric). On one hand, the number of non-edges between pairs of vertices \( j < j' \in [i] \) is \( \binom{i}{2} \).

On the other hand, a straightforward calculation shows that the number of edges between pairs of vertices \( j, j' \) where \( i + 1 \leq j < j' \) is \( \frac{1}{2} \binom{n}{2} - \binom{i}{2} - o(n^2) \). Thus, the number of edges that one needs
to add or remove if the threshold of Lemma 7.1 is set to \( i = \frac{1}{2} \binom{n}{2} - o(n^2) \) (regardless of the value of \( i \)), implying that \( d_1(G, \mathcal{H}) \geq 1/2 - o(1) \).

Conversely, it is trivial that \( d_1(H) \leq 1/2 \), as any graph \( G \) can be turned into either a complete or empty graph (which both satisfy \( \mathcal{H} \)) by adding or removing at most \( \frac{1}{2} \binom{n}{2} \) edges. Combined with the last lemma, we conclude that \( d_1(H) = 1/2 \).

### 7.2 Proof of Theorem 1.7

We continue to the proof of Theorem 1.7. Along this section, the ordered graphs that we consider are generally simple, and the notion of a hereditary property refers to a property of simple ordered graphs. To begin, we first establish several basic properties of hereditary properties in the ordered setting, starting with a discussion on their closure in the space of orderons.

Let \( H \) be a hereditary simple ordered graph property (\( H \) will be fixed throughout the section). Recall that a property is hereditary if it satisfies the following: if \( G \in H \) then every induced subgraph of \( G \) (vertex repetitions are not allowed) is also in \( H \).

In general, we define the closure \( \overline{P} \) of an ordered graph property \( P \) as the set of all orderons \( W \) for which there exists a sequence of graphs \( G_n \in P \) (with \( |V(G_n)| \to \infty \)) that converges to \( P \) in \( d_\Delta \). Note that \( \overline{P} \) is a closed set in \( \overline{W} \) with respect to \( d_\Delta \) (also note that it is generally not true that \( W \in P \) for any \( G \in P \)). The following characterization for the closure of a hereditary property will be useful in multiple occasions along this section as well as in Section 9.

**Lemma 7.4.** Let \( H \) be a hereditary property of ordered graphs. The following conditions are equivalent for an orderon \( W \).

- \( W \in \overline{H} \).
- \( t(F, W) = 0 \) for any ordered graph \( F \not\in H \).
- \( \Pr[G(k, W) \in H] = 1 \) for every \( k \).

**Proof:** The second and third conditions are clearly equivalent: for any fixed \( k \), the probability that \( G(k, W) \not\in H \) is the sum of \( t(F, W) \) over all \( F \not\in H \) on \( k \) vertices. We now show equivalence between the first condition and the other two. If \( W \in \overline{H} \) then there exists a sequence of ordered graphs \( G_n \in H \) with \( |V(G_n)| \to \infty \) that converges to \( W \) in \( d_\Delta \). Fix \( k \). For any \( n \), the density \( t(F, W_n) \) is bounded by the probability that, when picking \( k \) vertices from a set of \( n \) vertices with repetition, some vertex will be picked more than once (see Observation 8.1 for more details). This probability is bounded by \( \binom{k}{2}/n \to 0 \), that is \( t(F, W_n) \to 0 \) as \( n \to \infty \). Since \( G_n \to W \), it follows that \( t(F, W) = \lim t(F, W_n) = 0 \). The converse follows immediately from the third condition combined with our sampling theorem, which together show that with high probability, a large enough sample of \( W \) will both satisfy \( H \) and be arbitrarily close to \( W \) in \( d_\Delta \). 

We will need the definition of a flexible property, given below; this is an ordered analogue of flexible properties in the unordered setting, defined in [LS10].

**Definition 7.5** (support, flexibility). For an orderon \( W \in \overline{W} \) and a value \( 0 \leq \alpha \leq 1 \), the \( \alpha \)-support of \( W \), denoted \( \text{supp}_\alpha(W) \), is the set of all pairs \((u, w) \in [0, 1]^2\) for which \( W(u, v) = \alpha \).
We say that a property \( R \) is flexible if for any \( W,W' \in \mathcal{W} \) where \( \text{supp}_\alpha(W) \subseteq \text{supp}_\alpha(W') \) for \( \alpha = 0,1 \) and \( W \in R \) it holds that \( W' \in R \) as well. In particular, this means that the supports \( \text{supp}_\alpha(W) \) for \( \alpha = 0,1 \) fully determine whether \( W \) satisfies a given flexible property.

**Lemma 7.6.** The closure \( \overline{\mathcal{H}} \) of a hereditary property is flexible.

**Proof:** By Lemma 7.4, the closure of a hereditary property can be defined by conditions of the form 
\[ t(F,U) = 0 \] (for all graphs \( F \notin \mathcal{H} \)). For every fixed \( F \notin \mathcal{H} \), this condition means that for almost all vectors \( v \in [0,1]^{|V(F)|} \), at least one of the factors in \( \prod_{(i,j) \in E} U(v_i,v_j) \cdot \prod_{(i,j) \in E^c} (1 - U(v_i,v_j)) \) must be 0. This condition is preserved if values of \( W \) that are strictly between 0 and 1 are changed to any other value (including 0 and 1).

The next lemma shows, using the flexibility, that the distance from the closure of a hereditary property is a concave function over the space of orderons.

**Lemma 7.7.** Let \( \mathcal{H} \) be a hereditary ordered graph property. The distance \( d_1(\cdot, \overline{\mathcal{H}}) \) is a concave function over \( \mathcal{W} \).

**Proof:** Let \( W_1, W_2, U \in \mathcal{W} \) be three orderons satisfying \( U = \lambda W_1 + (1 - \lambda) W_2 \) for some \( 0 < \lambda < 1 \). It suffices to show that \( d_1(U, \overline{\mathcal{H}}) \geq \lambda d_1(W_1, \overline{\mathcal{H}}) + (1 - \lambda) d_1(W_2, \overline{\mathcal{H}}) \). The crucial observation is that \( \text{supp}_\alpha(U) = \text{supp}_\alpha(W_1) \cap \text{supp}_\alpha(W_2) \) for \( \alpha = 0,1 \).

Let \( U' \) be any orderon in \( \overline{\mathcal{H}} \). We show that there exist \( W'_1, W'_2 \in \overline{\mathcal{H}} \) so that
\[
\lambda d_1(W_1, W'_1) + (1 - \lambda) d_1(W_2, W'_2) \leq d_1(U, U').
\]

Indeed, for \( i = 1,2 \) pick \( W'_i \) as the unique orderon satisfying the following: \( W'_i(u,v) = 0 \) for \( (u,v) \in \text{supp}_0(U') \); \( W'_i(u,v) = 1 \) for \( (u,v) \in \text{supp}_1(U') \); and otherwise, \( W'_i(u,v) = W_i(u,v) \). By Lemma 7.6, \( W'_1, W'_2 \in \overline{\mathcal{H}} \), and it is easy to verify that \( W'_1, W'_2 \) satisfy the desired inequality. The proof of the lemma now follows by taking the infimum over all \( U', W'_1, W'_2 \in \mathcal{H} \).

Recall that an orderon \( W \) is called naive if its values do not depend on the second (“variability”) coordinate. That is, for every \( x,y \in [0,1] \) it holds that \( W((x,a), (y,b)) = W((x,a'), (y,b')) \) for any \( a,a',b,b' \in [0,1] \). Our next goal is to show, using the concavity, that for any hereditary property \( \mathcal{H} \) and orderon \( W \) there exists a naive orderon \( W' \) at least as far from \( \overline{\mathcal{H}} \) as \( W \).

**Lemma 7.8.** For any hereditary property \( \mathcal{H} \) and orderon \( W \) there exists a naive orderon \( W' \) so that \( d_1(W', \overline{\mathcal{H}}) \geq d_1(W, \overline{\mathcal{H}}) \).

**Proof:** To prove the lemma, we combine the concavity guaranteed by the previous lemma with some measure-theoretic and probabilistic tools. For any \( (x,y) \in [0,1]^2 \) define \( \mu^{x,y} \) as the expectation, over all \( (a,b) \in [0,1]^2 \), of \( W((x,a), (y,b)) \). We will prove that the unique \( W' : [0,1]^2 \rightarrow [0,1] \) for which \( W'((x,a), (y,b)) = \mu^{x,y} \) for all \( (x,y) \) satisfies the conditions of the lemma. (To show that \( W' \) is a naive orderon one needs to prove that \( W' \) as defined above is indeed a measurable function, but this will follow from the proof.)

Fix an integer \( n \) and pick \( A_n = \{ (x,y) \in [0,1]^2 : Q_n(x) \neq Q_n(y) \} \), where the function \( Q_n \) is as defined in the beginning of Section 3. Consider the family of measure-preserving bijections \( f_{\alpha_1,\alpha_2,\ldots,\alpha_n} \), where \( (\alpha_1,\ldots,\alpha_n) \in [0,1]^n \), defined as follows: For any \( (x,a) \in [0,1]^2 \), we set
\( f_{\alpha_1,\alpha_2,...,\alpha_n}(x,a) = (x,a+\alpha_{Q_n}(x)) \). Note that \( \text{Shift}(f_{\alpha_1,\alpha_2,...,\alpha_n}) = 0 \) always holds. As these bijections clearly do not change the \( L^1 \)-distance of \( W \) from \( \mathcal{H} \), we have that \( d_1(W, \mathcal{H}) = d_1(Wf_{\alpha_1,\alpha_2,...,\alpha_n}, \mathcal{H}) \) for any choice of \((\alpha_1,\ldots,\alpha_n) \in [0,1]^n\).

Pick \( n \) tuples \( \{(\alpha_1^i,\ldots,\alpha_n^i)\}_{i=1}^n \) uniformly at random. Set \( g_i = f_{\alpha_1^i,...,\alpha_n^i} \) and consider the orderon

\[
U_n = \frac{1}{n} \sum_{i=1}^n Wg_i = \frac{1}{n} \sum_{i=1}^n Wf_{\alpha_1^i,...,\alpha_n^i}.
\]

Since each \( g_i \) is a measure-preserving bijection with shift zero, we have that \( d_1(W, \mathcal{H}) = d_1(W_{g_i}, \mathcal{H}) \). As \( U_n \) is a convex combination of the \( W_{g_i} \), we get that

\[
d_1(U_n, \mathcal{H}) \geq \frac{1}{n} \sum_{i=1}^n d_1(W_{g_i}, \mathcal{H}) = d_1(W, \mathcal{H}).
\]

Now, given any fixed orderon \( U \in \{U_n, W_{g_1}, W_{g_2},\ldots, W_{g_n}\} \) and \((x,y) \in [0,1]^2\), define the random variable \( X_{U,x,y}^{x,y} \) as follows: pick \( (a,b) \in [0,1]^2 \) uniformly at random, and return the value \( U((x,a),(y,b)) - \mu^{x,y} \). Note that the expectation of \( X_{U,x,y}^{x,y} \) for any \( U \) as above is zero.

A basic probabilistic fact states that, for \( n \) i.i.d. bounded random variables \( Y_1,\ldots,Y_n \), the variance of \( \overline{Y} = (Y_1 + \cdots + Y_n)/n \) equals \( \text{Var}(Y_1)/n \), and tends to zero as \( n \to \infty \). Fix some \((x,y) \in A_n\), and for any \( i \in [n] \) take \( Y_i = X_{U,x,y}^{x,y} \). Finally pick \( \overline{Y} = X_{U,x,y}^{x,y} \), and note that the \( Y_i \) are indeed i.i.d., that \( |Y_i| \leq 1 \) (meaning that \( \text{Var}(Y_i) \leq 1 \)), and that \( \overline{Y} = \sum Y_i/n \). It follows that the variance of \( \overline{Y} = X_{U,x,y}^{x,y} \) is bounded by \( 1/n \).

Consequently, for fixed \((x,y) \in A_n\), the probability (ranging over the choices of \( g_1,\ldots,g_n \) and \( a,b \in [0,1]^2 \)) that \(|U_n((x,a),(y,b)) - \mu^{x,y}| \geq 1/n^{1/3} \) is bounded by \( 1/n^{1/3} \). Thus, the expected measure of the collection

\[
B_n \overset{\text{def}}{=} \{(x,a,y,b) : |U_n((x,a),(y,b)) - \mu^{x,y}| \geq 1/n^{1/3}\}
\]

of “bad” \((x,a,y,b)\) tuples is bounded by \( 1/n^{1/3} \) where the randomness is over the choices of \( g_1,\ldots,g_n \). In particular, there exists such a choice of \( U_n \) for which \(|B_n| \leq 1/n^{1/3}\); we henceforth fix this choice of \( U_n \).

As \(|[0,1]^2 \setminus A_n| = 1/n^2 \to 0 \) as \( n \to \infty \), and \( \int_{A_n} |U_n - U_{n'}| \leq 4/n^{1/3} \) for \( n' > n \), it follows that the sequence \( \{U_n\} \) is \( L^1 \)-Cauchy. Thus, it converges in \( L^1 \) to some limit \( U' \), and clearly the set of points \( \{(x,y) \in [0,1]^2 : U'(x,y) \neq \mu^{x,y}\} \) is of measure zero. In other words, \( U' \) is measurable and equals \( W' \) defined above almost everywhere, and thus \( W' \) is measurable as well and \( U_n \to W' \) in \( L^1 \). As \( d_1(U_n, \mathcal{H}) \geq d_1(W, \mathcal{H}) \) for any \( n \), the inequality still holds at the limit, that is, \( d_1(W', \mathcal{H}) \geq d_1(W, \mathcal{H}) \). This concludes the proof.

**Definition 7.9.** For any hereditary property \( \mathcal{H} \) let \( \Delta_\mathcal{H} \) denote the supremum of the distance \( d_1(W, \mathcal{H}) \) among all \( W \in \mathcal{W} \).

The following is an immediate consequence of the last lemma.

**Lemma 7.10.** For any hereditary property \( \mathcal{H} \) and \( \varepsilon > 0 \) there exists a naive orderon \( W' \) for which \( d_1(W', \mathcal{H}) \geq \Delta_\mathcal{H} - \varepsilon \).
With some abuse of notation, we henceforth view a naive orderon $W$ as a measurable symmetric function from $[0, 1]^2$ to $[0, 1]$ and for $x, y \in [0, 1]$ we denote by $W(x, y)$ the (unique) value of $W((x, a), (y, b))$ for $a, b \in [0, 1]$.

Recall the definition of a naive block orderon from Section 3 (in particular, the fact that the blocks are consecutive in terms of order). The next lemma asserts that there exist naive block orderons with a bounded number of blocks, that are almost the furthest away from $\overline{H}$.

**Lemma 7.11.** Let $H$ be a hereditary property and let $\varepsilon > 0$. There exists a naive block orderon $W$ with at most $M_H(\varepsilon)$ blocks so that $d_1(W, \overline{H}) \geq \Delta_H - \varepsilon$.

**Proof:** The proof relies on a fundamental fact in Lebesgue measure theory, stating that (finite) linear combinations of indicator functions of the form $I = I_{[a_i, b_i] \times [c_i, d_i]}$ are dense in the two-dimensional Lebesgue space $L^1[\mathbb{R}^2]$ (and specifically, in $L^1([0, 1]^2]$). Indeed, Theorem 2.4 (ii) in [SS05] states that for any function $W$ in $L^1[\mathbb{R}^2]$ and $\delta > 0$ there exists $N = N(W, \varepsilon) \geq 0$ and a step function $T$ of the form

$$T = \sum_{i=1}^{N} \alpha_i \cdot I_{[a_i, b_i] \times [c_i, d_i]},$$

where $\alpha_i \in \mathbb{R}$, so that $d_1(W, T) \leq \delta$. In our case, $W$ is a naive orderon, and is thus a symmetric function in $L^1([0, 1]^2]$. We note that, while Theorem 2.4 (ii) in [SS05] does not guarantee that the function $T$ approximating $W$ is symmetric, we can easily make $T$ symmetric—and in particular, a naive orderon—by replacing it with $(T + \overline{T})/2$, where $\overline{T} = \sum_{i=1}^{N} \alpha_i \cdot I_{[c_i, d_i] \times [a_i, b_i]}$.

We now describe how to construct a naive block orderon which is $L^1$-close to the above $T$. Recall the definition of $Q_n$ given in the beginning of Section 3, and for $M \geq 4N/\delta$, consider the following weighted ordered graph $G$ on $M$ vertices. Given $(i, j) \in [M]^2$, if $T$ is constant over all pairs $(x, y) \in [0, 1]^2$ for which $Q_M(x) = i$ and $Q_M(y) = j$, then we define $G(i, j)$ to be this constant value (and say that $(i, j)$ is good). Otherwise, $G(i, j)$ is defined arbitrarily. The number of pairs $(i, j)$ that are not good is at most $4MN$. Indeed, $(i, j)$ might not be good only if the square $[(i - 1)/M, i/M] \times [(j - 1)/M, j/M]$ intersects the boundary of $[a_\ell, b_\ell] \times [c_\ell, d_\ell]$ for some $\ell \in [N]$. However, the latter boundary intersects at most $4M$ such squares, and summing over all $\ell \in [N]$ we get the desired bound.

Now let $U$ be the naive block orderon over $M$ blocks defined by $U(x, y) = G(Q_M(x), Q_M(y))$ for any $(x, y) \in [0, 1]^2$. We get that $d_1(U, T) \leq 4MN/M^2 \leq \delta$, where the last inequality holds by our choice of $M$. It thus follows that $d_1(W, U) \leq d_1(W, T) + d_1(U, T) \leq 2\delta$.

To conclude the proof, let $H$ be a hereditary property, let $\varepsilon > 0$, and let $W$ be an arbitrary naive orderon satisfying that $d_1(W, \overline{H}) \geq \Delta_H - \varepsilon/3$ (the existence of such a $W$ is guaranteed by Lemma 7.10); specifically we can take such a $W$ that minimizes the quantity $N(W, \varepsilon)$, which was defined in the beginning of the proof. Now take $\delta = \varepsilon/3$ and let $U$ denote the naive block orderon over $M = \lceil 4N/\delta \rceil$ blocks which satisfies $d_1(U, W) \leq 2\delta = 2\varepsilon/3$. By the triangle inequality, we conclude that $d_1(U, \overline{H}) \geq \Delta_H - \varepsilon$, as desired. Note that $M$ depends only on $\varepsilon$ and $N$, which in turn depends only on $H$ and $\varepsilon$.

We now show how our results for orderons can be translated to finite graphs. Here we make use of several technical lemmas from [LS10]. The first lemma that we need is the following.
Lemma 7.12 (Lemma 3.13(a) in [LS10]). For any hereditary property $\mathcal{H}$ and simple ordered graph $G$, we have $d_1(G, \mathcal{H}) \leq d_1(W_G, \overline{\mathcal{H}}) \leq \Delta_\mathcal{H}$.

While the above lemma was stated in [LS10] for hereditary properties of unordered graphs, its proof in [LS10] only uses the flexibility of a closure of a hereditary property and a simple subgraph statistics argument, and translates as-is to our ordered setting.

The next lemma that we need from [LS10] is the following.

Lemma 7.13 (Lemma 2.8 in [LS10]). Suppose that $\|U_n - U\|_\square \to 0$ and $\|W_n - W\|_\square \to 0$ (where $U, U_n, W, W_n$ are naive orderons). Then $\liminf_{n \to \infty} d_1(W_n, U_n) \geq d_1(W, U)$.

Additionally, we need the following lemma, concerning the good behavior of sequences that converge to a naive block orderon.

Lemma 7.14. Fix an integer $M > 0$, and let $W$ be a naive $M$-block orderon. Also let $\{W_n\}$ be a sequence of orderons where $d_\triangle(W_n, W) \to 0$, and let $\{f_n\}$ be a sequence of shift functions with $\text{Shift}(f_n) \to 0$. Then it holds that $\|W_n^{f_n} - W\|_\square \to 0$.

Proof: Since $d_\triangle(W_n, W) \to 0$, by definition of the cut-shift distance there exists a sequence $g_n$ of shift functions with $\text{Shift}(g_n) \to 0$ so that $\|W_n^{g_n} - W\|_\square \to 0$. By applying the shift function $h_n = f_n \circ g_n^{-1}$ to both orderons in the expression, we get that $\|W_n^{f_n} - W^{h_n}\|_\square \to 0$. On the other hand, we shall show now that $d_1(W^{h_n}, W) \to 0$. Since the cut norm is always bounded by the $L^1$-distance, we can conclude that $\|W^{h_n} - W\|_\square \to 0$, implying by the triangle inequality that $\|W_n^{f_n} - W\|_\square \to 0$.

To show that $d_1(W^{h_n}, W) \to 0$, observe first that $\text{Shift}(h_n) \leq \text{Shift}(f_n) + \text{Shift}(g_n) \to 0$. For any pair $(x, y) \in [0, 1]^2$ where both $x$ and $y$ are not $\text{Shift}(h_n)$-close to the boundary of their block (i.e., are not of the form $i/M \pm \zeta$ where $\zeta \leq \text{Shift}(h_n)$), we thus have that $W(x, y) = W^{h_n}(x, y)$. We conclude that $d_1(W^{h_n}, W)$ is bounded by the total volume of pairs $(x, y)$ that are $\text{Shift}(h_n)$-close to the boundaries, which is bounded by $4M \cdot \text{Shift}(f_n)$, and tends to zero as $n \to \infty$. □

Lemma 7.12 gives us a global upper bound of $\Delta_\mathcal{H}$ on the distance of any ordered graph from a hereditary property $\mathcal{H}$. Our next main lemma, given below, provides an asymptotic lower bound. The statement of the lemma is analogous to (a special case of) Proposition 3.14 in [LS10], although the proof is slightly different (and makes use of Lemma 7.14).

Lemma 7.15. Let $\mathcal{P}$ be any ordered graph property and let $G_n \to W$ be an ordered graph sequence that converges (in $d_\triangle$) to a naive block orderon. Then

$$\liminf_{n \to \infty} d_1(G_n, \mathcal{P}) \geq d_1(W, \overline{\mathcal{P}}).$$

Proof: For any $n \in \mathbb{N}$ let $H_n \in \mathcal{P}$ be a graph with the same size as $G_n$ satisfying $d_1(G_n, H_n) = d_1(G_n, \mathcal{P})$ (this minimum is always attained as $d_1(G_n, \mathcal{P})$ is a minimum of finitely many values). By taking a subsequence of $G_n$ for which the distance to $\mathcal{P}$ converges to $\liminf d_1(G_n, \mathcal{P})$ and then taking a subsequence of it to ensure convergence of the corresponding subsequence of $H_n$, we may assume that $H_n \to U \in \overline{\mathcal{P}}$ (as usual, the convergence is in $d_\triangle$).
By the definition of the cut-shift distance, for any $H_n$, there exists a shift function $f_n$ so that $\|W_{H_n}^{f_n} - U\|_\square \to 0$ and furthermore $\text{Shift}(f_n) \to 0$. Applying Lemma 7.14 to the sequence $W_{G_n}$, which by the assumption of this lemma converges to the naive block orderon $W$, we conclude that $\|W_{G_n}^{f_n} - W\|_\square \to 0$. By applying Lemma 7.13 to the sequences $W_{G_n}^{f_n}$ and $W_{H_n}$ which converge in cut norm to $W$ and $U$ respectively, we conclude that

$$d_1(W, U) \leq \liminf_{n \to \infty} d_1(W_{G_n}^{f_n}, W_{H_n}) = \liminf_{n \to \infty} d_1(G_n, H_n) = \liminf_{n \to \infty} d_1(G_n, P),$$

where the first equality follows from the fact that shifting two orderons by the same shift does not change the $L^1$-distance between them, and the second equality follows from our choice of $H_n$. 

We are now ready to put it all together and prove Theorem 1.7.

**Proof of Theorem 1.7:** By Lemma 7.12 we know that $\overline{d_H} \leq \Delta_H$. Thus, it suffices to show the statement of the theorem with $\overline{d_H}$ replaced by $\Delta_H$. By Lemma 7.11, there exists a naive block orderon $W$, whose number of blocks $M$ is only a function of $\mathcal{H}$ and $\varepsilon$, for which $d_1(W, H) \geq \Delta_H - \varepsilon/2$. By Lemma 7.15, there exist $\delta > 0$ and $N$ so that any ordered graph $G$ on $n \geq N$ vertices with $d_\Delta(W, G, W) \leq \delta$ satisfies $d_1(G, H) \geq d_1(W, H) - \varepsilon/2 \geq \Delta_H - \varepsilon$. Pick $G$ according to the random model $G(n, W)$. From our sampling result, Theorem 1.4, the probability that $d_\Delta(W_{G(n,W)}, W) \leq \delta$ tends to one as $n \to \infty$. As the random model $G(n, W)$ is precisely a consecutive stochastic block model on $M$ blocks with parameters as in the statement of the theorem, the proof follows.

8 Parameter estimation

In this section we prove Theorem 1.9. Recall the definition of natural estimability from Subsection 1.2. First, observe that the random graph distributions $G|_k$ (no vertex repetitions) and $G(k, W_G)$ (allowing vertex repetitions) are very close in terms of variation distance.

**Observation 8.1.** Let $G$ be an ordered graph on $n$ vertices. For every fixed $k$ and a large enough $n$ the distribution $G|_k$ is arbitrarily close (in variation distance) to the distribution $G(k, W_G)$.

We now turn to the proof of the theorem. First we prove that natural estimability of $f$ implies convergence of $f(G_n)$ for any convergent $\{G_n\}$; then we prove that the latter condition implies the existence of a continuous functional on orderons that satisfies both items of the last condition of Theorem 1.9; and finally, we prove the other direction of both statements.

(1) $\implies$ (2): Let $\{G_n\}_{n \in \mathbb{N}}$ be a convergent sequence with $|V(G_n)| \to \infty$. Given $\varepsilon > 0$, let $k$ be such that for every ordered graph $G$ on at least $k$ nodes, $|f(G) - E_{G|_k}[f(G|_k)]| \leq \varepsilon$ (this can be done by setting $\delta = \varepsilon/M$ where $M$ is an upper bound on the values $f$). By the fact that $\{G_n\}_{n \in \mathbb{N}}$ is convergent, $t(F, W_{G_n})$ tends to a limit for all ordered graphs $F$ on $k$ vertices, which by Observation 8.1 implies that $\lim_{n \to \infty} \Pr_{G_n|_k}[G_n|_k = F] = \lim_{n \to \infty} t(F, W_{G_n}) = t(F, W)$. Therefore,

$$r_k \overset{\text{def}}{=} \lim_{n \to \infty} E_{G_n|_k}[f(G_n|_k)] = \sum_F \lim_{n \to \infty} \Pr_{G_n|_k}[G_n|_k = F] \cdot f(F) = \sum_F t(F, W) \cdot f(F).$$

Thus, for all sufficiently large $n$,

$$|f(G_n) - r_k| \leq |f(G_n) - E_{G_n|_k}[f(G_n|_k)]| + \varepsilon \leq 3\varepsilon,$$
which implies that \( \{f(G_n)\}_n \) is convergent.

(2) \( \implies \) (3): For a sequence \( \{G_n\}_{n \in \mathbb{N}} \) converging to \( W \), let \( \hat{f}(W) \) \( \overset{\text{def}}{=} \lim_{n \to \infty} f(G_n) \). Note that this quantity is well-defined: Given two ordered graph sequences \( \{G_n\}_{n \in \mathbb{N}} \) and \( \{H_n\}_{n \in \mathbb{N}} \) converging to \( W \), we can construct a new sequence \( \{S_n\}_{n \in \mathbb{N}} \), such that \( S_{2n} = G_n \) and \( S_{2n-1} = H_n \). By definition, the sequence \( \{S_n\}_{n \in \mathbb{N}} \) also converges to \( W \), and hence \( \lim_{n \to \infty} f(H_n) = \lim_{n \to \infty} f(G_n) = \lim_{n \to \infty} f(S_n) = \hat{f}(W) \).

To prove (3a), assume that \( \{W_n\}_{n \in \mathbb{N}} \subset W \) converges to \( W \). For every \( n \), we can apply Theorem 1.4 and obtain a sequence \( \{G_{n,k}\}_k \) such that \( \lim_{k \to \infty} W_{G_{n,k}} = W_n \). In addition, we can pick a subsequence of \( \{G_{n,k}\}_k \) such that

\[
d_\Delta(W_n, W_{G_{n,k}}) \leq 2^{-k} \quad \text{and} \quad \left| \hat{f}(W_n) - f(G_{n,k}) \right| \leq 2^{-k}.
\]

Now, since for every \( n \) the sequence \( G_{n,k} \) converges uniformly to \( W_n \) (as \( k \to \infty \)) and \( W_n \) converges to \( W \), we have that the diagonal sequence \( G_{n,n} \) converges to \( W \) as well. Therefore, by the fact that \( \hat{f} \) is well defined, we have that \( \lim_{n \to \infty} f(G_{n,n}) = \hat{f}(W) \). Therefore, for every \( \varepsilon > 0 \), we can find \( N \) such that for all \( n > N \), \( | f(G_{n,n}) - \hat{f}(W) | \leq \varepsilon / 2 \). Then, for every \( \varepsilon > 0 \) and all \( n \geq \max(N, \log(2 / \varepsilon)) \) we have

\[
\left| \hat{f}(W_n) - \hat{f}(W) \right| \leq \left| \hat{f}(W_n) - f(G_{n,n}) \right| + \left| f(G_{n,n}) - \hat{f}(W) \right| \leq \varepsilon / 2 + \varepsilon / 2 = \varepsilon,
\]

which concludes the proof for (3a).

To prove (3b): Assume towards a contradiction that (3b) does not hold. Let \( \{G_n\}_n \) be an ordered graph sequence with \( |V(G_n)| \to \infty \) such that \( | f(G_n) - \hat{f}(W_n) | > \varepsilon \). For each \( n \), consider the sequence \( \{G_n^{(j)}\}_{j \in \mathbb{N}} \) where \( G_n^{(j)} \) is the \( j \) blow-up of \( G_n \). Note that by the fact that for every \( j \) we have that \( W_{G_n^{(j)}} = W_{G_n} \), the sequence \( \{G_n^{(j)}\}_{j \in \mathbb{N}} \) converges to \( W_{G_n} \). Therefore, by the above construction of \( \hat{f} \), \( \{f(G_n^{(j)})\} \) converges to \( \hat{f}(W_{G_n}) \). Thus, for each \( n \) we can find \( j_n \in \mathbb{N} \) such that \( | f(G_n^{(j_n)}) - \hat{f}(W_{G_n}) | \leq \varepsilon / 2 \). Combined with the triangle inequality, this implies that \( | f(G_n) - f(G_n^{(j_n)}) | > \varepsilon / 2 \). By compactness we can assume that \( \{G_n\}_n \) has a subsequence \( \{G' _n\}_n \) that converges to \( W \). Construct a sequence \( \{H_n\}_n \) such that \( H_{2n-1} = G'_n \) and \( H_{2n} = G_n^{(j_n)} \). Note that the sequence \( \{H_n\}_n \) converges to \( W \) as well (since \( d_\Delta(W,G_n^{(j_n)}) = 0 \)), but \( \{f(H_n)\}_n \) does not converge, as it is not Cauchy, which is a contradiction to (2).

(3) \( \implies \) (2): Consider any convergent ordered graph sequence \( \{G_n\}_{n \in \mathbb{N}} \) such that \( |V(G_n)| \to \infty \), and let \( W \subset W \) be its limit. Then, \( d_\Delta(W_{G_n}, W) \to 0 \), and by the continuity of \( \hat{f} \), we have that \( \hat{f}(W_{G_n}) \to \hat{f}(W) \). Namely, for every \( \varepsilon > 0 \), we can find \( N \) such that for all \( n \geq N \),

\[
\left| \hat{f}(W_{G_n}) - \hat{f}(W) \right| \leq \varepsilon / 2.
\]

By assumption, we also have that for every \( \varepsilon > 0 \) there is a \( k(\varepsilon) \) such that for all \( |G_n| \geq k \), \( | f(G_n) - \hat{f}(W_{G_n}) | \leq \varepsilon / 2 \). This implies that for a large enough \( n \), \( | f(G_n) - \hat{f}(W) | \leq \varepsilon \), concluding the proof.

(2) \( \implies \) (1): Assume towards a contradiction that (1) does not hold. Namely, that there are \( \varepsilon > 0 \) and \( \delta > 0 \) such that for all \( k \), there is \( G \) on at least \( k \) vertices such that \( | f(G) - f(G_{k\{0\}}) | > \varepsilon \) with probability at least \( \delta \). Suppose we have a sequence \( \{G_k\}_k \) where \( |V(G_k)| = n(k) \to \infty \), and \( | f(G_k) - f(G_{k\{0\}}) | > \varepsilon \) with probability at least \( \delta \). By the compactness of \( \hat{W} \), we can select a
subsequence \( \{G'_k\} \) of \( \{G_k\}_k \) such that \( \{G'_k\}_k \) converges to some \( W \in \overline{W} \). Using Theorem 1.4 (along with a union bound on the confidence probabilities) and the assumption of (2), for every \( k \), let \( H_k = G'_k |_k \) be some specific induced subgraph such that \( d_\triangle(W_{G'_k}, W_{H_k}) = o_k(1) \) and \(|f(H_k) - f(G'_k)| > \varepsilon\).

Note that by the triangle inequality, the sequence \( \{H_k\}_k \) converges to \( W \). Let \( \{S_k\}_k \) be the sequence where \( S_{2t} = H_t \) and \( S_{2t-1} = G'_{t} \). Since both \( \{H_t\}_t \) and \( \{G'_t\}_t \) converge to \( W \), the sequence \( \{S_t\}_t \) also converges to \( W \). However, the sequence \( \{f(S_t)\}_t \) does not converge (as it is not Cauchy), which is a contradiction to (2).

9 Testability of hereditary ordered graph properties

This section contains a very detailed sketch of the proof of Theorem 1.10. The missing formalities (in the proof of Lemma 9.11) will be properly defined and explained in the next version of this paper.

Let \( \mathcal{H} \) be a hereditary property of simple ordered graphs, and recall that \( \overline{\mathcal{H}} \) denotes the closure of the property \( \mathcal{H} \). Our main technical lemma is as follows.

**Lemma 9.1.** For every \( U \in \overline{\mathcal{H}} \) and every \( \varepsilon > 0 \) there exists \( \delta(U, \varepsilon) > 0 \), so that if \( W \) is such that \( d_\triangle(W, U) \leq \delta(U, \varepsilon) \), then \( d_1(W, \overline{\mathcal{H}}) \leq \varepsilon \).

Note that the parameter \( \delta \) depends on the object \( U \). However, since \( \overline{\mathcal{H}} \) is a closed set in a compact space, it is also compact by itself, which implies the following removal lemma for orderons.

**Lemma 9.2.** For any \( \varepsilon > 0 \) there exists \( \delta(\varepsilon) > 0 \), such that if \( d_\triangle(W, \overline{\mathcal{H}}) \leq \delta \), then \( d_1(W, \overline{\mathcal{H}}) \leq \varepsilon \).

**Proof:** Fix \( \varepsilon > 0 \). Lemma 9.1 implies that for every \( U \in \overline{\mathcal{H}} \) there exists \( \delta(U, \varepsilon) > 0 \) such that if \( d_\triangle(W, U) \leq \delta(U, \varepsilon) \) then \( d_1(W, \overline{\mathcal{H}}) \leq \varepsilon \).

Assume that \( U' \in \overline{\mathcal{H}} \) and \( d_\triangle(U', U) \leq \eta \). Clearly \( \delta(U', \varepsilon) \geq \delta(U, \varepsilon) - \eta \) by the triangle inequality. So for a fixed \( \varepsilon \) and every \( U \) there is a ball around it with a guaranteed lower bound of \( \delta(U, \varepsilon)/2 \) on \( \delta(U', \varepsilon) \) for any \( U' \) in that ball. By compactness, we can cover \( \overline{\mathcal{H}} \) with a finite subset of this set of balls, obtaining a positive universal lower bound on \( \delta(U, \varepsilon) \) for every \( U \in \overline{\mathcal{H}} \).

Next, we describe how to derive the proof of Theorem 1.10. For this we need the following lemma.

**Lemma 9.3.** For any \( \varepsilon > 0 \) there are only finitely many ordered graphs \( H \in \mathcal{H} \) with \( d_1(W, \overline{\mathcal{H}}) > \varepsilon \).

**Proof:** Suppose that there is a sequence of ordered graphs \( \{G_n\}_{n \in \mathbb{N}} \) in \( \mathcal{H} \) such that \( d_1(W_{G_n}, \overline{\mathcal{H}}) \geq \varepsilon \). As the total number of ordered graphs with up to \( n \) vertices is bounded as a function of \( n \), we may assume that \( |V(G_n)| \to \infty \) as \( n \to \infty \). Furthermore, by compactness, we can assume that the sequence converges to some \( W \), that is, \( d_\triangle(W_{G_n}, W) \to 0 \). By definition of the closure \( \overline{\mathcal{H}} \), it follows that \( W \in \overline{\mathcal{H}} \). By Lemma 9.2, \( d_1(W_{G_n}, \overline{\mathcal{H}}) \to 0 \) as well, which is a contradiction.

**Proof of Theorem 1.10:** Fix \( \varepsilon, c > 0 \). By Lemma 9.2, there exists \( \delta > 0 \) such that any orderon \( W \) with \( d_1(W, \overline{\mathcal{H}}) \geq \varepsilon \) satisfies \( d_\triangle(W, \overline{\mathcal{H}}) \geq \delta \). On the other hand, by Lemma 9.3, there exists some \( k_1 \in \mathbb{N} \) so that any simple ordered graph \( H \in \mathcal{H} \) on at least \( k_1 \) vertices satisfies \( d_1(W_H, \overline{\mathcal{H}}) \leq \delta/2 \). Furthermore, by Theorem 1.4 and Observation 8.1, there exist integers \( k_2 \geq s \geq k_1 \) satisfying...
the following. For any graph $G$ on $n \geq k_2$ vertices, with probability at least $1 - c$ it holds that $d_\Delta(W_{G|\mathcal{I}_n}, W_G) < \delta/2$. Set $k = k_2$ in the statement of the theorem.

Let $G$ be a simple ordered graph on $n \geq k$ vertices with $d_1(G, \mathcal{H}) \geq \varepsilon$. By Lemma 7.12, we have $d_1(W_G, \mathcal{H}) \geq \varepsilon$. From the above paragraph we know that $d_\Delta(W_G, \mathcal{H}) \geq \delta$. Now let $H \sim G|\mathcal{I}_n$. Again by the above paragraph, with probability at least $1 - c$ it holds that $d_\Delta(W_H, W_G) < \delta/2$, which means by the triangle inequality that $d_\Delta(W_H, \mathcal{H}) > \delta/2$. As $s \geq k_1$, we conclude that $H \notin \mathcal{H}$. ■

9.1 Proof sketch of Lemma 9.1

In this subsection we sketch our proof of Lemma 9.1. We need to show that for every $U \in \mathcal{H}$ and every $\varepsilon > 0$ there exists $\delta(U, \varepsilon) > 0$, so that if $W$ is such that $d_\Delta(W, U) \leq \delta(U, \varepsilon)$, then $d_1(W, \mathcal{H}) \leq \varepsilon$. Along the proof, we will be using the following family of orderons, called layered strip orderons, in various occasions.

**Definition 9.4.** For $\ell \in \mathbb{N}$ define $\mathcal{I}_\ell = \{I_1^\ell, \ldots, I_{\ell}^\ell\}$ as a partition into blocks of the form $I_j^\ell = [(j-1)/\ell, j/\ell) \times [0,1]$ for any $j \in [\ell]$. A $\ell$-strip layered orderon is a step orderon whose steps refine $\mathcal{I}_\ell$, and in addition, for each $j \in [\ell]$, the sets contained in $I_j^\ell$ partition it into rectangles of the form $[(j-1)/\ell, j/\ell) \times [a,b]$. However, for points in the same block $(x,a), (y,b) \in I_j^\ell$, we allow the value of the orderon to also depend on whether $x < y$ or $y < x$.

Orderons are measurable functions, and similarly to Section 7, they can be approximated in $L^1$ by a step function whose steps are rectangles, which can be viewed in turn as a layered block orderon.

**Lemma 9.5.** Let $W \in \mathcal{W}$ be an orderon. Then, there exist $\ell \in \mathbb{N}$ and a function $W^{\mathcal{R}_\ell} \in \mathcal{W}$ which is an $\ell$-strip layered orderon satisfying $\|W - W^{\mathcal{R}_\ell}\|_1 \leq o_\ell(1)$.

The proof is identical to the first two paragraphs in the proof of Lemma 7.11, except that the functions we wish to approximate belong to $L^1([0,1]^2]$ (which is isomorphic, in relation to these arguments, to $L^1([0,1]^4]$), rather than $L^1([0,1]^2]$. Now, we take a slight detour and consider $U^\phi$ for some measure-preserving bijection $\phi : [0,1]^2 \rightarrow [0,1]^2$. Fix a partition $\mathcal{R}_\ell = \{P_1, \ldots, P_{|\mathcal{R}_\ell|}\}$ and define the shifted partition $\mathcal{R}_\ell^\phi$ as follows.

$$\mathcal{R}_\ell^\phi = \{P_1^\phi, \ldots, P_{|\mathcal{R}_\ell|}^\phi\}, \quad \text{where } P_i^\phi \overset{\text{def}}{=} \{\phi^{-1}(x,y) \mid (x,y) \in P_i\} \quad \text{for any } i \in [|\mathcal{R}_\ell|].$$

As $W^{\mathcal{R}_\ell}$ obtained in Lemma 9.5 is a good approximation of $W$ in $L^1$, and shifting both functions by the same measure-preserving bijection does not change the $L^1$-distance between them, we conclude that $(W^{\mathcal{R}_\ell})^\phi$ is a good approximation of $W^\phi$ in $L^1$.

The following observation on the cut norm will be useful for us.

**Observation 9.6.** Let $W \in \mathcal{W}$ be any orderon such that $\|W - U^\phi\|_\square \leq \gamma$. For every $i, j \in [|\mathcal{R}_\ell|]$, $\left| \int_{v_1,v_2 \in P_i \times P_j} U(v_1,v_2)dv_1dv_2 - \int_{v_1,v_2 \in P_i^\phi \times P_j^\phi} W(v_1,v_2)dv_1dv_2 \right| \leq \gamma$. 

39
Assuming that \( \text{Shift}(\phi) \) is small enough we can “trim” the area around the boundaries of the strips, and “stretch” (using a linear mapping) the area around the trimmed sections. By applying this procedure, we make sure that measure-preserving bijections do not cross strip boundaries. More precise details follow.

**Definition 9.7.** Suppose that \( W \in \mathcal{W} \) is an orderon, \( \ell \) is a positive integer, and \( \eta = \eta(\ell) \) is a parameter dependent on \( \ell \). Let \( \psi_{\ell, \eta} : [0, 1] \rightarrow [0, 1] \) be a function defined as follows.

- If for some integer \( 0 < r < \ell, \frac{r}{\ell} \leq x < \frac{r}{\ell} + 2\eta \) then \( \psi_{\ell, \eta}(x) = (x + \frac{r}{\ell})/2 + \eta \).
- If for some integer \( 0 < r < \ell, \frac{r}{\ell} - 2\eta < x < \frac{r}{\ell} \) then \( \psi_{\ell, \eta}(x) = (x + \frac{r}{\ell})/2 - \eta \).
- Otherwise, \( \psi_{\ell, \eta}(x) = x \).

The \((\ell, \eta)\)-trim of \( W \), denoted \( W_{\ell, \eta} \), is defined as \( W_{\ell, \eta}((x, a), (x', a')) = W((\psi_{\ell, \eta}(x), a), (\psi_{\ell, \eta}(x'), a')) \). As “trimming” and “stretching” operations only occur close to the strip boundaries, and in any other point \( W_{\ell, \eta} \) behaves the same as \( W \), we get the following.

**Lemma 9.8.** For \( \ell \in \mathbb{N} \), \( \eta(\ell) \leq 1/4\ell \) and \( W \in \mathcal{W} \), it holds that \( \|W - W_{\ell, \eta}\|_1 \leq 4\ell \eta \).

**Proof:** A real number \( x \in [0, 1] \) satisfies \( \psi_{\ell, \eta}(x) \neq x \) only if it lies in the set \( \bigcup_{r=1}^{\ell} [r/\ell - 2\eta, r/\ell + 2\eta] \), whose measure is less than \( 4\ell \eta \). Thus, the total measure of tuples \( (x, a, x', a') \in [0, 1]^4 \) for which \( W((x, a), (x', a')) \neq W_{\ell, \eta}((x, a), (x', a')) \) is bounded by \( 4\ell \eta \). \( \square \)

Consider the partition \( Q = \{Q_1, \ldots, Q_{|\mathcal{R}_\ell|}\} \) of \([0, 1]^2\) resulting from “trimming and stretching” \( W \), where \( Q_j = \{(x, a) \in [0, 1]^2 | (\psi_{\ell, \eta}(x), a) \in P_j^{\phi} \} \) is a measurable set for each \( j \in [|\mathcal{R}_\ell|] \). Roughly speaking, this partition is an “adjustment” of the partition \( \mathcal{R}_\ell^{\phi} \) to the orderon \( W_{\ell, \eta} \) resulting from trimming. Observe first that if a set of elements \( P \) is fully contained in some strip, then the set \( Q \) of all elements “trimmed and stretched” to fit into \( P \) is also contained in the same strip. Namely, we have the following.

**Observation 9.9.** If \( \text{Shift}(\phi) \leq 1/4\ell \), then for every \( j, t \in [|\mathcal{R}_\ell|] \times [\ell] \) such that \( P_j^\phi \) is contained in \( I_t^\ell \), the set \( Q_j \) is also contained in \( I_t^\ell \).

Next, observe that the (additive) affect of trimming and stretching on a set contained in a strip is small, as long as \( \eta \) is small enough. For our case, we will need the following.

**Observation 9.10.** Let \( \ell, \mathcal{R}_\ell, Q, \phi \) be as above, and write \( k = |\mathcal{R}_\ell| \). For any \( \varepsilon > 0 \) there exists a choice of \( \eta_0 > 0 \) that depends only on \( \ell, k, \varepsilon \), so that for each \( i \in [k] \) and any \( \eta < \eta_0 \),

\[
|\lambda(P_i^\phi) - \lambda(Q_i)| \leq \varepsilon,
\]

and

\[
\left| \int_{v_1,v_2 \in P_i^{\phi} \times P_j^{\phi}} W(v_1, v_2) dv_1 dv_2 - \int_{v_1,v_2 \in Q_i \times Q_j} W_{\ell, \eta}(v_1, v_2) dv_1 dv_2 \right| \leq \varepsilon.
\]

The following is our main technical lemma, about the possibility to “pixelate” any orderon in \( \mathcal{P} \).
Lemma 9.11. For any \( U \in \mathcal{H} \), and any \( \varepsilon > 0 \), there exists \( \ell = \ell(U, \mathcal{H}, \varepsilon) \) and a layered \( \ell \)-strip ordon \( U' \in \mathcal{H} \) so that \( d_1(U, U') \leq \varepsilon \).

We now provide a detailed sketch of the proof of Lemma 9.11 (with the full proof appearing in a later version of this paper; this is the only part in the proof of Lemma 9.1 that does not fully appear here). The following measure-theoretic lemma will be useful for the proof.

Lemma 9.12. Consider a probability space characterized by \( s \) random variables \( X_1, \ldots, X_s \), where \( X_i \) is independently and uniformly chosen from the interval \([c_i, d_i]\). Suppose that \( E \subseteq \prod_{i=1}^s [c_i, d_i] \) is a positive probability event (determined by the random variables \( X_1, \ldots, X_s \)), and let \( \delta > 0 \). Then, there exist \([c'_i, d'_i] \subseteq [c_i, d_i]\) for which the conditional probability of \( E \), where we constrain \( X_i \in [c'_i, d'_i] \) for every \( 1 \leq t \leq s \), is at least \( 1 - \delta \).

Proof sketch of Lemma 9.11: It will be convenient for us to work with orderons that have only a finite number of possible values (possibly dependent on \( \varepsilon \)) in their range. Picking \( \delta = \varepsilon/10 \), this can be done in an \( L^1 \)-efficient manner by rounding the value of any \( W(u, v) \), as long as \( 0 < W(u, v) < 1 \), to the closest multiple of \( \delta \) that is strictly between zero or one. Note that, by Lemma 7.6, \( W \in \mathcal{H} \) after the rounding if and only if it belonged to \( \mathcal{H} \) before the rounding. Thus, from now on we assume that \( U \) is a rounded ordon.

The starting point would be to use Lemma 9.5 to \( \delta \)-approximate \( U \) in \( L^1 \) by a (large enough) layered strip ordon \( Z \), whose set of parts is denoted by \( \mathcal{R} \). That is, the expected value of \( |U((x, a), (y, b)) - Z((x, a), (y, b))| \) over all tuples \( (x, a, y, b) \in [0, 1]^4 \) is bounded by \( \delta \).

Recall that each part in the underlying partition of \( Z \) is a rectangle of the form \([x_i, x_{i+1}] \times [a_j, a_{j+1}]\). For ease of discussion, we say that this part is the intersection of the block \( B_i = [x_i, x_{i+1}] \) and the layer \( L_i = [a_j, a_{j+1}] \). Denote the set of blocks (from first to last in terms of order) by \( \mathbb{B} = \{B_1, \ldots, B_b\} \) and the set of layers by \( \mathcal{L} = \{L_1, \ldots, L_l\} \). Consider a random variable \( X \) that does the following for any \( i \in [b] \): it picks one uniform index \( y_i \) from the block \( B_i \), then picks one uniform index \( b_{ij} \) from each layer \( L_j \), and returns the point ("vertex") \( g_{ij} = U(y_i, b_{ij}) \). This will form a \( b \times l \) grid \( \mathcal{G} = \{g_{ij}\}_{i \in [b], j \in [l]} \). We will be interested in the random model \( \mathcal{G}(U, b, l) \) that picks such a grid randomly, and returns the values of \( U \) induced on the elements of this grid, but \emph{without} using values that correspond to elements from the same block; this object is an ordered \((b \times l) \times (b \times l)\) tensor of values with (usually) "asterisks" within blocks. We call such an object a configuration. The expected \( L_1 \)-distance of a configuration generated this way from \( Z \) is at most \( \delta \), and we conclude that with probability at least \( 1/2 \), the distance is at most \( 2\delta \). We call such a configuration \emph{good} if it satisfies the latter condition on the distance from \( Z \). Note that at least one of the good configurations has positive probability to occur (since \( U \) is rounded, meaning that there is a finite number of possible configurations), and we henceforth fix this choice of configuration \( C \).

For the purpose of this sketch, suppose that \( \mathcal{H} \) is defined by a finite family of forbidden ordered subgraphs \( \mathcal{F} \), and let \( k \) be the size of the largest subgraph in \( \mathcal{F} \). The generalization to any hereditary property \( \mathcal{H} \) can be done by calculating an effective \( k \) through an Alon-Shapira type function of the property, \( b, \ell, \) and \( \varepsilon \), see [AS08]. Also take an integer \( r \) that is large enough as a function of \( k, \varepsilon, b, l \).

We apply Lemma 9.12 over the event of obtaining \( C \) in the above process, and take \( \delta \) sufficiently small (as a function of \( k, \varepsilon, b, l, r \)) in the statement of the lemma, to obtain \([y_i, z_i] \subset [x_i, x_{i+1}]\) and \([c_{ij}, d_{ij}] \subset [a_j, a_{j+1}]\) for every \( i \) and \( j \).
We now intend to take a “r-multigrid”, that is, take r possible options for each choice of an x from [y_i, z_i] and each choice of an a from [c_{ij}, d_{ij}], all independently. With probability at least 1/2 (by a union bound and our choice of δ in the application of Lemma 9.12), for any possible choice of one of the r options for each of the random variables, the resulting grid will have the configuration C.

Provided that r is large enough as a function of k, we can ensure, using a standard multipartite Ramsey-type lemma (which is in particular much simpler than the quantitative Ramsey lemma used in the combinatorial proof of [ABF17]) that for an r-multigrid as above, we can take from it a k-multigrid which satisfies an additional uniformity requirement where we consider the values of U inside the same block, so that the values between pairs of elements (y, b), (y', b') from the same block will depend only on the pair of layers that they lie in, and whether y < y' or y > y'. We completely ignore the case y = y', because this corresponds to a probability 0 event when sampling a subgraph from the orderon.

If we look at a subset of grid points where no x appears more than once, then its distribution is a conditioning of a uniformly random point sample from [0, 1]^2 over a positive probability event. Since the above r-multigrid will appear with positive probability, so will one of the options for a k-multigrid that is uniform inside blocks appear with positive probability.

Now, if this k-multigrid contains a subgraph F (where we are allowed to take elements from the same block or layer, but not elements with the same first coordinate) from the forbidden family F of H, then since this k-multigrid appears with positive probability, we have t(F, U) > 0. It follows by Lemma 7.4 that U ∉ H, which is a contradiction.

Otherwise, we can make U “imitate” the k-multigrid by setting U(u, v), for each pair u = (y, b) and v = (y', b') that lie in blocks b_1, b_2 and layers l_1, l_2 of the original partitioning R, as follows. If b_1 < b_2 or b_1 > b_2, we set U(u, v) to the unique value of the multigrid between (b_1, l_1) and (b_2, l_2). Otherwise, we distinguish between the case that y < y' and the case y > y' (again, the case y = y' corresponds to a zero-measure set and here the values can be set arbitrarily). If y < y', we set U(u, v) to be the unique value of the multigrid that lies between layers l_1 and l_2 within block b_1 corresponding the case where y < y'. Otherwise, we set U(u, v) to be the unique value of the multigrid, again between layers l_1 and l_2 within block b_1, but here we choose the value corresponding to the case where y > y'.

By our choice of the configuration C (as defined in the beginning of the proof), the L^1-cost of changing U according to this policy is 3δ, including a term of 2δ in order to imitate C between pairs of points in different blocks, and an additional cost of at most δ to edit the values for pairs of points within the same block. Clearly, the “imitation” creates a layered strip orderon with b strips and bl parts in total, concluding the proof of the lemma.

For the rest of the discussion, we will need several notions. Recall that I_ℓ is the partitioning of [0, 1]^2 into equally-sized strips of the form [(j - 1)/ℓ, j/ℓ] × [0, 1]. The first notion, that of clones, refers to pairs of partitions that refine I_ℓ in a similar way. The next two, decision functions and decisiveness, are related to structural changes that we want to impose on our orderons.

**Definition 9.13 (clone).** Given a partition P = {P_1, ..., P_m} which refines I_ℓ, we say that a partition P' = {P'_1, ..., P'_m}, also refining I_ℓ, is a clone of P if for any i ∈ [|P|], P_i and P'_i are contained in the same I_j ∈ I_ℓ.
Note that being a clone (for fixed $\ell$) is an equivalence relation: it is clearly reflexive, symmetric, and transitive.

**Definition 9.14** (decision function, implementation). A *decision function* with parameters $k, \ell \in \mathbb{N}$, where $\ell$ divides $k$, is a function $w: \{k\}^2 \to \{0, 1, *\}$ satisfying $w(i, j) = w(j, i)$ for any pair $(i, j) \in \{k\}^2$ with $\left\lceil \ell(i-1)/k \right\rceil \neq \left\lceil \ell(j-1)/k \right\rceil$. (For pairs $(i, j)$ where there is an equality in the last expression, it may hold that $w(i, j) \neq w(j, i)$.)

Let $W \in \mathcal{W}$ be an orderon, and let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a partition of $[0,1]^2$ which clones the stepping of a layered $\ell$-strip orderon. An *implementation* of the decision function $w$ on $W$ with respect to $\mathcal{P}$ is an orderon denoted $W_{\preceq \mathcal{wP}}$ and defined as follows. For every $(u, v) \in ([0,1]^2)^2$ such that $\pi_1(u) \prec \pi_1(v)$ and $(u, v) \in P_i \times P_j$,

1. If $w(i, j) \neq *$, we have $W_{\preceq \mathcal{wP}}(u, v) = w(i, j)$.
2. If $w(i, j) = *$, we have $W_{\preceq \mathcal{wP}}(u, v) = W(u, v)$.

When $\pi_1(u) = \pi_1(v)$, the value of $W_{\preceq \mathcal{wP}}(u, v)$ can be set arbitrarily.

**Definition 9.15** (decisiveness). Fix a partition $\mathcal{P} = \{P_1, \ldots, P_{|\mathcal{P}|}\}$ which clones the stepping of a layered strip orderon, a decision function $w : [|\mathcal{P}|]^2 \to \{0, 1, *\}$, and an orderon $W$. We say that $w$ is *decisive* with respect to $\mathcal{P}$ and $W$ if for any $(i, j) \in [|\mathcal{P}|]^2$ where $W$ induced on $\{(u, v) \in P_i \times P_j : \pi_1(u) < \pi_1(v)\}$ is almost everywhere zero, it holds that $w(i, j) = 0$, and if it is almost everywhere one when induced on this set, then $w(i, j) = 1$.

**Lemma 9.16.** Let $W \in \mathcal{W}$ be an $\ell$-strip layered orderon and let $\mathcal{P}$ be its stepping. Let $W' \in \mathcal{W}$ be another orderon, and let $\mathcal{P}'$ be a clone of $\mathcal{P}$. Then, if the decision function $w$ is decisive with respect to $\mathcal{P}$ and $W$, and if $W_{\preceq \mathcal{wP}} \in \overline{\mathcal{W}}$, then $W'_{\preceq \mathcal{wP'}} \in \overline{\mathcal{W}}$.

**Proof:** By Lemma 7.4, it suffices to show that for any fixed $k \in \mathbb{N}$ and any ordered graph $F$ on $k$ vertices for which $t(F, W) = 0$, it holds that $t(F, W') = 0$. Suppose towards a contradiction that there exists $F = (V, E)$ on $k$ vertices for which this is not the case. Then there exists a tuple $(i_1, i_2, \ldots, i_k) \in [|\mathcal{P}|]^k$ (possibly with repetitions) so that

\[
\int_{v \in P_{i_1}' \times P_{i_2}' \times \cdots \times P_{i_k}'} \left( \prod_{(i,j) \in E} W'_{\preceq \mathcal{wP'}}(v_i, v_j) \cdot \prod_{(i,j) \in \overline{E}} \left(1 - W'_{\preceq \mathcal{wP'}}(v_i, v_j)\right) \right) \cdot \prod_{1 \leq i < j \leq k} 1_{v_i \leq v_j} \, dv > 0.
\]

We show that in this case,

\[
\int_{v \in P_{i_1} \times P_{i_2} \times \cdots \times P_{i_k}} \left( \prod_{(i,j) \in E} W_{\preceq \mathcal{wP}}(v_i, v_j) \cdot \prod_{(i,j) \in \overline{E}} \left(1 - W_{\preceq \mathcal{wP}}(v_i, v_j)\right) \right) \cdot \prod_{1 \leq i < j \leq k} 1_{v_i \leq v_j} \, dv > 0,
\]

which means that $t(F, W_{\preceq \mathcal{wP}}) > 0$, leading to a contradiction.

\footnote{In particular, $\mathcal{P}$ has exactly $k/\ell$ parts in each $I'_j \in \mathcal{I}_d$: these are precisely the parts $P_i$ where $\left\lceil \ell(i-1)/k \right\rceil = j$.}
Indeed, for a $k$-tuple of subsets $(Q_1, \ldots, Q_k)$, define

$$B_{Q_1 \times \cdots \times Q_k} \overset{\text{df}}{=} \{(x_1, \ldots, x_k) \in Q_1 \times Q_2 \times \cdots \times Q_k \mid x_1 < \cdots < x_k\}.$$  

The crucial observation is that, since $B_{P'_1 \times \cdots \times P'_{ik}}$ has positive measure, so does $B_{P'_{ik}}$. This follows from the structure of layered strip orderons. Since the decision function is decisive, we know that if $W_{w'_{ik}} \in \mathcal{H}$ restricted to $P_i \times P_j$ equals zero or one, then the decision function would force all values of $W'_{w'_{ik}}$ restricted to $P_i' \times P_j'$ to equal zero or one, respectively. As the value of the first integral above is positive, the expression inside this integral is positive for a positive-measure subset of $B_{P'_1 \times \cdots \times P'_{ik}}$, and so the expression inside the second integral is positive over $B_{P'_1 \times \cdots \times P'_{ik}}$. □

The next lemma states that implementing the same decision function on different orderons with similar structural parameters induces a similar “$L^1$-cost”.

**Lemma 9.17.** For every $\ell, k \in \mathbb{N}$ and $\varepsilon > 0$ the following holds. Let $W, W' \in W$ be two orderons, let $P = (P_1, \ldots, P_k), P' = (P'_1, \ldots, P'_k)$ be two partitions of $[0, 1]^2$ which clone the stepping of some layered $\ell$-strip orderon. Suppose further that the following two conditions are satisfied.

1. For each $i \in [k]$, $|\lambda(P_i) - \lambda(P'_i)| \leq \varepsilon/3k^2$.
2. For all $i, j \in [k]^2$,

$$\left| \int_{v_1, v_2 \in P_i \times P_j} W(v_1, v_2) dv_1 dv_2 - \int_{v_1, v_2 \in P'_i \times P'_j} W'(v_1, v_2) dv_1 dv_2 \right| \leq \varepsilon/3k^2.$$ 

Finally, let $w$ be a decision function. Then,

$$|d_1(W_{w_{\ell}}, W) - d_1(W'_{w_{\ell}}, W')| \leq \varepsilon.$$

**Proof:** Write $\delta = \varepsilon/3k^2$. For any pair $(i, j) \in [k]^2$, we show that the $L^1$-cost of applying the decision $w(i, j)$ to $W$ restricted to $P_i \times P_j$ is at most an additive difference of $3\delta$ away from the cost of applying $w(i, j)$ to $W'$.

Indeed, if $w(i, j) = *$ then the cost difference is zero. If $w(i, j) = 0$, the $L^1$-cost for $W$ is $\int_{v_1, v_2 \in P_i \times P_j} W(v_1, v_2) dv_1 dv_2$ and for $W'$ it is $\int_{v_1, v_2 \in P'_i \times P'_j} W'(v_1, v_2) dv_1 dv_2$, which by our assumption differ by at most $\delta$. Similarly, for $w(i, j) = 1$ the costs are $\int_{v_1, v_2 \in P_i \times P_j} (1 - W(v_1, v_2)) dv_1 dv_2$ and $\int_{v_1, v_2 \in P'_i \times P'_j} (1 - W'(v_1, v_2)) dv_1 dv_2$, and they differ by at most $3\delta$. Summing over all pairs $(i, j) \in [k]^2$, the proof follows. □

We now explain how to complete the proof of Lemma 9.1.

**Proof of Lemma 9.1:** Fix $U \in \mathcal{H}$ and $\varepsilon > 0$. We need to show that if an orderon $W$ is $\delta$-close in cut-shift distance to $U$, for sufficiently small $\delta(\varepsilon)$, then $d_1(W, U) \leq \varepsilon$. By Lemma 9.11, there exists an $\ell$-strip layered orderon $U' \in \mathcal{H}$ for some sufficiently large $\ell = \ell(U, \mathcal{H}, \varepsilon)$ where $d_1(U', U) \leq \varepsilon/4$.

Let $R_\ell = (R_{1\ell}, \ldots, R_{|R_\ell|})$ be the partitioning of $U'$ into steps. We pick $\delta = \eta > 0$ small enough as a function of $\mathcal{H}, \varepsilon, U, |R_\ell|$ (specifically, along the proof, there are several statements that hold “for small enough $\eta$”; we pick $\eta$ that satisfies all of these requirements). 44
Consider first the case that $W = U^\phi$ where $\text{Shift}(\phi) \leq \eta$. We have that $d_1(W, (U')^\phi) = d_1(U, U') \leq \varepsilon/4$. On the other hand, consider the $(\ell, \eta)$-trim of $(U')^\phi$; denote this order on by $U''$, and note that $d_1((U')^\phi, U'') \leq \varepsilon/4$ for $\eta$ small enough, by Lemma 9.8. Now, each step $R_t \in \mathcal{R}_\ell$ is transformed by the procedure to a set $Q_t$ that lies in the same strip $I^\ell_t$ (by Observation 9.9), whose measure is very similar to that of $R_t$ (by Observation 9.10), and where the value of $U'$ over any pair $R_t \times R_j$ is equal to the value of $(U')^\phi$ over $Q_t \times Q_j$. Let $Q_{\ell} = \{Q_1, \ldots, Q_{|\mathcal{R}_\ell|}\}$ denote the collection of transformed parts. Note that this partition is a clone of $R_{\ell}$. Take the decision function $w$ where $w(i, j) = \alpha \in \{0, 1\}$ for any $(i, j)$ where $U'$ induced on $R_t \times R_j$ equals $\alpha$, and otherwise, $w(i, j) = \ast$, and note that $w$ is decisive with respect to $\mathcal{R}_\ell$ and $U'$. It follows by Lemma 9.16 that $U''_{w=Q_{\ell}} \in \mathcal{H}$ and also $W_{w=Q_{\ell}} \in \mathcal{H}$. Finally, by Lemma 9.17, we know that $d_1(U''_{w=Q_{\ell}}, U'') \leq \varepsilon/4$ (assuming again that $\eta$ is small enough). To summarize,

$$d_1(W, \mathcal{H}) \leq d_1(W, (U')^\phi) + d_1((U')^\phi, U'') + d_1(U'', U''_{w=Q_{\ell}}) \leq 3\varepsilon/4. \quad (11)$$

In particular, note that $U''_{w=Q_{\ell}}(u, v) = \alpha$ for any $\alpha \in \{0, 1\}$ and $(u, v) \in Q_t \times Q_j$ where $w(i, j) = \alpha$. That is, $U''_{w=Q_{\ell}}$ is generated from $W$ by applying the decision function $w$ on $W$ with respect to the partition $Q$, and then possibly making additional value changes in pairs $(u, v) \in Q_t \times Q_j$ where $w(i, j) = \ast$. Thus, by the triangle inequality, $d_1(W, W_{w=Q_{\ell}}) \leq 3\varepsilon/4$. Recall that $W_{w=Q_{\ell}} \in \mathcal{H}$.

For the general case, suppose that $W'$ is $\eta$-close in cut norm to the order on $W = U^\phi$ considered above, that is, $\|W' - W\|_\square \leq \eta$. By definition of the cut norm, we know that for any $(i, j) \in [|\mathcal{R}_\ell|]^2$,

$$\left| \int_{Q_t \times Q_j} W(v_1, v_2)dv_1dv_2 - \int_{Q_t \times Q_j} W'(v_1, v_2)dv_1dv_2 \right| \leq \eta \leq \frac{\varepsilon}{12|\mathcal{R}_\ell|^2}$$

provided that $\eta$ is small enough. Apply Lemma 9.17 with $\mathcal{P} = \mathcal{P}' = Q_{\ell}$ and parameters $k = |\mathcal{R}_\ell|$ and $\varepsilon/4$ to conclude that $d_1(W', W'_{w=Q_{\ell}}) \leq d_1(W, W_{w=Q_{\ell}}) + \varepsilon/4 \leq \varepsilon$. However, by Lemma 9.16, we know that $W'_{w=Q_{\ell}} \in \mathcal{H}$. This concludes the proof.

**References**


