Limits of Dense Simplicial Complexes

T. Mitchell Roddenberry   MITCH@RICE.EDU
Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005-1827, USA

Santiago Segarra   SEGARRA@RICE.EDU
Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005-1827, USA

Editor: Bharath Sriperumbudur

Abstract

We develop a theory of limits for sequences of dense abstract simplicial complexes, where a sequence is considered convergent if its homomorphism densities converge. The limiting objects are represented by stacks of measurable $[0, 1]$-valued functions on unit cubes of increasing dimension, each corresponding to a dimension of the abstract simplicial complex. We show that convergence in homomorphism density implies convergence in a cut-metric, and vice versa, as well as showing that simplicial complexes sampled from the limit objects closely resemble its structure. Applying this framework, we also partially characterize the convergence of nonuniform hypergraphs.

Keywords: Simplicial complex, graphon, stochastic topology, topological data analysis, graph limit

1. Introduction

The theory of graph limits has been one of the most fruitful developments in modern combinatorics, with applications ranging from extremal graph theory to statistical physics (Borgs et al., 2008, 2012; Lovász, 2012). Beyond combinatorics, graphs are undeniably of great interest in machine learning and data science applications, where they are used to reflect "geometry" in data (Bronstein et al., 2017). Moreover, graphs can be viewed as specific types of more general objects that capture higher-order relationships, such as simplicial complexes (Schaub et al., 2021). Given the fruitful applications of the theory of graphons in data science (Ruiz et al., 2020; Roddenberry et al., 2021; Navarro and Segarra, 2022; Coppini, 2022; Maskey et al., 2023), it is natural to ask how a similar theory can be developed for higher-order relational objects. Viewing graphs as (abstract) simplicial complexes of dimension one, we aim to develop such a theory for limiting objects of simplicial complexes. Indeed, recent work in stochastic topology has studied the topological properties of large random simplicial complexes, characterizing their connectivity and (co)homology (Bobrowski and Krioukov, 2022).

An example of such a random simplicial complex is a random geometric Čech complex. Let μ be a probability measure on $\mathbb{R}^d$ for some integer $d \geq 1$, and let $\epsilon > 0$ be a real number.

Denote the support of μ by $\mathcal{X} = \mathrm{supp}(\mu)$, and treat $\mathcal{X}$ as a metric probability space, with the metric inherited from $\mathbb{R}^d$. For some fixed $\epsilon > 0$, we consider the Čech complex of $\mathcal{X}$ in $\mathbb{R}^d$ with radius ϵ, which is defined as the infinite simplicial complex $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^d)$ composed of all finite subsets of $\mathcal{X}$ with diameter strictly less than ϵ.
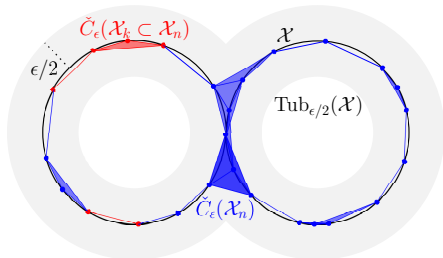

Figure 1: Čech complexes from a bouquet of circles in $\mathbb{R}^2$.

Denoting by $\mathrm{Tub}_{\epsilon/2}(\mathcal{X}, \mathbb{R}^d)$ the set of all points in $\mathbb{R}^d$ that are less than $\epsilon/2$-away from $\mathcal{X}$, the set of all open balls of radius $\epsilon/2$ centered at points in $\mathcal{X}$ is a cover of $\mathrm{Tub}_{\epsilon/2}(\mathcal{X}, \mathbb{R}^d)$ by contractible open sets. The nerve theorem then implies that the geometric realization of the Čech complex $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^d)$ is homotopy equivalent to $\mathrm{Tub}_{\epsilon/2}(\mathcal{X}, \mathbb{R}^d)$. As an example, let $\mathcal{X}$ be a bouquet of two unit circles in $\mathbb{R}^2$, that is, the union of two circles of unit radius centered at $(-1, 0)$ and $(1, 0)$ respectively, as pictured in Figure 1, and let μ be the uniform measure on $\mathcal{X}$. Clearly, the space $\mathrm{Tub}_{\epsilon/2}(\mathcal{X}, \mathbb{R}^2)$ (pictured in Figure 1 as the gray shaded region) is homotopy equivalent to $\mathcal{X}$ for any $\epsilon < 2$.

The field of topological data analysis (TDA) (Carlsson, 2009) has leveraged this idea to great success, replacing troublesome computations about the topological structure of complicated spaces with more tractable computations on approximations formed from point clouds, such as the Čech, or more commonly, Vietoris-Rips complex. In practice, TDA applications typically do not have access to the entire space $\mathcal{X}$; rather, a finite set of n points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ in $\mathcal{X}$ sampled i.i.d. according to the probability measure μ is available, from which the Čech complex $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^d)$ is formed. This Čech complex is used as an "estimate" of sorts for $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^d)$. If the space $\mathcal{X}$ is compact, one can show that as $n \to \infty$, the space $\mathcal{X}_n$ with the Euclidean metric almost surely converges to $\mathcal{X}$ in the Gromov-Hausdorff metric. This implies, for instance, convergence of the persistence diagrams of $\mathcal{X}_n$ to that of $\mathcal{X}$ in the bottleneck distance (Chazal et al., 2009). Returning to our example of the bouquet of circles, see in Figure 1 that a finite subset $\mathcal{X}_n \subset \mathcal{X}$ yields a Čech complex that preserves the topological features of the space, for example, having a free fundamental group of rank two. Observe that sampling more points would not introduce any additional topological features to the constructed complex.

However, convergence of this sort is not as easy to establish when $\mathcal{X}$ is not compact. For instance, the case when $\mathbb{X} = \mathbb{R}^d$ and μ is a radially symmetric power-law, exponential, or Gaussian distribution was considered by Adler et al. (2014), where they showed that under many circumstances, highly nontrivial homological features emerge in ways that vary with n, precluding the convergence results mentioned above. We consider an alternative way to describe the convergence of the Čech complexes $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^d)$ to $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^d)$. Observe that each $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^d)$ is a random induced subcomplex of $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^d)$, with nodes distributed i.i.d. according to the probability measure μ. This motivates a preliminary definition of convergence for simplicial complexes: we say that a sequence of finite simplicial complexes K_n is convergent if for all $k \geq 1$ and all finite simplicial complexes F with vertex set $[k]$, the probability that the induced subcomplex of k nodes chosen uniformly at random from K_n is equal to F is convergent. That is to say, the distribution of uniformly sampled

induced subcomplexes of any fixed size converges as $n \to \infty$. We indicate this further subsampling operation in red in Figure 1. In this case, for instance, it holds for all n that the probability of subsampling a cycle on $k < \lfloor \pi / \arcsin(\epsilon/2) \rfloor$ nodes from $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^2)$ is zero. The characterization of convergence via the distribution of induced substructures is closely related to the theory of graphons (Borgs et al., 2008). Following this approach, we develop an analogous limit theory for simplicial complexes. We consider a sequence of simplicial complexes to be convergent if their homomorphism densities converge, which we show to be equivalent to convergence in a cut-metric. The limiting object of such a convergent sequence is described in terms of a collection of symmetric kernel functions on the sequence of unit cubes[1] of increasing dimension; we dub these objects *complexons*.[2] These objects closely relate to graphons in the random complex models that they induce, with two representations that correspond to either the usual presentation of a simplicial complex as a collection of finite sets closed under restriction, or the presentation in terms of its facets. The main results of this work relate the cut-metric and homomorphism densities for the limit objects of large, dense simplicial complexes.

Theorem (Informal)**.** *If two complexons are close in the cut-metric, then they yield similar random simplicial complexes by sampling. Conversely, if they yield similar random simplicial complexes by sampling, then they are close in the cut-metric.*

This is stated more formally in Theorem 14. As part of the proof, it will also be shown that large simplicial complexes drawn from a complexon concentrate in the cut-metric, and thus have similar homomorphism densities to the complexon.

The paper is organized as follows. After covering preliminaries in Section 2, complexons and random sampling from complexons are defined in Section 3. The main results are stated in Section 4, with most proofs relegated to the appendix. Some remarks relating complexons to other notions in the literature, including the convergence of nonuniform hypergraphs under certain conditions, are made in Section 5.

1.1 Related Work

In graph limit theory, a key idea is to reduce extremely large graphs into simpler objects by treating them as distributions of random subgraphs. The study of such limits was initiated by Lovász and Szegedy (2006), and further developed by Borgs et al. (2008, 2012). These notions relate to exchangeable random arrays, considered originally by Aldous (1981); Hoover (1979) and related to random graphs by Diaconis and Janson (2007). The work of Elek and Szegedy (2012) extended this approach to dense, uniform hypergraphs by considering analogous notions of homomorphism convergence and cut-metrics, which was further studied by Zhao (2015). The general approach of relating homomorphism densities and some cut-metric has been applied to form limit theories for other combinatorial objects as well, such as partially ordered sets (Janson, 2011) and random cographs (Stufler, 2021).

The theory of large, random simplicial complexes for modeling higher-order interactions in network science has been quite active recently, with approaches from topological data analysis, dynamical systems, and signal processing (Schaub et al., 2021; Battiston and Petri,

---

1. Or, more generally, powers of a Borel probability space. See Section 5.3 for more details.
2. Following the concluding remarks of Bobrowski and Krioukov (2022).

2022). A series of recent papers considered the topological properties of large, random simplicial complexes (Costa and Farber, 2016, 2017a,b). Indeed, a type of limiting object for simplicial complexes similar to the Rado graph was considered by Farber et al. (2021), which arises as the limit of a dense random simplicial complex on countably many nodes, in a certain regime. In the context of TDA, the study of persistent homology of random subsamples of metric measure spaces has been considered in great detail by Chazal et al. (2014, 2015, 2017). In particular, the persistence homology of persistence barcodes of random subsets of compact metric measure spaces is considered. This is in contrast to our work in two ways. First, the simplicial complexes we consider are not required to be modeled by an underlying metric measure space. Second, complexons naturally model simplicial complexes where not only the nodes are random (in that the nodes themselves are positioned in space according to some distribution), but where the faces connecting them are random. Initial studies of random complexes in noncompact spaces have been undertaken by Adler et al. (2014); Owada and Bobrowski (2020), where the (persistent) homological structure of point clouds drawn from distributions of unbounded support in $\mathbb{R}^d$ are considered. In the observation of what they dub "topological crackle," the robustness of TDA methods to potentially unbounded noise is studied. Both of these works consider simplicial complexes determined via fixed rules applied to random sets of points, which varies from our study of complexes that are random even when conditioned on a fixed set of points. We refer the reader to the excellent survey paper of Bobrowski and Krioukov (2022) for more references on the emerging topic of large, random simplicial complexes.

There has also been a body of literature studying regularity lemmas and limit theories for hypergraphs (Rödl and Skokan, 2004; Gowers, 2006, 2007; Tao, 2006; Elek and Szegedy, 2012; Zhao, 2015; Balasubramanian, 2021). These works are concerned with studying *uniform hypergraphs*, where each hyperedge has some fixed cardinality. This stands in contrast to our work, as simplicial complexes are inherently nonuniform (unless of course, they are merely graphs, which are 2-uniform hypergraphs). We also develop a partial theory of limits of nonuniform hypergraphs in Section 5.4, framed by the upper and lower bounds of a hypergraphs by simplicial complexes.

2. Preliminaries

The notation $[n]$ denotes the set of integers $\{1, 2, \ldots, n\}$. A simplicial complex[3] K is a finite collection of finite sets that is closed under restriction, that is, the taking of subsets. The collection of sets in K with cardinality $d + 1$ is denoted by $K^{(d)}$, and we call that set the set of d-simplices. For a d-simplex σ, we say that $\dim \sigma = d$. The set $K^{(0)}$ is referred to as the set of nodes, with the alternative notation $\mathsf{V}(K)$, when convenient. We denote the cardinality of $\mathsf{V}(K)$ by $\mathsf{v}(K)$. Although $K^{(0)}$ is a set of singleton sets, we will find it useful sometimes to identify said singleton sets with their constituent elements. The sets $K^{(1)}$ and $K^{(2)}$ are sometimes referred to as the sets of edges and triangles, respectively. If K is such that for all $c > d$, $K^{(c)} = \emptyset$, then we say that K is a simplicial complex of dimension d. We denote the smallest such d by $\dim K$. A simple graph, for instance, is a simplicial complex of dimension one. It will also be useful to consider weighted simplicial complexes.

---

3. Strictly speaking, we are concerned with abstract simplicial complexes. However, we will never refer to the geometric realization, so we will omit the word *abstract*.

A weighted simplicial complex (H, ω) is taken to be the power set $H = 2^{\mathsf{V}(H)}$ for some finite $\mathsf{V}(H)$, and a function $\omega : H \to [0, 1]$, where ω is called the weight function. It is assumed that $\omega(\mathsf{V}(H)) = 1$ and $\omega(\emptyset) = 1$.

For a simplicial complex K, the set of facets of K is the collection of maximal sets in K, that is, the sets in K that are subsets of no other sets in K, denoted K°. Observe that the set of facets of a simplicial complex completely characterizes the simplicial complex, as the simplicial complex can be reconstructed by taking the set of facets and all of their subsets. The set of antifacets of K is the collection of minimal sets in $2^{\mathsf{V}(K)} \setminus K$, that is, the collections of nodes that do not constitute a simplex, but whose strict subsets do constitute a simplex, denoted $K_\circ$. Similarly, the set of antifacets of a simplicial complex completely characterize the simplicial complex, as the simplicial complex can be reconstructed by taking the set of all strict subsets of the antifacets. We define the facets of a weighted simplicial complex not by changing the underlying set, but by changing the weight function. For a weighted simplicial complex (H, ω), the faceted weighted simplicial complex (H°, ω°) is such that $H^\circ = H$, and for each $\sigma \in H^\circ$,

$$\omega^\circ(\sigma) = \prod_{\sigma' \subseteq \sigma} \omega(\sigma').$$

2.1 Homomorphism Densities

Much like in the theory of dense graph limits, we characterize simplicial complexes in terms of their homomorphism densities. For two simplicial complexes F, K, a homomorphism from F to K is a map $\phi : \mathsf{V}(F) \to \mathsf{V}(K)$ such that for any $\sigma \in F^{(d)}$, it holds that $\phi(\sigma) \in K^{(d)}$, that is, a homomorphism is a simplex-preserving map. An induced homomorphism is an injective homomorphism from F to K with the added condition that $\sigma \in F^{(d)}$ if and only if $\phi(\sigma) \in K^{(d)}$, that is, an injective homomorphism that also preserves non-simplices. We denote the number of homomorphisms and induced homomorphisms from F to K by $\mathsf{hom}(F, K)$ and $\mathsf{ind}(F, K)$, respectively.

Normalizing these quantities appropriately yields homomorphism densities. Homomorphism densities answer the following type of question: given a uniformly chosen (injective) random map $\phi : \mathsf{V}(F) \to \mathsf{V}(K)$, what is the probability that that map is an (induced) homomorphism? More precisely, we respectively define the homomorphism density and induced homomorphism density of F in K as

$$t(F, K) = \frac{\mathsf{hom}(F, K)}{\mathsf{v}(F)^{\mathsf{v}(K)}}$$

$$t_{\mathsf{ind}}(F, K) = \frac{\mathsf{ind}(F, K)}{P(\mathsf{v}(K), \mathsf{v}(F))},$$

where $P(n, k)$ is the number of k-permutation of n, that is, the number of injective maps from a set of cardinality k into a set of cardinality n.

2.2 The Cut Distance

The homomorphism densities characterize the distribution of subcomplexes in a simplicial complex. This may be thought of as looking at local structures in a simplicial complex. On the other hand, the cut-metric characterizes simplicial complexes in a global way. The

cut-norm is defined for a multidimensional matrix (Frieze and Kannan, 1999, Section 6) in the following way. For finite sets $X_1, X_2, \ldots, X_r$, let $A : \prod_j X_j \to \mathbb{R}$. For subsets $S_j \subseteq X_j$ for $j \in [r]$, put

$$A(S_1, \ldots, S_r) = \sum_{e \in \prod_j S_j} A(e).$$

The (normalized) cut-norm of A is defined as

$$\|A\|_\square = \frac{\max\left\{|A(S_1, \ldots, S_r)| : S_j \subseteq X_j \text{ for } j \in [r]\right\}}{\prod_j |X_j|}.$$

Let K_1, K_2 be simplicial complexes such that $\mathsf{V}(K_1) = \mathsf{V}(K_2)$. For some $d \geq 1$, let $A_{1,d} : \prod_{j=1}^{d+1} \mathsf{V}(K_1) \to \{0, 1\}$ be the indicator function of d-simplices in K_1, that is, $A_{1,d}(i_1, \ldots, i_{d+1}) = 1$ if and only if $\{i_1, \ldots, i_{d+1}\} \in K_1^{(d)}$. Define $A_{2,d}$ in the same way for K_2. The d-dimensional labeled cut-distance between K_1 and K_2 is defined as follows:

$$\mathrm{d}_{\square,d}(K_1, K_2) = \|A_{1,d} - A_{2,d}\|_\square.$$

We extend this definition to weighted simplicial complexes (K_3, ω) by putting $A_{3,d}(i_1, \ldots, i_{d+1}) = \omega(\{i_1, \ldots, i_{d+1}\})$, and using the normalized cut-norm for the metric as before. This allows us to compare weighted and unweighted simplicial complexes.

Observe that the d-dimensional labeled cut-distance finds the maximal collection of subsets $S_1, \ldots, S_{d+1} \subseteq \mathsf{V}(K_1)$ such that the number of d-simplices contained in $\prod_j S_j$ differs maximally between K_1 and K_2. More precisely, for arbitrary subsets $S_1, \ldots, S_{d+1} \subseteq \mathsf{V}(K_1)$, put

$$K_1^{(d)}(S_1, \ldots, S_{d+1}) = \left\{ \sigma \in K^{(d)} : \sigma \in \prod_j S_j \right\},$$

and similarly define[4] $K_2^{(d)}(S_1, \ldots, S_{d+1})$. Then, one can see that

$$\mathrm{d}_{\square,d}(K_1, K_2) = \frac{\max_{S_j \subseteq X_j : j \in [d+1]} \left| |K_1^{(d)}(S_1, \ldots, S_{d+1})| - |K_2^{(d)}(S_1, \ldots, S_{d+1})| \right|}{\mathsf{v}(K)^{d+1}}.$$

The d-dimensional labeled cut-distance describes how different two simplicial complexes may look when observing d-simplices across partitions of the node set.

The distances $\mathrm{d}_{\square,d}$ for $d \geq 1$ are each only able to characterize the difference between two complexes in a single dimension at a time. We define the labeled cut-distance by taking a weighted sum of all such distances. Namely, for a nonnegative, summable sequence $(\alpha_j)_{j \geq 1}$ of real numbers, put

$$\mathrm{d}_\square(K_1, K_2; (\alpha_j)) = \sum_{j=1}^\infty \alpha_j \mathrm{d}_{\square,j}(K_1, K_2).$$

Again, this definition naturally extends to include weighted simplicial complexes as well.

---

4. We abuse the notation $\sigma \in \prod_j S_j$ here, meaning that some tuple formed by permuting the elements of σ is contained in the Cartesian product.

The family of labeled cut-distances are dependent upon the assumption that the node sets of K_1 and K_2 are equal in size and correspond to one another. We now consider the case where two simplicial complexes have node sets that do not align. As a preliminary definition, for a nonnegative, summable sequence $(\alpha_j)_{j \geq 0}$, and two simplicial complexes K_1, K_2 such that $\mathsf{v}(K_1) = \mathsf{v}(K_2)$, put

$$\widehat{\delta}_\square(K_1, K_2; (\alpha_j)) = \min_{\phi:\mathsf{V}(K_2) \to \mathsf{V}(K_1)} \mathrm{d}_\square(K_1, \phi(K_2); (\alpha_j)),$$

where ϕ ranges over bijective maps from $\mathsf{V}(K_2)$ to $\mathsf{V}(K_1)$. That is to say, $\widehat{\delta}_\square$ first finds the optimal alignment between the two simplicial complexes, and then compares them via $\mathrm{d}_\square$.

With this in hand, we now extend the definition to handle simplicial complexes whose node sets differ in size. First, for an integer $m > 0$, define the m-blowup of a simplicial complex K as the simplicial complex mK where $\mathsf{V}(mK) = \mathsf{V}(K) \times [m]$, and a set $\{(v_i, j_i)\}_{i=1}^{d+1} \in (mK)^{(d)}$ if and only if $\{v_i\}_{i=1}^{d+1} \in K^{(d)}$, where we assume that $\{v_i\}_{i=1}^{d+1}$ is a set of cardinality $d+1$ (no duplicate nodes). Then, for two simplicial complex K_1, K_2 such that $\mathsf{v}(K_i) = n_i$, define the unlabeled cut-distance as

$$\delta_\square(K_1, K_2; (\alpha_j)) = \lim_{m \to \infty} \widehat{\delta}_\square(mn_2 K_1, mn_1 K_2; (\alpha_j)).$$

The unlabeled cut-distance blows both simplicial complexes up by a very large factor, so that it can then find a very fine alignment between them, amounting to a so-called fractional overlay of the node sets. Indeed, $\delta_\square$ can be equivalently defined in terms of a minimizing fractional overlay of the node sets (see Lovász, 2012, Eq. 8.10), but we omit that discussion here.

3. Complexons and Random Sampling

As in the theory of graph limits, the convergence of a sequence of simplicial complexes can be defined in terms of homomorphism densities. Let $K_1, K_2, \ldots$ be a sequence of simplicial complexes. We say that the sequence $(K_n)_{n \geq 1}$ is convergent if for all simplicial complexes F, it holds that $(t(F, K_n))_{n \geq 1}$ is a convergent sequence. The remainder of this paper is devoted to understanding the appropriate limiting objects for such convergent sequences.

The appropriate analogs for graphons in this case are functions on the disjoint union of unit cubes of dimension greater than or equal to 2. A complexon is a measurable[5] function

$$W : \bigsqcup_{d \geq 1} [0, 1]^{d+1} \to [0, 1],$$

such that W is measurable and totally symmetric. That is to say, the restriction of W to each $[0, 1]^{d+1}$ is measurable and symmetric in its coordinates. We find it convenient to assume that $W([0, 1]) = 1$ and $W(\emptyset) = 1$, that is, W evaluated on the "empty tuple" has unit value. We denote the set of all complexons by $\mathcal{W}$.

In the same way that a simplicial complex as a collection of sets closed under restriction can be equivalently described by its facets, we formulate a faceted complexon. For a

---

5. In this paper, "measurable" means Borel-measurable. The properties considered are not sensitive to differences on sets of measure zero, so one could also consider the Lebesgue measure.

complexon W, the faceted complexon W° is a totally symmetric measurable function $W^\circ :$ $\bigsqcup_{d \geq 1}[0,1]^{d+1} \to [0,1]$ obeying the following rule. For all $d \geq 1, (x_1, \ldots, x_{d+1}) \in [0,1]^{d+1}$, put

$$W^\circ(x_1, \ldots, x_{d+1}) = \prod_{\sigma \subseteq [d+1]} W(x_\sigma),$$

where x_σ denotes the coordinates of $(x_1, \ldots, x_{d+1})$ indexed by σ (this is well-defined, by the total symmetry of the complexon W). Prematurely interpreting the values taken by W as "probabilities" of simplices existing conditioned on the existence of their faces, the faceted complexon W° intuitively describes the probability of a simplex existing. For instance, if $W(x_1, x_2, x_3) = 1$, but $W(x_1, x_2) = W(x_1, x_3) = W(x_2, x_3) = 0.5$, then $W^\circ(x_1, x_2, x_3) = 0.125$, reflecting the closure of a simplicial complex under restriction.

3.1 Complexons as Random Simplicial Complex Models

A complexon induces a distribution of random simplicial complexes via a simple sampling procedure. Let W be a complexon, and $S_n = \{x_1, \ldots, x_n\}$ be a set of n points contained in $[0,1]$, for some integer $n \geq 1$. From W and S_n, define a weighted simplicial complex $H = W[S_n]$ so that $\mathsf{V}(H) = [n], H = 2^{[n]}$, and each $\sigma \in H$ has weight $\omega(\sigma) = W(x_\sigma)$. From the weighted simplicial complex H, we randomly sample an unweighted simplicial complex in the following way. Let K be a simplicial complex, and put $\mathsf{V}(K) = [n]$. Inductively for $d \geq 1$, for each $d + 1$-subset $\sigma \subseteq [n]$ such that every strict subset of σ is contained in K, include σ in K with probability $\omega(\sigma)$ (that is, the weight of the simplex σ in H). Upon termination, this yields a finite simplicial complex K. We call the distribution from which K is drawn $\mathbb{K}(H)$.

Noting that H is defined based on the complexon W via the set S_n, we randomize the set S_n to yield a more general random simplicial complex model. Let S_n be such that $x_1, \ldots, x_n$ are i.i.d. samples from the uniform distribution probability distribution on $[0,1]$. Then, the random simplicial complex sampled from $W[S_n]$ with random S_n is denoted $\mathbb{K}(n, W)$.

The distribution $\mathbb{K}(n, W)$ can be thought of in the same way as the lower model of Farber et al. (2022). Indeed, it can be constructed by taking the largest simplicial complex contained in a random hypergraph on node set $[n]$ whose hyperedges $\sigma \subseteq [n]$ exist independently each with probability $W(x_\sigma)$. Moreover, the faceted complexon W° can be used via the upper model of Farber et al. (2022) to yield the distribution $\mathbb{K}(n, W)$. Again, take a random hypergraph H on the node set $[n]$ whose hyperedges $\sigma \subseteq [n]$ exist independently each with probability $W^\circ(x_\sigma)$. Taking K to be the smallest simplicial complex that contains H, it can be shown that K is distributed according to $\mathbb{K}(n, W)$. We go in to more detail on the relationship between complexons and hypergraph limits in Section 5.4.

3.2 Homomorphisms in Complexons

Motivated by the random sampling model $\mathbb{K}(n, W)$, the homomorphism densities t and t_{ind} can be naturally generalized to complexons. Let F be a simplicial complex, and identify $\mathsf{V}(F) = [n]$ for some $n \geq 1$. The homomorphism density of F in a complexon W asks the following question: what is the probability that F is contained in the random simplicial complex $\mathbb{K}(n, W)$? Based on this question, we directly define the homomorphism density

of F in a complexon W. Put

$$t(F, W) = \int_{[0,1]^{\mathsf{V}(F)}} \prod_{\sigma \in F} W(x_\sigma) \mathrm{d}x. \tag{1}$$

The question of the homomorphism density can also be posed in terms of facets. Put

$$t(F^\circ, W^\circ) = \int_{[0,1]^{\mathsf{V}(F)}} \prod_{\sigma \in F^\circ} W^\circ(x_\sigma) \mathrm{d}x. \tag{2}$$

A simple comparison of the integrands of (1) and (2) yields the following result, further justifying the name "faceted complexon" for W°.

Lemma 1. *For any simplicial complex F and complexon W, it holds that $t(F, W) = t(F^\circ, W^\circ)$.*

In the case of graphons, the induced homomorphism density is computed in a fashion similar to (1), but with a product over all antifacets included in the integrand. For simplicial complexes, taking the product over all non-simplices would be redundant, as the non-existence of an edge, for instance, implies the non-existence of all higher-order simplices that would contain that edge. Treating a graph with no isolated nodes as a simplicial complex of dimension one, notice that the facets of a graph are its edges, and the antifacets are its non-edges. Based on this, we write the induced homomorphism density of F in W as follows:

$$t_{\mathsf{ind}}(F, W) = \int_{[0,1]^{\mathsf{V}(F)}} \prod_{\sigma \in F} W(x_\sigma) \prod_{\tau \in F_\circ} (1 - W(x_\tau)) \, \mathrm{d}x.$$

The following result can be shown by unrolling definitions, in order to relate the induced homomorphism densities and random samples of a complexon:

Lemma 2. *For a complexon W and a simplicial complex F such that $\mathsf{V}(F) = [n]$, it holds that*

$$t_{\mathsf{ind}}(F, W) = \mathbb{P}\{\mathbb{K}(n, W) = F\}.$$

In graph theory, we are typically concerned with properties of graphs that do not change under the relabeling of nodes. Similarly, in graph limit theory, we typically describe properties of complexons up to measure-preserving transformations of $[0, 1]$. For complexons, a property of this type holds for homomorphism densities. Let $\phi : [0, 1] \to [0, 1]$ be a measure-preserving transformation, and for a complexon W, define

$$W^\phi(x_1, \ldots, x_{d+1}) = W(\phi(x_1), \ldots, \phi(x_{d+1}))$$

for all $d \geq 1$ and $x \in [0, 1]^{d+1}$. The following result proceeds directly from the definition of the (induced) homomorphism density.

Lemma 3. *Let W be a complexon, and $\phi : [0, 1] \to [0, 1]$ a measure-preserving transformation. Then, for all simplicial complexes F,*

$$t(F, W) = t(F, W^\phi)$$
$$t_{\mathsf{ind}}(F, W) = t_{\mathsf{ind}}(F, W^\phi).$$

3.3 The Cut Distance

Similar to the family of cut-distances for simplicial complexes, we define cut-distances for complexons. Let W_1, W_2 be complexons, and $d \geq 1$ some integer. The d-dimensional labeled cut-distance between W_1 and W_2 is defined as follows:

$$\mathrm{d}_{\square,d}(W_1, W_2) = \sup_{S_1,\ldots,S_{d+1} \in \mathcal{B}[0,1]} \left| \int_{\prod_j S_j} (W_1(x) - W_2(x))\,\mathrm{d}x \right|,$$

where $\mathcal{B}[0,1]$ denotes the Borel σ-field on the interval $[0,1]$. This relates to the cut-norm defined for general measurable functions $U : \bigsqcup_{d \geq 1}[0,1]^{d+1} \to \mathbb{R}$:

$$\|U\|_{\square,d} = \sup_{S_1,\ldots,S_{d+1} \in \mathcal{B}[0,1]} \left| \int_{\prod_j S_j} U(x)\mathrm{d}x \right|,$$

so that $\mathrm{d}_{\square,d}(W_1, W_2) = \|W_1 - W_2\|_{\square,d}$. Alternatively, the cut-norm can be written as

$$\|U\|_{\square,d} = \sup_{f_1,\ldots,f_{d+1}:[0,1]\to[0,1]} \left| \int_{[0,1]^{d+1}} U(x)f_1(x_1)\cdots f_{d+1}(x_{d+1})\mathrm{d}x \right|,$$

where $f_1, \ldots, f_{d+1}$ are taken to be measurable functions from $[0,1]$ to $[0,1]$.

Following (Borgs et al., 2008, Eq. 7.2), restricting the sets considered in the d-dimensional labeled cut-distance to be pairwise disjoint only changes the distance by a constant factor depending on d.

Lemma 4. *Let $W_1, W_2 \in \mathcal{W}$ be complexons, and let $d \geq 1$. Put $\mathcal{S}$ as the collection of pairwise disjoint sets $S_1, \ldots, S_{d+1} \in \mathcal{B}[0,1]$. Then,*

$$\sup_{(S_1,\ldots,S_{d+1})\in\mathcal{S}} \left| \int_{\prod_j S_j} (W_1(x) - W_2(x))\mathrm{d}x \right| \leq \frac{1}{(d+1)^{d+1}} \mathrm{d}_{\square,d}(W_1, W_2).$$

The proof follows the presentation of Janson (2013, Lemma E.2).

Proof Let $\mathcal{A} = \{A_i\}_{i=1}^n$ be an n-equipartition of $[0,1]$ (that is, a partition of $[0,1]$ into n measurable subsets all with measure $1/n$). Let I be an element of $[d+1]^n$ chosen uniformly at random. For each $j \in [d+1]$, define $B_j = \bigcup_{m:I(m)=j} A_m$.

For any measurable sets $S_1, \ldots, S_{d+1}$, the sets $S_1 \cap B_1, \ldots, S_{d+1} \cap B_{d+1}$ are pairwise disjoint, so that

$$\mathbb{E}\left[\left| \int_{\prod_j S_j \cap B_j} (W_1(x) - W_2(x))\mathrm{d}x \right| \right] \leq \sup_{(S_1,\ldots,S_{d+1})\in\mathcal{S}} \left| \int_{\prod_j S_j} (W_1(x) - W_2(x))\mathrm{d}x \right|. \qquad (3)$$

Moreover,

$$\mathbb{E}\left[\int_{\prod_j S_j \cap B_j} (W_1(x) - W_2(x))\mathrm{d}x \right] = \sum_{i_1 \neq \ldots \neq i_{d+1} \in [n]} \int_{\prod_j S_j \cap A_{i_j}} \frac{W_1(x) - W_2(x)}{(d+1)^{d+1}}\mathrm{d}x. \qquad (4)$$

One can see that the right-hand side of (4) approaches the quantity

$$\frac{1}{(d+1)^{d+1}} \int_{\prod_j S_j} (W_1(x) - W_2(x)) \mathrm{d}x$$

as $n \to \infty$. Combining (3) and (4) concludes the proof. ∎

As the labeled d-dimensional cut-distance only considers the value of complexons on $[0,1]^{d+1}$, we define a cut-distance that considers all dimensions. Let $(\alpha_j)_{j \geq 1}$ be a nonnegative, summable sequence of real numbers. For complexons W_1, W_2, put

$$d_\square(W_1, W_2; (\alpha_j)_{j \geq 1}) = \sum_{j \geq 1} \alpha_j d_{\square,j}(W_1, W_2).$$

One can clearly see that for any complexons W_1, W_2,

$$d_\square(W_1, W_2; (\alpha_j)_{j \geq 1}) \leq \sum_{j \geq 1} \alpha_j \int_{[0,1]^{j+1}} |W_1(x) - W_2(x)| \mathrm{d}x. \tag{5}$$

We extend this definition to finite nonnegative sequences $\alpha_1, \ldots, \alpha_d$ by implicitly assuming that $\alpha_j = 0$ for $j > d$.

The labeled cut-distance implicitly assumes a correspondence between the domains of W_1 and W_2. Intuitively, this corresponds to coupling the distributions $\mathbb{K}(n, W_1)$ and $\mathbb{K}(n, W_2)$ (for any $n \geq 1$) in such a way that the points $x_1, \ldots, x_n \in [0,1]$ are always equal for $\mathbb{K}(n, W_1)$ and $\mathbb{K}(n, W_2)$.

To decouple the domains, we define the unlabeled cut-distance between W_1 and W_2 by taking the infimum over measure preserving transformations of the domain. Put, again for a (finite or infinite) nonnegative summable sequence $(\alpha_j)_{j \geq 1}$,

$$\delta_\square(W_1, W_2; (\alpha_j)_{j \geq 1}) = \inf_{\phi:[0,1] \to [0,1]} d_\square(W_1, W_2^\phi; (\alpha_j)_{j \geq 1}),$$

where ϕ ranges over measure-preserving transformations. Similar to (5), we have the inequality

$$\delta_\square(W_1, W_2; (\alpha_j)_{j \geq 1}) \leq \inf_{\phi:[0,1] \to [0,1]} \sum_{j \geq 1} \alpha_j \int_{[0,1]^{j+1}} |W_1(x) - W_2^\phi(x)| \mathrm{d}x.$$

The definition of the unlabeled cut-distance for simplicial complexes in terms of the limiting distance under blowups motivates an immediate representation of simplicial complexes as complexons, closely resembling the representation of graphs as graphons via "pixel pictures." For a simplicial complex K whose nodes are identified as $\mathsf{V}(K) = [n]$, define a complexon W_K in the following way. Define the sets $P_j = [(j-1)/n, j/n)$ for $j \in [n]$. Then, for each $\{j_1, \ldots, j_{d+1}\} \in K$, put $W_K(x_1, \ldots, x_{d+1}) = 1$ for all permutations of $(x_1, \ldots, x_{d+1}) \in \prod_{\ell=1}^{d+1} P_{j_\ell}$. Otherwise, put $W_K(x_1, \ldots, x_{d+1}) = 0$. If instead of a simplicial complex K we have a weighted simplicial complex (H, ω), we put $W_H(x_1, \ldots, x_{d+1}) = \omega(\{j_1, \ldots, j_{d+1}\})$.

11

For a simplicial complex K and a complexon W, we will find it convenient to use the notation $\mathrm{d}_\square(K, W)$ and $\delta_\square(K, W)$ to refer to $\mathrm{d}_\square(W_K, W)$ and $\delta_\square(W_K, W)$, respectively. Moreover, the complexon corresponding to a simplicial complex preserves the cut-distance and all homomorphism densities. The following lemma summarizes this, in a way similar to (Lovász, 2012, Eq. 7.2 and Lemma 8.9), so we omit the proof here.

Lemma 5. *For any simplicial complexes F, K and any nonnegative, summable sequence $(\alpha_j)_{j \geq 1}$,*

$$t(F, K) = t(F, W_K)$$
$$\delta_\square(F, K; (\alpha_j)) = \delta_\square(W_F, W_K; (\alpha_j)).$$

Moreover, if $\mathsf{V}(F) = \mathsf{V}(K)$, then, under any identification of $\mathsf{V}(F)$ with $[n]$,

$$\mathrm{d}_\square(F, K; (\alpha_j)) = \mathrm{d}_\square(W_F, W_K; (\alpha_j)).$$

Furthermore, if (H, ω) is a weighted simplicial complex such that $\mathsf{V}(F) = \mathsf{V}(H)$, then

$$\mathrm{d}_\square(F, H; (\alpha_j)) = \mathrm{d}_\square(W_F, W_H; (\alpha_j)).$$

4. Main Results

In this section, we characterize the space of complexons equipped with the cut-distance, and show how the cut-distance relates to the homomorphism densities. In particular, we show that they induce equivalent, compact topologies on the space of complexons. Our treatment closely follows that of Lovász and Szegedy (2006); Borgs et al. (2008).

4.1 The Counting Lemma

We begin by showing how closeness in the cut-distance yields closeness in the sense of homomorphism densities.

Lemma 6 (Counting Lemma). *Let F be a simplicial complex. Define the sequences $\alpha_j = |(F^\circ)^{(j)}|, \beta_j = |F^{(j)}|, \gamma_j = |F^{(j)}| + |(F_\circ)^{(j)}|$ for $j \geq 1$. Then, for complexons U, W, the following three inequalities hold:*

$$|t(F, U) - t(F, W)| \leq \delta_\square(U^\circ, W^\circ; (\alpha_j)_{j \geq 1})$$
$$|t(F, U) - t(F, W)| \leq \delta_\square(U, W; (\beta_j)_{j \geq 1})$$
$$|t_{\mathsf{ind}}(F, U) - t_{\mathsf{ind}}(F, W)| \leq \delta_\square(U, W; (\gamma_j)_{j \geq 1}).$$

This result is analogous to (Lovász and Szegedy, 2006, Lemma 4.1), with a similar proof as well.

Proof We establish the first inequality, as the other two follow from a similar argument. Order the elements of F° as $\{\sigma_1, \ldots, \sigma_m\}$, where $m = |F^\circ|$. For $x \in [0, 1]^{\mathsf{V}(F)}, t \in [m]$, define

$$X_t(x) = \left(\prod_{s < t} U^\circ(x_{\sigma_s}) \right) \cdot \left(\prod_{s > t} W^\circ(x_{\sigma_s}) \right) \cdot (U^\circ(x_{\sigma_t}) - W^\circ(x_{\sigma_t})).$$

One can check that

$$|t(F, U) - t(F, W)| = \left| \int_{[0,1]^{\mathsf{V}(F)}} \sum_{t=1}^{m} X_t(x) \mathrm{d}x \right|$$

$$= \left| \sum_{t=1}^{m} \int_{[0,1]^{\mathsf{V}(F) \setminus \sigma_t}} \int_{[0,1]^{\sigma_t}} X_t(x) \prod_{j \in \sigma_t} \mathrm{d}x_j \prod_{j \in \mathsf{V}(F) \setminus \sigma_t} \mathrm{d}x_j \right|.$$

For all $t \in [m]$, it holds that

$$\prod_{s < t} U^{\circ}(x_{\sigma_s}) \prod_{s > t} W^{\circ}(x_{\sigma_s}) \in [0, 1],$$

This implies via Lemma 1 that

$$|t(F, U) - t(F, W)| \leq \left| \sum_{t=1}^{m} \mathrm{d}_{\square, \dim \sigma_t}(U^{\circ}, W^{\circ}) \right|$$

$$= \mathrm{d}_{\square}(U^{\circ}, W^{\circ}; (\alpha_j)_{j \geq 1}).$$

Taking the infimum over measure-preserving transformations W^{ϕ} yields the first inequality, via Lemma 3. The proofs of the second and third inequalities proceed similarly. ∎

This result is perhaps not too surprising. Thinking of the cut-metric as describing the global differences between complexons, closeness globally forces closeness locally (that is, in the sense of homomorphism densities). More formally, for any simplicial complex F, the homomorphism density $t(F, \cdot) : \mathcal{W} \to [0, 1]$ is continuous with respect to the cut-distance $\delta_{\square}(\cdot, \cdot; (\alpha_j)_{j \geq 1})$ for any strictly positive, summable sequence $(\alpha_j)_{j \geq 1}$. This can be shown using the usual ϵ-δ definition of continuity, choosing for a given simplicial complex F and $\epsilon > 0$

$$\delta = \frac{\epsilon}{\dim F \cdot \max_j |F^{(j)}| \cdot \min_{k \in [\dim F]} \alpha_k}.$$

Note then that $\delta_{\square}(U, W; (\alpha_j)_{j \geq 1}) < \delta$ implies $|t(F, U) - t(F, W)| < \epsilon$. Similarly, since $t_{\mathsf{ind}}(F, W) = \mathbb{P}\{\mathbb{K}(n, W) = F\}$ for any simplicial complex F with $\mathsf{V}(F) = [n]$, closeness in the cut-metric forces two complexons to yield similar random sampling models, particularly for small n.

4.2 The Sampling Lemma

Before proving the inverse of Lemma 6, we establish a concentration result for samples $\mathbb{K}(n, W)$: namely, we show that sufficiently large samples drawn from a complexon will be close in the cut-distance. Formally,

Lemma 7 (Sampling Lemma). *Let W be a complexon. Let $n \geq 1$ be an integer, and $\alpha_1, \ldots, \alpha_d$ be a finite nonnegative sequence. Then, with probability at least $1 - \exp(-n/(2 \log_2 n))$, it holds that*

$$\delta_{\square}(W^{\circ}, \mathbb{K}(n, W); (\alpha_j)) \leq \frac{12 \cdot 2^d + 1}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j.$$

13

We leave the proof to Section A. As a corollary of Lemma 7, it can be shown via the Borel-Cantelli Lemma that every faceted complexon W° arises as the limit of a sequence of simplicial complexes. That is,

Corollary 8. *Let $\alpha_1, \ldots, \alpha_d$ be a finite nonnegative sequence. For a complexon W, the sequence of simplicial complexes $(\mathbb{K}(n, W))_{n \geq 1}$ converges to W° in the cut-distance $\delta_\square(\cdot, \cdot; (\alpha_j))$ with probability 1.*

4.3 The Inverse Counting Lemma

We are now ready to state the inverse of Lemma 6.

Lemma 9 (Inverse Counting Lemma). *Let U, W be two complexons, and suppose that for some $d \geq 1, n \geq 2$, for all simplicial complexes F on n nodes of dimension d,*

$$|t(F, U) - t(F, W)| \leq 0.999 \cdot 2^{-((1+n)^{d+2})}$$

Then, for all finite nonnegative sequences $\alpha_1, \ldots, \alpha_d$,

$$\delta_\square(U^\circ, W^\circ; (\alpha_j)) \leq \frac{24 \cdot 2^d + 2}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j.$$

We leave the proof to Section B. The bound in Lemma 9 is only useful for $\log_2 n \gtrsim 24 \cdot 2^d$, with homomorphism densities that are extremely close. However, it does demonstrate that the cut-distance, a quantity determined by a Lebesgue integral on a continuous domain over all possible Borel sets of certain dimension, can be characterized by a collection of quantities determined by finite objects. Keeping in mind that this bound is generic, holding for any possible pair of complexons, one would expect it to be much better if considered on a tamer family of complexons, for example, those that are stepfunctions on some coarse partition of $[0, 1]$. Indeed, we have the following result for stepfunctions.

Proposition 10. *Let U, W be two complexons. Suppose for some integer $m \geq 1$, U and W are both stepfunctions on respective equipartitions $\mathcal{P}_U, \mathcal{P}_W$, each with at most m steps. Furthermore, suppose that for some $d \geq 1, n \geq 2$, for all simplicial complexes F on n nodes of dimension d,*

$$|t(F, U) - t(F, W)| \leq 0.999 \cdot 2^{-((1+n)^{d+2})}.$$

Then, for all finite nonnegative sequences $\alpha_1, \ldots, \alpha_d$,

$$\delta_\square(U^\circ, W^\circ; (\alpha_j)) \leq \frac{\sqrt{m}(4d + 5) + 4}{\sqrt{n}} \sum_{j=1}^{d} \alpha_j.$$

We omit the proof, as the argument is a mere simplification of the proofs of Lemmas 7 and 9. By making the stronger assumption that the complexons of interest are stepfunctions on coarse equipartitions of $[0, 1]$, Proposition 10 presents an inverse counting lemma that holds for $n \gtrsim md^2$.

14

4.4 Topology of the Space of Complexons

Denote the set of all complexons by $\mathcal{W}$, and let $(\alpha_j)_{j \geq 1}$ be a nonnegative, summable sequence of real numbers. If two complexons $W_1, W_2 \in \mathcal{W}$ only differ on a set of measure zero, then $\mathrm{d}_\square(W_1, W_2; (\alpha_j)_{j \geq 1}) = 0$. Indeed, the labeled cut-distance is only a pseudometric on $\mathcal{W}$. Furthermore, the unlabeled cut-distance $\delta_\square$ can take value zero if there exists a measure-preserving bijection $\phi : [0, 1] \to [0, 1]$ such that W_1 and W_2^ϕ only differ on a set of measure zero, and is thus also a pseudometric. The following result states that, topologically speaking, the choice of (α_j) in this construction does not matter too much.

Lemma 11. *Let $(\alpha_j)_{j \geq 1}$ and $(\beta_j)_{j \geq 1}$ be strictly positive, summable sequences of real numbers. Then, the cut-metrics $\delta_\square(\cdot, \cdot; (\alpha_j)_{j \geq 1})$ and $\delta_\square(\cdot, \cdot; (\beta_j)_{j \geq 1})$ are topologically equivalent pseudometrics on $\mathcal{W}$.*

With Lemmas 6, 9 and 11 in mind, we endow $\mathcal{W}$ with a topology determined by any strictly positive, summable sequence $(\alpha_j)_{j \geq 1}$. In particular, we define the canonical topology on $\mathcal{W}$ as the topology determined by the pseudometric

$$\gamma(U, W; (\alpha_j)_{j \geq 1}) = \delta_\square(U^\circ, W^\circ; (\alpha_j)_{j \geq 1}),$$

for an arbitrary strictly positive, summable sequence $(\alpha_j)_{j \geq 1}$. Observe that the pseudometric γ behaves similarly to the unlabeled cut-distance $\delta_\square$, except it first applies the faceting map to the arguments, to reflect the fact that when considering homomorphism densities, the faceted complexon is a better descriptor – indeed, one can check that the faceting map is a contractive mapping with respect to the cut-distance $\delta_\square$. We also find it useful to define the δ-topology on $\mathcal{W}$ as the topology determined by the pseudometric $\delta_\square$, again for some strictly positive, summable sequence. The following result about the canonical topology on $\mathcal{W}$ is immediate.

Proposition 12. *Let $\mathcal{W}_0 \subset \mathcal{W}$ be the set of complexons that arise from simplicial complexes, that is, for every $W \in \mathcal{W}_0$, there is some simplicial complex K such that $W = W_K$. Then, $\mathcal{W}_0$ is dense in $\mathcal{W}$ in the canonical topology.*

Our primary result of this section establishes the compactness of $\mathcal{W}$ with respect to this topology.

Theorem 13 (Compactness). *The space $\mathcal{W}$ with the canonical topology is compact.*

We leave the proofs of Lemma 11, Proposition 12, and Theorem 13 to Section C. So far, we have established that the space of complexons with the canonical topology is "essentially finite," in that it can be approximated by finitely many (Theorem 13) finite simplicial complexes (Proposition 12) in the cut-metric. This implies, for instance, that the space of complexons $\mathcal{W}$ with the pseudometric γ is complete.

The canonical topology on $\mathcal{W}$ can also be defined based on homomorphism densities.

Theorem 14 (Equivalence of Cuts and Homomorphisms). *Let $(a_j)_{j \geq 1}$ be a strictly positive, summable sequence of real numbers. Noting that there are countably many simplicial*

complexes (up to isomorphism), let $(F_j)_{j \geq 1}$ be an enumeration of the set of all isomorphism classes of simplicial complexes. Define a pseudometric ρ on $\mathcal{W}$ so that for any $U, W \in \mathcal{W}$,

$$\rho(U, W; (a_j)) = \sum_{j=1}^{\infty} a_j \cdot |t(F_j, U) - t(F_j, W)|.$$

Then, the topology on $\mathcal{W}$ induced by ρ is equal to the canonical topology.

Proof We prove that a sequence of complexons $W_1, W_2, \ldots$ is convergent in the canonical topology if and only if it is convergent with respect to the pseudometric ρ.

(If). Let $\epsilon > 0$ be given, and suppose $W_m \to W$ in the canonical topology. Let M be such that $\sum_{j>M} a_j < \epsilon/2$. Then, by Lemma 6, there exists some m_0 such that for all $j \leq M, m > m_0$, it holds that

$$|t(F_j, W_m) - t(F_j, W)| < \frac{\epsilon}{2M \max_{k \leq M} a_k}.$$

It follows, then, that for all $m > m_0$, $\rho(W_m, W; (a_j)_{j \geq 1}) < \epsilon$. Since ϵ was given arbitrarily, this implies that $W_m \to W$ in the topology induced by ρ, as desired.

(Only if). Let $\epsilon > 0$ be given, and let $(\alpha_j)_{j \geq 1}$ be an arbitrary strictly positive, summable sequence. Suppose $W_m \to W$ in the topology induced by ρ. Let d be such that $\sum_{j>d} \alpha_j < \epsilon/2$. Put n such that

$$\frac{24 \cdot 2^d + 2}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j < \frac{\epsilon}{2}.$$

Noting that there are only finitely many isomorphism classes of simplicial complexes on n nodes of dimension d, let M be such that for every simplicial complex F on n nodes of dimension d, there is some $j \leq M$ such that $F \cong F_j$. By assumption, there is some m_0 such that for all $j \leq M, m > m_0$

$$|t(F_j, W_m) - t(F_j, W)| < 0.999 \cdot 2^{-((1+n)^{d+2})}.$$

By Lemma 9, this implies that

$$\delta_\square(W_m^\circ, W^\circ; (\alpha_j)_{j \geq 1}) \leq \frac{24 \cdot 2^d + 2}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j + \sum_{j>d} \alpha_j < \epsilon.$$

Since ϵ was given arbitrarily, this implies that $W_m \to W$ in the canonical topology, as desired. ∎

Observe that the pseudometric ρ in Theorem 14 is defined for an arbitrary strictly positive, summable sequence $(a_j)_{j \geq 1}$ and enumeration $(F_j)_{j \geq 1}$: this indicates that, similar to the δ-topology as described by Lemma 11, the topology induced by ρ is not dependent on the particular sequence $(a_j)_{j \geq 1}$ and enumeration $(F_j)_{j \geq 1}$. In other words, there is a "hom-topology" on $\mathcal{W}$ that is pseudometrizable by any such ρ. Moreover, one can see that a sequence $(W_m)_{m \geq 1}$ converges in the hom-topology if and only if for all simplicial complexes

F, $t(F, W_m)$ is a convergent sequence. Theorem 14, then, indicates that the canonical topology on $\mathcal{W}$ is the topology defined by convergence in homomorphism densities.

Indeed, by the definition of the canonical topology on $\mathcal{W}$, two complexons U, W are identified with each other if and only if they are indistinguishable *by sampling*, which is a property that holds modulo sets of measure zero across the faceted complexons, rather than the original complexons themselves. The identification of complexons by the canonical topology not only ignores differences on sets of measure zero, which have no consequence in sampling and homomorphism densities, but also enforces a "consistency structure across different dimensions" as suggested by Bobrowski and Krioukov (2022), by considering the faceted version.

5. Discussion and Remarks

By Theorem 14, homomorphism densities are sufficient to characterize complexons in the cut-metric. This yields a characterization of limit objects for sequences of simplicial complexes. Indeed, recall that sequence (K_n) for $n \geq 1$ of simplicial complexes is said to be convergent if for all simplicial complexes F, the sequence $(t(F, K_n))_{n \geq 1}$ is convergent. Equivalently, by Lemma 5, the sequence (K_n) is convergent if for all simplicial complexes F, the sequence $(t(F, W_{K_n}))$ is convergent. Applying Theorem 14, this is equivalent to convergence of the sequence (W_{K_n}) in the canonical topology. Since $\mathcal{W}$ is compact in the canonical topology (see Theorem 13), the limit of the sequence (W_{K_n}) is itself an element $W \in \mathcal{W}$. It is in this sense that we say W (or any equivalent complexon in the canonical topology) is "the limit of" the sequence of simplicial complexes (K_n) as $n \to \infty$.

In the remainder of the paper, we discuss how complexons relate to some other notions in the literature.

5.1 Limits of partially ordered sets

Limits of partially ordered sets (posets) in homomorphism density were considered by Janson (2011), where it was shown that any sequence of partially ordered sets whose homomorphism densities converge have a limit representable by a kernel on an ordered probability space. It was conjectured that all such kernels can be taken to be defined on the ordered probability space $[0, 1]$ with the Lebesgue measure, which was affirmed by Hladký et al. (2015). In the context of this work, one may wonder if limits of sequences of simplicial complexes could be understood as posets. Indeed, a simplicial complex K carries with it a strict partial order determined by set inclusion, yielding the partially ordered set $(K, \subset)$. Moreover, since the simplices have a notion of dimension, there is a rank function associated with $(K, \subset)$, yielding a graded poset $(K, \subset, \dim)$.

A simplicial homomorphism from a simplicial complex F into K is a poset homomorphism as considered by Janson (2011), but has some additional structure. Namely, a simplicial homomorphism $\phi : \mathsf{V}(F) \to \mathsf{V}(K)$ is rank-preserving as a poset homomorphism, in that $\dim(\sigma) = \dim(\phi(\sigma))$ for all $\sigma \in F$. General poset homomorphisms as considered by Janson (2011) do not require the rank to be preserved. It is interesting to see how the rank-preservation condition on these poset homomorphisms yields a significantly different limit structure, namely a complexon, as opposed to a kernel on an ordered probability space. Indeed, finite graded posets may be treated as abstract cell complexes, of which abstract

simplicial complexes are a special type – this suggests that the symmetry conditions of complexons may be relaxed in order to yield limit objects for more general combinatorial structures.

5.2 Local Consistent Random Simplicial Complexes

We have studied the limits of dense sequences of simplicial complexes via complexons, but this is not the only obvious limiting structure. Another such object for the graph case is a locally consistent random graph model (Lovász, 2012, Chapter 11.2). Analogous structures exist for simplicial complexes. A random complex model is a sequence $(\mu_n)_{n \geq 1}$ of probability measures such that, for all $n \geq 1$, μ_n is a probability measure on the set of simplicial complexes K such that $\mathsf{V}(K) = [n]$, and if K and K' are isomorphic, then $\mu_n(K) = \mu_n(K')$.

The random complex model (μ_n) is said to be consistent if for any simplicial complex K on nodes $[n - 1]$, we have

$$\mu_{n-1}(K) = \mu_n \left(\left\{ F : \mathsf{V}(F) = [n], F' = K \right\} \right),$$

where $F' = K$ indicates that the removal of the node n from F yields the simplicial complex K. Furthermore, if for all $n > 1$, and disjoint subsets $S, T \subseteq [n]$, the random simplicial complexes determined by taking the induced subcomplex on S, T of a simplicial complex sampled following μ_n are independent, we say that (μ_n) is local.

Indeed, any simplicial complex K yields a random complex model. Define, for $n \geq 1$, the probability measure $\mu_{K,n}$ such that for any simplicial complex F with $\mathsf{V}(F) = [n]$

$$\mu_{K,n}(F) = t_{\mathsf{ind}}(F, K).$$

The following result, following (Lovász, 2012, Theorem 11.7), links convergent sequences of simplicial complexes to local consistent random complex models.

Theorem 15. *Let a convergent sequence of simplicial complexes $K_1, K_2, \ldots$ with $\mathsf{v}(K_m) \to \infty$ as $m \to \infty$ be given. Define, for all $n \geq 1$ and all simplicial complexes F with $\mathsf{V}(F) = [n]$, the probability measure*

$$\mu_n(F) = \lim_{m \to \infty} \mu_{K_m,n}(F),$$

noting that the sequence $K_1, K_2, \ldots$ being convergent implies the limit exists. Then, the sequence $(\mu_n)_{n \geq 1}$ forms a local consistent random complex model. Conversely, every local consistent random complex model arises this way.

The proof follows that of (Lovász, 2012, Theorem 11.7) almost exactly, so we omit it here. Furthermore, an initial connection with complexons is stated in the following result (compared to Lemma 2).

Proposition 16. *Let W be a complexon. Define, for all $n \geq 1$, the probability measure for simplicial complexes F on nodes $[n]$*

$$\mu_n(F) = \mathbb{P} \left\{ \mathbb{K}(n, W) = F \right\} = t_{\mathsf{ind}}(F, W).$$

Then, the sequence $(\mu_n)_{n \geq 1}$ forms a local consistent random complex model.

Viewing the sampled simplicial complexes from a complexon as being equivalently sampled from a local consistent random complex model is a helpful perspective for establishing convergence of a complexon's samples in the cut-metric. In particular,

Proposition 17. *Let $(\mu_n)_{n\geq 1}$ be a local consistent random complex model. For each $n \geq 1$, independently draw a simplicial complex K_n on nodes $[n]$ according to the distribution μ_n. Then, with probability 1,*

 (i) *The sequence (K_n) is convergent*

 (ii) *For all n, $\lim_{m\to\infty} \mu_{K_m,n} = \mu_n$.*

In particular, if $(\mu_n = \mathbb{K}(n, W))_{n\geq 1}$ for some complexon W, we have that a sequence of increasingly large simplicial complexes sampled from a complexon converges to W° with probability 1.

We also omit the proof, as it resembles that of (Lovász, 2012, Lemma 11.8) with very little modification. Propositions 16 and 17 together indicate that for a given complexon W and a sequence of increasingly large simplicial complexes $(K_n \sim \mathbb{K}(n, W))_{n\geq 1}$, $K_n \to W^\circ$ with probability 1. That is to say, we can (almost surely) generate a sequence converging to an arbitrary complexon by simply taking a sequence of large simplicial complexes sampled from it. Indeed, this recovers Corollary 8.

5.2.1 Homogeneous random complexes

Some simple instances of local consistent random complex models have been described in the literature. We go through the dense variants of these models, specifying both the corresponding complexon and the random complex model.

Linial and Meshulam (2006); Meshulam and Wallach (2009) considered the homological connectivity of random simplicial complexes generated in the following way. For some real number $p \in [0,1]$ and integers $d > 1, n \geq 1$, a random simplicial complex $K_{n,p}$ is formed on the node set $[n]$ by including all $d-1$ simplices on $[n]$, and including each element of $\binom{[n]}{d+1}$ independently with probability p. This model is known as the "Linial-Meshulam model." One can define a complexon whose samples are distributed in this way. Define $W \in \mathcal{W}$ so that for all $x \in [0,1]^{c+1}$ for $c < d$, we have $W(x) = 1$, and for all $x \in [0,1]^{d+1}$ we have $W(x) = p$. It is clear to see that for all n, the random simplicial complex $\mathbb{K}(n, W)$ follows the Linial-Meshulam model.

If the Linial-Meshulam assumes a fully-connected $(d-1)$-skeleton and fills in d-simplices at random, the random flag, or clique, complex (Kahle, 2009) does the opposite. The flag complex, again for some $p \in [0,1]$, yields a random complex on nodes $[n]$ by beginning with an Erdős-Rényi random graph G_n with edge probability p, and then taking the maximal simplicial complex whose 1-skeleton is equal to G_n. This can also be expressed with an appropriately constructed complexon. Define $W \in \mathcal{W}$ so that for all $x \in [0,1]^2$, $W(x) = p$, and otherwise $W(x) = 1$. Then, the random simplicial complex $\mathbb{K}(n, W)$ is a random flag complex.

As a generalization of both of these, Costa and Farber (2016) considers random simplicial complexes parameterized by a sequence of coefficients $p_1, p_2, p_3, \ldots$ contained in $[0,1]$. The multiparameter, or Costa-Farber, random simplicial complex is distributed according to

$\mathbb{K}(n, W)$, where W is such that for all $x \in [0,1]^{d+1}$, $W(x) = p_d$. Further work showed that in the dense, or medial, regime of $p \le p_j \le P$ for some $0 < p \le P < 1$ for all j, the dth homology group of such a random complex in a given dimension is nontrivial for n in a logarithmically-sized interval, being trivial elsewhere with high probability (Farber and Mead, 2020). That is to say, as $n \to \infty$, the homology group of a fixed dimension for complexes sampled according to $\mathbb{K}(n, W)$ vanishes with probability tending to one. For some complexons, particularly those bounded away from zero in correspondence with the medial regime studied by Farber and Mead (2020), it is seemingly difficult then to speak of the "homological structure of a complexon" in general. As an example, consider a simplicial complex K on n nodes with a nonzero first Betti number. Then, consider some simplicial complex on $n' \gg n$ nodes sampled from W_K, which will likely have a much larger first Betti number. Thus, the Betti numbers are in some sense destroyed when going from a simplicial complex to its corresponding complexon.

These three models are all instances of homogeneous random simplicial complexes, where the probability of a simplex existing is dependent only on its dimension. For a complexon W, this corresponds to W being constant on $[0,1]^{d+1}$ for each $d \ge 1$.

5.3 Other Random Simplicial Complex Models

As described in Section 3.1, a random simplicial complex sampled from a complexon following $\mathbb{K}(n, W)$ essentially picks nodes $[n]$ with "latent positions" $x_1, \ldots, x_n$ uniformly distributed in the unit interval. That is to say, the random complex structure is dependent on each node's position in the space $[0, 1]$. In the ensuing construction, no topological properties of $[0, 1]$ were leveraged, suggesting that the definition of a complexon can be extended to nodes sampled from any Borel probability space. Let $(\Omega, \mathcal{F}, \mu)$ be a Borel probability space, which we denote by Ω for short. A complexon on Ω is a measurable function

$$W : \bigsqcup_{d \ge 1} \Omega^{d+1} \to [0, 1].$$

All of the basic constructions so far generalize immediately to complexons on Ω: sampling, homomorphism densities, cut-metrics, and so on. In particular, if $(\Omega, \mathcal{F}, \mu)$ is a Borel probability space, then a complexon W on Ω can be transformed into a complexon W on $[0, 1]$ in a way that preserves all homomorphism densities. That is to say, complexons on $[0, 1]$ are sufficient for describing complexons on Borel probability spaces in general, which arise, for instance, when Ω is a metric space endowed with the Borel σ-field. For example, it is well-known that every probability measure on the Borel σ-field of $\mathbb{R}^d$ yields a Borel probability space. We outline the details of this in Section D, with reference to (Lovász, 2012, Section 13.1).

Defining complexons over alternative probability spaces is useful for describing random simplicial complexes tied to a natural underlying space. We consider here dense random geometric Čech complexes (Bobrowski and Krioukov, 2022). Let μ be a probability measure on $\mathbb{R}^d$ for some integer $d \ge 1$, and let $\mathcal{X}$ denote the support of μ. For some $\epsilon > 0$, define a complexon W_ϵ on $\mathcal{X}$ so that, for any points $x_1, \ldots, x_{d+1} \in \mathcal{X}$,

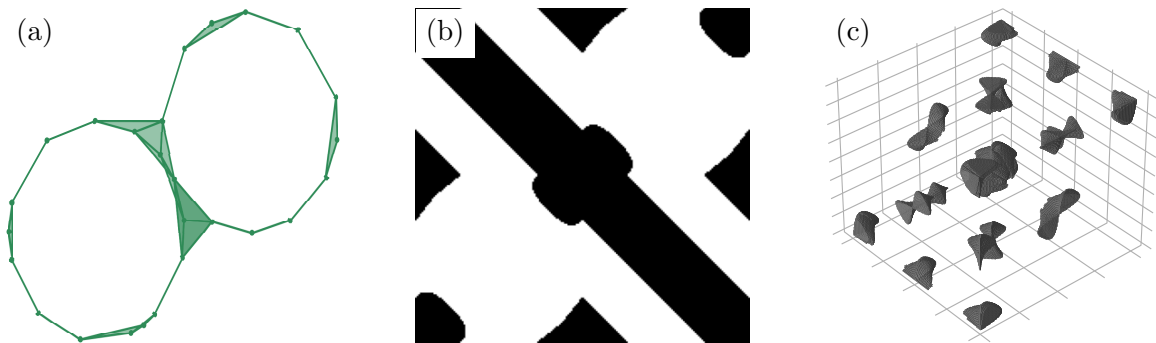$$W_\epsilon(x_1, \ldots, x_{d+1}) = \mathbf{1}(\bigcap_{j=1}^{d+1} B_{\epsilon/2}(x_j) \neq \emptyset),$$

Figure 2: **(a)** Čech complex formed from points uniformly sampled from a bouquet of two circles in $\mathbb{R}^2$. **(b)** The complexon W_ϵ as defined on $[0,1]^2$ (dark regions indicate a value of 1). **(c)** The complexon W_ϵ as defined on $[0,1]^3$ (dark regions indicate a value of 1).

where $B_{\epsilon/2}(x_j)$ indicates the open ball in $\mathbb{R}^d$ of radius $\epsilon/2$ centered about x_j. That is to say, $W_\epsilon(x_1, \ldots, x_{d+1}) = 1$ if and only if the collection of points $\{x_1, \ldots, x_{d+1}\}$ has diameter less than ϵ.

Going back to our definition of the random simplicial complex $\mathbb{K}(n, W_\epsilon)$, let $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ be such that $x_1, \ldots, x_n$ are i.i.d. samples from the probability distribution μ. Form a simplicial complex K such that $\mathsf{V}(K) = [n]$. Inductively for $d \geq 1$ until termination, for each $d+1$-subset $\sigma \subseteq [n]$ such that every strict subset of σ is contained in K, include σ in K with probability $W_\epsilon(x_\sigma)$. The random simplicial complex K formed in this way is, much like before, said to be sampled from the distribution $\mathbb{K}(n, W_\epsilon)$. Moreover, under our particular definition of W_ϵ, it is clear that the resulting simplicial complex K is the Čech complex $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^d)$. In this way, $\mathbb{K}(n, W_\epsilon)$ is the distribution of Čech complexes with fixed radius ϵ formed by sampling n i.i.d. points in $\mathcal{X}$ from the probability distribution μ.

Revisiting our example of the Čech complex formed from the bouquet of two circles in Section 1 (see Figure 1), we can now couch our understanding of convergent subsampled Čech complexes in the language of complexons. As before, let $\mathcal{X}$ be the bouquet of two unit circles in $\mathbb{R}^2$. For any $n \geq 1$, let $\mathcal{X}_n$ be a collection of n i.i.d. samples from the uniform distribution on $\mathcal{X}$. An instance of the random Čech complex $\check{C}_\epsilon(\mathcal{X}_n, \mathbb{R}^2)$ is pictured in Figure 2 (a).

We can equivalently describe the random Čech complex as a complexon. For the purpose of illustration, parameterize the space $\mathcal{X}$ as a curve $c : [0,1] \to \mathbb{R}^2$, defined as

$$c(t) = \begin{cases} (\cos(4\pi t) - 1, \sin(4\pi t)) & t < 1/2 \\ (1 - \cos(4\pi t), \sin(4\pi t)) & t \geq 1/2. \end{cases}$$

The uniform measure μ on $\mathcal{X}$ is merely the pushforward measure of the uniform measure on $[0,1]$ via the map c. Moreover, c is injective almost everywhere, so we can represent the complexon modeling $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^2)$ as a standard complexon on $[0,1]$, up to a set of measure

21

zero. The complexon W_ϵ as defined on $[0, 1]^2$ and $[0, 1]^3$ is pictured in Figure 2 (b,c). Indeed, W_ϵ can be thought of as the indicator function for simplices in the Čech complex $\check{C}_\epsilon(\mathcal{X}, \mathbb{R}^2)$.

5.4 Convergence of Nonuniform Hypergraphs

An open problem in the theory of graph homomorphisms and graph limits is the characterization of nonuniform hypergraphs with unbounded edges (Lovász, 2008). As noted in Section 3.1, the distribution $\mathbb{K}(n, W)$ can be thought of in terms of "lower sampling" following (Farber et al., 2022), where a random hypergraph H is drawn according to W, and then taking the largest simplicial complex contained in that hypergraph. On the other hand, the "upper sampling" approach (Farber et al., 2022) can be reproduced by drawing a random hypergraph H according to W°, and then taking the smallest simplicial complex that contains that hypergraph. Both of these approaches yield the same distribution $\mathbb{K}(n, W)$.

   The viewpoint of simplicial complexes formed from the upper or lower bounds of hypergraphs sheds some light on limits of dense, nonuniform hypergraphs. For a hypergraph H, let $\lfloor H \rfloor$ be the largest simplicial complex contained in H, and $\lceil H \rceil$ the smallest simplicial complex containing H. We say that a sequence of hypergraphs $(H_n)_{n\geq 1}$ is upper-lower-convergent (or, for brevity, UL-convergent) if the following conditions hold simultaneously for some complexon W:

   (i) $(\lceil H_n \rceil)_{n\geq 1}$ is convergent with limiting complexon W

   (ii) $(\lfloor H_n \rfloor)_{n\geq 1}$ is convergent with limiting complexon W°.

If this is the case, we say that $H_n \to W$ as $n \to \infty$. Since for all n, $\lceil H_n \rceil \in \mathcal{W}^\circ$, by Lemmas 33 and 34, it holds that $W \in \mathcal{W}^\circ$.

   Based on Lemma 7, we know that large simplicial complexes sampled from a complexon are close to the faceted complexon in the cut-metric, with high probability. Based on the relationship between faceted complexons and upper/lower sampling, we establish a similar approach to show that sequences of hypergraphs sampled from faceted complexons are UL-convergent. Let W be a complexon, and $n \geq 1$ an integer. Similar to the definition of $K \sim \mathbb{K}(n, W)$, we define a random hypergraph H as follows. First, take n i.i.d. samples $x_1, \ldots, x_n$ uniformly at random in the unit interval $[0, 1]$. Let H be a hypergraph with $\mathsf{V}(H) = [n]$. Then, for each $\sigma \in 2^{[n]}$, include $\sigma \in H$ with probability $W(x_\sigma)$. We call the distribution from which H is drawn $\mathbb{H}(n, W)$.

Lemma 18. *For a complexon W and integer $n \geq 1$, it holds that*

$$\mathbb{K}(n, W) \overset{D}{=} \lfloor \mathbb{H}(n, W) \rfloor \overset{D}{=} \lceil \mathbb{H}(n, W^\circ) \rceil,$$

where $\overset{D}{=}$ denotes equality in distribution.

   We now consider the relationship between UL-convergence, simplicial homomorphisms, and hypergraph homomorphisms. For a hypergraph H, define a complexon W_H in the same way as one does for simplicial complexes, that is, by taking indicator functions of tuples

defined on a partition of $[0, 1]$. From this, we inherit all the usual notions of homomorphisms and cut-metrics for hypergraphs. For a hypergraph H and a complexon W, define

$$t(H, W) = \int_{[0,1]^{\mathsf{V}(H)}} \prod_{\sigma \in H} W(x_\sigma) \prod_{j \in \mathsf{V}(H)} \mathrm{d}x_j.$$

With the notion of a homomorphism for hypergraphs, it is natural to ask whether the convergence of a sequence of hypergraphs in all hypergraph homomorphisms implies UL-convergence. Indeed, convergence of all hypergraph homomorphism densities is enough to establish convergence of the sequence $(\lfloor H_n \rfloor)$, but not to establish convergence of $(\lceil H_n \rceil)$.

Counterexample 1. *For some $p \in [0, 1]$ define $W_p \in \mathcal{W}$ so that $W(x) = p$ for $x \in [0, 1]^3$, and $W(x) = 0$ otherwise. Define a sequence of hypergraphs $(H_{n,p})_{n \geq 1}$ where each $H_{n,p}$ is a random hypergraph distributed according to $\mathbb{H}(n, W_p)$. One can check that with probability 1, for any hypergraph G, $t(G, H_{n,p})$ is convergent, so that $(H_{n,p})$ is a convergent sequence in the sense of homomorphisms.*

Setting $p = 1$, it is obvious that $(H_{n,1})$ is not UL-convergent, since $\lfloor H_{n,1} \rfloor$ is the empty simplicial complex on n nodes, and $\lceil H_{n,1} \rceil$ is the complete 2-dimensional simplicial complex on n nodes. Then, one can see that $\lceil H_{n,1} \rceil \to W$, where $W(x) = 1$ for $x \in [0, 1]^{d+1}$ when $d \in \{1, 2\}$, and $W(x) = 0$ otherwise. On the other hand, $\lfloor H_{n,1} \rfloor \to 0$. It is not true that $W^\circ = 0$ in this case, so UL-convergence does not hold.

In the above counterexample, the hypergraphs $(H_{n,1})$ were drawn from a complexon $W_1 \in \mathcal{W} \setminus \mathcal{W}^\circ$, that is, "non-faceted" complexons. In light of the UL-limit of a convergent hypergraph sequence necessarily being an element of $\mathcal{W}^\circ$, this may imply some utility of the extra structure imposed by $\mathcal{W}^\circ$. This is noted in the following proposition.

Proposition 19. *Let $W \in \mathcal{W}^\circ$ be given, and take a sequence $(H_n)_{n \geq 1}$ of random hypergraphs $H_n \sim \mathbb{H}(n, W)$. Then, (H_n) is UL-convergent with limit W with probability 1.*

Proof For the sequence (H_n), note that for each n, $\lfloor H_n \rfloor$ is distributed according to $\mathbb{K}(n, W)$ by Lemma 18. By Corollary 8, this implies that $(\lfloor H_n \rfloor) \to W^\circ$ with probability 1.

Similarly, since $W \in \mathcal{W}^\circ$, there is some $U \in \mathcal{W}$ such that $W = U^\circ$. Then, for each n, $\lceil H_n \rceil$ is distributed according to $\mathbb{K}(n, U)$, again by Lemma 18. Thus, $(\lceil H_n \rceil) \to W$ with probability 1. ∎

Proposition 19 and Counterexample 1 together indicate that considering the upper and lower simplicial complexes of a sequence of hypergraphs may help establish a theory of nonuniform hypergraphs, but fails in cases where the limiting object is not a faceted complexon.

Acknowledgments

Appendix A. Proof of Lemma 7

A.1 Equipartitioning

Let $\mathcal{P}$ be a partition of $[0,1]$ into finitely many measurable sets. For a complexon W, define the projection $W_{\mathcal{P}}$ as follows. For $\{p_j \in P_j : j \in [d+1]\}$, where each P_j is an element of $\mathcal{P}$, put

$$W_{\mathcal{P}}(p_1, \ldots, p_{d+1}) = \int_{\prod_{j=1}^{d+1} P_j} W(x_1, \ldots, x_{d+1}) \mathrm{d}x.$$

The complexon $W_{\mathcal{P}}$ is an instance of a stepfunction on $\mathcal{P}$, that is, $W_{\mathcal{P}}(x_1, \ldots, x_{d+1})$ is only dependent on which sets in $\mathcal{P}$ the points $x_1, \ldots, x_{d+1}$ are contained in. We establish a result stating that any complexon can be approximated in the cut-distance by a piecewise constant complexon with respect to an equipartition of the unit interval. The following is a weak regularity lemma, stating that any complexon can be approximated by a stepfunction over an appropriate partition. It is a trivial extension of (Frieze and Kannan, 1999, Theorem 12).

Theorem 20. *Let W be a complexon. For given integers $n > 1, d \geq 1$, there is a partition $\mathcal{P}$ of $[0,1]$ into n measurable sets such that there exists a stepfunction U on $\mathcal{P}$ such that for all $1 \leq j \leq d$,*

$$\mathrm{d}_{\square,j}(W, U) \leq \sqrt{\frac{d+1}{\log_2 n}}.$$

Following (Lovász, 2012), we strengthen Theorem 20 to hold for equipartitions, where an equipartition of a measure space is a partition such that each set has equal measure.

Lemma 21 (Equipartitioning). *Let W be a totally symmetric, measurable function. For given integers $n > 1, d \geq 1$, there exists an n-equipartition $\mathcal{P}$ of $[0,1]$ such that for all finite, nonnegative sequences of real numbers $\alpha_1, \ldots, \alpha_d$,*

$$\mathrm{d}_{\square}(W, W_{\mathcal{P}}; (\alpha_j)) \leq \frac{\sum_{j=1}^{d} \alpha_j \left(4\sqrt{d+1} + 2^j\right)}{\sqrt{\log_2 n}}.$$

Proof Put $m = \lceil n^{1/4} \rceil$. Let $\mathcal{Q} = \{Q_j\}_{j=1}^{m}$ be an m-partition as guaranteed by Theorem 20, with corresponding stepfunction U. Partition each class Q_j into sets of measure $1/n$, with at most one exceptional set of measure less than $1/n$ for each class. Notice that the measure of the union of the exceptional sets is bounded by m/n. Take the union of the exceptional sets, and partition it into sets of measure $1/n$, yielding an equipartition $\mathcal{P}$.

Take the common refinement $\mathcal{R} = \mathcal{P} \wedge \mathcal{Q}$. Note that $\mathcal{P}$ and $\mathcal{R}$ only differ over those classes in $\mathcal{P}$ formed from exceptional sets. Thus, for all $j \geq 1$, the restrictions of $W_{\mathcal{P}}$ and $W_{\mathcal{R}}$ to $[0,1]^{j+1}$ only differ on a set of measure at most $2^j(m/n)$. This implies that

$$\mathrm{d}_{\square,j}(W_{\mathcal{R}}, W_{\mathcal{P}}) \leq 2^j \frac{m}{n}.$$

Via an easily shown variation of (Lovász, 2012, Lemma 9.12), we see that

$$\mathrm{d}_{\square,j}(W, W_{\mathcal{R}}) \leq 2\mathrm{d}_{\square,j}(W, U),$$

which implies via the triangle inequality that

$$\mathrm{d}_{\square,j}(W, W_{\mathcal{P}}) \leq 2\mathrm{d}_{\square,j}(W, U) + 2^j \frac{m}{n}.$$

One can then show via straightforward calculation that

$$\mathrm{d}_{\square,j}(W, W_{\mathcal{P}}) \leq \frac{4\sqrt{d+1} + 2^j}{\sqrt{\log_2 n}}.$$

This yields the bound

$$\mathrm{d}(W, W_{\mathcal{P}}; (\alpha_j)) = \sum_{j=1}^{d} \alpha_j \mathrm{d}_{\square,j}(W, W_{\mathcal{P}}) \leq \frac{\sum_{j=1}^{d} \alpha_j \left(4\sqrt{d+1} + 2^j\right)}{\sqrt{\log_2 n}},$$

as desired. ∎

A.2 Sample Concentration

We show that "nice" parameters of simplicial complexes sampled from a complexon concentrate around their expectation.

Definition 22 (Reasonably Smooth Parameter). *Let f be a function mapping simplicial complexes to the real numbers. Suppose that for any simplicial complexes F, G such that $\mathsf{V}(F) = \mathsf{V}(G)$ and whose structure varies only on faces incident to a single node, that is, there exists some $v \in \mathsf{V}(F)$ such that for all $\sigma \in F$,*

$$\sigma \in F \triangle G \Rightarrow v \in \sigma,$$

we have

$$|f(F) - f(G)| \leq 1.$$

Under these conditions, we say that f is a reasonably smooth parameter.[6]

We will need the following result (Lovász, 2012, Corollary A.15), which is a corollary of Azuma's inequality.

Proposition 23. *Let $(\Omega, \mathcal{A}, \pi)$ be a probability space. Let $f : \Omega^n \to \mathbb{R}$ be a measurable function such that*

$$|f(x) - f(y)| \leq 1$$

whenever x and y differ in one coordinate only. Let X be a random point in Ω^n distributed according to the product measure. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left\{f(X) \geq \mathbb{E}[f(X)] + \epsilon n\right\} \leq \exp\left(-\epsilon^2 n/2\right).$$

---

6. $F \triangle G$ denotes the symmetric difference between sets.

Proposition 23 gives us a generic, one-sided tail bound for well-behaved functions on sequences of i.i.d. random variables. We now construct the random simplicial complex $\mathbb{K}(n, W)$ in a way that fits this schema.

Let W be a complexon, and let $n \geq 1$ be given. Let $\Omega = \prod_{d=0}^{n-1}[0, 1]^{([n]^d)}$ with the uniform probability measure. Take i.i.d. samples $x_j = (\alpha_{j,0}, \ldots, \alpha_{j,n-1})$ from Ω for $j \in [n]$. We construct a simplicial complex F from the points $\{x_j\}_{j=1}^n$ in the following way.

For all $\sigma \subseteq [n]$, put $\sigma \in F$ if and only if the following condition holds. For any $\sigma' \subseteq \sigma$, denote the ordered elements of σ' by $1 \leq i_1 \leq \ldots \leq i_{d+1} \leq n$. Then, we require for all such σ'

$$\alpha_{i_1,d}(i_2, \ldots, i_{d+1}) \leq W(\alpha_{i_1,0}, \ldots, \alpha_{i_{d+1},0}).$$

Under this model, a random simplicial complex F distributed according to $\mathbb{K}(n, W)$ is taken as a function of the i.i.d. sequence x_j and the complexon W. Moreover, changing any coordinate x_j in the sequence only affects simplices that contain j. We can now state the main result of this section.

Theorem 24 (Sample Concentration). *Let f be a reasonably smooth parameter. Then, for a complexon W, an integer $n \geq 1$, and real number $t \geq 0$, we have*

$$\mathbb{P}\left\{f(\mathbb{K}(n, W)) \geq \mathbb{E}[f(\mathbb{K}(n, W))] + \sqrt{2tn}\right\} \leq \exp(-t).$$

Proof Follows from Proposition 23, treating $f(\mathbb{K}(n, W))$ as the composition of functions of the sequence of i.i.d. random variables $\{x_j \in \Omega\}_{j=1}^n$. ∎

A.3 Concentration of Norm

The following lemma will be the most challenging part of proving Lemma 7.

Lemma 25 (Norm Concentration). *Let $U : [0, 1]^{d+1} \to [-1, 1]$, and let X be a random ordered n-subset of $[0, 1]$. Let $U[X]$ be the matrix in $[-1, 1]^{n^{d+1}}$ determined by evaluating U on X^{d+1}. Then with probability at least $1 - 4\exp(-n^{1/(d+1)}/2)$,*

$$-3r(n, d+1) \leq \|U[X]\|_{\square,d} - \|U\|_{\square,d} \leq \frac{5(d+1)^2}{\sqrt{n^{1/(d+1)}}},$$

where

$$r(n, d+1) = 1 - \frac{P(n, d+1)}{n^{d+1}}.$$

We begin with the following result, which is a special case of (Alon et al., 2003, Lemma 3).

Lemma 26. *Let $B : [n]^{d+1} \to [-1, 1]$. Let $S_1, \ldots, S_{d+1} \subseteq [n]$, and for each $j \in [d+1]$, let Q_j be a random subset of $[n]^{[d+1]\setminus\{j\}}$ with cardinality q. Then,*

$$B(S_1, \ldots, S_{d+1}) \leq \mathbb{E}_{Q_j}\left[B\left(S_1, \ldots, S_{j-1}, (Q_j \cap \prod_{i \neq j} S_i)^+, S_{j+1}, \ldots, S_{d+1}\right)\right] + \frac{n^{d+1}}{\sqrt{q}},$$

where

$$B(S_1, \ldots, S_{d+1}) = \sum_{v_1 \in S_1} \cdots \sum_{v_{d+1} \in S_{d+1}} B(v_1, \ldots, v_{d+1}),$$

and $(Q_j \cap \prod_{i \neq j} S_i)^+$ *is the collection of elements* $v \in [n]$ *such that,*

$$\sum_{\mathbf{q} \in Q_j \cap \prod_{i \neq j} S_i} B(q_1, \ldots, q_{j-1}, v, q_{j+1}, \ldots, q_{d+1}) > 0.$$

Lemma 26 can be iterated over all dimensions of the matrix B in order to yield an expectational bound on a variation of the cut-norm for a matrix. We find it useful to consider the one-sided cut-norm, defined as follows.

Definition 27. *Let* $B : [n]^{d+1} \to \mathbb{R}$ *and* $U : [0,1]^{d+1} \to \mathbb{R}$. *Define*

$$\|B\|_{\square,d}^+ = \frac{\max_{S_1, \ldots, S_{d+1} \subseteq [n]} B(S_1, \ldots, S_{d+1})}{n^{d+1}}$$

$$\|U\|_{\square,d}^+ = \sup_{S_1, \ldots, S_{d+1} \in \mathcal{B}[0,1]} \int_{\prod_j S_j} U(x) \mathrm{d}x.$$

Observing that $\|B\|_{\square,d} = \max\{\|B\|_{\square,d}^+; \| - B\|_{\square,d}^+\}$, our goal is to find bounds for $\|U[X]\|_{\square,d}^+ - \|U\|_{\square,d}^+$ and apply them twice, yielding the desired result.

Lemma 28. *Let* $B : [n]^{d+1} \to [-1,1]$. *Let* $\{Q_j\}_{j=1}^{d+1}$ *each be random subsets of* $[n]^{[d+1]\setminus\{j\}}$ *with cardinality* q. *Then,*

$$\|B\|_{\square,d}^+ \leq \frac{1}{n^{d+1}} \mathbb{E}_{Q_1, \ldots, Q_{d+1}} \left[\max_{R_j \subseteq Q_j} B(R_1^+, \ldots, R_{d+1}^+) \right] + \frac{d+1}{\sqrt{q}}.$$

Proof Fix $S_1, \ldots, S_{d+1} \in [n]$. Applying Lemma 26 repeatedly over each coordinate $j \in [d+1]$ yields

$$B(S_1, \ldots, S_{d+1}) \leq \mathbb{E}_{Q_1, \ldots, Q_{d+1}} \left[\max_{R_j \subseteq Q_j} B(R_1^+, \ldots, R_{d+1}^+) \right] + \frac{(d+1)n^{d+1}}{\sqrt{q}}.$$

Taking the supremum of the left-hand side over all possible sets $S_1, \ldots, S_{d+1}$ and normalizing by n^{d+1} yields the desired bound. ∎

The following lemma immediately precedes our desired result.

Lemma 29. *Let* $U : [0,1]^{d+1} \to [-1,1]$, *and let* X *be a random ordered* n-*subset of* $[0,1]$. *Then with probability at least* $1 - 2\exp(-n^{1/(d+1)}/2)$,

$$-3r(n, d+1) \leq \|U[X]\|_{\square,d}^+ - \|U\|_{\square,d}^+ \leq \frac{5(d+1)^2}{\sqrt{n^{1/(d+1)}}}.$$

27

Proof For a random n-subset $X \subseteq [0,1]$, let $B = U[X]$.

(Lower bound) For any measurable subsets $S_1, \ldots, S_{d+1} \subseteq [0,1]$ we have

$$\|B\|_{\Box,d}^+ \geq \frac{1}{n^{d+1}} U(S_1 \cap X, \ldots, S_{d+1} \cap X),$$

where $U(S_1 \cap X, \ldots, S_{d+1} \cap X)$ is used as shorthand for

$$U(S_1 \cap X, \ldots, S_{d+1} \cap X) = \sum_{x_1 \in S_1 \cap X} \cdots \sum_{x_{d+1} \in S_{d+1} \cap X} U(x_1, \ldots, x_{d+1}).$$

On average, we see that

$$\begin{aligned}
\mathbb{E}_X \left[\|B\|_{\Box,d}^+ \right] &\geq \frac{1}{n^{d+1}} \mathbb{E}_X \left[U(S_1 \cap X, \ldots, S_{d+1} \cap X) \right] \\
&= \frac{P(n, d+1)}{n^{d+1}} \int_{\prod_j S_j} U(x) \mathrm{d}x + r(n, d+1) \cdot g(U; S_1, \ldots, S_{d+1}),
\end{aligned}$$

where g is some function taking values with magnitude at most 1. Thus, since the sets $S_1, \ldots, S_{d+1}$ were given arbitrarily, we can choose them to yield the following inequality.

$$\begin{aligned}
\mathbb{E}_X \left[\|B\|_{\Box,d}^+ \right] &\geq \int_{\prod_j S_j} U(x) \mathrm{d}x - 2r(n, d+1) \\
&\geq \|U\|_{\Box,d}^+ - 2r(n, d+1).
\end{aligned}$$

Applying Proposition 23 yields

$$\|B\|_{\Box,d}^+ - \|U\|_{\Box,d}^+ \geq -3r(n, d+1)$$

with probability at least $1 - \exp(-r^2(n, d+1)/2)$.

(Upper bound) Let $Q_1, \ldots, Q_{d+1}$ be, for each j, random q-subsets of $[n]^{[d+1] \setminus \{j\}}$. We will bound the expectation of $\|B\|_{\Box,d}^+$ by first fixing some parameters. Fix the sets $R_j \subseteq Q_j \subseteq [n]^d$, and define $Q = \bigcup_j \{Q_{j,1}, \ldots, Q_{j,d}\}$, so that $Q \subseteq [n]$. Fix those values of X_i for which $i \in Q$. Define for each $j \in [d+1]$ the set

$$Y_j = \left\{ y \in [0,1] : \sum_{(i_1, \ldots, i_d) \in R_j} U(X_{i_1}, \ldots, X_{i_{j-1}}, y, X_{i_j}, \ldots, X_{i_d}) > 0 \right\}.$$

Let $X' = \{X_i : i \in [n] \setminus Q\}$. For every collection of indices $\{i_j \in [n] \setminus Q\}_{j=1}^{d+1}$, the contribution of $U(X_{i_1}, \ldots, X_{i_{d+1}})$ to $\mathbb{E}_{X'}[B(R_1^+, \ldots, R_{d+1}^+)]$ is at most

$$\int_{\prod_j Y_j} U(y_1, \ldots, y_{d+1}) \mathrm{d}y \leq \|U\|_{\Box,d}^+.$$

There are fewer than n^{d+1} such choices of indices. The remaining terms where $i_j \in Q$ for at least one $j \in [d+1]$ contribute at most $(2n|Q|)^d \leq (2(d+1)nq)^d$ in absolute value, yielding the bound

$$\mathbb{E}_{X'} \left[B(R_1^+, \ldots, R_{d+1}^+) \right] \leq n^{d+1} \|U\|_{\Box,d}^+ + 2^d \left((d+1)nq \right)^d.$$

Note that $B(R_1^+, \ldots, R_{d+1}^+)$ is a function of the random variables X'. Moreover, $B(R_1^+, \ldots, R_{d+1}^+)$ varies by at most $4n^d$ when a single coordinate of X' is changed. Thus, by Proposition 23, we have for a given $\epsilon > 0$, with probability at least $1 - \exp(-\epsilon^2 n/2)$,

$$B(R_1^+, \ldots, R_{d+1}^+) \leq \mathbb{E}_{X'}\left[B(R_1^+, \ldots, R_{d+1}^+)\right] + 4\epsilon n^{d+1}.$$

The number of possible choices of sets $R_1, \ldots, R_{d+1}$ is at most $2^{q(d+1)}$, so that the above bound holds for all $\{R_j \subset Q_j\}_{j=1}^{d+1}$ with probability at least $1 - 2^{q(d+1)} \exp(-\epsilon^2 n/2)$. This property is preserved when taking the expectation over $Q_1, \ldots, Q_{d+1}$, so that, by Lemma 28,

$$\|B\|_{\square,d}^+ \leq \frac{1}{n^{d+1}} \mathbb{E}_{Q_1,\ldots,Q_{d+1}}\left[\max_{R_j \subseteq Q_j} B(R_1^+, \ldots, R_{d+1}^+)\right] + \frac{d+1}{\sqrt{q}}$$

$$\leq \|U\|_{\square,d}^+ + \frac{(2(d+1)q)^d}{n} + 4\epsilon + \frac{d+1}{\sqrt{q}}$$

with probability at least $1 - 2^{q(d+1)} \exp\left(-\epsilon^2 n/2\right)$.

Put $q = \lfloor n^{1/(d+1)} \rfloor / (2(d+1))$ and $\epsilon = \sqrt{2(d+1)/q}$, noting that $q \geq 1$ for n large enough to make the desired bound nontrivial. This yields the bound

$$\|B\|_{\square,d}^+ - \|U\|_{\square,d}^+ \leq \frac{5(d+1)^2}{\sqrt{n^{1/(d+1)}}}$$

with probability at least $1 - \exp(-n^{1/(d+1)}/2)$.

(Conclusion of Proof) Combining the lower and upper bounds, we see that

$$-3r(n, d+1) \leq \|B\|_{\square,d}^+ - \|U\|_{\square,d}^+ \leq \frac{5(d+1)^2}{\sqrt{n^{1/(d+1)}}}$$

with probability at least $1 - 2\exp(-n^{1/(d+1)}/2)$, as desired. ∎

Proof [Proof of Lemma 25] Apply Lemma 29 to both U and $-U$. Since $\|U\|_{\square,d} = \max\{\|U\|_{\square,d}^+, \|-U\|_{\square,d}^+\}$, this yields the desired tail bound. ∎

A.4 Sampling Lemma

With these results established, we now prove Lemma 7. We will first need the following result.

Lemma 30. *Let (H, ω) be a weighted simplicial complex on n nodes. Let $\mathbb{K}(H)$ be the random simplicial complex drawn from H. For every $p \in [0, 1]$, we have*

$$\mathbb{E}\left[d_{\square,d}\left(\mathbb{K}(H), H^\circ\right)\right] \leq \frac{3(d+1)^p + 1}{n^p}.$$

Proof Let $\epsilon > 0$ be given. For $\{i_j \in [n]\}_{j=1}^{d+1}$, define the random variable $X_{\mathbf{i}} = \mathbf{1}(\mathbf{i} \in \mathbb{K}(H))$. Let $S_1, \ldots, S_{d+1}$ be pairwise disjoint subsets of $[n]$, and let $A_{K,d}$ and $A_{H^\circ,d}$ be the (weighted)

d-dimensional adjacency matrices associated to $\mathbb{K}(H)$ and H°, respectively. We call such a collection of sets *bad* if

$$|(A_{K,d} - A_{H^\circ,d})(S_1, \ldots, S_{d+1})| > \epsilon \left(\frac{n}{d+1}\right)^{d+1},$$

Note that this is only possible if $\left|\prod_j S_j\right| > \epsilon(n/(d+1))^{d+1}$. Then, by the Chernoff-Hoeffding inequality,

$$\mathbb{P}_{\mathbb{K}(H)} \left\{\left|\sum_{\mathbf{i} \in \prod_j S_j} (X_{\mathbf{i}} - \mathbb{E}[X_{\mathbf{i}}])\right| > \epsilon \left(\frac{n}{d+1}\right)^{d+1}\right\} \leq 2 \exp\left(-2\epsilon \left(\frac{n}{d+1}\right)^{d+1}\right).$$

There are $(d+2)^n$ such collections of disjoint subsets, so the probability of a bad collection existing is bounded by

$$2 \cdot (d+2)^n \exp\left(-2\epsilon \left(\frac{n}{d+1}\right)^{d+1}\right).$$

By Lemmas 4 and 5, the nonexistence of bad subsets implies $\mathrm{d}_{\square,d}(\mathbb{K}(H), H^\circ) \leq \epsilon$. Therefore, we have

$$\mathbb{E}\left[\mathrm{d}_{\square,d}(\mathbb{K}(H), H^\circ)\right] \leq \epsilon + 2 \cdot (d+2)^n \exp\left(-2\epsilon \left(\frac{n}{d+1}\right)^{d+1}\right).$$

It is sufficient to consider the case when $n^p > 3(d+1)^p + 1$, otherwise the bound is trivial. In particular, taking $\epsilon = 3\left(\frac{d+1}{n}\right)^p$ yields the desired bound. ∎

Proof [Proof of Lemma 7] Let integers $d \geq 1, n > 1$ be given. We bound the expected cut-distance for any finite, nonnegative sequence $\alpha_1, \ldots, \alpha_d$ as follows. For some partition $\mathcal{P}$ of $[0,1]$ and a random n-subset $S \subseteq [0,1]$:

$$\mathbb{E}\left[\delta_{\square}(W^\circ, \mathbb{K}(W[S]); (\alpha_j))\right] \leq \mathrm{d}_{\square}(W^\circ, W_{\mathcal{P}}^\circ; (\alpha_j)) \tag{6a}$$

$$+ \mathbb{E}\left[\delta_{\square}(W_{\mathcal{P}}^\circ, W_{\mathcal{P}}^\circ[S]; (\alpha_j))\right] \tag{6b}$$

$$+ \mathbb{E}\left[\mathrm{d}_{\square}(W_{\mathcal{P}}^\circ[S], W^\circ[S]; (\alpha_j))\right] \tag{6c}$$

$$+ \mathbb{E}\left[\mathrm{d}_{\square}(W^\circ[S], \mathbb{K}(W[S]); (\alpha_j))\right], \tag{6d}$$

by the triangle inequality. We now bound each of the above summands individually.

Bound (6a): Let $m = \lceil n^{1/4} \rceil$. By Lemma 21, there is an m-equipartition $\mathcal{P} = \{P_j\}_{j=1}^m$ of $[0,1]$ such that

$$\mathrm{d}_{\square}(W^\circ, W_{\mathcal{P}}^\circ; (\alpha_j)) \leq \frac{2 \sum_{j=1}^d \alpha_j \left(4\sqrt{d+1} + 2^j\right)}{\sqrt{\log_2 n}}$$

for all finite, nonnegative sequences $\alpha_1, \ldots, \alpha_d$, where $W_{\mathcal{P}}^\circ$ denotes the projection of W° to a stepfunction on $\mathcal{P}$ (as opposed to the faceted complexon corresponding to $W_{\mathcal{P}}$).

Bound (6b): Let $H = W_{\mathcal{P}}^\circ[S]$, and denote the corresponding complexon by W_H°. Following the proof of (Lovász, 2012, Lemma 10.16), observe that both W_H° and $W_{\mathcal{P}}^\circ$ are stepfunctions

on m-partitions of $[0,1]$, with the same function values on corresponding steps. Each step of $W_{\mathcal{P}}^{\circ}$ has measure $1/m$, while each step of W_H° has measure $|P_i \cap S|/n = 1/m + r_i$ for $i \in [m]$, where r_i denotes some residual. It is easy to see that

$$\delta_{\square}(W_{\mathcal{P}}^{\circ}, W_H^{\circ}; (\alpha_j)) \leq \sum_{i=1}^{m} |r_i| \sum_{j=1}^{d} \alpha_j(j+1).$$

By the estimate in the proof of (Lovász, 2012, Lemma 10.16), we have

$$\mathbb{E}\left[\delta_{\square}(W_{\mathcal{P}}^{\circ}, W_H^{\circ}; (\alpha_j))\right] \leq \mathbb{E}\left[\sum_{i=1}^{m} |r_i|\right] \sum_{j=1}^{d} \alpha_j(j+1) < \frac{\sum_{j=1}^{d} \alpha_j(j+1)}{n^{3/8}}.$$

Bound (6c): By Lemma 25, we have

$$|\mathrm{d}_{\square,d}(W^{\circ}[S], W_{\mathcal{P}}^{\circ}[S]) - \mathrm{d}_{\square,d}(W^{\circ}, W_{\mathcal{P}}^{\circ})| \leq \frac{5(d+1)^2}{\sqrt{n^{1/(d+1)}}}$$

with probability at least $1 - 8\exp(-n^{1/(d+1)}/2)$. This implies that

$$\mathbb{E}\left[|\mathrm{d}_{\square,d}(W^{\circ}[S], W_{\mathcal{P}}^{\circ}[S]) - \mathrm{d}_{\square,d}(W^{\circ}, W_{\mathcal{P}}^{\circ})|\right] \leq \frac{5(d+1)^2 + 5}{\sqrt{n^{1/(d+1)}}}.$$

Applying the triangle inequality yields

$$\begin{aligned}
\mathbb{E}\left[\mathrm{d}_{\square,d}(W^{\circ}[S], W_{\mathcal{P}}^{\circ}[S])\right] &\leq \mathbb{E}\left[|\mathrm{d}_{\square,d}(W^{\circ}[S], W_{\mathcal{P}}^{\circ}[S]) - \mathrm{d}_{\square,d}(W^{\circ}, W_{\mathcal{P}}^{\circ})|\right] \\
&\qquad + \mathrm{d}_{\square,d}(W^{\circ}, W_{\mathcal{P}}^{\circ}) \\
&\leq \frac{5(d+1)^2 + 5}{\sqrt{n^{1/(d+1)}}} + \frac{2^d + 4\sqrt{d+1}}{\sqrt{\log_2 n}} \\
&\leq \frac{2^{d+2}}{\sqrt{\log_2 n}},
\end{aligned}$$

where the final inequality holds by assuming that n is large enough so that the bound in Lemma 7 is nontrivial. By linearity of expectation, this yields

$$\mathbb{E}\left[\mathrm{d}_{\square}(W^{\circ}[S], W_{\mathcal{P}}^{\circ}[S]; (\alpha_j))\right] \leq \frac{\sum_{j=1}^{d} \alpha_j\left(2^{j+2}\right)}{\sqrt{\log_2 n}}.$$

Bound (6d): By Lemma 30, we have

$$\mathbb{E}\left[\mathrm{d}_{\square,d}(W^{\circ}[S], \mathbb{K}(W[S]))\right] \leq \frac{3(d+1)^{3/8} + 1}{n^{3/8}},$$

which implies by linearity of expectation

$$\mathbb{E}\left[\mathrm{d}_{\square}(W^{\circ}[S], \mathbb{K}(W[S]); (\alpha_j))\right] \leq \frac{\sum_{j=1}^{d} \alpha_j\left(3(j+1)^{3/8} + 1\right)}{n^{3/8}}.$$

Bounding expected cut-distance: Combining the respective bounds of (6a) to (6d) yields the following:

$$\mathbb{E}\left[\delta_\square(W^\circ, \mathbb{K}(W[S]); (\alpha_j))\right] \leq \sum_{j=1}^{d} \alpha_j \left(\frac{8\sqrt{d+1}+6\cdot 2^j}{\sqrt{\log_2 n}} + \frac{j+2+3(j+1)^{3/8}}{n^{3/8}}\right).$$

Since $\sum_{j=1}^{d} \alpha_j$ is always a bound for $\delta(W^\circ, \mathbb{K}(W[S]); (\alpha_j))$, it is sufficient to consider n sufficiently large so that the bound in Lemma 7 is nontrivial. In this regime, we have

$$\mathbb{E}\left[\delta_\square(W^\circ, \mathbb{K}(W[S]); (\alpha_j))\right] \leq \frac{12 \cdot 2^d}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j.$$

Observing that $f(F) = \mathsf{v}(F) \cdot \delta(F, W^\circ; (\alpha_j))/\sum_{j=1}^{d} \alpha_j$ is a reasonably smooth parameter for any (nonzero) finite, nonnegative sequences $\alpha_1, \ldots, \alpha_d$, Theorem 24 implies

$$\delta_\square(W^\circ, \mathbb{K}(W[S]); (\alpha_j)) \leq \frac{12 \cdot 2^d + 1}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j$$

with probability at least $1 - \exp(-n/(2\log_2 n))$. Since $\mathbb{K}(W[S])$ is identically distributed to $\mathbb{K}(n, W)$, this yields the desired bound. ∎

Appendix B. Proof of Lemma 9

Given our interest in cut-distances that only consider the first d dimensions, we denote by $\mathbb{K}_d(n, W)$ the d-skeleton of the random simplicial complex $\mathbb{K}(n, W)$. Equivalently, if we define W_d as the complexon that is equal to W in all dimensions less than or equal to d, and 0 otherwise, $\mathbb{K}_d(n, W)$ is equal to $\mathbb{K}(n, W_d)$ in distribution.

We state a weaker version of Lemma 31 in terms of total variation distances first, which we then strengthen to be in terms of homomorphism densities.

Lemma 31. *Let U, W be two complexons, and suppose that for some $d \geq 1, n \geq 2$, we have*

$$\mathrm{d_{var}}(\mathbb{K}_d(n, U), \mathbb{K}_d(n, W)) < 1 - 2\exp\left(-\frac{n}{2\log_2 n}\right),$$

where $\mathrm{d_{var}}$ denotes the total variation distance between probability distributions. Then, for all finite nonnegative sequences $\alpha_1, \ldots, \alpha_d$,

$$\delta_\square(U^\circ, W^\circ; (\alpha_j)) \leq \frac{24 \cdot 2^d + 2}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j.$$

Proof We prove this bound via the probabilistic method. By assumption, we can couple $\mathbb{K}_d(n, U)$ and $\mathbb{K}_d(n, W)$ so that with probability at least $2\exp(-n/(2\log_2 n))$,

$$\mathbb{K}_d(n, U) = \mathbb{K}_d(n, W). \tag{7}$$

By Lemma 7, we have with probability at least $1 - \exp(-n/(2\log_2 n))$,

$$\delta_\square(U^\circ, \mathbb{K}_d(n, U); (\alpha_j)) \leq \frac{12 \cdot 2^d + 1}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j. \tag{8}$$

Similarly, with probability at least $1 - \exp(-n/(2\log_2 n))$,

$$\delta_\square(W^\circ, \mathbb{K}_d(n, W); (\alpha_j)) \leq \frac{12 \cdot 2^d + 1}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j. \tag{9}$$

With positive probability, then, (7) to (9) hold simultaneously, which implies that

$$
\begin{aligned}
\delta_\square(U^\circ, W^\circ; (\alpha_j)) &\leq \delta_\square(U^\circ, \mathbb{K}_d(n, U); (\alpha_j)) \\
&\quad + \delta_\square(\mathbb{K}_d(n, U), \mathbb{K}_d(n, W); (\alpha_j)) \\
&\quad + \delta_\square(\mathbb{K}_d(n, W), W^\circ; (\alpha_j)) \\
&\leq \frac{24 \cdot 2^d + 2}{\sqrt{\log_2 n}} \sum_{j=1}^{d} \alpha_j,
\end{aligned}
$$

as desired. ∎

The inverse counting lemma follows, by showing that closeness in homomorphism densities implies the condition of Lemma 31.

Proof [Proof of Lemma 9] For convenience, put $f(n, d) = 0.999 \cdot 2^{-((1+n)^{d+2})}$. Let F be a finite, simplicial complex on n nodes of dimension d. By the inclusion-exclusion principle, the induced homomorphism density can be written as

$$t_{\mathsf{ind}}(F, W_d) = \sum_{G \subseteq F_\circ : \dim G \leq d} (-1)^{|G|} t(F \cup G, W),$$

with a similar expression holding for the complexon U. Since each $F \cup G$ yields a simplicial complex of dimension d, we have $|t(F \cup G, U) - t(F \cup G, W)| \leq f(n, d)$ for all such F, G, by assumption. Observe that for any simplicial complex F on n nodes of dimension d, the number of antifacets of F of dimension at most d is bounded by $\binom{n}{d+1}$. This implies that

$$|t_{\mathsf{ind}}(F, U_d) - t_{\mathsf{ind}}(F, W_d)| \leq 2^{\binom{n}{d+1}} \cdot f(n, d).$$

Then, by Lemma 2, for any such F,

$$|\mathbb{P}\{\mathbb{K}_d(n, U) = F\} - \mathbb{P}\{\mathbb{K}_d(n, W) = F\}| \leq 2^{\binom{n}{d+1}} \cdot f(n, d),$$

keeping in mind that there exists at least one F such that strict inequality holds. Therefore, assuming n is sufficiently large so that the bound on $\delta_\square(U^\circ, W^\circ; (\alpha_j))$ is nontrivial (that is,

less than $\sum_j \alpha_j$), summing over all simplicial complexes F on $[n]$ of dimension d yields the bound

$$
\begin{aligned}
d_{\text{var}}(\mathbb{K}_d(n, U), \mathbb{K}_d(n, W)) &= \sum_F |\mathbb{P}\{\mathbb{K}(n, U) = F\} - \mathbb{P}\{\mathbb{K}(n, W) = F\}| \\
&< 2^{\sum_{j=1}^{d+1} \binom{n}{j}} 2^{\binom{n}{d+1}} \cdot f(n, d) \\
&\leq 1 - 2\exp\left(-\frac{n}{2\log_2 n}\right).
\end{aligned}
$$

The lemma follows from Lemma 31. ∎

Appendix C. Proofs from Section 4.4

Proof [Proof of Lemma 11] Let $\epsilon > 0$ and $U \in \mathcal{W}$ be given. Let $B_\epsilon(U; (\alpha_j))$ be the open ball centered at U of radius ϵ with respect to the pseudometric $\delta_\square(\cdot, \cdot; (\alpha_j))$. Similarly, let $B_\epsilon(U; (\beta_j))$ be defined in the same way using the sequence (β_j) to define the pseudometric.

By assumption, there exists an M such that

$$
\sum_{j>M} \alpha_j < \epsilon/2
$$

$$
\sum_{j>M} \beta_j < \epsilon/2.
$$

Put

$$
\epsilon' = \frac{\epsilon \cdot \min_{j \leq M} \beta_j}{2M \cdot \max_{j \leq M} \alpha_j},
$$

noting that $\epsilon' > 0$, since both sequences are strictly positive.

Let $W \in B_{\epsilon'}(U; (\beta_j))$ be given. Then, there exists a measure-preserving bijection $\phi : [0, 1] \to [0, 1]$ such that

$$
d_\square(U, W^\phi; (\beta_j)_{j \geq 1}) < \epsilon'.
$$

This implies that for all $j \leq M$,

$$
d_{\square, j}(U, W^\phi) \leq \frac{\epsilon}{2M\alpha_j}.
$$

One can then check that this implies $d_\square(U, W^\phi; (\alpha_j)_{j \geq 1}) < \epsilon$, so that $B_{\epsilon'}(U; (\beta_j)) \subseteq B_\epsilon(U; (\alpha_j))$. Applying the same argument symmetrically and for arbitrary values of ϵ yields the fact that the topologies on $\mathcal{W}$ induced by both metrics are equal, as desired. ∎

We use a similar technique to prove Proposition 12.
Proof [Proof of Proposition 12] We show that every complexon $W \in \mathcal{W}$ arises as the limit of a sequence in $\mathcal{W}_0$ in the canonical topology. By Lemma 11, we can choose an arbitrary strictly positive, summable sequence $(\alpha_j)_{j \geq 1}$ with which to define the pseudo-metric $\delta_\square(\cdot, \cdot; (\alpha_j))$. Let $\epsilon > 0$ be given, and consider the sequence of random simplicial

complexes $(\mathbb{K}(n, W))_{n \geq 1}$. Let d be such that $\sum_{j>d} \alpha_j < \epsilon/2$. By Corollary 8, the sequence $\mathbb{K}(n, W)$ converges to W° with respect to the metric $\delta_\square(\cdot, \cdot; (\alpha_j)_{j \leq d})$ with probability 1. Therefore, there is some m_0 such that for all $m > m_0$, it holds (with probability 1) that $\delta_\square(\mathbb{K}(n, W), W^\circ; (\alpha_j)_{j \leq d}) < \epsilon/2$. Thus, we also have for all $m > m_0$ that $\delta_\square(\mathbb{K}(n, W), W^\circ; (\alpha_j)_{j \geq 1}) < \epsilon$. Observe that $W_{\mathbb{K}(n, W)} = W^\circ_{\mathbb{K}(n, W)}$. Since ϵ was given arbitrarily, this establishes the convergence of $\mathbb{K}(n, W)$ to W in the canonical topology. Of course, by Lemma 5, each $\mathbb{K}(n, W)$ can be equivalently represented by an element of $\mathcal{W}_0$, establishing the desired result. ∎

Before proceeding with the proof of Theorem 13, we state the following useful result, which is a corollary of Theorem 20.

Corollary 32. *Let W be a complexon, and let $1 \leq m < n, d \geq 1$. For every m-partition $\mathcal{Q}$ of $[0, 1]$, there is an n-partition $\mathcal{P}$ refining $\mathcal{Q}$ such that, for all $0 \leq j \leq d$,*

$$\mathrm{d}_{\square, j}(W, W_{\mathcal{P}}) \leq \sqrt{\frac{d+1}{\log_2 n/m}}.$$

Lemma 33. *The space $\mathcal{W}$ with the δ-topology is compact.*

Proof By Lemma 11, the δ-topology on $\mathcal{W}$ is the same for all strictly positive, summable sequences $(\alpha_j)_{j \geq 1}$. Thus, we can assume without loss of generality that $\alpha_j = 2^{-j}$ for all $j \geq 1$, so that $\sum_{j \geq 1} \alpha_j = 1$.

To prove compactness, we show that every sequence $W_1, W_2, \ldots \in \mathcal{W}$ has a convergent subsequence. Following the proof of (Lovász and Szegedy, 2007, Theorem 5.1), for every $n \geq 1$, we can construct via Theorem 20 partitions $\mathcal{P}_{n,k}$ of $[0, 1]$ for $k \geq 1$, with corresponding stepfunctions $W_{n,k} = (W_n)_{\mathcal{P}_{n,k}}$ that satisfy

(i) $\mathrm{d}_\square(W_n - W_{n,k}; (\alpha_j)) \leq 1/k$

(ii) $\mathcal{P}_{n,k}$ is refined by $\mathcal{P}_{n,k+1}$ for all n, k

(iii) $|\mathcal{P}_{n,k}| = m_k$ depends only on k.

To do so, assume that for some $k \geq 0$, $\mathcal{P}_{n,k}$ is an m_k-partition of $[0, 1]$ satisfying conditions (i) and (iii). Put

$$\epsilon = \frac{1}{2(k+1)}$$
$$M_\epsilon = \lceil -\log_2 \epsilon + 1 \rceil$$
$$m_{k+1} = m_k \cdot 2^{4(k+1)^2(M_\epsilon + 1)}.$$

Applying Corollary 32 with $d = M_\epsilon$, there exists an m_{k+1}-partition $\mathcal{P}_{n,k+1}$ satisfying (i) and (ii). Moreover, m_{k+1} does not depend on the complexon W_n, so that $\mathcal{P}_{n,k+1}$ also satisfies (iii).

For the base case, let $\mathcal{P}_{n,1}$ be the trivial (indiscrete) partition, which satisfies conditions (i) and (iii) with $m_k = 1$. By induction, then, there exists a sequence of partitions $\{\mathcal{P}_{n,k}\}_{k \geq 1}$ satisfying conditions (i)–(iii).

Following the approach of (Lovász and Szegedy, 2007, Theorem 5.1), we can rearrange the points of $[0,1]$ for each n by a measure-preserving bijection so that every partition in every $\mathcal{P}_{n,k}$ is an interval, while preserving the properties (i)–(iii). Having done this, we replace the sequence $(W_n)_{n\geq 1}$ by a subsequence so that for every k, the sequence $W_{n,k}$ converges almost everywhere to a stepfunction U_k with m_k steps as $n \to \infty$.

To do so, select a subsequence of W_n such that the length of the ith interval in $P_{n,1}$ converges for all $i \in [m_1]$. Then, take a further subsequence such that the value of $W_{n,1}$ converges on the product of the i_1^{th} and i_2^{th} interval, for $i_1, i_2 \in [m_1]$. Repeat this for the products of the intervals indexed by $i_1, \ldots, i_{d+1} \in [m_1]$ for all $d \geq 1$. It follows that $W_{n,1}$ converges almost everywhere to a limit stepfunction U_1 with m_1 steps.

Repeat this procedure for all $k > 1$, taking further and further subsequences so that for all k, $W_{n,k} \to U_k$ almost everywhere, where each U_k is a stepfunction on m_k steps. This yields the desired subsequence.

The proof proceeds identically to that of (Lovász and Szegedy, 2007, Theorem 5.1). Let $\mathcal{P}_k$ denote the partition of $[0,1]$ into the steps of U_k. By the condition (ii) of the array $W_{n,k}$, one can show that for every $k < l$, the partition $\mathcal{P}_{n,l}$ refines $\mathcal{P}_{n,k}$, and furthermore that $\mathcal{P}_l$ refines $\mathcal{P}_k$. Moreover, it holds that

$$U_k = (U_l)_{\mathcal{P}_k}. \tag{10}$$

We now seek to establish the almost everywhere convergence of the sequence U_k. Define a function $g : \bigsqcup_{d\geq 1}[0,1]^{d+1} \to [0,1]$ so that $g([0,1]^{d+1}) = \alpha_d$, for all $d \geq 1$. Define a probability measure μ on $\bigsqcup_{d\geq 1}[0,1]^{d+1}$ such that, for any Borel set $\mathcal{A} \subseteq \bigsqcup_{d\geq 1}[0,1]^{d+1}$,

$$\mu(\mathcal{A}) = \int_{\mathcal{A}} g \mathrm{d}\lambda,$$

where λ is the Lebesgue measure. Let X be a random point in $\bigsqcup_{d\geq 1}[0,1]^{d+1}$ chosen according to the measure μ. Then, the equation (10) implies that the sequence $U_1(X), U_2(X), \ldots$ is a martingale with bounded terms. By the martingale convergence theorem (see, for instance, Williams, 1991), this sequence is convergent with probability 1. That is to say, the sequence of functions $U_1, U_2, \ldots$ converges almost everywhere, with respect to the measure μ. Since μ is absolutely continuous with respect to λ, this implies convergence almost everywhere with respect to the Lebesgue measure. Denote the pointwise limit of this sequence by U. By the dominated convergence theorem, we have

$$\lim_{k\to\infty} \sum_{j=1}^{\infty} \alpha_j \int_{[0,1]^{j+1}} |U(x) - U_k(x)| \mathrm{d}x = \lim_{k\to\infty} \int |U - U_k| \mathrm{d}\mu = 0,$$

that is, $U_k \xrightarrow{L^1} U$ with respect to the probability measure μ.

We now show that the subsequence W_n converges to U in the cut-distance. Let $\epsilon > 0$ be given arbitrarily. Then, there is some $k > 3/\epsilon$ such that

$$\sum_{j=1}^{\infty} \alpha_j \int_{[0,1]^{j+1}} |U(x) - U_k(x)| \mathrm{d}x = \int |U - U_k| \mathrm{d}\mu < \epsilon/3.$$

For this k, there is an n_0 such that for all $n > n_0$,

$$\sum_{j=1}^{\infty} \alpha_j \int_{[0,1]^{j+1}} |U_k(x) - W_{n,k}(x)| \mathrm{d}x = \int |U_k - W_{n,k}| \mathrm{d}\mu < \epsilon/3,$$

again by dominated convergence. Then, for all $n > n_0$,

$$
\begin{aligned}
\delta_\square(U, W_n; (\alpha_j)_{j \geq 1}) \leq\ & \delta_\square(U, U_k; (\alpha_j)) \\
& + \delta_\square(U_k, W_{n,k}; (\alpha_j)) \\
& + \delta_\square(W_{n,k}, W_n; (\alpha_j)) \\
\leq\ & \int |U - U_k| \mathrm{d}\mu \\
& + \int |U_k - W_{n,k}| \mathrm{d}\mu \\
& + \delta_\square(W_{n,k}, W_n; (\alpha_j)) \\
\leq\ & \epsilon,
\end{aligned}
$$

thus proving convergence of the subsequence, as desired. ∎

We define the map $(\cdot)^\circ : \mathcal{W} \to \mathcal{W}$ that sends any $W \in \mathcal{W}$ to W°, that is, the faceting map.

Lemma 34. *The faceting map $(\cdot)^\circ : \mathcal{W} \to \mathcal{W}$ is continuous with respect to the δ-topology on $\mathcal{W}$.*

Proof By Lemma 11, we can consider an arbitrary strictly positive summable sequence $(\alpha_j)_{j \geq 1}$, so that the cut-distance $\delta_\square(\cdot, \cdot; (\alpha_j)_{j \geq 1})$ induces the δ-topology on $\mathcal{W}$. Let $W_1, W_2, \ldots$ be a convergent sequence $\mathcal{W}$ with respect to the cut-metric, with limiting complexon W. By Lemma 6, this implies that for all $n \geq 1$, the sequence of distributions $\mathbb{K}(n, W_m)$ converges to $\mathbb{K}(n, W)$ in the total variation distance as $m \to \infty$.

We now show that $W_1^\circ, W_2^\circ, \ldots \to W^\circ$ in the cut-metric. Let $\epsilon > 0$ be given arbitrarily. Let M_ϵ be an integer such that $\sum_{j > M_\epsilon} \alpha_j < \epsilon/2$, and put

$$n = \left\lceil 2^{\left(8 \cdot 5^{M_\epsilon + 1}/\epsilon\right)^2} \right\rceil.$$

By Lemma 7, we have that for any complexon W,

$$\delta_\square(W^\circ, \mathbb{K}(n, W); (\alpha_j)_{j \geq 1}) < \frac{\epsilon}{2}$$

with probability at least $1 - \exp(-n/(2 \log_2 n))$. Since the sequence $\mathbb{K}(n, W_m) \to \mathbb{K}(n, W)$ converges in the total variation distance as $m \to \infty$, this implies that for some sufficiently large m_0, for each $m > m_0$, there is some simplicial complex F such that simultaneously

$$
\begin{aligned}
\delta_\square(W_m^\circ, F; (\alpha_j)_{j \geq 1}) &< \frac{\epsilon}{2} \\
\delta_\square(W^\circ, F; (\alpha_j)_{j \geq 1}) &< \frac{\epsilon}{2}.
\end{aligned}
$$

37

By the triangle inequality, this implies $\delta_\square(W_m^\circ, W^\circ; (\alpha_j)_{j\geq 1}) < \epsilon$ for all $m > m_0$. Thus, $W_1^\circ, W_2^\circ, \ldots \to W^\circ$ in the cut-distance, so that the map $(\cdot)^\circ$ is continuous, as desired. ∎

Defining $\mathcal{W}^\circ$ as the image of $\mathcal{W}$ under $(\cdot)^\circ$, Lemmas 33 and 34 imply that $\mathcal{W}^\circ$ is a compact subspace of $\mathcal{W}$ with respect to the δ-topology. The compactness of the space $\mathcal{W}$ with the canonical topology follows.

Proof [Proof of Theorem 13] From the pseudometric space $(\mathcal{W}, \delta_\square)$, where $\delta_\square$ is defined in terms of an arbitrary strictly positive, summable sequence, take the metric identification $(\overline{\mathcal{W}}, \overline{\delta_\square})$, and define the δ-topology on $\overline{\mathcal{W}}$ be the topology induced by the metric $\overline{\delta_\square}$. Since the metric identification preserves compactness $\overline{\mathcal{W}}$ is a compact space under the δ-topology, by Lemma 33. Moreover, taking $\overline{\mathcal{W}^\circ}$ to be the image of $\overline{\mathcal{W}}$ under the map $(\cdot)^\circ$, we have that $(\overline{\mathcal{W}^\circ}, \overline{\delta_\square})$ is a compact metric space by Lemma 34. In other words, $(\cdot)^\circ : \overline{\mathcal{W}} \to \overline{\mathcal{W}^\circ}$ is a quotient map.

It is clear that the canonical topology on $\mathcal{W}$ is equal to the quotient topology by the map $(\cdot)^\circ : \overline{\mathcal{W}} \to \overline{\mathcal{W}^\circ}$. Since the space $\overline{\mathcal{W}}$ with the quotient topology is compact, the canonical topology is compact, as desired. ∎

Appendix D. Complexons on Borel Probability Spaces

We consider here how to transform complexons on Borel probability spaces to complexons on $[0, 1]$ in a way that preserves reasonable notions of sampling and homomorphism density. One can extend the following discussion to more general spaces as discussed in (Lovász, 2012, Section 13.1,Appendix A.3), but we restrict our attention to Borel probability spaces for convenience.

A probability space $(\Omega, \mathcal{F}, \mu)$ is said to be a Borel probability space if for some sequence of constants $\{m_n\}_{n\geq 0}$, it is isomorphic (up to sets of measure zero) to a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu})$ such that

1. $\tilde{\Omega} = [0, m_0] \cup \bigcup_{n\geq 1}\{p_n\}$, where each $\{p_n\}$ is considered as an atom

2. $\tilde{\mu}$ is the Lebesgue measure on the Borel sets of the closed interval $[0, m_0]$, and for each $n \geq 1$ we have $\tilde{\mu}(p_n) = m_n$

3. $\tilde{\mathcal{F}}$ is the Borel σ-field on $\tilde{\Omega}$.

For instance, any probability measure on the Borel σ-field of a separable complete metric space yields a Borel probability space.

With this definition, we now outline a procedure, as described by (Lovász, 2012, Section 13.1), for describing a complexon on a Borel probability space by a complexon on $[0, 1]$. Let a Borel probability space $(\Omega, \mathcal{F}, \mu)$ be given with corresponding constants $\{m_n\}_{n\geq 0}$. Let $\phi : \tilde{\Omega} \to \Omega$ be the bijective isomorphism guaranteed by the definition of a Borel probability space. For a complexon W_Ω on Ω, define a complexon W on $[0, 1]$ in the following way. For each $n \geq 0$, define the interval

$$I_n = [\sum_{\ell < n} m_\ell, \sum_{\ell \leq n} m_\ell),$$

so that $I_0 = [0, m_0)$, $\bigcup_{n \geq 0} I_n = [0, 1)$, and each I_n has Lebesgue measure m_n.

For any $x \in [0, 1)$, define the corresponding $\tilde{x} \in \tilde{\Omega}$ so that

$$\tilde{x} = \begin{cases} x & x \in I_0 \\ p_n & x \in I_n. \end{cases}$$

Then, define the complexon W so that for any points $x_1, \ldots, x_{d+1} \in [0, 1)$

$$W(x_1, \ldots, x_{d+1}) = W_\Omega \left(\phi(\tilde{x}_1), \ldots, \phi(\tilde{x}_{d+1}) \right).$$

Since we characterize complexons in a way that does not depend on sets of measure zero, W can be trivially extended to points in $[0, 1]$.

To justify this definition, we will show that this preserves a reasonable notion of homomorphism density. Let F be a simplicial complex. The (induced) homomorphism density of F in W_Ω is defined as (see (1))

$$t(F, W_\Omega) = \int_{\Omega^{\mathsf{V}(F)}} \prod_{\sigma \in F} W_\Omega(x_\sigma) \mathrm{d}\mu^{\mathsf{V}(F)}(x)$$

$$t_{\mathsf{ind}}(F, W_\Omega) = \int_{\Omega^{\mathsf{V}(F)}} \prod_{\sigma \in F} W_\Omega(x_\sigma) \prod_{\tau \in F_\circ} (1 - W_\Omega(x_\tau)) \, \mathrm{d}\mu^{\mathsf{V}(F)}(x),$$

where $\mu^{\mathsf{V}(F)}$ indicates the usual product measure on $\Omega^{\mathsf{V}(F)}$. By simple unrolling of definitions, one can check that

$$t(F, W_\Omega) = t(F, W)$$
$$t_{\mathsf{ind}}(F, W_\Omega) = t_{\mathsf{ind}}(F, W),$$

for all simplicial complexes F. Thus, it is justifiable to use W as a representative of the complexon W_Ω, particularly for the purposes of modeling sampling random simplicial complexes from W_Ω. We refer to (Lovász, 2012, Section 13.1) for a further discussion on how this can be extended to a broader class of probability spaces, as well as some typical conditions under which a probability space is guaranteed to be Borel.

References

Robert J Adler, Omer Bobrowski, and Shmuel Weinberger. Crackle: The homology of noise. *Discrete & Computational Geometry*, 52(4):680–704, 2014.

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Noga Alon, W Fernandez De La Vega, Ravi Kannan, and Marek Karpinski. Random sampling and approximation of MAX-CSPs. *Journal of Computer and System Sciences*, 67(2):212–243, 2003.

Krishnakumar Balasubramanian. Nonparametric modeling of higher-order interactions via hypergraphons. *The Journal of Machine Learning Research*, 22(1):6464–6498, 2021.

Federico Battiston and Giovanni Petri. *Higher-Order Systems*. Springer, 2022.

Omer Bobrowski and Dmitri Krioukov. *Random Simplicial Complexes: Models and Phenomena*, pages 59–96. Springer, 2022. Referenced quotations come from the preprint version `https://arxiv.org/abs/2105.12914v1`.

Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Annals of Mathematics*, pages 151–219, 2012.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46 (2):255–308, 2009.

Frédéric Chazal, David Cohen-Steiner, Leonidas J Guibas, Facundo Mémoli, and Steve Y Oudot. Gromov-Hausdorff stable signatures for shapes using persistence. In *Computer Graphics Forum*, volume 28, pages 1393–1403. Wiley Online Library, 2009.

Frédéric Chazal, Marc Glisse, Catherine Labruère, and Bertrand Michel. Convergence rates for persistence diagram estimation in topological data analysis. In *International Conference on Machine Learning*, pages 163–171. PMLR, 2014.

Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Subsampling methods for persistent homology. In *International Conference on Machine Learning*, pages 2143–2151. PMLR, 2015.

Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance to a measure and kernel distance. *The Journal of Machine Learning Research*, 18(1):5845–5884, 2017.

Fabio Coppini. A note on Fokker–Planck equations and graphons. *Journal of Statistical Physics*, 187(2):15, 2022.

A Costa and M Farber. Large random simplicial complexes, I. *Journal of Topology and Analysis*, 8(03):399–429, 2016.

A Costa and M Farber. Large random simplicial complexes, II; the fundamental group. *Journal of Topology and Analysis*, 9(03):441–483, 2017a.

A Costa and M Farber. Large random simplicial complexes, III the critical dimension. *Journal of Knot Theory and Its Ramifications*, 26(02):1740010, 2017b.

Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*, 2007.

Gábor Elek and Balázs Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Advances in Mathematics*, 231(3-4):1731–1772, 2012.

Michael Farber and Lewis Mead. Random simplicial complexes in the medial regime. *Topology and its Applications*, 272:107065, 2020.

Michael Farber, Lewis Mead, and Lewin Strauss. The Rado simplicial complex. *Journal of Applied and Computational Topology*, 5(2):339–356, 2021.

Michael Farber, Lewis Mead, and Tahl Nowik. Random simplicial complexes, duality and the critical dimension. *Journal of Topology and Analysis*, 14(01):1–31, 2022.

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

W Timothy Gowers. Quasirandomness, counting and regularity for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):143–184, 2006.

W Timothy Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007.

Jan Hladkỳ, András Máthé, Viresh Patel, and Oleg Pikhurko. Poset limits can be totally ordered. *Transactions of the American Mathematical Society*, pages 4319–4337, 2015.

Douglas N Hoover. Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2:275, 1979.

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

Svante Janson. Graphons, cut norm and distance, couplings and rearrangements. *NYJM Monographs*, 4, 2013.

Matthew Kahle. Topology of random clique complexes. *Discrete mathematics*, 309(6):1658–1671, 2009.

Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

László Lovász. Graph homomorphisms: Open problems. 2008. URL `http://www.cs.elte.hu/lovasz/problems.pdf`.

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

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

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

Sohir Maskey, Ron Levie, and Gitta Kutyniok. Transferability of graph neural networks: an extended graphon approach. *Applied and Computational Harmonic Analysis*, 63:48–83, 2023.

Roy Meshulam and Nathan Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures & Algorithms*, 34(3):408–417, 2009.

Madeline Navarro and Santiago Segarra. Joint network topology inference via a shared graphon model. *IEEE Transactions on Signal Processing*, 70:5549–5563, 2022.

Takashi Owada and Omer Bobrowski. Convergence of persistence diagrams for topological crackle. *Bernoulli*, 26(3):2275–2310, 2020.

T Mitchell Roddenberry, Madeline Navarro, and Santiago Segarra. Network topology inference with graphon spectral penalties. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5390–5394. IEEE, 2021.

Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures & Algorithms*, 25(1):1–42, 2004.

Luana Ruiz, Luiz Chamon, and Alejandro Ribeiro. Graphon neural networks and the transferability of graph neural networks. *Advances in Neural Information Processing Systems*, 33:1702–1712, 2020.

Michael T Schaub, Yu Zhu, Jean-Baptiste Seby, T Mitchell Roddenberry, and Santiago Segarra. Signal processing on higher-order networks: Livin'on the edge... and beyond. *Signal Processing*, 187:108–149, 2021.

Benedikt Stufler. Graphon convergence of random cographs. *Random Structures & Algorithms*, 59(3):464–491, 2021.

Terence Tao. A variant of the hypergraph removal lemma. *Journal of combinatorial theory, Series A*, 113(7):1257–1280, 2006.

David Williams. *Probability with martingales*. Cambridge university press, 1991.

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