RESEARCH ARTICLE

WILEY

Enumeration and randomized constructions of hypertrees

Nati Linial | Yuval Peled

Department of Computer Science, Hebrew University, Jerusalem, Israel

Correspondence
Yuval Peled, School of Computer Science and Engineering, The Hebrew University, The Edmond J. Safra Campus - Givat Ram Jerusalem 9190401, Israel.
Email: yuvalp@cs.huji.ac.il

Abstract

Over 30 years ago, Kalai proved a beautiful d-dimensional analog of Cayley's formula for the number of n-vertex trees. He enumerated d-dimensional hypertrees weighted by the squared size of their $(d - 1)$-dimensional homology group. This, however, does not answer the more basic problem of unweighted enumeration of d-hypertrees, which is our concern here. Our main result, Theorem 1.4, significantly improves the lower bound for the number of d-hypertrees. In addition, we study a random 1-out model of d-complexes where every $(d - 1)$-dimensional face selects a random d-face containing it, and show that it has a negligible d-dimensional homology.

KEYWORDS

Homology, Hypertrees, Simplicial Complexes, The Probabilistic Method

1 | INTRODUCTION

Trees are among the most fundamental objects in discrete mathematics and computer science, as documented in innumerable theoretical and applied studies. As part of our ongoing research in *high-dimensional combinatorics*, we study here *high-dimensional trees*. In graph theory, a *tree* is characterized by being *connected* and *acyclic*. Since both these properties are topological in nature, it makes sense to consider them in higher dimensional simplicial complexes as well. This was indeed done over 30 years ago in a beautiful paper by Kalai [13].

From a topological perspective, a graph is a one-dimensional simplicial complex. Also, connectivity and cycles in graphs are expressible in the language of *simplicial homology*. Namely, connectivity is the vanishing of the zeroth homology and cycles are elements of the graph's first homology. Kalai's

definition applies the same line of thought to d-dimensional complexes regarding the $(d - 1)$st and dth homology:

Definition 1.1 A d-hypertree T is a d-dimensional simplicial complex with a full $(d - 1)$-skeleton such that both $H_{d-1}(T; \mathbb{Q}) = 0$ and $H_d(T; \mathbb{Q}) = 0$.

Recall that a d-dimensional simplicial complex (d-complex in short) has a full $(d - 1)$-skeleton if it contains all the faces of dimension less than d that are spanned by its vertex set. In this paper, unless stated otherwise, all d-complexes are assumed to have a full $(d-1)$-skeleton, and we sometimes identify a d-complex with its set of d-dimensional faces. Note also that it makes sense to consider a similar notion of a hypertree where $\mathbb{Q}$ is replaced by a different commutative ring of coefficients. However, this would yield a different class of complexes, and unless stated otherwise, we stick to $\mathbb{Q}$-acyclic complexes, that is, d-hypertrees over $\mathbb{Q}$.

The notion of a d-hypertree is expressible as well in terms of elementary linear algebra, and specifically the *boundary operator* (or matrix) ∂_d that maps a d-face to the linear sum of its $(d - 1)$-subfaces. A d-hypertree is a set of d-faces whose corresponding columns in ∂_d form a basis for the column space of ∂_d. For example, the d-dimensional star, which is comprised of the d-faces that contain a specific vertex, is a d-hypertree. Clearly, an n-vertex d-hypertree has exactly $\binom{n-1}{d}$ d-dimensional faces.

One of the earliest nontrivial discoveries about trees is Cayley's formula, which states that the number of trees on n labeled vertices is n^{n-2}. Kalai found a beautiful generalization of this formula in higher dimensions. Let $\mathcal{T}_{n,d}$ denote the family of n-vertex d-hypertrees.

Theorem 1.2 *Let $d < n$ be integers. Then,*

$$\sum_{T \in \mathcal{T}_{n,d}} |H_{d-1}(T; \mathbb{Z})|^2 = n^{\binom{n-2}{d}}.$$

Recall that $H_{d-1}(T; \mathbb{Z})$, the *torsion* of T, is a finite group for every d-hypertree T. For $d = 1$, Kalai's formula reduces to Cayley's formula, since the integral zeroth homology is always torsion-free. In fact, Kalai's argument is a high-dimensional extension of the Matrix-Tree Theorem.

In the one-dimensional case of graphs, trees can also be defined by the combinatorial notion of collapsibility. An elementary 1-collapse in a graph is the removal of a leaf (= a vertex of degree one) and the unique edge that contains it. An n-vertex graph with $n - 1$ edges is a tree if and only if it is collapsible, that is, it can be reduced to a single vertex by a sequence of elementary 1-collapses.

This definition of trees generalizes to a notion of *d-collapsible d-hypertrees*. A $(d - 1)$-face τ in a simplicial complex X is said to be *exposed* if there is exactly one d-face σ in X that contains it. In the *elementary d-collapse* on τ, we remove τ and σ from X. Note that the remaining complex is homotopy equivalent to X. We say that X is *d-collapsible* if it is possible to eliminate all its d-faces by a series of elementary collapses. In particular, d-collapsibility implies the vanishing of the dth homology. Therefore, a d-collapsible d-complex with $\binom{n-1}{d}$ d-faces is a d-hypertree. Such a complex is called a *d-collapsible d-hypertree*.

While all one-dimensional trees are collapsible, it is conjectured that in higher dimension the situation is entirely different. We denote by $C_{n,d}$ the set of all n-vertex d-collapsible hypertrees.

Conjecture 1.3 *For every $d \geq 2$, asymptotically almost none of the d-hypertrees are d-collapsible. Namely, $|C_{n,d}|/|\mathcal{T}_{n,d}| \rightarrow 0$, as $n \rightarrow \infty$.*

Kalai's formula has motivated several other results of torsion-related weighted enumeration of hypertrees [3, 7, 16]. But, despite its remarkable beauty, the formula leaves a substantial gap regarding the question of *unweighted enumeration of (labeled) d-hypertrees*. Here are the bounds that are mentioned in [13]:

$$\left(\frac{n}{d+1}\right)^{\binom{n-1}{d}} < |\mathcal{T}_{n,d}| < \left(\frac{e}{d+1} \cdot n\right)^{\binom{n-1}{d}}.$$

As mentioned, the number of d-faces in a d-hypertree is $\binom{n-1}{d}$, and the upper bound only considers the number of ways to select them from among the total of $\binom{n}{d+1}$. The lower bound follows from the identity in Theorem 1.2 and an upper bound of the size of the torsion of a d-hypertree. Our analysis in Section 3 yields an elementary proof of this lower bound.

In fact, the following simple inductive construction, suggested to us by Gil Kalai, yields a better lower bound even for collapsible hypertrees. Let S be a $(d-1)$-collapsible hypertree with vertex set $[n]$, and let $vS = \{v\tau \mid \tau \in S\}$ be a simplicial cone over S, where v is a new vertex. Let T be a d-collapsible hypertree on $[n]$. The union $T \cup (vS)$ is a d-collapsible hypertree on $[n] \cup \{v\}$. Indeed, extend first the $(d-1)$-collapse of S to the cone vS and then collapse T.

Consequently, $|C_{n+1,d}| \geq |C_{n,d}||C_{n,d-1}|$, and we may proceed by induction on n and d to show that for every $d \geq 2$,

$$|C_{n,d}| \geq \left(\frac{n}{e^{\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{d}}}\right)^{\binom{n-2}{d}} \geq \left(\frac{e^{1-\gamma}}{d+1} \cdot n\right)^{\binom{n-2}{d}}, \tag{1}$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. For the induction step one needs to show that

$$\left(\frac{n}{e^{\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{d}}}\right)^{\binom{n-2}{d}} \left(\frac{n}{e^{\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{d-1}}}\right)^{\binom{n-2}{d-1}} \geq \left(\frac{n+1}{e^{\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{d}}}\right)^{\binom{n-1}{d}}.$$

To prove this step and to derive the right inequality in Equation (1) use the fact that $e > (1 + 1/t)^t$ for every positive integer t.

Our main theorem improves the lower bound on $|\mathcal{T}_{n,d}|$.

Theorem 1.4 *Let $t = t_d^*$ be the unique root in $(0, 1)$ of*

$$(d+1)(1-t) + (1+dt)\ln t = 0,$$

and

$$0 > \alpha_d = \frac{1}{d+1} \int_0^{t_d^*} \frac{(1-(1-y)^{d+1}) \cdot \log(1-(1-y)^{d+1}) \cdot (1-y+d \cdot y \log y)}{y(1-y)^{d+1}} dy.$$

Then,

$$|\mathcal{T}_{n,d}| \geq \left((1-o_n(1))\frac{e^{1+\alpha_d}}{d+1} \cdot n\right)^{\binom{n-1}{d}}.$$

Remark 1.5   1. Theorem 1.4 offers an exponential improvement over Equation (1) in every dimension. For example, for $d = 2$ the lower bound is improved from approximately $(0.606 \cdot n)^{n^2/2}$ to $(0.751 \cdot n)^{n^2/2}$.

2. In addition, note that $\alpha_d \to 0$ as d grows. Therefore, in contrast to the bound of Equation (1), this lower bound on $|\mathcal{T}_{n,d}|$ approaches Kalai's trivial *upper bound* of $(en/(d+1))^{\binom{n-1}{d}}$ as d grows.

It is interesting to speculate on whether Theorem 1.4 can help us prove Conjecture 1.3, at least for large d. Namely, could it be that $|\mathcal{C}_{n,d}|$ is even smaller than the lower bound of $|\mathcal{T}_{n,d}|$ in the theorem? In particular, it is conceivable that $|\mathcal{C}_{n,d}|$ does not approach the trivial upper bound as d grows. This discussion naturally suggests the following quantitative version of Conjecture 1.3.

Question 1.6 *Does it hold that $|\mathcal{C}_{n,d}|/|\mathcal{T}_{n,d}| = e^{-\Omega(n^d)}$?*

Put together, these two questions ask whether $|\mathcal{C}_{n,d}|/|\mathcal{T}_{n,d}|$ tends to zero and if so, whether the convergence is as fast as $e^{-\Omega(n^d)}$.

As we observe next, the upper bound of $|\mathcal{T}_{n,d}|$ can be slightly improved.

Theorem 1.7 *For every dimension $d \geq 2$ there exists $\varepsilon_d > 0$ such that $|\mathcal{T}_{n,d}| < \left(\frac{e-\varepsilon_d}{d+1}n\right)^{\binom{n-1}{d}}$.*

We now turn to discuss the model of *random 1-out d-dimensional simplicial complexes*, that generalizes the well-studied model of random 1-out graphs. It is reasonable to discuss such d-complexes here, since random 1-out graphs have been useful for sampling and enumerating spanning trees. For example, we mention Wilson's "cycle-popping" algorithm to uniformly sample spanning trees [20], and Frieze's [9] estimate of the number of spanning trees in almost-regular dense graphs. Random k-out graphs are also studied in several additional combinatorial contexts [6, 10, 18].

The random 1-out n-vertex d-dimensional simplicial complex $S_d(n, 1)$ is sampled as follows. We start with a full $(d-1)$-skeleton. Each $(d-1)$-dimensional face τ independently chooses a vertex $v \notin \tau$ to form the d-face $v\tau$. Here we do not allow multifaces, and every d-face that is chosen more than once is counted only once.

As is often the case with the study of large combinatorial objects, we have a rather limited supply of interesting d-acyclic (ie, having a trivial dth homology) complexes, and d-hypertrees in particular. The reason is that it is typically hard to analyze the boundary matrices of complexes that arise from combinatorial and probabilistic constructions. An obvious exception is the case where the complex is d-collapsible. In this case a simple purely combinatorial argument can show it to be d-acyclic. Notable noncollapsible examples are the sum complexes, introduced in [14], whose boundary operator is structured enough to be amenable to analysis. In [15] we used the theory of local weak convergence to bound the dimension of the d-homology of random Linial-Meshulam complexes. This approach can work only when the (bipartite) incidence graph of $(d-1)$-faces versus d-faces is locally a tree, in the sense of local weak convergence. Here we use similar techniques to show that $S_d(n, 1)$ typically has a small top dimensional homology.

Theorem 1.8 *For every $d \geq 2$, the dimension of the homology $H_d(S_d(n, 1); \mathbb{Q})$ is a.a.s. $o(n^d)$.*

We also show that almost all the d-cycles in $S_d(n, 1)$ can be eliminated by removing each d-face independently with probability ε, for an arbitrarily small $\varepsilon > 0$. We denote this random complex by $S_d(n, 1 - \varepsilon)$, and note the following alternative way to sample it: Initially, each $(d-1)$-face is made *active* independently with probability $1 - \varepsilon$. Each active $(d-1)$-face then selects a random vertex to form a d-face as in $S_d(n, 1)$.

It turns out that a random d-complex in the $S_d(n, 1-\varepsilon)$ model is almost d-acyclic, in the sense that the only d-cycles it has are $\partial\Delta_{d+1}$, that is, a boundary of $(d+1)$-simplex. Such d-cycles appear with positive probability that is bounded away from 1.

Theorem 1.9 *Fix an integer $d \geq 2$ and $\varepsilon > 0$, and let S be a random complex from $S_d(n, 1-\varepsilon)$. Then, a.a.s. $H_d(S; \mathbb{Q})$ is generated by $\partial\Delta_{d+1}$'s, the number of which is Poisson-distributed with a bounded parameter.*

It is interesting to investigate whether these results can yield nontrivial lower bounds for $|\mathcal{T}_{n,d}|$. In the graph-theoretic case, Frieze [9] showed that the number of spanning trees in a dense almost-regular n-vertex graph is, up to a factor of $(1 - o(1))^n$, at least as large as the number of choices made when generating a 1-out subgraph. The reason is that, with high probability, the random 1-out subgraph has only $O(\sqrt{n})$ cycles. This fact implies that many 1-out subgraphs S have the following two properties: (i) S can be transformed into a spanning tree by changing a small number of edges, and (ii) there are only few 1-out choices that yield the graph S. Indeed, the different 1-out choices that yield S are related to each other by reversing the orientations of its cycles. Frieze's assertion follows directly.

As it turns out, the higher dimensional counterpart of this argument is quite different. Property (i) remains valid in higher dimensions, since by Theorem 1.8 a typical 1-out d-complex can be turned into a d-hypertree by changing a small number of d-faces. However, property (ii) fails for $d \geq 2$. Indeed, the number of 1-out choices that yield a given d-complex S is not expressed by its homology and by evaluating determinants. Rather, it is determined by the *permanent* of the inclusion matrix of $(d-1)$-faces vs. d-faces of the complex. Unfortunately, the current upper bound we have for this permanent yields a worse lower bound for $|\mathcal{T}_{n,d}|$ than the one stated in Theorem 1.4. Whether or not sharper estimates on the permanent can yield improved bounds remains presently open.

The rest of the paper is organized as follows. Section 2 contains some general background in simplicial combinatorics and basic facts on d-hypertrees. In Section 3 we prove the theorems regarding the enumeration of d-hypertrees, and Section 4 is dedicated to the homology of the 1-out random d-complex. Finally, we present various open questions in Section 5.

2 | BACKGROUND

2.1 | Simplicial combinatorics

A *simplicial complex* is comprised of a *vertex set V* and a collection X of subsets of V that is closed under taking subsets. Namely, if $\sigma \in X$ and $\tau \subseteq \sigma$, then $\tau \in X$ as well. We usually refer to X as the simplicial complex and call its members *faces* or *simplices*. The *dimension* of the simplex $\sigma \in X$ is defined as $|\sigma| - 1$. A d-dimensional simplex is also called a d-simplex or a d-face for short. The dimension $\dim(X)$ of the complex X is defined as $\max \dim(\sigma)$ over all faces $\sigma \in X$. A d-dimensional simplicial complex is also referred to as a d-complex. The set of j-faces in X is denoted by $F_j(X)$. For $t < \dim(X)$, the *t-skeleton* of X is the simplicial complex that consists of all faces of dimension $\leq t$ in X, and X is said to have a *full t-dimensional skeleton* if it contains all the faces of dimension $\leq t$ that can be induced by its vertex set. In this paper we usually work with a d-complex that has a full $(d-1)$-skeleton.

For a face σ, the permutations on σ's vertices are split in two *orientations*, according to the permutation's sign. The *boundary operator* $\partial = \partial_d$ maps an oriented d-simplex $\sigma = (v_0, \ldots, v_d)$ to the formal sum $\sum_{i=0}^{d} (-1)^i(\sigma^i)$, where $\sigma^i = (v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_d)$ is an oriented $(d-1)$-simplex. We fix some commutative ring R and linearly extend the boundary operator to free R-sums of simplices.

We denote by $\partial_d(X)$ the d-dimensional boundary operator of a d-complex X. Over the reals, the upper $(d-1)$-dimensional *Laplacian* of X is $L = \partial_d(X)\partial_d(X)^*$.

When X is finite, we consider the $|F_{d-1}(X)| \times |F_d(X)|$ matrix form of ∂_d by choosing arbitrary orientations for $(d-1)$-simplices and d-simplices. Note that changing the orientation of a d-simplex (resp. $d-1$-simplex) results in multiplying the corresponding column (resp. row) by -1.

Let X be a d-dimensional simplicial complex. An element in the right kernel of the boundary operator $\partial_d(X)$ is called a d-*cycle*. Since X is d-dimensional, the dth homology group $H_d(X; R)$ (or vector space when R is a field) of a d-complex X equals to the space $\ker(\partial_d(X))$ of d-cycles. If $H_d(X; R)$ is trivial we say that X is d-*acyclic*.

An element in the (right) image of ∂_d is called a d-*boundary*. The $(d-1)$st homology group $H_{d-1}(X; R)$ is the quotient group $\ker(\partial_{d-1})/\mathrm{Im}(\partial_d)$. Namely, the quotient of the $(d-1)$-cycles (=kernel of $\partial_{d-1}(X)$) and the d-boundaries.

Let us restrict the discussion to an n-vertex d-complex X with full $(d-1)$-skeleton and the ring of rationals. Denote by $|X| = |F_d(X)|$ the number of top dimensional faces, $\beta_d(X)$ and $\beta_{d-1}(X)$ the dimensions of the corresponding homology groups (=Betti numbers), and $\mathrm{rank}(X)$ the rank of the operator $\partial_d = \partial_d(X)$. Note that the Betti numbers and the rank are equal when we work over the rationals or the reals. By the rank-nullity theorem, $\beta_d(X) + \mathrm{rank}(X) = |X|$. In addition, $\beta_{d-1}(X) + \mathrm{rank}(X) = \binom{n-1}{d}$, since the space of the $(d-1)$-cycles is $\binom{n-1}{d}$-dimensional (it is spanned by all the $(d-1)$-boundaries containing a specific vertex). These observations imply the following fact.

Fact 2.1 *Consider the three properties:*

1. $|X| = \binom{n-1}{d}$.
2. $\beta_d(X) = 0$.
3. $\beta_{d-1}(X) = 0$.

Then, any two properties imply the third. In such case, X is a d-hypertree.

Finally, the *homological shadow* $\mathrm{SH}(X)$, of a d-complex X, is the set of d-simplices σ such that $\partial\sigma$ is a d-boundary of X. In other words, the complement $\overline{\mathrm{SH}}(X)$ is the set of d-simplices σ for which $rank(X) < rank(X \cup \{\sigma\})$. The set $\overline{\mathrm{SH}}(X)$ plays a natural role in the construction of d-hypertrees, since it is comprised of those d-faces that can be added to the d-acyclic complex X without creating a d-cycle.

2.2 | Linial-Meshulam complexes

The Linial-Meshulam complex $Y_d(n, p)$ is a random n-vertex d-dimensional simplicial complex with a full $(d-1)$-skeleton where every d-face appears independently with probability p. The topological invariants and combinatorial properties of these complexes have been intensively studied in recent years. This includes their homology groups, homotopy groups, collapsibility, embeddability, and spectral properties. Here we use our previous paper [15] that concerns the phase transition of this random simplicial complex, the threshold probability for d-acyclicity, and the emergence of a giant shadow. We need to briefly recall the pertinent results. Let $t = t_d^*$ be the unique root in $(0, 1)$ of

$$(d+1)(1-t) + (1+d \cdot t)\ln t = 0,$$

and let

$$c_d^* := \frac{-\ln t_d^*}{(1 - t_d^*)^d}.$$

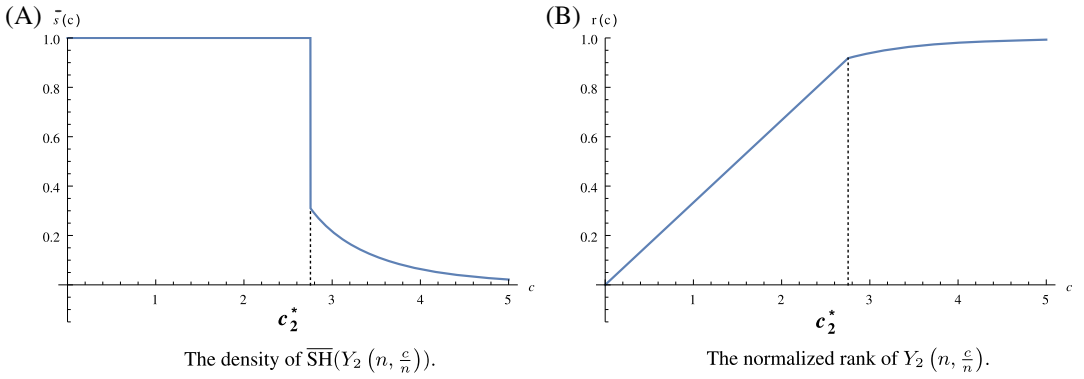**(A)** $\bar{s}(c)$

The density of $\overline{SH}(Y_2\left(n, \frac{c}{n}\right))$.

(B) $r(c)$

The normalized rank of $Y_2\left(n, \frac{c}{n}\right)$.

FIGURE 1 Plot of the functions $\bar{s}$ and r for dimension $d = 2$ [Colour figure can be viewed at wileyonlinelibrary.com]

Standard calculus arguments show that $t_d^* = (1 - o_d(1))e^{-(d+1)}$ as $d \to \infty$, hence $c_d^* = d + 1 - o_d(1)$.

For $c > c_d^*$, let t_c be the smallest positive root of $t = e^{-c(1-t)^d}$. Consider the functions $s, r : \mathbb{R}_{\geq 0} \to [0, 1]$,

$$\bar{s}(c) = \begin{cases} 1 & c \leq c_d^* \\ 1 - (1 - t_c)^{d+1} & c > c_d^* \end{cases}, \quad r(c) = \begin{cases} \frac{c}{d+1} & c \leq c_d^* \\ \frac{c}{d+1}(1 - (1 - t_c)^{d+1}) + (1 - t_c) - ct_c(1 - t_c)^d & c > c_d^* \end{cases}$$

The function r is strictly monotone, and we denote its inverse by r^{-1}.

Theorem 2.2 *Let* $Y = Y_d\left(n, \frac{c}{n}\right)$. *Then,*

$$\text{(I)} \quad \lim_{n \to \infty} \frac{1}{\binom{n-1}{d}} \mathbb{E}[\text{rank}(Y)] = r(c).$$

$$\text{(II)} \quad \textit{If } c \neq c_d^*, \lim_{n \to \infty} \frac{1}{\binom{n}{d+1}} \mathbb{E}[|\overline{SH}(Y)|] = \bar{s}(c).$$

In addition, for every $\varepsilon > 0$, *the probability that either the normalized rank or the density of the shadow's complement deviate from their expectation by more than* ε *tends to 0 as* $n \to \infty$.

Note that the functions $\bar{s}$ and r appear, with small variations in [15]. Namely, $1 - \bar{s}(c)$ is the density of the shadow of $Y_d\left(n, \frac{c}{n}\right)$, and $\frac{c}{d+1} - r(c)$ is its normalized d-dimensional Betti number, since the number of d-faces is $\approx c/(d + 1)\binom{n-1}{d}$.

2.3 | Local weak convergence and d-trees

We turn to describe the notion of local weak convergence of d-complexes. This concept is best described in the framework of rooted graphs. A rooted graph (G, r) is comprised of a graph G and a root vertex r. Two rooted graphs (G, r) and (G', r') are considered isomorphic, denoted by $(G, r) \cong (G', r')$, if there is a root-preserving isomorphism between them.

Associated with a d-dimensional complex X, is the bipartite inclusion graph $G(X)$ between X's set of $(d - 1)$-faces $F_{d-1}(X)$ and its d-faces $F_d(X)$. A rooted d-complex (X, o) is comprised of a d-complex X and a $(d - 1)$-face o that is marked as its root. The graph $G(X)$ of a rooted d-complex (X, o) is a

rooted graph. For every integer $k \geq 0$, we denote by $(G(X), o)_k$ the rooted subgraph of $G(X)$ that is induced by vertices of *distance* at most k from o in $G(X)$.

Let X_n be a sequence of random d-complexes and (X, o) be a random rooted d-complex. Formally speaking, X_n is a sequence of distributions on d-complexes and (X, o) is a distribution on rooted d-complexes. We say that (X, o) is the *local weak limit* of X_n if for every integer $k > 0$ and every rooted graph (G, r),

$$\Pr_{(X_n, o_n)} [(G(X_n), o_n)_k \cong (G, r)] \xrightarrow{n \to \infty} \Pr_{(X, o)} [(G(X), o)_k \cong (G, r)],$$

where the root o_n is sampled uniformly at random from $F_{d-1}(X_n)$.

We next define the concept of a *d-tree*. Do bear in mind that this is not to be confused with the notion of a d-hypertree. A d-tree (B, o) is a rooted d-complex that can be viewed as a (possibly infinite) d-dimensional branching process. Initially the complex consists of the $(d-1)$-face o. At every step $k \geq 0$, every $(d-1)$-face τ of distance $2k$ in the inclusion graph from o picks a nonnegative number $m = m_\tau$ of new vertices $v_1, \ldots, v_m$, and adds the d-faces $v_1 \tau, \ldots, v_m \tau$ to B.

We observe that the graph $G(B)$ is a *rooted tree*, with o as the root. Every vertex of odd depth (=distance from o) in $G(B)$ corresponds to a d-face and has exactly d children. Every vertex τ of even depth has m_τ children. In fact, every rooted tree in which every vertex of odd depth has precisely d children can be realized as an inclusion graph of a d-tree. Therefore, for every $(d-1)$-face $\tau \in B$ we refer to the *d-subtree rooted at τ* as the rooted complex that contains τ and all its descendant faces. In particular, if the root o is contained in m d-faces $\sigma_1, \ldots, \sigma_m$, we denote by $B_{j,i}$ the d-subtree rooted at the ith $(d-1)$-face of σ_j, for every $1 \leq j \leq m$ and $1 \leq i \leq d$.

The concept of local weak convergence is useful for us, since under some assumptions, the Betti numbers of a convergent sequence of finite complexes X_n can be bounded by a parameter of its local weak limit. In addition, when the local weak limit is a d-tree, this parameter is expressible by some inductive formula. This approach is described in [15], and we briefly mention the pertinent parts of that work.

Let X be a (possibly infinite) d-complex. We are interested in the spectral measure of its upper $(d-1)$-dimensional Laplacian. This Laplacian L is a symmetric operator acting on the Hilbert space $\mathcal{H} = \ell^2(F_{d-1}(B))$. For finite X, we can apply the spectral theorem to L which is symmetric, and therefore self-adjoint. However, for infinite X the situation is more subtle. The Laplacian L is only densely defined on the subspace of functions with finite support. It has a unique extension $\hat{L}$ to $\mathcal{H}$ which may be (but is not necessarily) self-adjoint. We say that a complex is self-adjoint if the extension $\hat{L}$ is a self-adjoint operator. For example, if B is a random rooted d-tree such that the expected degree of its $(d-1)$-faces is bounded, then it is almost surely self-adjoint. When X is self-adjoint, we can apply the *spectral theorem* and obtain the *spectral measure* $\mu_{X,\tau}$ of $\hat{L}$ with respect to the characteristic vector of a $(d-1)$-face τ.

A key parameter that we study is $x_B := \mu_{B,o}(\{0\})$. In words, $\mu_{B,o}$ is the spectral measure of the (extended) Laplacian of a d-tree B with respect to its root o. The reason that we are interested in the measure of the atom $\{0\}$ is that for a finite complex X, the sum $\sum_{\tau \in F_{d-1}(X)} \mu_{X,\tau}(\{0\})$ is equal to the dimension of the kernel of X's Laplacian, which is very close to $\beta_{d-1}(X)$. Here are two key lemmas from [15] that we use.

Lemma 2.3 *Let X_n be a random n-vertex d-complex with a full $(d-1)$-skeleton, and (B, o) be a random rooted d-tree that is almost surely self-adjoint. If X_n locally weakly converges to (B, o) then,*

$$\limsup_{n \to \infty} \frac{1}{\binom{n}{d}} \mathbb{E}_{X_n}[\beta_{d-1}(X_n)] \leq \mathbb{E}_B[x_B]$$

We can, in fact, say more. Namely, the expected spectral measure of X_n's Laplacian with respect to a uniformly random root weakly converges to the expected spectral measure of B's Laplacian. The following lemma enables us to compute the seemingly complicated parameter x_B.

Lemma 2.4 *Let B be a self-adjoint d-tree. Suppose that the root o is contained in m d-faces and consider the rooted d-subtrees $B_{j,i}$, $1 \leq j \leq m$, $1 \leq i \leq d$, as defined above. If there exists a $1 \leq j \leq m$ such that $x_{B_{j,1}} = \cdots = x_{B_{j,d}} = 0$, then $x_B = 0$. Otherwise,*

$$\frac{1}{x_B} = 1 + \sum_{j=1}^{m} \frac{1}{\sum_{i=1}^{d} x_{B_{j,i}}}.$$

3 | ENUMERATION OF d-HYPERTREES

3.1 | Proof of Theorem 1.4

We start with the following extremal question: What is the largest possible shadow of an n-vertex d-acyclic complex X with a given number of d-faces? Equivalently, what is the least possible number of ways that X can be extended to a d-acyclic complex with $(|X| + 1)$ d-faces? Although the following claim is not tight, it suffices for our purposes and its proof is fairly simple. It is possible to derive a tight bound using shifting methods [5].

Claim 3.1 *If X is an n-vertex d-acyclic complex, then $|\overline{SH}(X)| \geq \binom{n}{d+1} \left(1 - \frac{|X|}{\binom{n-1}{d}} \right)$.*

Proof Denote $Y = SH(X)$ and observe that $\mathrm{rank}(Y) = |X|$ since X is d-acyclic. There are clearly exactly $(d + 1)|Y|$ pairs (v, σ) where $v \in \sigma$ is a vertex and $\sigma \in Y$ a d-face. In addition, consider the collection of all d-faces of Y that contain v. This is an acyclic subcomplex, since even the d-complex with *all* $\binom{n-1}{d}$ d-faces that contain v is acyclic. Therefore, the number of such d-faces is at most $\mathrm{rank}(Y) = |X|$. It follows that $(d + 1)|Y| \leq n \cdot |X|$, as claimed. ∎

As we observe next, Claim 3.1 yields a simple proof of Kalai's lower bound.

Corollary 3.2 ([13]) $|\mathcal{T}_{n,d}| \geq \left(\frac{n}{d+1} \right)^{\binom{n-1}{d}}.$

Proof Let us construct a d-hypertree starting with a full $(d - 1)$-skeleton and adding one d-face at a time. By Claim 3.1 there are at least $\binom{n}{d+1} \left(1 - \frac{i}{\binom{n-1}{d}} \right)$ possible choices for our ith step. This argument counts every d-hypertree $\binom{n-1}{d}!$ times. Therefore,

$$\binom{n-1}{d}! \cdot |\mathcal{T}_{n,d}| \geq \prod_{i=0}^{\binom{n-1}{d}-1} \binom{n}{d+1} \left(1 - \frac{i}{\binom{n-1}{d}} \right).$$

The claim follows directly. ∎

We turn to prove Theorem 1.4 in a way that refines the previous argument. Consider the *random d-acyclic complex process* on n vertices $\mathbf{T}_d(n) = T_0 \subset T_1 \subset \cdots \subset T_{\binom{n-1}{d}}$, where T_0 is the full

$(d-1)$-skeleton on n vertices. In every step $1 \leq i \leq \binom{n-1}{d}$, a d-face σ_i is sampled uniformly at random from $\overline{\text{SH}}(T_{i-1})$ and added to the complex, that is, $T_i = T_{i-1} \cup \{\sigma_i\}$. Since this process produces a random ordered d-hypertree, its support size is $\binom{n-1}{d}! \cdot |\mathcal{T}_{n,d}|$, and its entropy does not exceed the logarithm of this number. But this entropy can actually be computed:

$$H(\mathbf{T}_d(n)) = \sum_{i=0}^{\binom{n-1}{d}-1} H(\sigma_{i+1} \mid T_i) = \sum_{i=0}^{\binom{n-1}{d}-1} \mathbb{E}_{T_i}[\log \left(|\overline{\text{SH}}(T_i)| \right)]. \tag{2}$$

The first equality is the chain rule for entropy, and the second equality follows since σ_{i+1} is selected uniformly from $\overline{\text{SH}}(T_i)$.

The random process $\mathbf{T}_d(n)$ is closely related to the Linial-Meshulam model. Indeed, consider a random ordering $\vec{\sigma} = \sigma_1, \ldots, \sigma_{\binom{n}{d+1}}$ of all d-faces over n vertices. The d-faces $\sigma_1, \ldots, \sigma_M$ along with a full $(d-1)$-skeleton on n vertices constitute the complex $Y_d(n, M)$. Therefore, $\vec{\sigma}$ can be also used to define the complex $Y_d(n, p)$ which equals to $Y_d(n, \hat{M})$, where $\hat{M}$ is $\text{Bin}\left(\binom{n}{d+1}, p \right)$-distributed. Let us call an index $1 \leq j \leq \binom{n}{d+1}$ *critical* if $\sigma_j \in \overline{\text{SH}}(Y_d(n, j-1))$. The complex $\mathbf{T}_d(n)$ equals to $\{\sigma_j \mid j \text{ is critical}\}$ with a full $(d-1)$-skeleton, and moreover, T_i is its subcomplex containing the first i critical faces $\sigma_{j_1}, \ldots, \sigma_{j_i}$ in $\vec{\sigma}$.

Recall that the function $r = r(c)$ depicts the normalized rank of $Y_d\left(n, \frac{c}{n} \right)$ as defined in Section 2. If $i = i(n)$ is an integer and $c > 0$ real such that $r(c) - i/\binom{n-1}{d}$ is positive and bounded away from zero, we may assume that a.a.s. $T_i \subset Y_d\left(n, \frac{c}{n} \right)$. Indeed, in such case $Y_d\left(n, \frac{c}{n} \right)$ a.a.s. has rank greater than i, and therefore contains more than i critical faces. In particular, a.a.s.,

$$|\overline{\text{SH}}(T_i)| \geq \left| \overline{\text{SH}} \left(Y_d \left(n, \frac{c}{n} \right) \right) \right|. \tag{3}$$

As explained in Section 2, the size $\left| \overline{\text{SH}} \left(Y_d \left(n, \frac{c}{n} \right) \right) \right|$ is concentrated at $\bar{s}(c) \binom{n}{d+1}$. In the proof below we use this accurate estimation *for every $c > 0$* to deduce the bound stated in Theorem 1.4. However, note that even a substantially simpler argument already yields a pretty good bound. Namely, since $\bar{s}(c) = 1$ for $c < c_d^*$, it follows that each of the first $\frac{c_d^* - \varepsilon}{d+1} \binom{n-1}{d}$ summands in Equation (2) equals $(1 - o(1)) \log \binom{n}{d+1}$, for $\varepsilon > 0$ arbitrarily small. The other summands in that equation can be easily bounded by the worst-case analysis of Claim 3.1. It turns out that for every dimension $d \geq 2$, this simple argument improves the bound of Equation (1). In addition, since $c_d^* = d + 1 - o_d(1)$, it gives a bound of the form

$$|\mathcal{T}_{n,d}| \geq \left(\frac{e^{1-\varepsilon_d}}{d+1} \cdot n \right)^{\binom{n-1}{d}},$$

where $0 < \varepsilon_d \to 0$ as d grows. We do not go into further details of this argument, since we derive below a better lower bound by a more careful analysis.

We denote

$$\lambda_i := \mathbb{E}_{T_i} \left[\log \left(|\overline{\text{SH}}(T_i)| \right) \right] - \log \binom{n}{d+1} \leq 0, \tag{4}$$

and split the summation in Equation (2) into four parts.

Claim 3.3 *Let $\varepsilon > 0$.*

1. *Subcritical: If $0 \le i \le \left(\frac{c_d^*}{d+1} - \varepsilon\right)\binom{n-1}{d}$ then*

$$\lambda_i \ge (1 - \varepsilon)\log(1 - \varepsilon) + \varepsilon \log\left(1 - \frac{c_d^*}{d+1}\right).$$

2. *Transition: If $\left(\frac{c_d^*}{d+1} - \varepsilon\right)\binom{n-1}{d} < i \le \left(\frac{c_d^*}{d+1} + \varepsilon\right)\binom{n-1}{d}$ then,*

$$\lambda_i \ge \log\left(1 - \frac{c_d^*}{d+1} - \varepsilon\right).$$

3. *Superctitical: If $\left(\frac{c_d^*}{d+1} + \varepsilon\right)\binom{n-1}{d} < i \le (1 - 2\sqrt{\varepsilon})\binom{n-1}{d}$ then,*

$$\lambda_i \ge (1 - \varepsilon)\log\left(\bar{s}\left(r^{-1}\left(\frac{i}{\binom{n-1}{d}}\right)\right)\right) - (1 - \varepsilon)\cdot 2\sqrt{\varepsilon} + \varepsilon \log(2\sqrt{\varepsilon}).$$

4. *Rearguard: If $(1 - 2\sqrt{\varepsilon})\binom{n-1}{d} < i < \binom{n-1}{d}$ then,*

$$\lambda_i \ge \log\left(1 - \frac{i}{\binom{n-1}{d}}\right).$$

Proof For the Transition and Rearguard ranges, the inequality is just that of Claim 3.1. For the Subcritical range, we fix some c with $c_d^* - \varepsilon(d+1) < c < c_d^*$ and apply Equation (3). This implies that with probability at least $1 - \varepsilon$, there holds $|\overline{\mathrm{SH}}(T_i)| > \binom{n}{d+1}(1 - \varepsilon)$, since $Y_d\left(n, \frac{c}{n}\right)$ has a.a.s. a shadow of vanishingly small density. By Claim 3.1, $|\overline{\mathrm{SH}}(T_i)| > \binom{n}{d+1}(1 - c_d^*/(d+1))$ always holds. Therefore,

$$\lambda_i \ge (1 - \varepsilon)\log(1 - \varepsilon) + \varepsilon \log\left(1 - c_d^*/(d+1)\right).$$

We turn to consider the Supercritical range. Let $\left(\frac{c_d^*}{d+1} + \varepsilon\right)\binom{n-1}{d} < i \le (1 - 2\sqrt{\varepsilon})\binom{n-1}{d}$ and $c_i = r^{-1}\left(\frac{i}{\binom{n-1}{d}}\right) > c_d^*$. By using Claim 3.1 as we did in the Subcritical item, it suffices to show that there exists $c > 0$ such that (i) $r(c) - i/\binom{n-1}{d} \gg 0$ and (ii) a.a.s.,

$$\log\left|\overline{\mathrm{SH}}\left(Y_d\left(n, \frac{c}{n}\right)\right)\right| \ge \log\binom{n}{d+1} + \log(\bar{s}(c_i)) - 2\sqrt{\varepsilon}.$$

Since $\bar{s}$ is continuous when $c > c_d^*$ and since r is strictly monotone, we may choose $c > c_i$ such that both $\bar{s}(c) > \bar{s}(c_i) - \varepsilon$ and condition (i) is satisfied. In addition, a.a.s. $\left|\overline{\mathrm{SH}}\left(Y_d\left(n, \frac{c}{n}\right)\right)\right| \ge \binom{n}{d+1}(\bar{s}(c) - \varepsilon)$, and therefore,

$$\log\left|\overline{\mathrm{SH}}\left(Y_d\left(n, \frac{c}{n}\right)\right)\right| - \log\binom{n}{d+1} > \log(\bar{s}(c_i) - 2\varepsilon) \ge \log(\bar{s}(c_i)) - 4\varepsilon/\bar{s}(c_i),$$

where the last inequality is by straightforward analysis. By Claim 3.1, $\bar{s}(c_i) \ge 1 - i/\binom{n-1}{d} \ge 2\sqrt{\varepsilon}$, which concludes the proof. ∎

Let us estimate the sum $\sum_{i=0}^{\binom{n-1}{d}-1} \lambda_i$. Note that this is a sum of nonpositive numbers that we want to estimate from below. We use the notation $o_\varepsilon(1)$ for terms that vanish (possibly from below) as $\varepsilon \to 0$. Since every index in the Subcritical regime contributes $o_\varepsilon(1)$ to the sum, the contribution of the entire range is bounded from below by $\binom{n-1}{d} \cdot o_\varepsilon(1)$. The same lower bound applies as well to the Transition regime, which contains $2\varepsilon\binom{n-1}{d}$ bounded terms. The entire contribution of the Rearguard regime is at least

$$\log\left(\frac{\left(2\sqrt{\varepsilon}\binom{n-1}{d}\right)!}{\binom{n-1}{d}^{\binom{n-1}{d}}}\right) = \binom{n-1}{d} \cdot o_\varepsilon(1).$$

In addition, the last two terms in our bound for each λ_i in the Supercritical regime are also of order $o_\varepsilon(1)$. By Equations (2) and (4), and these estimations, we conclude that

$$H(\mathbf{T}_d(n)) \geq \binom{n-1}{d}\left(\log\binom{n}{d+1} + (1-\varepsilon)\frac{1}{\binom{n-1}{d}}\sum_{i=\left(\frac{c_d^*}{d+1}+\varepsilon\right)\binom{n-1}{d}}^{\left(1-2\sqrt{\varepsilon}\right)\binom{n-1}{d}} \log\left(\bar{s}\left(r^{-1}\left(\frac{i}{\binom{n-1}{d}}\right)\right)\right) + o_\varepsilon(1)\right).$$

On the other hand, the entropy cannot exceed the logarithm of the cardinality of the support, that is,

$$H(\mathbf{T}_d(n)) \leq \log|\mathcal{T}_{d,n}| + \binom{n-1}{d}\left(\log\binom{n-1}{d} - 1\right).$$

Letting $\varepsilon \to 0$ and $\alpha_d := \int_{c_d^*/d+1}^1 \log \bar{s}(r^{-1}(x))\mathrm{d}x$ yields

$$|\mathcal{T}_{n,d}| \geq \left((1-o_n(1))\frac{e^{1+\alpha_d}}{d+1} \cdot n\right)^{\binom{n-1}{d}}.$$

In order to complete the proof, we need to establish the integral form for α_d as stated in the theorem. Consider the function t that maps every $c > c_d^*$ to the smallest positive root of $t = e^{-c(1-t)^d}$. The derivative of t w.r.t. c is $t' = -\frac{t(1-t)^{d+1}}{1-t+dt\ln t}$ [15]. In addition, recall that for $c > c_d^*$, $\bar{s}(c) = 1 - (1-t(c))^{d+1}$ and a straightforward computation yields that $r'(c) = \bar{s}(c)/(d+1)$. The desired integral form is obtained by the change of variables $y = t(r^{-1}(x))$.

3.2 | Proof of Theorem 1.7

Kalai's upper bound $|\mathcal{T}_{n,d}| < \left(\frac{e}{d+1} \cdot n\right)^{\binom{n-1}{d}}$ accounts for *all* n-vertex d-complexes with $\binom{n-1}{d}$ d-faces. We slightly improve this bound by estimating the probability that a uniformly sampled complex with these parameters is a d-hypertree. Theorem 1.7 immediately follows from the following lemma.

Lemma 3.4 *The probability that $Y := Y_d\left(n, \binom{n-1}{d}\right)$ is a d-hypertree is $\exp(-\Omega(n^d))$.*

Proof Let $\sigma_1, \sigma_2, \ldots$ be an infinite sequence of d-faces on a set of n vertices, each chosen *independently* uniformly at random. Consider the n-vertex complex $\tilde{Y}_d(n, M)$ that has a full

$(d-1)$-skeleton and the first M d-faces in the sequence, where repetitions of the same face get removed. The complex Y is equivalent to $\tilde{Y}_d(n, \bar{M})$ where $\bar{M}$ is the (random) index for which the prefix $\sigma_1, \ldots, \sigma_{\bar{M}}$ contains exactly $\binom{n-1}{d}$ distinct d-faces.

Let $\varepsilon > 0$ be a small constant. If $M = (1 + \frac{\varepsilon}{d+1})\binom{n-1}{d}$ then Y is contained in $\tilde{Y}_d(n, M)$ with probability $1 - \exp(-\Theta(n^d))$ by a standard measure concentration argument. Let us denote by R the rank of $\tilde{Y}_d(n, M)$.

The probability that Y is a d-hypertree is bounded by $\exp(-\Theta(n^d))$ plus the probability that $\tilde{Y}_d(n, M)$ has full rank, that is, $R = \binom{n-1}{d}$. In addition, $\tilde{Y}_d(n, M)$ is a.a.s. contained in $Y_d(n, p)$ for $p = \frac{d+1+2\varepsilon}{n}$, since a $\text{Bin}\left(\binom{n}{d+1}, p\right)$ random variable is a.a.s. greater than M. Therefore, the expectation $\mathbb{E}[R]$ is bounded, up to an additive error term of $o(n^d)$, by the expected rank of $Y_d(n, p)$ which equals $\binom{n-1}{d} \cdot r(d+1+2\varepsilon)$. Since $r(d+1) < 1$, we can choose ε so that $\mathbb{E}[R] < (1-\delta)\binom{n-1}{d}$ for some constant $\delta > 0$. The proof is concluded by observing that R is a 1-Lipschitz function that depends on $M = (1 + \frac{\varepsilon}{d+1})\binom{n-1}{d}$ independent variables. By Azuma's inequality [17],

$$\Pr\left[R = \binom{n-1}{d}\right] \leq \exp\left(-\frac{\delta^2}{2\left(1 + \frac{\varepsilon}{d+1}\right)}\binom{n-1}{d}\right).$$

∎

4 | THE RANDOM 1-OUT d-COMPLEX

The random 1-out d-complex $S_d(n, 1)$ is an n-vertex d-dimensional complex with a full $(d-1)$-skeleton, in which every $(d-1)$-face τ selects, independently uniformly at random, a d-face σ_τ that contains it. To wit, τ selects a uniform random vertex $v_\tau \notin \tau$ and $\sigma_\tau = v_\tau \tau$. The selection process is done independently, but we remove multiply selected d-faces to maintain a simplicial complex. The purpose of this section is to show that the d-dimensional homology of $S_d(n, 1)$ is a.a.s. of dimension $o(n^d)$. Moreover, the complex can be made very close to d-acyclic by a random sparsification. The upper bound on the dimension of the homology is proved using the spectral measure of the local weak limit as presented in Section 2.

4.1 | The local weak limit of the random 1-out process

We first describe the local weak limit of $S = S_d(n, 1)$. That is, the limiting distribution of local neighborhoods of a root $(d-1)$-face in S. This local weak limit is a random d-tree that we denote by $\mathbf{B}_d$. The number of children m_τ of a $(d-1)$-face τ in $\mathbf{B}_d$ is independently distributed, but the $(d-1)$-faces and d-faces in $\mathbf{B}_d$ come in two types—(A) and (B). The type of every $(d-1)$-face τ always coincides with that of its selected d-face σ_τ. The type indicates whether, when exposing the neighborhood of the root, the selected d-face σ_τ appeared (A) after or (B) before the selecting $(d-1)$-face τ. For instance, the root is always of type (A).

For a $(d-1)$-face τ of type (A), the number of children m_τ is $1 + \text{Poi}(d)$ distributed. One descendant d-face is σ_τ whose type is also (A), whence all its $(d-1)$-subfaces also have type (A). The other $\text{Poi}(d)$ descendant d-faces have type (B). These are the d-faces that contain τ but were selected by some other $(d-1)$-face. Therefore, a type (B) d-face $\sigma = \sigma_\rho$ has precisely one descending $(d-1)$-face ρ of type (B) whereas all others have type (A). On the other hand, for τ a $(d-1)$-face of type (B), the distribution of m_τ is $\text{Poi}(d)$, with all descending d-faces of type (B) (Figure 2).
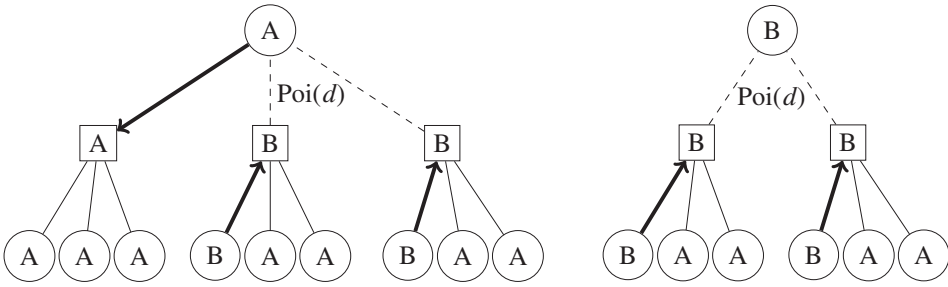
FIGURE 2 Illustration of the inclusion graph of the d-tree $\mathbf{B}_d$ for $d = 3$. Circles represent $(d - 1)$-faces, and squares are d-faces, each vertex is marked with the face's type. The thick arrows are included only to explain how the selection process of $S_d(n, 1)$ is reflected in the d-tree. A dashed line demonstrate that the number of type (B) d-faces descending from a $(d - 1)$-face is Poi(d) randomly distributed. Since the root has type (A), the information in this figure completely determines the distribution of $\mathbf{B}_d$

To generate a d-tree $\mathbf{B}_d$, we start with a root $(d - 1)$-face of type (A), generate its descendant d-faces, and keep track of the types of the new $(d - 1)$-faces. At each step $k > 0$, we generate the descendants of the $(d - 1)$-faces at distance $2k$ from the root according to their type.

For $d = 1$, $\mathbf{B}_1$ is a well-known random tree model. Type (A) vertices form an infinite rooted path, and every such vertex "grows" a Galton-Watson Poi(1) branching process with vertices of type (B). This is known to be the local weak limit of a uniform spanning tree [11].

Claim 4.1 *$\mathbf{B}_d$ is the local weak limit of $S_d(n, 1)$.*

Proof It is easy to observe that a.a.s. all the $(d - 1)$-faces in S have degree $O(\log n)$. Indeed, up to negligible duplications, the degree of a $(d - 1)$-face τ is one plus the number of $(d - 1)$-faces that chose a d-face containing τ, which is $\text{Bin}\left((n - d)d, \frac{1}{n-d}\right)$-distributed.

Fix some root $(d - 1)$-face $o \in S$ and an integer $k > 0$. Both the number of $(d - 1)$-faces and vertices that appear in a bounded-radius neighborhood of o are a.a.s. polylogarithmic in n. The local d-tree structure is disrupted only if in generating S, a $(d - 1)$-face at a bounded distance from o selects a vertex in its bounded radius neighborhood. This, however, is very unlikely to occur, so that only with probability $o_n(1)$ is the k-radius neighborhood not a d-tree.

It remains to show that the probability of every fixed d-tree of depth at most k tends to its probability in the distribution $\mathbf{B}_d$. Since the k-local neighborhood of the root $o \in S$ is a.a.s. a d-tree, we can expose it as in a generative process of a d-tree. By the construction of $\mathbf{B}_d$, there is only one difference between the exposure process of o's neighborhood in S and the generative process of $\mathbf{B}_d$. Namely, the independent Poisson(d) variables in $\mathbf{B}_d$ are replaced by possibly dependent Binomial variables in S.

Suppose we have already exposed some part of o's neighborhood in S, and we are about to expose the descendant d-faces of some $(d - 1)$-face τ. The situation varies according to whether τ's selected d-face σ_τ has already been exposed, but in this respect there is no difference between the two processes. The difference is that in $\mathbf{B}_d$, τ has a Poi(d) independently distributed number of descendant d-faces of type (B). These correspond to the d-faces in S that contain τ, that were selected by some other $(d - 1)$-face but have not yet been exposed. In S, this number is distributed binomially, where the number of trials is a.a.s. $nd - o(n)$ and the success probability is $(1 - o(1))/n$. In particular, as $n \to \infty$, this number tends to a Poi(d) variable. Note that in both parameters of this binomial distribution, the error term may depend on the already exposed neighborhood of o, so the different degrees could be dependent. However, since this dependency a.a.s. only affects the error term, the joint distribution of all these numbers tends to the distribution of independent Poisson variables. ∎

4.2 | Proof of Theorem 1.8

Let $S = S_d(n, 1)$. Theorem 1.8 would follow if we can show that $\mathbb{E}[\beta_{d-1}(S)] = o(n^d)$. Indeed, as we saw in Section 2, $\beta_d(S) = \beta_{d-1}(S) + \left(|S| - \binom{n-1}{d}\right)$, and a.a.s. $|S| = \binom{n}{d} - o(n^d)$ since there are only $o(n^d)$ duplications. In addition, by Markov's inequality, if $\mathbb{E}[\beta_d(S)] = o(n^d)$ then a.a.s. $\beta_d(S) = o(n^d)$.

Recall that we denote by $x_{B,o} \in [0, 1]$ the spectral measure of the Laplacian of a d-tree B with respect to the characteristic vector of its root o, measured at the atom $\{0\}$. By Lemma 2.3,

$$\limsup_{n \to \infty} \frac{1}{\binom{n}{d}} \mathbb{E}[\beta_{d-1}(S)] \le \mathbb{E}[x_{\mathbf{B}_d}],$$

since S locally weakly converges to $\mathbf{B}_d$. Therefore, Theorem 1.8 follows from the following lemma.

Lemma 4.2 $\mathbb{E}[x_{\mathbf{B}_d}] = 0$.

Proof We denote the random variable $X = x_{\mathbf{B}_d}$. In addition, let $\mathbf{B}_d^{(B)}$ denote a random d-tree that is generated exactly like $\mathbf{B}_d$ except that the root is of type (B), and we denote the random variable $Y = x_{\mathbf{B}_d^{(B)}}$. In addition, we denote $a := \Pr[X > 0]$ and $b := \Pr[Y > 0]$. We also use $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ to denote i.i.d. copies of X and Y respectively.

By Lemma 2.4, X has the distribution obtained by the following process. We independently sample $X_1, \ldots, X_d$, a Poi(d) distributed number m, and for every $1 \le j \le m$, we also sample Y_j and $X_{j,1}, \ldots, X_{j,d-1}$. The random variable X equals 0 if $X_1 = \cdots = X_d = 0$ or $Y_j = X_{j,1} = \cdots = X_{j,d-1} = 0$ for some j. It is otherwise computed by the formula in Lemma 2.4:

$$X = \left(1 + \frac{1}{X_1 + \cdots + X_d} + \sum_{j=1}^{m} \frac{1}{Y_j + X_{j,1} + \cdots + X_{j,d-1}}\right)^{-1}.$$

Clearly, a similar distributional equation can be derived for Y.

First, we use these distributional equations in order to derive relations between a and b. The probability that $Y_j = X_{j,1} = \cdots = X_{j,d-1} = 0$ for some given j is $(1-b)(1-a)^{d-1}$. Therefore, the probability that for every $1 \le j \le m$ this does not hold, where $m \sim$ Poi(d), equals to $e^{-d(1-b)(1-a)^{d-1}}$. In addition, the probability that $X_1 = \cdots = X_d = 0$ equals to $(1-a)^d$. Therefore,

$$a = (1 - (1-a)^d)e^{-d(1-b)(1-a)^{d-1}}.$$

By a similar argument, $b = e^{-d(1-b)(1-a)^{d-1}}$, so that

$$a = (1 - (1-a)^d)b. \tag{5}$$

Let P be a random variable whose distribution is that of $X_1 + \cdots + X_d$, and let $Q, Q_1, Q_2, \ldots$ be random variables whose distribution is that of $Y + X_1 + \cdots + X_{d-1}$. In addition, let m be a Poi(d) distributed random variable. We use the distributional equations derived by Lemma 2.4 to compute the expectation of X.

$$\mathbb{E}[X] = \mathbb{E}\left[\frac{\mathbf{1}_{\{P>0, \; \forall j \in [m] : \; Q_j > 0\}}}{1 + P^{-1} + \sum_{j=1}^{m} Q_j^{-1}}\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{P>0, \ \forall j\in[m]: \ Q_j>0\}}\left(1 - \frac{P^{-1} + \sum_{j=1}^m Q_j^{-1}}{1 + P^{-1} + \sum_{j=1}^m Q_j^{-1}}\right)\right]$$

$$= a - \mathbb{E}\left[\frac{P^{-1}\cdot\mathbf{1}_{\{P>0, \ \forall j\in[m]: \ Q_j>0\}}}{1 + P^{-1} + \sum_{j=1}^m Q_j^{-1}}\right] - \mathbb{E}\left[\sum_{i=1}^m \frac{Q_i^{-1}\cdot\mathbf{1}_{\{P>0, \ \forall j: \ Q_j>0\}}}{1 + P^{-1} + Q_i^{-1} + \sum_{\substack{j=1\\j\neq i}}^m Q_j^{-1}}\right]. \tag{6}$$

Let us separately expand the two expectations.

$$\mathbb{E}\left[\frac{P^{-1}\cdot\mathbf{1}_{\{P>0, \ \forall j\in[m]: \ Q_j>0\}}}{1 + P^{-1} + \sum_{j=1}^m Q_j^{-1}}\right] = \mathbb{E}\left[\frac{\frac{1}{1+\sum_{j=1}^m Q_j^{-1}}}{\frac{1}{1+\sum_{j=1}^m Q_j^{-1}} + P}\cdot\mathbf{1}_{\{P>0, \ \forall j\in[m]: \ Q_j>0\}}\right] \tag{7}$$

$$= \mathbb{E}\left[\frac{Y\cdot\mathbf{1}_{\{X_1+\cdots+X_d>0, \ Y>0\}}}{Y + X_1 + \cdots + X_d}\right]. \tag{8}$$

For the equality (7) we multiply both the nominator and the denominator by $P/(1 + \sum_{j=1}^m Q_j^{-1})$. The distributional equation of Y yields Equation (8).

The second expectation in (6) requires a little more work.

$$\mathbb{E}\left[\sum_{i=1}^m \frac{Q_i^{-1}\cdot\mathbf{1}_{\{P>0, \ \forall j: \ Q_j>0\}}}{1 + P^{-1} + Q_i^{-1} + \sum_{\substack{j=1\\j\neq i}}^m Q_j^{-1}}\right] = \mathbb{E}_m\left[m\cdot\mathbb{E}\left[\frac{Q^{-1}\cdot\mathbf{1}_{\{P>0, \ Q>0 \ , \ \forall j: \ Q_j>0\}}}{1 + P^{-1} + Q^{-1} + \sum_{j=1}^{m-1} Q_j^{-1}}\right]\right] \tag{9}$$

$$= d\cdot\mathbb{E}\left[\frac{Q^{-1}\cdot\mathbf{1}_{\{P>0, \ Q>0 \ , \ \forall j: \ Q_j>0\}}}{1 + P^{-1} + Q^{-1} + \sum_{j=1}^m Q_j^{-1}}\right] \tag{10}$$

$$= d\cdot\mathbb{E}\left[\frac{\frac{1}{1+P^{-1}+\sum_{j=1}^m Q_j^{-1}}}{Q + \frac{1}{1+P^{-1}+\sum_{j=1}^m Q_j^{-1}}}\cdot\mathbf{1}_{\{P>0, \ Q>0 \ , \ \forall j: \ Q_j>0\}}\right] \tag{11}$$

$$= d\cdot\mathbb{E}\left[\frac{X\cdot\mathbf{1}_{\{Y+X_1+\cdots+X_{d-1}>0, \ X>0\}}}{Y + X_1 + \cdots + X_{d-1} + X}\right] \tag{12}$$

$$= \mathbb{E}\left[\sum_{i=1}^d \frac{X_i\cdot\mathbf{1}_{\{Y+X_1+\cdots+X_{i-1}+X_{i+1}+\cdots+X_d>0, \ X_i>0\}}}{Y + X_1 + \cdots + X_d}\right]. \tag{13}$$

Equation (9) follows by linearity of expectation and the symmetry of the Q_i's. To derive (10) we note that for every function $\varphi : \mathbb{N} \to \mathbb{R}$, $\mathbb{E}_m[m\cdot\varphi(m-1)] = d\cdot\mathbb{E}_m[\varphi(m)]$ where m is Poi(d) distributed. Equations (11) and (12) are obtained similarly to (7) and (8). The last equation (13) is derived by linearity of expectation and the symmetry of the X_i's.

Let us return to the main computation of $\mathbb{E}[X]$ by plugging in (8) and (13) into (6).

$$\mathbb{E}[X] = a - \mathbb{E}\left[\frac{Y\cdot\mathbf{1}_{\{X_1+\cdots+X_d>0, \ Y>0\}} + \sum_{i=1}^d X_i\cdot\mathbf{1}_{\{Y+X_1+\cdots+X_{i-1}+X_{i+1}+\cdots+X_d>0, \ X_i>0\}}}{Y + X_1 + \cdots + X_d}\right]$$

$$= a - b(1 - (1-a)^d) - (1-b)(1 - (1-a)^d - da(1-a)^{d-1}). \tag{14}$$

Equation (14) is derived by the following observation. To compute the expectation in the preceding line we sample $Y, X_1, ..., X_d$ independently. If two or more of these variables are positive, the contribution to

the expectation is 1. Otherwise the contribution is 0. Therefore, the expectation equals the probability that two or more of these variables are nonzero.

To conclude the proof, we recall that $a = b(1 - (1 - a)^d)$ by (5), hence $E[X] \leq 0$. ∎

4.3 | The random $(1 - \varepsilon)$-out d-complex

We turn to prove Theorem 1.9. Let us denote by $Z(n, m, d)$ the number of n-vertex inclusion-minimal d-cycles whose number of d-faces is m. In ([4], Theorem 4.1), it is proved that there exists some constant $\delta = \delta(d) > 0$ such that

$$\sum_{m=d+2}^{\delta n^d} Z(n, m, d) \left(\frac{d+1}{n} \right)^m = o(1).$$

The probability of any complex with m d-faces to be a subcomplex of S is at most $\left(\frac{d+1}{n} \right)^m$ since the d-faces of $S = S_d(n, 1)$ are nonpositively correlated. Therefore, a.a.s., the d-cycles in S are either of size at least δn^d (=large) or are boundaries of a $(d + 1)$-dimensional simplex.

Fix some $\varepsilon > 0$, and let $S' = S_d(n, 1 - \varepsilon)$. Recall that S' is obtained from S by removing every d-face independently with probability ε. In order to analyze the d-homology of S', we remove the d-faces of S sequentially. That is, we consider a sequence of complexes $S = S_0 \supset S_1 \supset \cdots \supset S_M = S'$, where a.a.s. $M = \Theta(n^d)$. In every step i of this process, $\beta_d(S_i) < \beta_d(S_{i-1})$ if and only if the d-face we removed participated in a d-cycle of S_i. If S_i contains a large d-cycle, that is not a boundary of a $(d + 1)$-simplex, the Betti number decreases with probability bounded away from zero. Since $\beta_d(S_0) = o(n^d)$ and $M = \theta(n^d)$, the probability that any large d-cycle survives the process is negligible.

5 | DISCUSSION AND OPEN QUESTIONS

The study of hypertrees raises many open questions. Here are some that concern enumeration and randomized constructions.

- There is no reason to believe that the bounds in Theorem 1.4 are tight. In fact, it is known [2] that a similar argument for $d = 1$ does not yield a bound of $((1 - o(1)n)^n$. However, as mentioned in the introduction, it is conceivable that Conjecture 1.3 can be answered using Theorem 1.4 coupled with nontrivial upper bounds for the number of d-collapsible hypertrees. It is also interesting to improve our lower bound for $|C_{n,d}|$ which at the moment is supported on nonevasive complexes that have a very irregular vertex-degree sequence [12].

- Is it possible to efficiently sample d-hypertrees uniformly at random? It is suggestive to do this using rapidly mixing Markov chains (eg, [19]), possibly the *base-exchange* Markov Chain $\Omega = \Omega_{n,d}$ that is of interest for matroids in general [8]. The states of Ω are all the n-vertex d-hypertrees. To proceed from a d-hypertree T, we select a d-face $\sigma \notin T$ uniformly at random, and replace T by $T \setminus \{\tau\} \cup \{\sigma\}$, where τ is a random d-face in the unique d-cycle of $T \cup \{\sigma\}$. The stationary distribution of Ω is uniform, but we do not know whether it is rapidly mixing.

- Can the random 1-out complex help us improve our estimates for the number of d-hypertrees? The problem boils down to bounding the typical *permanent* of such a d-complex T. Namely, the number of injective functions from $F_d(T)$ to $F_{d-1}(T)$ that map every d-face to one of its subfaces. In other words, the permanent of T is the number of maximum matchings in T's inclusion graph. We wonder if the typical permanent of $S_d(n, 1)$ can be bounded in terms of its local weak limit $\mathbf{B}_d$ [1].

- We know even less about random generation of d-collapsible hypertrees. It is possible to restrict the base-exchange chain $\Omega_{n,d}$ to d-collapsible hypertrees, but we do not even know whether the restricted chain is connected, not to speak of rapid mixing. There is also an interesting greedy-random process that suggests itself, where we sequentially add a random d-face to the current complex provided that d-collapsibility is not violated. How many d-faces does this process acquire before it halts? What is the combinatorial structure of the final complex?

- Other types of hypertrees such as contractible, $\mathbb{Z}$-hypertrees, and $\mathbb{F}_2$-hypertrees can be considered in all these contexts. For instance, one can ask whether Theorem 1.8 also holds over $\mathbb{F}_2$ coefficients. In particular, applying a first moment method on the $\mathbb{F}_2$-cohomology of $S_2(n, 1)$ yields the following interesting question. For a fixed graph G and a pair of vertices i, j, let

$$\varphi_{i,j}(G) := \frac{1}{n-2} \left| \{k \notin \{i,j\} \ : \ i,j,k \text{ span an even number of edges in } G\} \right|.$$

In words, $\varphi_{i,j}$ is the probability that the selection of the edge ij in $S_2(n, 1)$ does not exclude G from being a cocycle of the complex. Prove that

$$\sum_{G \text{ an } n\text{-vertex graph}} \left[\prod_{i,j} \varphi_{i,j} \right] = 2^{o(n^2)}.$$

It is conceivable that this sum is of order $2^{\Theta(n)}$.

ACKNOWLEDGMENTS

REFERENCES

1. M. Abért, P. Csikvári, P. Frenkel, and G. Kun, *Matchings in Benjamini-Schramm convergent graph sequences*, Trans. Amer. Math. Soc. **368** (2016), no. 6, 4197–4218.

2. L. Addario-Berry, *Partition functions of discrete coalescents: From Cayley's formula to Frieze's ζ(3) limit theorem*. XI Symposium on Probability and Stochastic Processes, Springer, Birkhauser, Basel, pp. 1–45, 2015.

3. R.M. Adin, *Counting colorful multi-dimensional trees*, Combinatorica. **12** (1992), no. 3, 247–242.

4. L. Aronshtam, N. Linial, T. Łuczak, and R. Meshulam, *Collapsibility and vanishing of top homology in random simplicial complexes*, Discrete Comput. Geom. **49** (2013), no. 2, 317–334.

5. A. Björner and G. Kalai, *An extended Euler-Poincaré theorem*, Acta Math. **161** (1988), no. 1, 279–303.

6. T. Bohman and A. Frieze, *Hamilton cycles in 3-out*, Random Structures Algorithms. **35** (2009), no. 4, 393–417.

7. A. Duval, C. Klivans, and J. Martin, *Simplicial matrix-tree theorems*, Trans. Amer. Math. Soc. **361** (2009), no. 11, 6073–6114.

8. T. Feder and M. Mihail, *Balanced matroids*. Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing, ACM, New York, NY, pp. 26–38, 1992.

9. A. Frieze, *On the number of perfect matchings and Hamilton cycles in ε-regular non-bipartite graphs*, Electron. J. Combin. **7** (2000), no. R57, 2.

10. A.M. Frieze, *Maximum matchings in a class of random graphs*, J. Combin. Theory Ser. B. **40** (1986), no. 2, 196–212.

11. G.R. Grimmett, *Random labelled trees and their branching networks*, J. Aust. Math. Soc. Ser. A. **30** (1980), no. 2, 229–237.

12. J. Kahn, M. Saks, and D. Sturtevant, *A topological approach to evasiveness*, Combinatorica. **4** (1984), no. 4, 297–306.

13. G. Kalai, *Enumeration of $\mathbb{Q}$-acyclic simplicial complexes*, Israel J. Math. **45** (1983), no. 4, 337–351.

14. N. Linial, R. Meshulam, and M. Rosenthal, *Sum complexes—A new family of hypertrees*, Discrete Comput. Geom. **44** (2010), no. 3, 622–636.

15. N. Linial and Y. Peled, *On the phase transition in random simplicial complexes*, Ann. Math. **184** (2016), no. 3, 745–773.

16. R. Lyons, *Random complexes and ℓ^2-Betti numbers*, J. Topol. Anal. **1** (2009), no. 02, 153–175.

17. C. McDiarmid, *On the method of bounded differences*, Surv. Combin. **141** (1989), no. 1, 148–188.

18. E. Shamir and E. Upfal, *One-factor in random graphs based on vertex choice*, Discrete Math. **41** (1982), no. 3, 281–286.

19. A. Sinclair and M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. Comput. **82** (1989), no. 1, 93–133.

20. D.B. Wilson, *Generating random spanning trees more quickly than the cover time*. Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, ACM, New York, NY, pp. 296–303, 1996.