Random Simplicial Complexes - Around the Phase Transition

Nathan Linial[*]   Yuval Peled[†]

November 27, 2016

Abstract

This article surveys some of the work done in recent years on random simplicial complexes. We mostly consider higher-dimensional analogs of the well known phase transition in $G(n, p)$ theory that occurs at $p = \frac{1}{n}$. Our main objective is to provide a more streamlined and unified perspective of some of the papers in this area.

1 Introduction

There are at least two different perspectives from which our subject can be viewed. We survey some recent developments in the emerging field of *high-dimensional combinatorics*. However, these results can be viewed as well as part of an ongoing effort to apply *the probabilistic method in topology*. The systematic study of random graphs was started by Erdős and Rényi in the early 1960's and had a major impact on discrete mathematics, computer science and engineering. Since graphs are one-dimensional simplicial complexes, why not develop an analogous theory of d-dimensional random simplicial complexes for all $d \geq 1$? To this end, an analog of Erdős and Rényi's $G(n, p)$ model, called $Y_d(n, p)$, was introduced in [12]. Such a simplicial complex Y is d-dimensional, it has n vertices and has a full $(d - 1)$-dimensional skeleton. Each d-face is placed in Y independently with probability $p = p(n)$. Note that $Y_1(n, p)$ is identical with $G(n, p)$. Throughout the paper we consider a fixed dimension $d > 1$ and investigate the asymptotic topological properties of $Y_d(n, p)$. We say that a property holds asymptotically-almost-surely (a.a.s.) if its probability tends to 1 as n tends to infinity.

One of the most natural questions to ask in any model of random graphs concerns graph *connectivity*. As Erdős and Rényi famously showed, the threshold for graph connectivity in $G(n, p)$ is $p = \frac{\ln n}{n}$. To draw the analogy from a topological perspective, one should seek the threshold for the vanishing of the $(d - 1)$-st homology. This indeed was the motivating problem in [12]. As that paper showed, and together with subsequent work [17] this threshold in $Y_d(n, p)$ is $p = \frac{d \ln n}{n}$. Here the coefficients can come from any fixed finite abelian group. The same question for *integral* $(d - 1)$-st homology has attracted considerable attention and the answer is believed to be the same. This was recently confirmed for $d = 2$ [15], and is not yet fully resolved for higher dimensions (but see [10]). The threshold for the vanishing of the fundamental group of $Y_2(n, p)$ is fairly well (but still not perfectly) understood [7, 11].

Since we tend to work by analogy with the $G(n, p)$ theory, it is a very challenging problem to seek a high-dimensional counterpart to the *phase transition* that occurs at $p = \frac{1}{n}$. It is here that the random graph a.a.s. acquires cycles. Namely, for every $0 < c < 1$ there is a $0 < q = q(c) < 1$ such that a graph in $G(n, \frac{c}{n})$ is a forest with probability $q + o_n(1)$, but for $p \geq \frac{1}{n}$, a $G(n, p)$ graph has, a.a.s. at least one cycle. These notions

d	**2**	**3**	**4**	**5**	**10**	**100**	**1000**
γ_d	2.455	3.089	3.509	3.822	4.749	7.555	10.175
c_d	2.754	3.907	4.962	5.984	$11 - 10^{-3.73}$	$101 - 10^{-41.8}$	$1001 - 10^{-431.7}$

Table 1: The critical constants γ_d and c_d.

have natural analogs in higher-dimensional complexes that suggest what is being sought. However, even more famously, a *giant* connected component with $\Omega(n)$ vertices emerges at $p = \frac{1}{n}$. Since there is no natural notion of connected components at dimensions $d > 1$, it is not even clear what to ask. Finding the correct framework for asking this question and discovering the answer is indeed one of the main accomplishments of the research that we survey here.

Another reason that makes the high-dimensional scenario more complicated than the graph-theoretic picture is that there are several natural analogs for acyclicity. A $(d-1)$-face τ in a d-dimensional complex Y is *free* if it is contained in exactly one d-dimensional face σ of Y. In the corresponding *elementary collapse* step, which is a special case of homotopy equivalence, τ and σ are removed from Y. We say that Y is *d-collapsible* if it is possible to eliminate all its d-faces by a series of elementary collapses. Otherwise, the maximal subcomplex of Y in which all $(d-1)$-faces are contained in at least two d-faces is called the *core* of Y. Note that a graph (i.e., a 1-dimensional complex) is 1-collapsible if and only if it is acyclic, i.e., a forest.

A d-dimensional complex Y is said to be *d-acyclic* if its d-th homology group vanishes. Namely, if the d-dimensional boundary matrix $\partial_d(Y)$ has a trivial right kernel. Unless otherwise stated, we consider this matrix over the reals. The real d-Betti number of Y is $\beta_d(Y; \mathbb{R}) := \dim H_d(Y; \mathbb{R}) = \dim(\ker \partial_d(Y))$.

Whereas acyclicity and 1-collapsibility are equivalent for graphs, this is no longer the case for d-dimensional complexes. Clearly, a d-collapsible simplicial complex has a trivial d-th homology, but the reverse implication does not hold in dimension $d \geq 2$.

In this view, there are now two potentially separate thresholds to determine in $Y_d(n, p)$: For d-collapsibility and for the vanishing of the d-th homology. These questions were answered and the respective thresholds were determined in a series of four papers. A lower bound on the threshold for d-collapsibility was found in [6] and a matching upper bound was proved in [4]. An upper bound on the threshold for the vanishing of the d-th homology was found in [5], with a recent matching lower bound for real homology [14]. We conjecture that the same bound holds for all coefficient rings, but this remains open at present. In the present article we note an error in the proof that is presented in [4], and we indicate how to overcome it. The main theorem in the paper is correct, and here we also derive some additional information on the face numbers of the core.

The purpose of this paper is to survey these results and present the main ingredients of the proofs. In particular, we highlight the key role of the local structure of random complexes in all these proofs.

Both thresholds are of the form $p = \frac{c}{n}$. Namely there is a constant $c = \gamma_d$ corresponding to the d-collapsibility threshold and $c = c_d$ for acyclicity. As functions of the dimension d, the constants γ_d and c_d differ substantially. Our results allow us to numerically compute them to desirable accuracy (See Table 1).

We briefly refer to a d-dimensional complexes as a d-complex. Before stating the theorems, a small technical remark is in order. An obvious obstacle for a d-complex Y to be either d-collapsible or d-acyclic is that it contains the boundary of a $(d+1)$-simplex $\partial \Delta_{d+1}$, i.e. all the $d+2$ d-faces that are spanned by some $d+2$ vertices. In the random complex $Y_d\left(n, \frac{c}{n}\right)$, these objects appear with probability bounded away from both zero and one, and it is easy to see that their number is Poisson distributed with a constant expectation. In particular, $Y_d\left(n, \frac{c}{n}\right)$ is $\partial \Delta_{d+1}$-free with positive probability. There are several ways to go around this technical difficulty. In [6] a model of random complexes conditioned on being $\partial \Delta_{d+1}$-free was considered, which allowed a cleaner form for the theorems. Here we work with the simple binomial model, and consequently must mention these simplices.

We turn to the main theorem. Let $d \geq 2$ be an integer, $c > 0$ real and denote the core of $Y = Y_d\left(n, \frac{c}{n}\right)$ by $\tilde{Y}$. We define γ_d as the minimum of the function $\psi(x) := -\frac{\ln x}{(1-x)^d}$, $0 < x < 1$. Furthermore, we let x_* be the

unique root in $(0,1)$ of $(d+1)(1-x) + (1+dx)\ln x = 0$, and $c_d := \psi(x_*)$.

In addition, for an integer k and $\lambda > 0$ real, we let $\Psi_k(\lambda) := \Pr[\text{Poi}(\lambda) \geq k]$, and $t = t(c,d)$ be the smallest positive root of

$$t = e^{-c(1-t)^d}, \tag{1}$$

or equivalently $1 - t = \Psi_1(c(1-t)^d)$.

Theorem 1.1. *Let $d \geq 2$ be an integer, $c > 0$ real, and $Y = Y_d\left(n, \frac{c}{n}\right)$.*

(I) *The collapsible regime: If $c < \gamma_d$ then a.a.s. either Y is d-collapsible or its core is comprised of $O_d(1)$ vertex disjoint $\partial\Delta_{d+1}$'s.*

(II) *The intermediate regime: If $\gamma_d < c < c_d$ then a.a.s.*

(a) *Y is not d-collapsible. Moreover, its core $\tilde{Y}$ contains a constant fraction of the $(d-1)$-faces:*

$$f_{d-1}(\tilde{Y}) = \Psi_2(c(1-t)^d)\binom{n}{d}(1+o(1)) \quad , \quad f_d(\tilde{Y}) = \frac{c}{n}\binom{n}{d+1}(1-t)^{d+1}(1+o(1)). \tag{2}$$

In particular, $f_d(\tilde{Y}) < f_{d-1}(\tilde{Y})$.

(b) *Either Y is d-acyclic or $H_d(Y; \mathbb{R})$ is generated by $O_d(1)$ vertex disjoint $\partial\Delta_{d+1}$'s.*

(III) *The cyclic regime: If $c > c_d$ then a.a.s. $H_d(Y; \mathbb{R})$ is non-trivial. Furthermore, $f_{d-1}(\tilde{Y})$ and $f_d(\tilde{Y})$ still satisfy equation (2), but in this regime $f_{d-1}(\tilde{Y}) < f_d(\tilde{Y})$ and*

$$\beta_d(Y) = \left(\frac{c}{d+1}(1-t)^{d+1} - (1-t) + ct(1-t)^d\right)\binom{n}{d}(1+o(1)) = \left(f_d(\tilde{Y}) - f_{d-1}(\tilde{Y})\right)(1+o(1)).$$
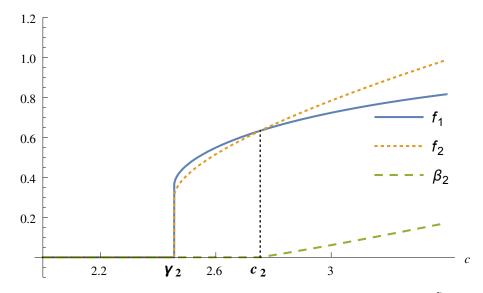

Figure 1: Illustration of Theorem 1.1 for $d = 2$. Here f_1, f_2 are the face numbers of $\tilde{Y}$ and β_2 is the second Betti number. The functions are normalized by $\binom{n}{2}$, and $n \to \infty$.

It is not hard to determine the asymptotic behaviour (in d) of these expressions, namely,

$$c_d = (d+1)(1 - e^{-(d+1)}) + O_d(d^3 e^{-2d})$$

and

$$\gamma_d = (1 + o_d(1))\ln d.$$

3

Consequently, there is a wide range of the parameter $p = p(d, n)$ for which almost all the complexes in $Y_d(n, p)$ are acyclic and non-collapsible.

Note that $f_d(\tilde{Y}) - f_{d-1}(\tilde{Y}) > 0$ implies that $H_d(Y; \mathbb{R}) \neq 0$. Moreover, this difference of the face numbers is a lower bound for the d-th Betti number. Theorem 1.1 shows that for all $0 \leq p \leq 1$, and up to the appearance of $\partial \Delta_{d+1}$'s these two conditions are typically equivalent and the lower bound is asymptotically tight. This is clearly a probabilistic statement which does not hold in general.

We turn to deal with the emergence of the giant component, a subject on which there exists an extensive body of literature. As mentioned above, there is no obvious high-dimensional counterpart to the notion of connected components, and we need a conceptual idea in order to even get started. The notion of *shadows*, introduced in [13], offers a way around this difficulty. The idea is to tie connected components with cycles, which do have natural high dimensional counterparts. The shadow of a graph G is the set of those edges that are not in G, whose addition to G creates a new cycle. It turns out that the giant component emerges exactly when the shadow of the evolving random graph acquires positive density. In particular, for $c > 1$ the shadow of $G(n, \frac{c}{n})$ has density $(1 - t)^2 + o(1)$, where t is the unique root in $(0, 1)$ of $t = e^{-c(1-t)}$ (See Figure 2a).

This suggests how we should define $\mathrm{SH}_{\mathbb{R}}(Y)$, the shadow of Y, a d-dimensional complex with full $(d-1)$-skeleton. Namely, it is the following set of d-faces:

$$\mathrm{SH}_{\mathbb{R}}(Y) = \{\sigma \notin Y : H_d(Y; \mathbb{R}) \text{ is a proper subspace of } H_d(Y \cup \{\sigma\}; \mathbb{R})\}.$$

In words, a d-face belongs to $\mathrm{SH}_{\mathbb{R}}(Y)$ if it is not in Y and its addition to Y creates a new d-cycle.

We are considering throughout the vanishing of the d-th homology and d-collapsibility. These two notions capture acyclicity from an algebraic, respectively, combinatorial, perspective. For $d = 1$ the two coincide, but they differ widely for $d \geq 2$. This dual perspective carries over to two notions of shadows. A d-face σ that does not belong to a d-complex Y is in Y's $\mathbb{R}$-shadow if its addition to Y increases the d-homology. It is in the *C-shadow* of Y if its addition to Y increases the core. Again, these notions coincide for $d = 1$, and the $\mathbb{R}$-shadow is always contained in the C-shadow, but for $d \geq 2$ they may differ.

The notion of shadows lets us compare the phase transitions of random graphs and random complexes of higher dimensions, and a substantial qualitative difference reveals itself. While the density of the shadow of $G(n, p)$ undergoes a smooth transition around $p = 1/n$, when $d \geq 2$ both the C-shadow and the $\mathbb{R}$-shadow of $Y_d\left(n, \frac{c}{n}\right)$ undergo *discontinuous* first-order phase transitions at the critical points γ_d and c_d respectively.

Theorem 1.2. *Let $d \geq 2$ be an integer, $c > 0$ real, and $Y = Y_d\left(n, \frac{c}{n}\right)$.*

 (I) The collapsible regime: If $c < \gamma_d$ then a.a.s. $|SH_{\mathbb{R}}(Y)| \leq |SH_C(Y)| = \Theta(n)$.

 (II) The intermediate regime: If $\gamma_d < c < c_d$ then a.a.s. $|SH_{\mathbb{R}}(Y)| = \Theta(n)$, and

$$|SH_C(Y)| = \binom{n}{d+1}((1 - t)^{d+1} + o(1)).$$

 (III) The cyclic regime: If $c > c_d$ then a.a.s. the size of both $SH_{\mathbb{R}}(Y)$ and $SH_C(Y)$ is $\binom{n}{d+1}((1-t)^{d+1} + o(1))$.

An essential idea that is common to all these results is that in the range $p = \Theta(\frac{1}{n})$ many of the interesting properties of $Y_d(n, p)$ can be revealed by studying its *local* structure. Initially, this seemed as merely a useful tool in studying the threshold for d-collapsibility, and in establishing an upper bound on the threshold of the vanishing of the d-th homology. However, in obtaining a lower bound on this threshold, it became apparent that this idea should be viewed in the wider context of *local weak limits*. This framework was introduced by Benjamini and Schramm [8] and Aldous and Steele [3]. In recent years, this approach was used in deriving new asymptotic results in various fields of mathematics (e.g. [1, 16]).

The study of d-collapsibility in random complexes was significantly influenced by work on k-cores in random hypergraphs and specifically the works by Molloy [18] and Riordan [20]. Also, the proof of Theorem 1.1

4

(a) Density of the shadow of $G(n, \frac{c}{n})$.

(b) Density of the C-shadow and $\mathbb{R}$-shadow of $Y_2\left(n, \frac{c}{n}\right)$.
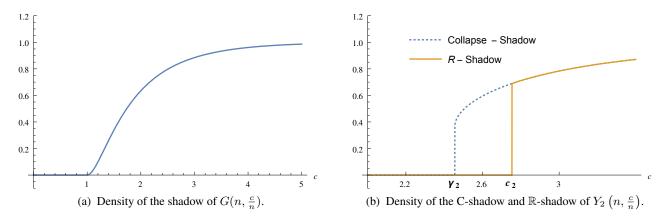
Figure 2: Illustration of Theorem 1.2 for $d = 2$, and comparison to the density of the shadow of a random graph.

makes substantial use of tools from the paper of Bordenave, Lelarge and Salez [9] on the rank of the adjacency matrix of random graphs.

The rest of the paper is organized as follows. Section 2 gives some necessary background material about simplicial complexes. In Section 3 we introduce the concept of a Poisson d-tree which is the local weak limit of random simplicial complexes. The main ingredients of the proofs of the main theorems are presented in Sections 4 and 5, that respectively addressing the subjects of collapsibility and acyclicity. Concluding remarks and open questions are presented in Section 6.

2   Preliminaries

A simplicial complex Y is a collection of subsets of its *vertex set* V that is closed under taking subsets. Namely, $\sigma \in Y$ and $\tau \subseteq \sigma$ imply that $\tau \in Y$ as well. Members of Y are called *faces* or *simplices*. The *dimension* of the simplex $\sigma \in Y$ is defined as $|\sigma| - 1$, and $\dim(Y)$ is defined as $\max \dim(A)$ over all faces $A \in Y$. A d-dimensional simplex is also called a d-simplex or a d-face, and a d-dimensional simplicial complex is also referred to as a d-complex. The set of j-faces in Y is denoted by Y_j, and the face numbers by $f_j(Y) := |Y_j|$. For $t < \dim(Y)$, the *t-skeleton* of Y is the simplicial complex that consists of all faces of dimension $\leq t$ in Y, and Y is said to have a *full t-dimensional skeleton* if its t-skeleton contains all the t-faces of V. In this paper, the *degree* $d_Y(\tau)$ of a face τ in a complex Y is the number of $\dim(Y)$-faces that contain it. A face of degree zero is said to be *exposed*. Although we are directly interested only in finite complexes, infinite ones do play a role here, but we consider only *locally-finite* complexes in which every face has a finite degree. We occasionally use the bipartite incidence graph between $(d-1)$-faces and d-faces of a d-complex Y. This allows us, in particular, to speak about *distances* among such faces.

The permutations on the vertices of a face σ are split in two *orientations*, according to the permutation's sign. The *boundary operator* $\partial = \partial_d$ maps an oriented d-simplex $\sigma = (v_0, ..., v_d)$ to the formal sum $\sum_{i=0}^{d}(-1)^i(\sigma^i)$, where $\sigma^i = (v_0, ...v_{i-1}, v_{i+1}, ..., v_d)$ 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 $\partial_d(Y)$ the d-dimensional boundary operator of a d-complex Y.

When Y is finite, we consider the $f_{d-1}(Y) \times f_d(Y)$ 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.

The d-th homology group $H_d(Y; R)$ (or vector space if R is a field) of a d-complex Y is the (right) kernel of its boundary operator ∂_d. Most of the homology groups in this paper are considered over $\mathbb{R}$. An element in $H_d(Y; \mathbb{R})$ is called a *d-cycle*, and the whole group is called the *d-cycle space* of Y. The *d-th Betti number* $\beta_d(Y; \mathbb{R})$ of a complex Y is defined to be the dimension of $H_d(Y; \mathbb{R})$.

Recall the concept of elementary collapse as defined in the introduction. A d-collapse phase is a procedure in which all the possible elementary d-collapses take place at once. In case more than one $(d-1)$-face can collapse some d-face, one of them is chosen by some predetermined arbitrary criteria. Given a d-complex Y, the complex $R_k(Y)$ is the complex that is obtained from Y after k phases of d-collapse. Similarly, $R_\infty(Y)$ is obtained after all possible d-collapse steps are carried out. A d-*core* (or core, for brevity) is a d-complex in which all the $(d-1)$-faces are of degree ≥ 2. The d-*core of* Y is the maximal d-core subcomplex of Y. Note that the d-core of Y is obtained from $R_\infty(Y)$ by removing the exposed $(d-1)$-faces.

3 Poisson d-tree

The concept of a Poisson d-tree process was introduced in [6] and turned out to be extremely useful in the study of random simplicial complexes. It can be viewed as a high-dimensional counterpart of the Poisson Galton-Watson process which plays a key role in the study of the giant component in $G(n, p)$ graphs.

A *rooted* d-*tree* is a pair (T, o) where T is a d-complex and o is some $(d-1)$-face of T. A d-tree is generated by the following process. Initially the complex consists of the $(d-1)$-face o. At every step $k \geq 0$, every $(d-1)$-face τ of distance k from o picks a non-negative number $m = m_\tau$ of new vertices $v_1, ..., v_m$, and adds the d-faces $v_1\tau, ..., v_m\tau$ to T.

We use some self-explanatory terminology in our study of d-trees. A *leaf* is a $(d-1)$-face with no *descendant* d-faces. If (T, o) is a rooted d-tree and T' is a subtree of T which contains the root o we refer to (T', o) as a *rooted subtree*. If τ is a $(d-1)$-face of T, the *branch of* T *rooted at* τ is the subtree that contains τ and all its descendants. A $(d-1)$-face τ is an *ancestor* of a $(d-1)$-face τ' if τ' belongs to the *branch rooted at* τ. The *depth* of a $(d-1)$-face is its distance from the root, and the depth of the d-tree is the maximal depth of any of its $(d-1)$-faces.

A *Poisson* d-*tree with parameter* c, denoted by $T_d(c)$, is a rooted d-tree in which all the numbers m_τ throughout this generative process are i.i.d. $Poi(c)$-distributed. The rooted subtree of $T_d(c)$ that consists of the first k generations of this process is denoted $T_{d,k}(c)$.

The most important fact about $T_d(c)$ in this context is that it approximates the local neighborhood of a $(d-1)$-face in $Y_d\left(n, \frac{c}{n}\right)$. This fact is well-known and very useful in the Erdős-Rényi random graphs.

Lemma 3.1. [6] *For every fixed integer* $k > 0$, *the* k-*neighborhood of a fixed* $(d-1)$-*face* τ *in* $Y = Y_d\left(n, \frac{c}{n}\right)$ *converges in distribution to the* k-*neighborhood of the root of* $T_d(c)$ *as* $n \to \infty$.

Proof. First, we observe that a.a.s. the k-neighborhood of a fixed $(d-1)$-face τ is a d-tree. Indeed, the violation of this statement requires that at least $k + 1$ d-faces are spanned in Y by τ's fixed vertices and k additional ones. The fact that this event is negligible follows from a first moment argument. In addition, conditioned on the k-neighborhood being a d-tree, the number m_η of new vertices that a $(d-1)$-face η adds to this d-tree is $\mathrm{Bin}\left(n - o(n), \frac{c}{n}\right)$-distributed, which tends to Poisson with parameter c as $n \to \infty$. ∎

This lemma easily implies convergence of $Y_d\left(n, \frac{c}{n}\right) \to T_d(c)$ in the sense of local weak convergence introduced by Benjamini and Schramm [8] and Aldous and Steele [3]. Here is a brief explanation of this concept. A rooted d-complex is a pair (Y, τ) of a d-complex and some $(d-1)$-face in it. We denote by $(Y, \tau)_k$ the τ-rooted subcomplex of (Y, τ) comprised of all the d-faces of distance at most k from τ and their subfaces. Let $\mathcal{Y}_d$ be the set of all (isomorphism types of) rooted d-complexes, equipped with the metric

$$\mathrm{dist}((Y, \tau), (Y', \tau')) = \inf\left\{\frac{1}{t+1} \; : \; (Y, \tau)_k \cong (Y', \tau')_k\right\}.$$

It can be easily verified that $(\mathcal{Y}_d, \mathrm{dist})$ is a separable and complete metric space, which comes, as usual, equipped with its Borel σ-algebra (See [2]). The fact that $Y_d\left(n, \frac{c}{n}\right)$ converges to $T_d(c)$ means that for every bounded and continuous function $f : \mathcal{Y}_d \to \mathbb{R}$,

$$\mathbb{E}_{Y=Y_d\left(n,\frac{c}{n}\right)}[f(Y, \tau)] \xrightarrow[n\to\infty]{} \mathbb{E}_{T=T_d(c)}[f(T, o)],$$

for every fixed $(d-1)$-face τ and the root o of T.

As we explain below, this fact will be applied directly to a function of particular interest in this context, namely, the degree of the root τ after k phases of τ-rooted collapse. In addition, it will be used in combination with the spectral theorem to bound the Betti numbers of $Y_d\left(n, \frac{c}{n}\right)$ with the spectral measure of the Poisson d-tree.

3.1 Rooted collapse

Let Y be a d-complex and τ some $(d-1)$-face of Y. A τ-*rooted collapse* of Y is a d-collapse process in which we forbid to collapse τ. Let k be a non-negative integer. The complex obtained from Y after k phases in the τ-rooted collapse process is denoted by $R_k(Y, \tau)$.

In the case $Y = Y_d\left(n, \frac{c}{n}\right)$, the degree $d_{R_k(Y,\tau)}(\tau)$ turns out to be relevant to several different questions. We approximate it using $\delta_k := d_{R_k(T,o)}(o)$, where $T = T_d(c)$, the Poisson d-tree with root o.

Lemma 3.2. *With the above notations*

$$\mathbb{E}[d_{R_k(Y,\tau)}(\tau)] \xrightarrow[n\to\infty]{} \mathbb{E}[\delta_k].$$

Note that this is not a direct corollary of the local weak convergence. Even though the function $d_{R_k(Y,\tau)}(\tau)$ is continuous, being dependent only on some fixed neighborhood of the root, it is not bounded. Nevertheless, we allow ourselves to omit the proof, since this difficulty can be bypassed by a simple calculus trick. Namely, by considering the function $\min\{d_{R_k(Y,\tau)}(\tau), A\}$, where A is a sufficiently large constant.

Lemma 3.3. *Let $c > 0$ and $(t_k)_{k \geq -1}$ a sequence of real numbers defined by*

$$t_{-1} = 0 \;, \quad t_{k+1} = e^{-c(1-t_k)^d}, \; \forall k \geq 0.$$

Then, δ_k is Poisson distributed with parameter $c(1 - t_{k-1})^d$, for every $k \geq 0$.

We refer throughout the paper to the sequences t_k of real numbers and δ_k of random variables that are defined here without denoting the underlying parameter $c > 0$ that is clear from the context.

Proof. By induction on k. The case $k = 0$ is trivial since δ_0 is Poisson distributed with parameter c. For the induction step, let us consider the distribution of δ_k. A d-face σ that contains the root o survives k phases of rooted collapse if and only if each of its $(d-1)$-faces τ other than the root (there are d such τ's) is contained in a d-face other then σ after $k-1$ phases. This occurs if and only if τ has a positive degree in the branch of T rooted at τ after $k-1$ phases of τ-rooted collapse. Since this subtree is also a Poisson d-tree with parameter c, this occurs, by the induction hypothesis, with probability $\Pr[\delta_{k-1} > 0] = 1 - t_{k-1}$. Moreover, different branches of the tree are independent, so these events for different τ's and σ's are independent. Namely, the distribution of δ_k is a Binomial distribution with $\delta_0 = \text{Poi(c)}$ trials and success probability $(1 - t_{k-1})^d$. By a standard computation in probability theory, this implies that δ_k is Poisson distributed with parameter $c(1 - t_{k-1})^d$. $\blacksquare$

We say that *a rooted d-tree is collapsible* if its root gets exposed in the rooted collapse process. For instance, the previous lemma shows that the probability that $T_d(c)$ is collapsed after k phases is t_k, and the probability that $T_d(c)$ is collapsible is $t = t(c, d)$.

4 d-collapsibility

The behavior of the sequence (t_k) of Lemma 3.3 changes quite substantially when $c = \gamma_d$. A simple calculus exercise tells us that $\lim_{k\to\infty} t_k = t(c, d)$, which is defined in (1). In addition, $t = t(c, d)$ equals to 1 if $c < \gamma_d$, and if $c > \gamma_d$ then t is strictly smaller then 1.

In other words, if $c < \gamma_d$ then the root of $T_d(c)$ gets exposed after k collapse phases with probability $1 - o_k(1)$. Moreover, the expected degree of the root is $o_k(1)$. On the other hand, if $c > \gamma_d$ then, for arbitrarily large k, with probability bounded away from zero some d-faces that contain the root will survive the collapse process.

How do these facts reflect on the behavior of the random simplicial complex $Y = Y_d\left(n, \frac{c}{n}\right)$ under d-collapse phases? Many parameters of $R_k(Y)$ can be understood almost directly from the τ-rooted collapse of Y, where τ is a typical $(d-1)$-face. Moreover, the Poisson d-tree plays a key role here since k phases of τ-rooted collapse depend only on the k-neighborhood of τ. Consequently, as k grows, almost all $(d-1)$-faces of Y will either collapse or become exposed in $R_k(Y)$ if $c < \gamma_d$. On the other hand, if $c > \gamma_d$, a constant fraction of the $(d-1)$-faces survive k phases of d-collapse, but only very few of them remain with degree 1, giving the collapse process a slim chance to continue much further. In fact, the fraction of the $(d-1)$-faces that are contained in Y's core is asymptotically approximated by the probability that the root of $T_d(c)$ has degree ≥ 2 after infinitely many collapse phases.

While the transition from the Poisson d-tree to the random simplicial complex is straightforward in the subcritical regime, in the supercritical regime we follow an involved argument of Riordan [20] for k-cores of random graphs. The reader is encouraged to read the introduction of Riordan's paper for an intuitive discussion of the proof method.

4.1 The collapsible regime - Theorem 1.1 (I)

Let $Y = Y_d\left(n, \frac{c}{n}\right)$, $c < \gamma_d$, and let τ be some $(d-1)$-face in Y.

$$
\begin{aligned}
\mathbb{E}[f_d(R_\infty(Y))] &\leq \mathbb{E}[f_d(R_k(Y))] \\
&= \frac{1}{d+1}\mathbb{E}\left[\sum_{\tau \in Y_{d-1}} \mathbf{1}_{\tau \in R_k(Y)} \cdot d_{R_k(Y)}(\tau)\right] \\
&= \frac{1}{d+1}\binom{n}{d}\mathbb{E}\left[\mathbf{1}_{\tau \in R_k(Y)} \cdot d_{R_k(Y)}(\tau)\right] \qquad (3) \\
&\leq \frac{1}{d+1}\binom{n}{d}\mathbb{E}\left[d_{R_k(Y,\tau)}(\tau)\right] \qquad (4) \\
&= \frac{1}{d+1}\binom{n}{d}(1+o(1))\mathbb{E}[\delta_k] \\
&= \frac{1}{d+1}\binom{n}{d}(1+o(1))c(1-t_{k-1})^d.
\end{aligned}
$$

Identity (3) is obtained by considering some fixed $(d-1)$-face τ, using linearity of expectation and symmetry. The subsequent inequality (4) is due to the fact that in the τ-rooted collapse process fewer collapses occur than in d-collapse phases, whence an inequality $d_{R_k(Y)}(\tau) \leq d_{R_k(Y,\tau)}(\tau)$ between τ's degrees after k phases in either d-collapse processes. The following equations are straightforward applications of the lemmas in Section 3.

Consequently, $f_d(R_\infty(Y)) = o(n^d)$ a.a.s. for every $c < \gamma_d$.

The argument which completes the proof says that for every $c > 0$, the complex $Y_d\left(n, \frac{c}{n}\right)$ has no core subcomplex with $o(n^d)$ d-faces, other than vertex disjoint $\partial\Delta_{d+1}$'s. This is proved in Theorem 4.1 of [6], concerning inclusion-minimal core complexes. It turns out that a slight modification of that proof yields a more general conclusion.

A d-complex whose d-faces are comprised of a vertex-disjoint union of boundaries of $(d+1)$-simplices is called here a d-*gravel*.

Lemma 4.1. *For every integer $d \geq 2$ and real $c > 0$ real there is $\alpha > 0$ such that a.a.s. the following holds. Let $Y = Y_d\left(n, \frac{c}{n}\right)$, then either $R_\infty(Y)$ is a d-gravel or $f_d(R_\infty(Y)) > \alpha n^d$.*

Proof. Let $m_1 := (d^3 \log n)^d$. Our first goal is to show that every core d-subcomplex C of Y with $f_d(C) \leq m_1$ is a d-gravel. A simple first moment argument yields that Y cannot contain two intersecting copies of $\partial\Delta_{d+1}$, nor can it contain more than $\log\log n$ copies of $\partial\Delta_{d+1}$. A core d-complex C that is comprised of exactly l vertex disjoint $\partial\Delta_{d+1}$'s and m additional d-faces is said to have *type* (l, m).

Let C have type (l, m). We partition its vertex set into $S \dot\cup T$, where S is the set of the vertices in some $\partial\Delta_{d+1}$ of C. Let $T' \subseteq T$ be those vertices in T of degree $d + 1$ in C (i.e., such a vertex is in exactly $d + 1$ of C's d-faces). Since C is a core, the degree of every vertex in C is at least $d + 1$.

In addition, every d-face of C with two or more vertices of degree $d + 1$ is included in a $\partial\Delta_{d+1}$. Recall the notion of a *link of a vertex* v in a simplicial complex Y. Namely, $\mathrm{lk}_Y(v) = \{\tau \in Y : v \dot\cup \tau \in Y\}$. In particular, the link of a vertex in a d-core is a $(d - 1)$-core. Therefore, if a vertex has degree $d + 1$ in C then its link is a $\partial\Delta_d$. Suppose that the vertices $v, u \in C$ have degree $d + 1$ in C and are contained in a common d-face $\sigma = (u, v, x_1, ..., x_{d-1})$. Their links are $\partial\Delta_d$'s so there exist vertices $u', v' \notin \sigma$ such that $\mathrm{lk}_C(u) = \partial(u', v, x_1, ..., x_{d-1})$ and $\mathrm{lk}_C(v) = \partial(u, v', x_1, ..., x_{d-1})$. This can occur only if $u' = v'$ and the claim follows.

As a result, every non-gravel d-face contains at most one vertex of T', so that $m \geq |T'|(d + 1)$. Counting incidences of vertices and d-faces in C yields

$$\big(m + l(d + 2)\big) \cdot (d + 1) \geq (|S| + |T|)(d + 1) + (|T| - |T'|).$$

But $|S| = l(d + 2)$, and a simple manipulation of these inequalities gives $|T| \leq m \cdot \frac{d+3}{d+4}$. We can assume w.l.o.g. that $m \leq m_1$, $l \leq \log\log n$ and derive the following upper bound on the number of type (l, m) core d-complexes with at most n vertices

$$n^{l(d+2)+\frac{d+3}{d+4}m} \cdot (|S| + |T|)^{(d+1)m} = n^{l(d+2)} \cdot \left[n^{\frac{d+3}{d+4}} \cdot O(\log^{d(d+1)} n)\right]^m.$$

The first term counts the choices for S, T, and the second the choice of non-gravel d-faces. We conclude that a.a.s. Y contains no core of type (l, m) with $l < \log\log n$ and $0 < m \leq m_1$. This is because any subcomplex of type (l, m) appears in Y with probability $(c/n)^{l(d+2)+m}$.

The proof of Theorem 4.1 in [6] yields a constant $\alpha = \alpha(c, d)$ such that a.a.s. $Y_d\left(n, \frac{c}{n}\right)$ has no inclusion-minimal subcomplex that is a core with $m_1 \leq m \leq \alpha n^d$ d-faces. In fact, the argument presented at [6] only uses the fact that a minimal core C is *connected* in the sense that between every two $(d - 1)$-faces τ, τ' in C there is a path alternating between $(d - 1)$-faces and d-faces of C with an inclusion relation. However, since every core is a union of connected cores, this means that there are no cores of size m in Y. It follows that the only possible cores that Y can contain have type $(l, 0)$, i.e., it is a d-gravel. ∎

4.2 The core of $Y_d\left(n, \frac{c}{n}\right)$ - Theorem 1.1 (II.a)

The proof of this theorem closely follows the argument of Riordan [20]. We fix the dimension d and refer to $r(c) := 1 - t(c, d)$ as a function of c. For brevity we denote $r^+(c) := \Psi_2(cr(c)^d)$. Note that both $r(c)$ and $r^+(c)$ are continuous, bounded away from 0 and increasing when $c > \gamma_d$. Our main goal is to show that for every $\tilde c > \gamma_d$ and $\varepsilon > 0$, $f_{d-1}(\tilde Y) > (r^+(\tilde c) - \varepsilon)\binom{n}{d}$, where $\tilde Y$ is the d-core of $Y_d\left(n, \frac{\tilde c}{n}\right)$.

This is motivated by the fact that $r^+(\tilde c)$ is the probability that the root's degree is ≥ 2 after every finite number of rooted collapse phases in $T_d(\tilde c)$. In other words, this is the probability that the root survives the non-rooted collapse process. Although this argument is simple and appealing, the actual proof is substantially more involved. Our strategy is to define some carefully crafted property $\mathcal{A}$ of d-trees of depth $\log\log n \ll S = S(n) \ll \log n$ such that the following two statements hold a.a.s. First, the subset $A \subset Y_{d-1}$ of $(d - 1)$-faces τ such that Y contains a τ-rooted d-tree with property $\mathcal{A}$ is of density at least $(r^+(\tilde c) - \varepsilon)$. Second, for every $\tau \in A$ there exists a d-tree $T_\tau \subset Y$ in which τ's degree is at least 2 and every leaf also belongs to A. Consequently, no $(d - 1)$-face in A can be collapsed.

We refer throughout the proof to certain properties of rooted d-trees (T, o), and occasionally write $T \in \mathcal{P}$ to say that T has property $\mathcal{P}$. Every property $\mathcal{P}$ of rooted d-trees induces a property of $(d-1)$-faces in a d-complex Y. Namely, we say that the $(d-1)$-face τ *has property* $\mathcal{P}$ if Y contains a d-tree rooted on τ that has property $\mathcal{P}$.

Here are some relevant properties: $\mathcal{D}_{\leq L}$ means T has depth $\leq L$, and $\mathcal{D}_{<\infty}$ means that it has finite depth. Let $\mathcal{P}_1$ and $\mathcal{P}$ be properties of finite (resp. general) d-trees. A d-tree (T, o) has property $\mathcal{P}_1 \circ \mathcal{P}$ when: (i) T has a finite rooted subtree T' with property $\mathcal{P}_1$, and (ii) For every leaf τ of T', the branch of T rooted at τ has property $\mathcal{P}$. For example, we consider the property that T has depth $k+1$ and it does not collapse in k phases, i.e., $\mathcal{B}_k := \{T \in \mathcal{D}_{\leq k+1} \mid \delta_k(T) > 0\}$ and note that $\mathcal{B}_k = \mathcal{B}_0 \circ \mathcal{B}_{k-1}$. Property $\mathcal{B}$ means that T does not collapse at finite time.

We also define for $k \geq 0$, the properties $\mathcal{R}_k$ which are stronger than $\mathcal{B}_k$ as follows

$$\mathcal{R}_0 = \{T \in \mathcal{D}_{\leq 1} \mid \delta_0(T) \geq 2\}, \quad \mathcal{R}_k = \mathcal{R}_0 \circ \mathcal{B}_{k-1} \cup \mathcal{B}_0 \circ \mathcal{R}_{k-1}, \ k > 0.$$

The difference between $\mathcal{B}_k$ and $\mathcal{R}_k$ is this: $T \in \mathcal{B}_k$ means that T has depth $k+1$ and every non-leaf $(d-1)$-face has at least one descendant d-face. In defining $\mathcal{R}_k$ we add the requirement that along every root-to-leaf path we encounter at least one $(d-1)$-face with ≥ 2 descendant d-faces.

Finally, we introduce a stochastic version of $\mathcal{P}$, a property of finite rooted d-trees (T, o). For some $0 \leq p \leq 1$ we *mark* each leaf of T independently with probability p, and remove every d-face that contains any unmarked leaf. We say that the event $\mathcal{M}(\mathcal{P}, p, T)$ holds if the remaining d-tree has property $\mathcal{P}$. Marking is a convenient way of capturing the following phenomenon: We let each leaf in a finite d-tree grow a Poisson d-tree and we only ask whether or not this "tail" has some desired property. This is expressed in the simple identity

$$\Pr[\mathcal{M}(\mathcal{P}_1, p, T_{d,k}(c))] = \Pr[T_d(c) \in \mathcal{P}_1 \circ \mathcal{P}] \tag{5}$$

where d, k are integers, $c > 0$, $\mathcal{P}_1$ is a property of depth-k trees and $\mathcal{P}$ is a property of probability p for Poisson d-trees with parameter c. For instance,

$$\Pr[\mathcal{M}(\mathcal{B}_k, r(c), T_{d,k+1}(c))] = \Pr[T_d(c) \in \mathcal{B}] = r(c)$$

The following lemma can be viewed as a variation on this identity. It shows that although property $\mathcal{R}_k$ is stronger than $\mathcal{B}_k$, the two are almost equally likely in a Poisson d-tree.

Lemma 4.2. *For every $c > c_1 > \gamma_d$ there is a sufficiently large k such that*

$$\Pr[\mathcal{M}(\mathcal{R}_k, r(c_1), T_{d,k+1}(c))] > r(c_1).$$

Proof. Consider the following probabilistic experiment where we randomize thrice. Initially we generate the first $k+1$ generations of $T_d(c)$. Then we do the random marking that yields the d-tree T. Finally we remove each d-face of T independently with probability c_1/c. We denote the component of the root by T'. Note that T' is distributed like a $T_{d,k+1}(c_1)$ to which random $r(c_1)$-marking is applied. In particular, $\Pr[T' \in \mathcal{B}_k] = r(c_1)$, hence we need to prove that $\Pr[T \in \mathcal{R}_k] > \Pr[T' \in \mathcal{B}_k]$. Since $T' \in \mathcal{B}_k$ implies that $T \in \mathcal{B}_k$ it suffices to show that $\Pr[W] > \Pr[L]$, where

$$W = [T \in \mathcal{R}_k, \ T' \notin \mathcal{B}_k] \ \text{and} \ L = [T \in \mathcal{B}_k \setminus \mathcal{R}_k, \ T' \in \mathcal{B}_k].$$

To this end we show that $\Pr[L] \to 0$ as $k \to \infty$ whereas $\Pr[W]$ stays bounded away from 0.

Indeed, if $T \in \mathcal{B}_k \setminus \mathcal{R}_k$, then there exists some d-face of depth k whose removal violates property $\mathcal{B}_k$. This d-face survives in T' with probability $(c_1/c)^k$, so that $\Pr[L] < (c_1/c)^k$. On the other hand W contains the event that $\delta_0(T) = 2$, $T \in \mathcal{R}_0 \circ \mathcal{B}_{k-1}$ and $\delta_0(T') = 0$ whose probability is positive and independent of k. ∎

A d-tree T of depth $L+1$ is *(p, η)-rigid* if $\Pr[\mathcal{M}(\mathcal{R}_L, p, T)] > 1 - \eta$.

Lemma 4.3. *For every $c > c_1 > \gamma_d$ and $\eta > 0$ and for a large enough integer L there holds*

$$\Pr[T_{d,L+1}(c) \text{ is } (r(c), \eta)\text{-rigid}] \geq r(c_1).$$

Proof. Below we assume that k is large enough, as required in Lemma 4.2. We claim that

$$\Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}] = \Pr[\mathcal{M}(\mathcal{R}_k, r(c), T_{d,k+1}(c))] \geq \Pr[\mathcal{M}(\mathcal{R}_k, r(c_1), T_{d,k+1}(c))] > r(c_1).$$

The equality is a special case of identity (5). The first inequality follows by a simple monotonicity consideration and the fact that $r(c) > r(c_1)$. The last inequality comes from Lemma 4.2. The condition of (p, η)-rigidity is stronger the smaller η is, so we fix it to satisfy

$$\Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}] - r(c_1) > \eta.$$

For fixed k the conditions $T_{d,k+l+2}(c) \in \mathcal{R}_k \circ \mathcal{B}_l$ become more strict as l grows and their conjunction over all $l \geq 0$ is exactly the condition $T_d(c) \in \mathcal{R}_k \circ \mathcal{B}$. Therefore, we can and will choose l large enough so that

$$\Pr[T_{d,k+l+2}(c) \in \mathcal{R}_k \circ \mathcal{B}_l] \leq \Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}] + \eta^2.$$

For a d-tree T of depth $L + 1 = k + l + 2$, let $\phi(T) := \Pr[\mathcal{M}(\mathcal{R}_k \circ \mathcal{B}_l, r(c), T)]$. We denote by $\mathcal{L}$ the property that T is $(r(c), \eta)$-rigid. The expectation of $\phi(T)$, where T is $T_{d,L+1}(c)$ distributed, equals the probability $\Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}]$. In addition, $\phi(T) = 0$ if $T \notin \mathcal{R}_k \circ \mathcal{B}_l$ and $\phi(T) \leq 1 - \eta$ if $T \notin \mathcal{L}$ since $\mathcal{R}_k \circ \mathcal{B}_l$ implies $\mathcal{R}_L$. Therefore

$$\Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}] \leq \Pr[T_{d,L+1}(c) \in (\mathcal{R}_k \circ \mathcal{B}_l) \cap \mathcal{L}] + (1 - \eta)\Pr[T_{d,L+1}(c) \in (\mathcal{R}_k \circ \mathcal{B}_l) \setminus \mathcal{L}],$$

whence

$$\eta \cdot \Pr[T_{d,L+1}(c) \in \mathcal{R}_k \circ \mathcal{B}_l \setminus \mathcal{L}] \leq \Pr[T_{d,L+1}(c) \in \mathcal{R}_k \circ \mathcal{B}_l] - \Pr[T_d(c) \in \mathcal{R}_k \circ \mathcal{B}].$$

Putting everything together we conclude that $\Pr[T_{d,L+1}(c) \in \mathcal{L}] \geq \Pr[T_{d,L+1}(c) \in \mathcal{R}_k \circ \mathcal{B}_l] - \eta > r(c_1)$, as stated. ∎

We set all the parameters that appear in the discussion below. Recall that our goal is to show that for every $\tilde{c} > \gamma_d$ and $\varepsilon > 0$, $f_{d-1}(\tilde{Y}) > (r^+(\tilde{c}) - \varepsilon)\binom{n}{d}$. Let $\gamma_d < c_1 < c < \tilde{c}$ such that $r^+(c_1) \geq r^+(\tilde{c}) - \varepsilon/2$ and $\Psi_1(\tilde{c}r(c_1)^d) > r(c)$. Again we choose k large enough to make Lemma 4.2 hold, and we fix some $0 < \eta < d^{-2(k+1)}/8$. Also L is so large that Lemma 4.3 holds. Recall that $\mathcal{L}$ denotes the property $(r(c), \eta)$-rigidity of d-trees of depth $L + 1$.

Consider an integer $S = S(n)$ such that $\log\log n \ll S \ll \log n$, and we define for $0 \leq s \leq S$, the properties $\mathcal{A}_s$ by

$$\mathcal{A}_0 = \mathcal{L}, \quad \mathcal{A}_s = \mathcal{R}_k \circ \mathcal{A}_{s-1}, \ 0 < s \leq S.$$

Note that $\mathcal{A}_S$ depends on the first $Q := (k+1)S + L + 1$ generations. Finally, we define two key properties:

$$\mathcal{A} = \mathcal{R}_0 \circ \mathcal{D}_{\leq L} \circ \mathcal{A}_S, \quad \mathcal{P} = \mathcal{R}_0 \circ \mathcal{D}_{<\infty} \circ \mathcal{A}.$$

In words, $T \in \mathcal{A}$ means that T has a rooted subtree of depth $\leq L + 1$ in which the root's degree is ≥ 2 and the branch of T rooted at each of its leaves has property $\mathcal{A}_S$. $T \in \mathcal{P}$ means that T has a finite rooted subtree in which the root's degree is ≥ 2 and the branch of T rooted at each of its leaves has property $\mathcal{A}$.

It follows by induction that $\Pr[T_{d,(k+1)s+L+1} \in \mathcal{A}_s] > r(c_1)$ for every $s \geq 0$. Indeed, Lemma 4.3 yields the case $s = 0$, and the inductive step follows from Lemma 4.2. Also, since $\mathcal{R}_0 \circ \mathcal{A}_S$ implies $\mathcal{A}$, the probability that $T_d(c)$ has a rooted subtree that satisfies $\mathcal{A}$ is at least

$$\Pr[T_{d,Q+1}(c) \in \mathcal{R}_0 \circ \mathcal{A}_S] \geq \Psi_2(cr(c_1)^d) \geq \Psi_2(c_1 r(c_1)^d) = r^+(c_1). \tag{6}$$

Indeed, the probability that in a d-face which contains the root all the branches rooted at a $(d-1)$-face of depth 1 has $\mathcal{A}_S$ is at least $r(c_1)^d$. Hence, the number of such d-faces is $\mathrm{Poi}(cr(c_1)^d)$ distributed.

We come to the most significant step in the proof: We show, by a first-moment argument, that a.a.s. every $(d-1)$-face in $Y = Y_d\left(n, \frac{\tilde{c}}{n}\right)$ that has property $\mathcal{A}$ also has $\mathcal{P}$. This implies that every such $(d-1)$-face τ is contained in a d-tree $T_\tau \subset Y$ in which (i) τ is of degree ≥ 2, and (ii) every leaf of T_τ has property $\mathcal{A}$. Consequently, the union of these d-trees $\{T_\tau \mid \tau$ has property $\mathcal{A}\}$ is contained in the core $\tilde{Y}$.

Claim 4.4. *For every $s \geq 0$, and every $T \in \mathcal{A}_s$,*

$$\Pr[\mathcal{M}(\mathcal{B}_{(k+1)s} \circ \mathcal{R}_L, r(c), T)] \geq 1 - \frac{2^{-2^s}}{4d^{2(k+1)}}.$$

Proof. We proceed by induction on s. Our definition of $\mathcal{L}$ and the choice of η yield the case $s = 0$. Let $s \geq 0$ be an integer, and $T \in \mathcal{A}_{s+1} = \mathcal{R}_k \circ \mathcal{A}_s$. Let $T' \subset T$ be a *minimal* d-tree of depth $k+1$ such that $T' \in \mathcal{R}_k$ and every branch of T rooted at its leaves has the property $\mathcal{A}_s$. By a straightforward computation, T' has exactly $2d^{k+1}$ leaves. By induction, after the marking process the branch rooted at every leaf of T' fails to have property $\mathcal{B}_{(k+1)s} \circ \mathcal{R}_L$ independently with probability at most $\frac{2^{-2^s}}{4d^{2(k+1)}}$. Let us refer to such a leaf as *bad*, and remove from T' every d-face that contains a bad leaf. Since initially T' had the property $\mathcal{R}_k$, it now has property $\mathcal{B}_k$ unless at least 2 d-faces were removed. But this can only occur if at least 2 leaves are bad, an event of probability at most

$$\binom{2d^{k+1}}{2}\left(\frac{2^{-2^s}}{4d^{2(k+1)}}\right)^2 \leq \frac{2^{-2^{s+1}}}{4d^{2(k+1)}}.$$

Namely, the tree T has the property $\mathcal{B}_k \circ \mathcal{B}_{(k+1)s} \circ \mathcal{R}_L = \mathcal{B}_{(k+1)(s+1)} \circ \mathcal{R}_L$ with the desired probability. ∎

This leads to the following key lemma.

Lemma 4.5. *For every fixed $(d-1)$-face τ in $Y = Y_d\left(n, \frac{\tilde{c}}{n}\right)$, the probability that τ has property $\mathcal{A}$ but does not have $\mathcal{P}$ is $o(n^{-d})$.*

Proof. It is easy to show that with probability $o(n^{-d})$ the $(1 + L + 2Q)$-neighborhood of τ consists of at most $n^{1/3}$ vertices, and we condition on this event. Suppose that τ has the property $\mathcal{A}$, and consider $\tau \in T \subset Y$ such that $T \in \mathcal{A}$. In particular, there exists a d-tree $T' \subset T$ rooted at τ of depth at most $L+1$ such that $d_{T'}(\tau) = 2$ and every branch $T''_\pi \subset T$ rooted at a leaf π of T' has property $\mathcal{A}_S$. Denote by $X \subset Y_{d-1}$ the union of the leaves of T''_π over all the leaves π of T'.

We now expose an additional subset of the $(Q + 1)$-neighborhoods of the $(d-1)$-faces of X with the following precaution. When we reach some $(d-1)$-face ρ and query whether a d-face that contains it belongs to Y, we only consider d-faces of the form $v\rho$ where v is a vertex that does not belong to T nor did it appear in the exposing process upto the current query. In this manner, every $\rho \in X$ is the root of a d-tree $\tilde{T}_\rho \subset Y$ in which every $(d-1)$-face has at least $\mathrm{Bin}(n - n^{1/3}, \frac{\tilde{c}}{n})$ descendants. Therefore, the probability that $\tilde{T}_\rho \in \mathcal{B}_0 \circ \mathcal{A}_S$ is at least

$$\Pr[T_{d,Q}(\tilde{c}) \in \mathcal{B}_0 \circ \mathcal{A}_S] - o(1) \geq \Psi_1(\tilde{c}r(c_1)^d) - o(1) > r(c),$$

and these events are independent over $\rho \in X$. Therefore we can consider the event $\tilde{T}_\rho \in \mathcal{B}_0 \circ \mathcal{A}_S$ as an alternative for marking the leaves of d-trees T''_π and plug it in Claim 4.4. Since the number of leaves of T' is bounded, the claim implies that with probability $1 - O\left(2^{-2^S}\right) = 1 - o(n^{-d})$, after the described exposure of the additional neighborhoods, all the subtrees rooted in these leaves have the property $\mathcal{B}_{(k+1)S} \circ \mathcal{R}_L \circ \mathcal{B}_0 \circ \mathcal{A}_S$. But recall that $\mathcal{R}_L$ means that in every path from the root of the d-tree to a $(d-1)$-face of depth $L+1$, there is a $(d-1)$-face with at least two descendants. In other words, $\mathcal{R}_L \circ \mathcal{B}_0 \circ \mathcal{A}_S$ implies $\mathcal{D}_{\leq L} \circ \mathcal{R}_0 \circ \mathcal{D}_{\leq L} \circ \mathcal{A}_S = \mathcal{D}_{\leq L} \circ \mathcal{A}$. In particular, all the leaves of T' have the property $\mathcal{D}_{<\infty} \circ \mathcal{A}$, and since $d_{T'}(\tau) = 2$, it follows that τ has the property $\mathcal{P}$. ∎

12

We are now ready to prove the theorem.

Proof of Theorem 1.1 (II.a). Let $N_{\mathcal{A}}$ denote the number of $(d-1)$-faces in $Y = Y_d\left(n, \frac{\tilde{c}}{n}\right)$ that have property $\mathcal{A}$. By the previous discussion, we know that $f_{d-1}(\tilde{Y}) \geq N_{\mathcal{A}}$. We approximate the expectation of $N_{\mathcal{A}}$ by $\Pr_{\tilde{c}}[\mathcal{A}]$, the probability that $T_d(\tilde{c})$ has a rooted subtree that satisfies $\mathcal{A}$,

$$\mathbb{E}[N_{\mathcal{A}}] = (\Pr_{\tilde{c}}[\mathcal{A}] + o(1))\binom{n}{d} \geq r^+(c_1)\binom{n}{d} \geq (r^+(\tilde{c}) - \varepsilon/2)\binom{n}{d}$$

by Equation (6). Since $\mathcal{A}$ depends only on the $O(S)$-neighborhood of the $(d-1)$-face, and two $(d-1)$-faces have non-disjoint neighborhoods with negligible probability, it follows that $\mathbb{E}[N_{\mathcal{A}}^2] = \mathbb{E}[N_{\mathcal{A}}]^2(1 + o(1))$. By the second moment method $\Pr[N_{\mathcal{A}} < (r^+(\tilde{c}) - \varepsilon)\binom{n}{d}] = o(1)$. The upper bound is much simpler. Let N_k denote the number of $(d-1)$-faces that survive (= did not collapse nor become free) the first k phases of collapse. Clearly, $f_{d-1}(\tilde{Y}) \leq N_k$ for every k. Similarly to $N_{\mathcal{A}}$, this property depends on the k-neighborhood of a $(d-1)$-face and by the same argument as before, N_k is concentrated around its expectation. The expectation of N_k can be bounded by the Poisson d-tree as follows.

$$\mathbb{E}[N_k] \leq (\Pr_{\tilde{c}}[\delta_{k-1} \geq 2] + o(1))\binom{n}{d} = (\Psi_2(\tilde{c}(1 - t_{k-2})^d) + o(1))\binom{n}{d},$$

and since $t_k \to t$, we obtain that $\Psi_2(\tilde{c}(1 - t_{k-2})^d)$ tends to $r^+(\tilde{c})$ as $k \to \infty$.

We turn to prove that a.a.s. the number of d-faces in the core $f_d(\tilde{Y}) = r(\tilde{c})^{d+1}\frac{\tilde{c}}{n}\binom{n}{d+1}(1 + o(1))$. Let $M_{\mathcal{A}}$ denote the number of d-faces in Y all of whose $(d-1)$-faces have property $\mathcal{A}$. Clearly, $f_d(\tilde{Y}) \geq M_{\mathcal{A}}$ since no $(d-1)$-face with property $\mathcal{A}$ is collapsed. In addition, since this is a local property it suffices, as before, to compute the expectation of $M_{\mathcal{A}}$. The probability that a d-simplex σ belongs to Y is $\frac{\tilde{c}}{n}$, and we can expose a subset T of its neighborhood in the same careful fashion as done in the proof of Lemma 4.5. The probability that all the d-trees growing from σ's $(d-1)$-faces have property $\mathcal{A}_S$ is at least $\Pr_{\tilde{c}}[\mathcal{A}_S]^{d+1} - o(1) > r(c_1)^{d+1}$. If this occurs, then every $(d-1)$-face $\tau \subset \sigma$ has property $\mathcal{A}$ by letting τ be the root of the d-tree $T \cup \{\sigma\}$. Consequently $\mathbb{E}[M_{\mathcal{A}}] \geq r(c_1)^{d+1}\frac{\tilde{c}}{n}\binom{n}{d+1}$. The upper bound is proved similarly, by showing that the probability that a d-face survives the first k collapse phases tends to $r(\tilde{c})^{d+1}$ as k grows. ∎

4.3 C-shadow of $Y_d\left(n, \frac{c}{n}\right)$

Here we prove the parts of Theorem 1.2 that deal with the C-shadow of $Y_d\left(n, \frac{c}{n}\right)$. Namely, we show that for $c < \gamma_d$, the C-shadow of $Y = Y_d\left(n, \frac{c}{n}\right)$ has size $\Theta(n)$, and for $c > \gamma_d$ its size is

$$|\text{SH}_C(Y)| = \binom{n}{d+1}((1 - t)^{d+1} + o(1)).$$

Both statements follow directly from the previous proofs. Regarding the range $c < \gamma_d$, a simple second moment calculation shows that a.a.s. there are $\Theta(n)$ sets of $d + 2$ vertices in Y that span all but one of the d-faces in the boundary of a $(d + 1)$-simplex. The missing d-face in every such configuration is obviously in the C-shadow. On the other hand, if the C-shadow is large, viz., $|\text{SH}_C(Y)| \gg n$, then for every $c < c' < \gamma_d$, with probability bounded away from zero, the core of $Y_d\left(n, \frac{c'}{n}\right)$ contains a complex that is not the boundary of $(d + 1)$-simplex. But this contradicts Theorem 1.1(I).

We prove the supercritical case $c > \gamma_d$ in much the same way that we calculated the number of d-faces in the core. Namely, for the lower bound we count d-simplices not in Y all of whose $(d-1)$-faces have property $\mathcal{A}$. For the upper bound we count d-simplices that if added into Y do not survive k phases of collapse. As before, both properties are local and by a second moment argument are concentrated around their means, which are computed by Poisson d-tree approximations.

5 d-acyclicity

In the previous section we saw that the threshold γ_d for d-collapsibility in $Y_d\left(n, \frac{c}{n}\right)$ coincides with the threshold in which rooted collapsibility in $T_d(c)$ almost surely eliminates all the d-faces containing the root. In the case of d-acyclicity, the correspondence is similar but more intricate. In fact, the threshold c_d for d-acyclicity coincides with *two* seemingly separate thresholds of $T_d(c)$'s parameters. These will used to bound the d-acyclicity threshold from below and above respectively. Since both occur at $c = c_d$ it follows that these bounds are tight. Furthermore, if $c > c_d$, these two parameters yield upper and lower bounds for $\beta_d(Y, \mathbb{R})$ which are tight upto small order error terms. Finally, the tight estimation for $\beta_d(Y, \mathbb{R})$ allows us to compute the density of the *shadow* of Y.

If a d-complex Y has more d-faces than $(d-1)$-faces, then $\beta_d(Y, \mathbb{R}) \geq f_d(Y) - f_{d-1}(Y) > 0$. For $Y = Y_d\left(n, \frac{c}{n}\right)$, this happens only when $c > d + 1$, but we can say a bit more. Even though Y and its core $\tilde{Y}$ have the same d-th Betti number, it turns out that there is a wider range of the parameter c for which $\tilde{Y}$ has more d-faces than $(d-1)$-faces. In fact, one can show that $f_d(\tilde{Y}) > f_{d-1}(\tilde{Y})$ if and only if $c > c_d$, using the expressions for these face numbers in Theorem 1.1(II.a). However, it is significantly easier to prove the same lower bound on $\beta_d(Y)$ by analyzing $T_d(c)$ as follows. Let $S_k(Y)$ be obtained by removing all exposed $(d-1)$-faces in $R_k(Y)$. The average degree of the $(d-1)$ faces in $S_k(Y)$ is approximated using the conditional expectation

$$\mathbb{E}[\delta_k \mid \delta_k > 0 \ \wedge \ \delta_{k-1} > 1].$$

In words, this is the expected degree of the root of $T_d(c)$ after k phases of rooted collapses, conditioned on the fact that its degree remains strictly greater than 1 throughout the collapse process. We claim that this captures the average degree of $(d-1)$-faces in $S_k(Y)$. A $(d-1)$-face τ of Y belongs to $S_k(Y)$ if and only if its degree after $k-1$ phases of τ-rooted collapse is larger than 1 and stays positive after one more phase. Indeed, as long as $d_\tau > 1$ the τ-rooted collapse and non-rooted collapse are identical. A difference occurs when $d_\tau = 1$, at which point the rooted collapse continues as usual, but the non-rooted collapse eliminates τ. If the average degree of $(d-1)$-faces exceeds $d + 1$, this yields, via a simple double-counting argument, a positive lower bound for $\beta_d(Y)$. A simple calculus exercise then shows that this condition holds if and only if $c > c_d$. Namely,

$$\lim_{k \to \infty} \mathbb{E}[\delta_k \mid \delta_k > 0 \ \wedge \ \delta_{k-1} > 1] > d + 1 \quad \Longleftrightarrow \quad c > c_d.$$

The most substantial role of local weak convergence is in proving the lower bound on the d-acyclicity threshold. We analyze $T = T_d(c)$ using tools from spectral theory and functional analysis. For this reason we are still unable to resolve this question over finite fields of coefficients. As further detailed below, we define $x_T := \pi_{L(T), e_o}(\{0\})$ to be the measure of the atom $\{0\}$ according to the *spectral measure* π of the *Laplacian* $L(T)$ of T with respect to the characteristic vector e_o of the root o. Local weak convergence implies that x_T is *an upper bound on the normalized dimension* of the left kernel Z of $\partial_d(Y)$. Note that if $Y_d\left(n, \frac{c}{n}\right)$ is d-acyclic then $\dim Z$ equals $(1 + o(1))\binom{n}{d}\left(1 - \frac{c}{d+1}\right)$, and otherwise it is greater. Indeed, the proof shows that

$$\mathbb{E}_{T=T_d(c)}[x_T] = 1 - \frac{c}{d+1} \quad \Longleftrightarrow \quad c < c_d,$$

and if $c > c_d$, this expectation is greater than $1 - \frac{c}{d+1}$.

5.1 Acyclicity beyond collapsibility - Theorem 1.1(II.b)

To prove that a random complex is d-acyclic beyond the d-collapsibility threshold, we cannot restrict ourselves to purely combinatorial arguments. It is not a-priori clear that the local weak limit of a random complex holds enough information to prove such a statement. Surprisingly, perhaps, this is the case when we work over $\mathbb{R}$. In this section we describe the main ingredients of this method, which appears in [14], where complete proofs can be found.

Let $Y = Y_d\left(n, \frac{c}{n}\right)$. The primary goal in the proof is to find a tight upper bound for $\lim_{n\to\infty} \frac{1}{\binom{n}{d}}\mathbb{E}[\beta_d(Y)]$. It turns out more useful to work with the corresponding Laplace operator $L(Y) = \partial_d(Y)\partial_d(Y)^*$. We consider its kernel Z which coincides with the left kernel of $\partial_d(Y)$. Let $P_Z : \mathbb{R}^{Y_{d-1}} \to Z$ be the orthogonal projection to the space Z. By linear algebra,

$$\dim Z = \sum_{\tau \in Y_{d-1}} \|P_Z(e_\tau)\|^2,$$

where e_τ is the unit vector of τ.

The *spectral theorem* from functional analysis offers a new perspective of $\|P_Z(e_\tau)\|^2$. Associated with every self-adjoint operator L on a Hilbert space $\mathcal{H}$, and a vector $\psi \in \mathcal{H}$ is the *spectral measure of L with respect to ψ*. It is a real measure denoted $\pi_{L,\psi}$ which satisfies

$$\langle F(L)\psi, \psi \rangle = \int_{\mathbb{R}} F(x)d\pi_{L,\psi}(x),$$

for every measurable function $F : \mathbb{R} \to \mathbb{C}$. The operator $F(L)$ is uniquely defined by extending the action of polynomials on the operator L.

If $\mathcal{H}$ is finite-dimensional, $\pi_{L,\psi}$ is a discrete measure supported on the spectrum of L and $\pi_{L,\psi}(\lambda) = \|P_\lambda\psi\|^2$, where P_λ is the orthogonal projection to the λ-eigenspace.

We use this theorem with the measure $\pi_{L(Y),e_\tau}$. Here Y is a d-complex and $\mathcal{H} = \ell^2(Y_{d-1})$. The self adjoint operator is the Laplacian $L(Y)$, and e_τ is the characteristic vector of some $(d-1)$-face of Y.

In particular, with $Y = Y_d\left(n, \frac{c}{n}\right)$ and Z as before, $\|P_Z(e_\tau)\|^2$ is simply the measure of the atom $\{0\}$ according to the spectral measure $\pi_{L(Y),e_\tau}$.

The difficulty with applying the spectral theorem to the Poisson d-tree is that the degrees in this tree may be unbounded. We must, therefore, consider the subtleties of the theory of unbounded operators [19]. Briefly, the Laplacian $L(T)$ of an infinite d-tree T is a symmetric operator, directly defined on the dense subset of finitely supported vectors of $\mathcal{H} = \ell^2(T_{d-1})$. The symmetric densely-defined operator $L(T)$ has a unique extension to $\mathcal{H}$. This extension need not be self-adjoint, and when it does we say that the tree T is self-adjoint. In such cases the spectral theorem can be applied on $L(T)$. It can be shown that a Poisson d-tree is, almost surely, self-adjoint.

We employ the useful property that spectral measures are continuous with respect to local weak convergence. Since $Y = Y_d\left(n, \frac{c}{n}\right)$ converges in local weak convergence to the Poisson d-tree $T = T_d(c)$, which is almost-surely self-adjoint, we conclude that the expected measure $\mathbb{E}_Y[\pi_{L(Y),e_\tau}]$ weakly converges to the expected measure $\mathbb{E}_T[\pi_{L(T),e_o}]$, where o is the root of T. In particular, by measuring the closed set $\{0\}$,

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P_Z(e_\tau)\|^2\right] \le \mathbb{E}[x_T],$$

where $x_T := \pi_{L(T),e_o}(\{0\})$. Consequently,

$$\mathbb{E}[\dim Z] \le (1 + o_n(1))\binom{n}{d}\mathbb{E}[x_T].$$

By the Rank-Nullity Theorem from linear algebra,

$$\dim Z - \beta_d(Y) = f_{d-1}(Y) - f_d(Y),$$

and we conclude that

$$\frac{1}{\binom{n}{d}}\mathbb{E}[\beta_d(Y)] \le \mathbb{E}[x_T] - 1 + \frac{c}{d+1} + o_n(1).$$

There remains the problem of bounding the expectation $\mathbb{E}_T[x_T]$ without directly computing the operator's kernel. This difficulty is bypassed using the recursive structure of d-trees to derive a simple recursion formulas on these spectral measures, as in the following lemma.

Let T be a self-adjoint d-tree with root o, and let $\sigma_1, ..., \sigma_m$ be the d-faces that contain the root. For $1 \le j \le m$ and $1 \le r \le d$ we denote by $\tau_{j,r}$ the $(d-1)$-faces of σ_j other than the root. Also, $T_{j,r}$ denotes the branch rooted at $\tau_{j,r}$.

Lemma 5.1. $x_T = 0$ *if there exists some* $1 \leq j \leq m$ *such that* $x_{T_{j,1}} = \ldots = x_{T_{j,d}} = 0$. *Otherwise,*

$$x_T = \left(1 + \sum_{j=1}^{m} \left(\sum_{r=1}^{d} x_{T_{j,r}} \right)^{-1} \right)^{-1}.$$

Proof sketch. Consider the following bounded family of measurable functions $\{H_s : \mathbb{R} \to \mathbb{C} \mid s \in \mathbb{R} \setminus \{0\}\}$,

$$H_s(x) = \frac{is}{x + is}.$$

Note that H_s approaches the Kronecker delta function $\delta_{x,0}$ as $s \to 0$. Given a self-adjoint d-tree T with root o, we define

$$h_T(s) := \int_{\mathbb{R}} H_s(x) d\pi_{L(T),e_o}(x).$$

In particular, $x_T = \lim_{s \to 0} h_T(s)$. The proof is concluded by showing that the functions $h_{T_{j,r}}$'s and h_T satisfy the formula

$$h_T(s) \left(1 + \sum_{j=1}^{m} \left(is + \sum_{r=1}^{d} h_{T_{j,r}}(s) \right)^{-1} \right)^{-1} = 1, \tag{7}$$

and letting $s \to 0$.

We turn to describe the derivation of equation (7). Denote $L := L(T)$, $\tilde{L} := \bigoplus_{j,r} L(T_{j,r})$ and M the Laplacian of the subcomplex of T which contains $\sigma_1, \ldots, \sigma_m$ and their subfaces. In particular, $L = M \oplus \tilde{L}$. The operators $H_s(L)$ and $H_s(\tilde{L})$ are scalar multiples of the *resolvents* $R := (L + is \cdot I)^{-1}$ and $\tilde{R} := (\tilde{L} + is \cdot I)^{-1}$. In particular, $h_T(s) = is \cdot \langle Re_o, e_o \rangle$. Simple observations about the d-tree structure yields that (i) $\tilde{R}e_o = \frac{1}{is} e_o$ and (ii) $\left\langle \tilde{R}e_{\tau_{j,r}}, e_{\tau_{j',r'}} \right\rangle = 0$ when $(j,r) \neq (j',r')$. In addition, we use the Second Resolvent Identity which states that $RM\tilde{R} = \tilde{R} - R$. Here the operator M has a very concrete and usable form, being the Laplacian of a d-complex which consists of m distinct d-faces with a common $(d-1)$-subface o. Equation (7) is obtained using simple algebraic manipulations by comparing terms of the form $\left\langle (RM\tilde{R})e_\tau, e_{\tau'} \right\rangle = \left\langle (\tilde{R} - R)e_\tau, e_{\tau'} \right\rangle$, when τ, τ' are $(d-1)$-faces of T of distance at most 1 from the root o. ∎

The remaining step of the argument is an application of the recursive formula in Lemma 5.1 to the Poisson d-tree.

Lemma 5.2. *Let* $T = T_d(c)$ *be a rooted Poisson d-tree with parameter c. Then,*

$$\mathbb{E}[x_T] \leq \max \left\{ t + ct(1-t)^d - \frac{c}{d+1} \left(1 - (1-t)^{d+1} \right) \mid t \in [0,1], \, t = e^{-c(1-t)^d} \right\}.$$

Note that this maximum is taken over a finite set, due to the condition $t = e^{-c(1-t)^d}$.

Proof. Let T be a Poisson d-tree with root degree m and $\{T_{j,r} \mid 1 \leq j \leq m, 1 \leq r \leq d\}$ its subtrees as above. The parameters $x_T, \{x_{T_{j,r}}\}$ can be considered as random variables when T is $T_d(c)$-distributed. The random variables $\{x_{T_{j,r}}\}$ are i.i.d. and are distributed like x_T since all the subtrees $T_{j,r}$ are independent Poisson d-trees. In addition, these variables satisfy the equation of Lemma 5.1.

These observations suggest the following equivalent description of $\mathcal{D}$, the distribution of the random variable x_T. First sample a $Poi(c)$-distributed integer m, and $x_{T_{j,r}} = \mathcal{D}$ i.i.d for every $1 \leq j \leq m$ and $1 \leq r \leq d$. Given these samples, the value of x_T is determined by Lemma 5.1.

In particular, if we let $t := \Pr(x_T > 0)$, then t satisfies the equation

$$t = \sum_{m=0}^{\infty} \frac{e^{-c} c^m}{m!} (1 - (1-t)^d)^m = e^{-c(1-t)^d}. \tag{8}$$

16

Let X be a $\mathcal{D}$-distributed random variable,

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

Here $S_1, S_2, \ldots, S_m$ are random variables whose distribution is that of a sum of d i.i.d. $\mathcal{D}$-distributed variables. By expressing the probability $\Pr[\mathbf{1}_{\{\forall j \in [m],\, S_j > 0\}}]$ as t, exploiting the symmetry between the different S_j's and using basic properties of the Poisson distribution, we are able to express this expectation in terms of t,

$$\mathbb{E}[X] = t + ct(1-t)^d - \frac{c}{d+1}\left(1 - (1-t)^{d+1}\right),$$

as was claimed. ∎

It requires only basic calculus to conclude the following.

1. For $c < c_d$, the maximum of

$$t + ct(1-t)^d - \frac{c}{d+1}\left(1 - (1-t)^{d+1}\right), \quad \text{s.t.} \quad t = e^{-c(1-t)^d}$$

is attained at $t = 1$. Consequently, $\mathbb{E}[\beta_d(Y)] \leq o(n^d)$.

2. For $c > c_d$, the maximum is attained at $t = t(c,d)$. Consequently,

$$\mathbb{E}[\beta_d(Y)] \leq (1 + o_n(1))\binom{n}{d}\left(\frac{c(1-t)^{d+1}}{d+1} - 1 + t + ct(1-t)^d\right). \tag{9}$$

The proof of Theorem 1.1 (II.b) is concluded by the following standard probabilistic argument. Let $c < c_d$ and $\varepsilon > 0$. The absence of small non-collapsible subcomplexes in Y (Lemma 4.1) implies that it has no small d-cycles except $\partial\Delta_{d+1}$'s. The computation above shows that the dimension of the cycle space is $o(n^d)$. Therefore, removing $\mathrm{Bin}\left(\binom{n}{d+1}, \varepsilon/n\right) = \Theta(n^d)$ random d-faces eliminates all large (non $\partial\Delta_{d+1}$) d-cycles with high probability, hence $Y_d\left(n, \frac{c-\varepsilon}{n}\right)$ is a.a.s. d-acyclic except $\partial\Delta_{d+1}$'s.

5.2 The cyclic regime - Theorem 1.1(III)

The key idea of [5] for the computation of a matching upper bound for the d-acyclicity threshold uses an analysis of the collapse process. Let Y be a d-complex and k some positive integer. Recall that $R_k(Y)$ is the simplicial complex obtained from Y by k phases of d-collapse and $S_k(Y)$ is its subcomplex obtained by removing the exposed $(d-1)$-faces. Clearly, $\beta_d(Y) = \beta_d(S_k(Y))$ since d-collapsing and removing exposed faces does not affect the right kernel of ∂_d. The final ingredient of the strategy is the observation that $\beta_d(S_k(Y)) \geq f_d(S_k(Y)) - f_{d-1}(S_k(Y))$, and the fact that the parameter $f_d(S_k(Y)) - f_{d-1}(S_k(Y))$ can be studied from the local weak limit.

Let $Y = Y_d\left(n, \frac{c}{n}\right)$ where $c > c_d$, and $Y' := S_k(Y)$ for a sufficiently large integer k.

$$\begin{aligned}
\mathbb{E}[\beta_d(Y)] &\geq \mathbb{E}[f_d(Y') - f_{d-1}(Y')] \\
&= \mathbb{E}\left[\sum_{\tau \in Y'_{d-1}} \left(\frac{d_{Y'}(\tau)}{d+1} - 1\right)\right] \\
&= \binom{n}{d}\Pr[\tau \in Y']\left(\frac{\mathbb{E}\left[d_{Y'}(\tau)|\tau \in Y'\right]}{d+1} - 1\right). \tag{10}
\end{aligned}$$

For the last equation we can, due to symmetry, consider a fixed τ and apply the Law of Total Expectation to the event $\{\tau \in Y'\}$. As mentioned above, the $(d-1)$-face τ of Y belongs to Y' if and only if in the τ-rooted collapse process of Y, the degree of τ is greater than 1 after $(k-1)$ phases and positive after k steps. By approximating the τ-rooted collapse process of Y with the rooted collapse process on $T_d(c)$ we obtain that up to $1 + o_n(1)$ factor,

$$\Pr[\tau \in Y'] \geq \Pr[\delta_k > 1] = 1 - t_k - c(1 - t_{k-2})^d t_{k-1}. \tag{11}$$

Furthermore, upto $1 + o_n(1)$ factor,

$$
\begin{aligned}
\mathbb{E}\left[d_{Y'}(\tau) | \tau \in Y'\right] &= \mathbb{E}[\delta_k \mid \delta_k > 0 \wedge \delta_{k-1} > 1] \\
&\geq \sum_{j=2}^{\infty} j \cdot \Pr[\delta_k = j \mid \delta_k > 0 \wedge \delta_{k-1} > 1] \\
&\geq \sum_{j=2}^{\infty} j \cdot \frac{\Pr[\delta_k = j]}{\Pr[\delta_{k-1} > 1]} \\
&= \frac{c(1 - t_{k-1})^d (1 - t_k)}{1 - t_{k-1} - c(1 - t_{k-2})^d t_{k-1}}.
\end{aligned} \tag{12}
$$

By combining Inequalities (10),(11) and (12), and letting $k \to \infty$, we obtain that

$$\mathbb{E}[\beta_d(Y)] \geq (1 + o_n(1)) \binom{n}{d} \left(\frac{c(1-t)^{d+1}}{d+1} - 1 + t + ct(1-t)^d \right).$$

This bound starts to be meaningful for $c > c_d$, where this expression matches the upper bound from (9). Since β_d is 1-Lipschitz, a straightforward application of Azuma's inequality yields that a.a.s. β_d deviates from its expectation by only $o(n^d)$. In particular, this shows a matching upper bound for the threshold of d-acyclicity.

5.3 $\mathbb{R}$-Shadow of $Y_d\left(n, \frac{c}{n}\right)$

The behavior of the $\mathbb{R}$-shadow when $c < c_d$ is studied similarly to the C-shadow in the collapsible regime (See Section 4.3).

We turn to the range $c > c_d$. Here we do not give a proof, but only a general intuitive explanation. An accurate analysis of the measure concentration can be found in [14]. Recall that

$$\frac{1}{\binom{n}{d}} \mathbb{E}[\beta_d(Y; \mathbb{R})] \xrightarrow[n \to \infty]{} g_d(c) := \frac{c}{d+1}(1-t)^{d+1} - (1-t) + ct(1-t)^d.$$

A simple technical claim shows that for every $c > c_d$, the limit function $g_d(c)$ is differentiable with respect to the variable c and its derivative equals to $\frac{1}{d+1}(1-t)^{d+1}$.

It turns out to be more convenient to work here with a d-dimensional analog of the so-called *evolution of random graphs*. Let $Y_d(n, m)$ be a random simplicial complex with n vertices, a complete $(d-1)$-skeleton and m uniformly random d-faces. $Y' = Y_d(n, m+1)$ can be sampled by the following procedure. First sample $Y = Y_d(n, m)$ and then add a random d-face which does not belong to Y. Therefore, the following equation holds in expectation,

$$\beta_d(Y'; \mathbb{R}) - \beta_d(Y; \mathbb{R}) = \frac{1}{\binom{n}{d+1}} |\mathrm{SH}_{\mathbb{R}}(Y)|.$$

Letting $m = \mathrm{Bin}\left(\binom{n}{d+1}, \frac{c}{n}\right)$ yields that $Y = Y_d\left(n, \frac{c}{n}\right)$, and its real d-Betti number is $\binom{n}{d} \cdot g_d(c) + o(n^d)$. The previous equation suggests, at least intuitively, the following relation between the growth of β_d and the $\mathbb{R}$-shadow. Namely,

$$\binom{n}{d} \left(g_d\left(c + \frac{d+1}{\binom{n}{d}}\right) - g_d(c) \right) \approx \frac{1}{\binom{n}{d+1}} |\mathrm{SH}_{\mathbb{R}}(Y)|,$$

and by letting $n \to \infty$,

$$\frac{1}{\binom{n}{d+1}} |\mathrm{SH}_{\mathbb{R}}(Y)| \approx (d+1)g_d'(c) = (1-t)^{d+1}.$$

This argument can be made rigorous by incrementing Y with εn^d random d-faces at a time rather than one by one, and applying standard measure concentration inequalities.

6 Concluding remarks

Although this article is mostly a review of previous work, it does contain several new results, e.g. Theorem 1.1(I) that deals with the collapsible regime is a little stronger than the original result of [6]. Other notable new results concern the asymptotic densities of the core and the C-shadow of $Y_d\left(n, \frac{c}{n}\right)$ for $c > \gamma_d$, improving the main theorem of [4] which says that $Y_d\left(n, \frac{c}{n}\right)$ is a.a.s. non-collapsible for $c > \gamma_d$. As mentioned above, although the main result in that paper is correct, there is an error in the proof, which we are able to remedy here using the techniques of Riordan [20].

The results surveyed here can be viewed from several perspectives which suggest different problems for future research.

From the combinatorial perspective, the phase transition in the density of the shadow of $Y_d\left(n, \frac{c}{n}\right)$ is of great interest. We conjecture that that the $\mathbb{R}$-shadow grows from linear in n to a giant (order $\Theta(n^{d+1})$) in a single step in a random evolution of simplicial complexes. This starkly contrasts with the gradual growth of the giant component in random graphs. It is of particular interest to understand the structure of the critical complex. Numerical experiments suggest that its $(d-1)$-homology group has torsion of size $\exp(\Theta(n^d))$, but we do not know a proof of this yet.

On the topological side, it would be very interesting to better understand the giant d-cycles which appear in $Y_d\left(n, \frac{c}{n}\right)$ when $c > c_d$. Provably, they consist of $\Theta(n^d)$ d-faces, but numerical experiments suggest that they lie in an unknown territory in the realm of homological d-cycles. Unlike closed manifolds, in which the degree of all $(d-1)$-faces equals to 2, it seems that in these d-cycles the average degree approaches $d+1$, which is the largest possible for a minimal d-cycle. Namely, these d-cycles are in some sense the opposite of manifolds.

In addition, the random complexes $Y_d\left(n, \frac{c}{n}\right)$, where $\gamma_d < c < c_d$ have the nice property of being d-acyclic but not d-collapsible. What other interesting topological or combinatorial properties do they have?

There is much more to study about random simplicial complexes in the regime $p = c/n$. In particular, the question regarding the vanishing of the top homology over finite fields is still open and presently out of reach. It is interesting to resolve whether homology thresholds over other fields can also be read off some parameter of the Poisson d-tree.

References

[1] David Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4):381–418, 2001.

[2] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab*, 12(54):1454–1508, 2007.

[3] David Aldous and Michael Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

[4] Lior Aronshtam and Nathan Linial. The threshold for collapsibility in random complexes. *arXiv preprint arXiv:1307.2684*, 2013.

[5] Lior Aronshtam and Nathan Linial. When does the top homology of a random simplicial complex vanish? *Random Structures & Algorithms*, 2013.

[6] Lior Aronshtam, Nathan Linial, Tomasz Łuczak, and Roy Meshulam. Collapsibility and vanishing of top homology in random simplicial complexes. *Discrete & Computational Geometry*, 49(2):317–334, 2013.

[7] Eric Babson, Christopher Hoffman, and Matthew Kahle. The fundamental group of random 2-complexes. *Journal of the American Mathematical Society*, 24(1):1–28, 2011.

[8] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. In *Selected Works of Oded Schramm*, pages 533–545. Springer, 2011.

[9] Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs. *Annals of Probability*, 39(3):1097–1121, 2011.

[10] Christopher Hoffman, Matthew Kahle, and Elliot Paquette. The threshold for integer homology in random d-complexes. *arXiv preprint arXiv:1308.6232*, 2013.

[11] Dániel Korándi, Yuval Peled, and Benny Sudakov. A random triadic process. *SIAM Journal on Discrete Mathematics*, 30(1):1–19, 2016.

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

[13] Nathan Linial, Ilan Newman, Yuval Peled, and Yuri Rabinovich. Extremal problems on shadows and hypercuts in simplicial complexes. *arXiv preprint arXiv:1408.0602*, 2014.

[14] Nathan Linial and Yuval Peled. On the phase transition in random simplicial complexes. *Annals of Mathematics, accepted for publication*, 2016.

[15] Tomasz Łuczak and Yuval Peled. Integral homology of random simplicial complexes. *arXiv preprint arXiv:1607.06985*, 2016.

[16] Russell Lyons. Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing*, 14(04):491–522, 2005.

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

[18] Michael Molloy. Cores in random hypergraphs and boolean formulas. *Random Structures & Algorithms*, 27(1):124–135, 2005.

[19] Michael Reed and Barry Simon. *Methods of modern mathematical physics: Functional analysis*, volume 1. Gulf Professional Publishing, 1980.

[20] Oliver Riordan. The k-core and branching processes *Combinatorics, Probability & Computing*, 17(1):111–136, 2008.