Collapsibility and Vanishing of Top Homology in Random Simplicial Complexes
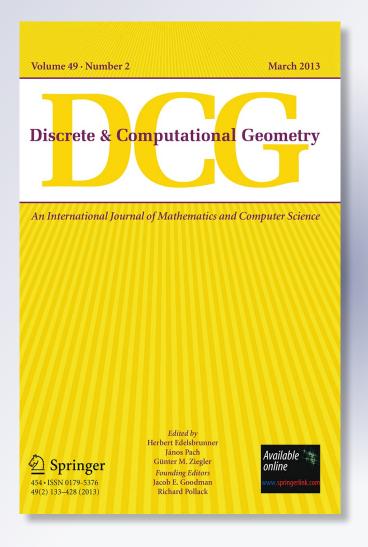
Lior Aronshtam, Nathan Linial, Tomasz Łuczak & Roy Meshulam

Springer

Springer

Collapsibility and Vanishing of Top Homology in Random Simplicial Complexes

Lior Aronshtam · Nathan Linial · Tomasz Łuczak · Roy Meshulam

Abstract   Let Δ_{n-1} denote the $(n-1)$-dimensional simplex. Let Y be a random d-dimensional subcomplex of Δ_{n-1} obtained by starting with the full $(d-1)$-dimensional skeleton of Δ_{n-1} and then adding each d-simplex independently with probability $p = \frac{c}{n}$. We compute an explicit constant γ_d, with $\gamma_2 \simeq 2.45$, $\gamma_3 \simeq 3.5$, and $\gamma_d = \Theta(\log d)$ as $d \to \infty$, so that for $c < \gamma_d$ such a random simplicial complex either collapses to a $(d-1)$-dimensional subcomplex or it contains $\partial\Delta_{d+1}$, the boundary of a $(d+1)$-dimensional simplex. We conjecture this bound to be sharp. In addition, we show that there exists a constant $\gamma_d < c_d < d+1$ such that for any $c > c_d$ and a fixed field $\mathbb{F}$, asymptotically almost surely $H_d(Y; \mathbb{F}) \neq 0$.

Keywords   Random complexes · Simplicial homology · Collapsibility

1 Introduction

Let $G(n, p)$ denote the probability space of graphs on the vertex set $[n] = \{1, \ldots, n\}$ with independent edge probabilities p. It is well known [2] that if $c \geq 1$ then a graph $G \in G(n, \frac{c}{n})$ a.a.s. contains a cycle, while for a constant $c < 1$

L. Aronshtam · N. Linial (✉)
Department of Computer Science, Hebrew University, Jerusalem 91904, Israel
e-mail: nati@cs.huji.ac.il

L. Aronshtam
e-mail: liobsar@gmail.com

T. Łuczak
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland
e-mail: tomasz@amu.edu.pl

R. Meshulam
Department of Mathematics, Technion, Haifa 32000, Israel
e-mail: meshulam@math.technion.ac.il

$$\lim_{n \to \infty} \Pr\big[G \in G\big(n, \tfrac{c}{n}\big) : G \text{ acyclic}\big] = \sqrt{1-c} \cdot \exp\big(\tfrac{c}{2} + \tfrac{c^2}{4}\big). \qquad (1)$$

In this article, we consider the analogous question for d-dimensional random complexes. There are two natural extensions to the notion of an acyclic graph. Namely, the vanishing of the dth homology, and collapsibility to a $(d-1)$-dimensional subcomplex. These are the two questions we consider here. We provide an upper bound on the threshold for the vanishing of the dth homology and a lower bound (which we believe to be tight) for the threshold for collapsibility.

For a simplicial complex Y, let $Y^{(i)}$ denote the i-dimensional skeleton of Y. Let $Y(i)$ be the set of i-dimensional simplices of Y and let $f_i(Y) = |Y(i)|$. Let Δ_{n-1} denote the $(n-1)$-dimensional simplex on the vertex set $V = [n]$. For $d \geq 2$, let $Y_d(n, p)$ denote the probability space of complexes $\Delta_{n-1}^{(d-1)} \subset Y \subset \Delta_{n-1}^{(d)}$ with probability measure

$$\Pr(Y) = p^{f_d(Y)}(1-p)^{\binom{n}{d+1} - f_d(Y)}.$$

Let $\mathbb{F}$ be an arbitrary fixed field and let $H_i(Y) = H_i(Y; \mathbb{F})$ and $H^i(Y) = H^i(Y; \mathbb{F})$ denote the ith homology and cohomology groups of Y with coefficients in $\mathbb{F}$. Let $\beta_i(Y) = \dim_{\mathbb{F}} H_i(Y) = \dim_{\mathbb{F}} H^i(Y)$. Kozlov [4] proved the following

Theorem 1.1 (Kozlov) *For any function $\omega(n)$ that tends to infinity*

$$\lim_{n \to \infty} \Pr[Y \in Y_d(n, p) : H_d(Y) \neq 0] = \begin{cases} 1 & p = \frac{\omega(n)}{n} \\ 0 & p = \frac{1}{\omega(n)n}. \end{cases}$$

It is easy to see that if np is bounded away from zero, then the probability that $Y \in Y_d(n, p)$ contains the boundary of a $(d+1)$-simplex does not tend to zero. Thus, the second part of the above statement cannot be improved. Concerning the first part of the statement, as was already observed by Cohen et al. for $d = 2$ (Theorem 6 in [3]), a simple Euler characteristic argument shows that if $p = \frac{c}{n}$, where $c > d+1$ then a.a.s. $H_d(Y) \neq 0$. Our first result is a further improvement on the upper bound in Theorem 1.1. Let

$$g_d(x) = (d+1)(x+1)e^{-x} + x(1 - e^{-x})^{d+1}$$

and let c_d denote the unique positive solution of the equation $g_d(x) = d+1$. It is easy to check that $g_d(d+1) > d+1$ and so $c_d < d+1$. A direct calculation yields that $c_d = d+1 - \Theta(\frac{d}{e^d})$.

Theorem 1.2 *For a fixed $c > c_d$*

$$\lim_{n \to \infty} \Pr\big[Y \in Y_d\big(n, \tfrac{c}{n}\big) : H_d(Y) \neq 0\big] = 1. \qquad (2)$$

Remark In the 2-dimensional case, Theorem 1.2 implies that if $c > c_2 \simeq 2.783$ then $Y \in Y_2(n, \frac{c}{n})$ a.a.s. satisfies $H_2(Y) \neq 0$. Simulations indicate that the actual threshold is somewhat lower (around 2.75).

We next turn to collapsibility. A $(d-1)$-dimensional simplex $\tau \in \Delta_{n-1}(d-1)$ is a *free face* of a complex $Y \subset \Delta_{n-1}^{(d)}$ if it is contained in a unique $\sigma \in Y(d)$. Let $R(Y)$ denote the complex obtained by removing all free $(d-1)$-faces of Y together with the d-simplices that contain them. We say that $R(Y)$ is obtained from Y by a d-collapse step. Let $R_0(Y) = Y$ and for $i \geq 1$ let $R_i(Y) = R(R_{i-1}(Y))$. We say that Y is d-*collapsible* if $\dim R_\infty(Y) < d$. Cohen et al. [3] proved the following

Theorem 1.3 (Cohen, Costa, Farber and Kappeler) *If* $\omega(n) \to \infty$ *then* $Y \in Y_2(n, \frac{1}{\omega(n)n})$ *is a.a.s. 2-collapsible.*

Our second result refines Theorem 1.3 and the lower bound in Theorem 1.1 as follows. Let

$$u_d(\gamma, x) = \exp(-\gamma(1-x)^d) - x.$$

For small positive γ, the only solution of $u_d(\gamma, x) = 0$ is $x = 1$. Let γ_d be the infimum of the set of all non-negative γ's for which the equation $u_d(\gamma, x) = 0$ has a solution $x < 1$. More explicitly, $\gamma_d = (dx(1-x)^{d-1})^{-1}$, where x satisfies $\exp(-\frac{1-x}{dx}) = x$. It is not hard to verify that this yields

$$\gamma_d = \log d + O(\log \log d).$$

For $\Delta_{n-1}^{(d-1)} \subset Y \subset \Delta_{n-1}^{(d)}$ let $s(Y)$ denote the number of $(d+1)$-simplices in Δ_{n-1} whose boundary is contained in Y. If $c > 0$ is fixed and $p = \frac{c}{n}$ then a straightforward application of the method of moments (see, e.g., Theorem 8.3.1 in [2]) shows that $s(Y)$ is asymptotically Poisson with parameter

$$\lambda = \lim_{n \to \infty} E[s] = \lim_{n \to \infty} \binom{n}{d+2} \left(\frac{c}{n}\right)^{d+2} = \frac{c^{d+2}}{(d+2)!}.$$

The next result asserts that if $c < \gamma_d$ and $p = \frac{c}{n}$ then $s(Y) > 0$ is a.a.s. the only obstruction for d-collapsibility. Let $\mathcal{F}_{n,d}$ denote the family of all $\Delta_{n-1}^{(d-1)} \subset Y \subset \Delta_{n-1}^{(d)}$ such that $s(Y) = 0$.

Theorem 1.4 *Let* $c < \gamma_d$ *be fixed. Then in the probability space* $Y_d(n, \frac{c}{n})$

$$\lim_{n \to \infty} \Pr[Y \text{ is } d\text{-collapsible}|Y \in \mathcal{F}_{n,d}] = 1. \tag{3}$$

Remark We have calculated $\gamma_2 \simeq 2.455$, and computer simulations suggest that this is indeed the actual threshold for collapsibility for random complexes in $\mathcal{F}_{n,2}$. Also, $\gamma_3 \simeq 3.089$, $\gamma_4 \simeq 3.508$, and $\gamma_{100} \simeq 7.555$.

Clearly, if Y is d-collapsible then Y is homotopy equivalent to a $(d-1)$-dimensional complex, and in particular $H_d(Y) = 0$. Hence, for a fixed $c < \gamma_d$ and $p = \frac{c}{n}$ the following hold:

$$\lim_{n \to \infty} \Pr[H_d(Y) = 0|Y \in \mathcal{F}_{n,d}] = 1$$

and

$$\lim_{n \to \infty} \Pr[H_d(Y) = 0] = \lim_{n \to \infty} \Pr[\mathcal{F}_{n,d}] = \exp(-\lambda) = \exp\left(-\frac{c^{d+2}}{(d+2)!}\right).$$

The article is organized as follows. In Sect. 2, we prove Theorem 1.2. In Sect. 3, we analyze a random d-tree process that underlies our proof that for $c < \gamma_d$, a random $Y \in Y_d(n, \frac{c}{n}) \cap \mathcal{F}_{n,d}$ is a.a.s. d-collapsible. Another main ingredient of the proof is an upper bound on the number of minimal non d-collapsible complexes given in Sect. 4. In Sect. 5, we combine these results to derive Theorem 1.4. We conclude in Sect. 6 with some comments and open problems.

2 The Upper Bound

Let $Y \in Y_d(n, p)$. Then $\beta_i(Y) = 0$ for $0 < i < d - 1$ and $f_i(Y) = \binom{n}{i+1}$ for $0 \le i \le d - 1$. The Euler–Poincaré relation

$$\sum_{i \ge 0} (-1)^i f_i(Y) = \sum_{i \ge 0} (-1)^i \beta_i(Y)$$

therefore implies

$$\beta_d(Y) = f_d(Y) - \binom{n-1}{d} + \beta_{d-1}(Y). \tag{4}$$

The inequality $\beta_d(Y) \ge f_d(Y) - \binom{n-1}{d}$ already implies that if $c > d + 1$ then a.a.s. $\beta_d(Y) \ne 0$. As mentioned above, this was observed in the 2-dimensional case by Cohen et al. [3]. The idea of the proof of Theorem 1.2 is to improve this estimate by providing a non-trivial lower bound on $E[\beta_{d-1}]$. For $\tau \in \Delta_{n-1}(d - 1)$ let

$$\deg_Y(\tau) = |\{\sigma \in Y(d) : \tau \subset \sigma\}|$$

and let

$$A_\tau = \{Y \in Y_d(n, p) : \deg_Y(\tau) = 0\}.$$

For $\sigma \in Y(d)$ let L_σ be the subcomplex of $\sigma^{(d-1)}$ given by

$$L_\sigma = \sigma^{(d-2)} \cup \{\tau \in \sigma^{(d-1)} : \deg_Y(\tau) > 1\}.$$

Let $P_{n,d}$ denote the family of all pairs (σ, L), such that $\sigma \in \Delta_{n-1}(d)$ and $\sigma^{(d-2)} \subset L \subset \sigma^{(d-1)}$. For $(\sigma, L) \in P_{n,d}$ let

$$B_{\sigma,L} = \{Y \in Y_d(n, p) : \sigma \in Y, L_\sigma = L\}.$$

The space of i-cocycles of a complex K is as usual denoted by $Z^i(K)$. The space of relative i-cocycles of a pair $K' \subset K$ is denoted by $Z^i(K, K')$ and will be identified with the subspace of i-cocycles of K that vanish on K'. Let $z^i(K) = \dim Z^i(K)$ and

$z^i(K, K') = \dim Z^i(K, K')$. For a $(d-1)$-simplex $\tau = [v_0, \ldots, v_{d-1}]$, let 1_τ be the indicator $(d-1)$-cochain of τ (i.e., $1_\tau(\eta) = \text{sgn}(\pi)$ if $\eta = [v_{\pi(0)}, \ldots, v_{\pi(d-1)}]$ and is zero otherwise). If $\tau \in \Delta_{n-1}(d-1)$ then $Z^{d-1}(\tau)$ is the 1-dimensional space spanned by 1_τ. If $(\sigma, L) \in P_{n,d}$ and $f_{d-1}(L) = j$, then $z^{d-1}(\sigma, L) = d - j$. Indeed, suppose $\sigma = [v_0, \ldots, v_d]$ and for $0 \le i \le d$ let $\tau_i = [v_0, \ldots, \widehat{v_i}, \ldots, v_d]$. If $L(d-1) = \{\tau_i\}_{i=d-j+1}^{d}$ then $\{1_{\tau_0} - 1_{\tau_i}\}_{i=1}^{d-j}$ forms a basis of $Z^{d-1}(\sigma, L)$.

Claim 2.1 *For any* $Y \in Y_d(n, p)$

$$Z^{d-1}(Y) \supset \bigoplus_{\{\tau \in \Delta_{n-1}(d-1) : Y \in A_\tau\}} Z^{d-1}(\tau) \oplus \bigoplus_{\{(\sigma, L) \in P_{n,d} : Y \in B_{\sigma, L}\}} Z^{d-1}(\sigma, L).$$

Proof The containment is clear. To show that the right-hand side is a direct sum, note that nontrivial cocycles in different summands must have disjoint supports and are therefore linearly independent. □

Let

$$a(Y) = |\{\tau \in \Delta_{n-1}(d-1) : \deg_Y(\tau) = 0\}|$$

and for $0 \le j \le d$ let

$$\alpha_j(Y) = |\{(\sigma, L) \in P_{n,d} : Y \in B_{\sigma, L}, f_{d-1}(L) = j\}|.$$

Note that $\alpha_j(Y)$ is the number of d-faces of Y that contain exactly $d+1-j$ $(d-1)$-faces of degree 1. By Claim 2.1

$$z^{d-1}(Y) \ge u(Y) \overset{\text{def}}{=} a(Y) + \sum_{j=0}^{d} \alpha_j(Y)(d - j). \tag{5}$$

As $\beta_{d-1}(Y) = \dim H^{d-1}(Y) = z^{d-1}(Y) - \binom{n-1}{d-1}$, it follows from (4) and (5) that

$$\beta_d(Y) \ge v(Y) \overset{\text{def}}{=} f_d(Y) + u(Y) - \binom{n}{d}. \tag{6}$$

Theorem 1.2 will thus follow from

Theorem 2.2 *Let* $c > c_d$ *and let* $p = \frac{c}{n}$. *Then*

$$\lim_{n \to \infty} \Pr[Y \in Y_d(n, p) : v(Y) \le 0] = 0.$$

Proof First note that

$$E[f_d] = \binom{n}{d+1} p = \frac{c}{(d+1)!} n^d - O(n^{d-1}),$$

$$E[a] = \binom{n}{d}(1-p)^{n-d} = \frac{e^{-c}}{d!}n^d - O(n^{d-1}),$$

and for $0 \leq j \leq d$

$$
\begin{aligned}
E[\alpha_j] &= \binom{n}{d+1}\binom{d+1}{j}p(1-p)^{(n-d-1)(d+1-j)}(1-(1-p)^{n-d-1})^j \\
&= \frac{n^d c}{(d+1)!}\binom{d+1}{j}e^{-c(d+1-j)}(1-e^{-c})^j - O(n^{d-1}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E[u] &= E[a] + \sum_{j=0}^{d} E[\alpha_j](d-j) \\
&= \frac{n^d e^{-c}}{d!} + \frac{n^d c}{(d+1)!}\sum_{j=0}^{d}\binom{d+1}{j}e^{-c(d+1-j)}(1-e^{-c})^j(d-j) - O(n^{d-1}) \\
&= \frac{n^d}{(d+1)!}((1+c)(d+1)e^{-c} - c(1-(1-e^{-c})^{d+1})) - O(n^{d-1}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E[v] &= E[f_d] + E[u] - \binom{n}{d} \\
&= \frac{n^d}{(d+1)!}\left(c + (1+c)(d+1)e^{-c} - c\left(1-(1-e^{-c})^{d+1}\right) - (d+1)\right) - O(n^{d-1}) \\
&= \frac{n^d}{(d+1)!}(g_d(c) - d - 1) - O(n^{d-1}).
\end{aligned}
$$

Since $c > c_d$ it follows that for sufficiently large n

$$E[v] \geq \varepsilon n^d, \tag{7}$$

where $\varepsilon > 0$ depends only on c and d. To show that v is a.a.s. positive, we use the following consequence of Azuma's inequality due to McDiarmid [6].

Theorem 2.3 *Suppose $f : \{0, 1\}^m \to \mathbb{R}$ satisfies $|f(x) - f(x')| \leq T$ if x and x' differ in at most one coordinate. Let $\xi_1, \ldots, \xi_m$ be independent $0, 1$ valued random variables and let $F = f(\xi_1, \ldots, \xi_m)$. Then for all $\lambda > 0$*

$$\Pr[F \leq E[F] - \lambda] \leq \exp\left(-\tfrac{2\lambda^2}{T^2 m}\right). \tag{8}$$

Let $m = \binom{n}{d+1}$ and let $\sigma_1, \ldots, \sigma_m$ be an arbitrary ordering of the d-simplices of Δ_{n-1}. Identify $Y \in Y_d(n, p)$ with its indicator vector $(\xi_1, \ldots, \xi_m)$, where $\xi_i(Y) = 1$ if $\sigma_i \in Y$ and $\xi_i(Y) = 0$ otherwise. Note that if Y and Y' differ in at most one d-simplex then $|a(Y) - a(Y')| \leq d + 1$ and $|\alpha_j(Y) - \alpha_j(Y')| \leq d + 1$ for all $0 \leq j \leq d$. It follows that $|v(Y) - v(Y')| \leq T = 2d^3$. Applying McDiarmid's inequality (8) with $F = v$ and $\lambda = E[v]$ it follows that

$$\Pr[v \leq 0] \leq \exp\left(-\tfrac{2E[v]^2}{(2d^3)^2 m}\right) \leq \exp\left(-C_2 n^{d-1}\right)$$

for some $C_2 = C_2(c, d) > 0$.

Remark The approach used in the proof of Theorem 1.2 can be extended as follows. For a fixed ℓ, let $Z_{(\ell)}^{d-1}(Y) \subset Z^{d-1}(Y)$ denote the subspace spanned by $(d-1)$-cocycles $\phi \in Z^{d-1}(Y)$ such that $|\mathrm{supp}(\phi)| \leq \ell$. Let

$$\theta_{d,\ell}(x) = \lim_{n \to \infty} \frac{E[\dim Z_{(\ell)}^{d-1}(Y)]}{\binom{n}{d}},$$

where the expectation is taken in the probability space $Y_d(n, \tfrac{x}{n})$. For example, it was shown in the proof of Theorem 1.2 that $\theta_{d,1}(x) = e^{-x}$ and

$$\theta_{d,2}(x) = (1+x)e^{-x} - \frac{x}{d+1}\left(1 - \left(1 - e^{-x}\right)^{d+1}\right).$$

Let $x = c_{d,\ell}$ denote the unique positive root of the equation

$$x + (d+1)\theta_{d,\ell}(x) = d + 1.$$

The following fact is implicit in the proof of Theorem 1.2.

Proposition 2.4 *For any fixed* $c > c_{d,\ell}$

$$\lim_{n \to \infty} \Pr\left[Y \in Y_d\left(n, \frac{c}{n}\right) : H_d(Y) \neq 0\right] = 1. \tag{9}$$

Let $\tilde{c}_d = \lim_{\ell \to \infty} c_{d,\ell}$. It seems likely that $\frac{\tilde{c}_d}{n}$ is the exact threshold for the vanishing of $H^d(Y)$. This is indeed true in the graphical case $d = 1$.

Proposition 2.5

$$\tilde{c}_1 = 1.$$

Proof For a subtree $K = (V_K, E_K)$ on the vertex set $V_K \subset [n]$ let A_K denote all graphs $G \in G(n, p)$ that contain K as an induced subgraph and contain no edges in the

cut $(V_K, \overline{V_K})$. The space of 0-cocycles $Z^0(K)$ is 1-dimensional and is spanned by the indicator function of V_K. As in Claim 2.1, it is clear that for $G \in G(n, p)$ and a fixed ℓ

$$Z^0(G) \supset \bigoplus_{\{K : |V_K| \leq \ell \text{ and } G \in A_K\}} Z^0(K).$$

Hence, for $p = \frac{x}{n}$

$$E[\dim Z^0_{(\ell)}(G)] = \sum \{\Pr[A_K] : K \text{ is a tree on } \leq \ell \text{ vertices}\}$$

$$= \sum_{k=1}^{\ell} \binom{n}{k} k^{k-2} \left(\frac{x}{n}\right)^{k-1} \left(1 - \frac{x}{n}\right)^{k(n-k)+\binom{k-1}{2}}$$

$$\sim n \sum_{k=1}^{\ell} \frac{k^{k-2}}{k!} x^{k-1} e^{-xk} = \frac{n}{x} \sum_{k=1}^{\ell} \frac{k^{k-2}}{k!} (xe^{-x})^k.$$

Let $S(z) = \sum_{k=1}^{\infty} \frac{k^{k-2}}{k!} z^k$ be the exponential generating function for the number of trees. Then

$$\lim_{\ell \to \infty} \theta_{1,\ell}(x) = \lim_{\ell,n \to \infty} \frac{E[\dim Z^0_{(\ell)}(G)]}{n} = \frac{S(xe^{-x})}{x}.$$

Therefore, $\tilde{c}_1 = \lim_{\ell \to \infty} c_{1,\ell}$ is the solution of the equation

$$x + \frac{2S(xe^{-x})}{x} = 2. \tag{10}$$

Let $R(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k$ be the exponential generating function for the number of rooted trees. It is classically known [7] that $R(z) = z \exp(R(z))$, and that $S(z) = R(z) - \frac{1}{2} R(z)^2$. It follows that $R(e^{-1}) = 1$ and $S(e^{-1}) = \frac{1}{2}$. Hence, $\tilde{c}_1 = 1$ is the unique solution of (10). $\qquad \square$

3 The Random d-Tree Process

A simplicial complex T on the vertex set V with $|V| = \ell \geq d$ is a *d-tree* if there exists an ordering $V = \{v_1, \ldots, v_\ell\}$ such that $\text{lk}(T[v_1, \ldots, v_i], v_i)$ is a $(d-1)$-dimensional simplex for all $d+1 \leq i \leq \ell$. Let G_T denote the graph with vertex set $T(d-1)$, whose edges are the pairs $\{\tau_1, \tau_2\}$ such that $\tau_1 \cup \tau_2 \in T(d)$. Let $\text{dist}_T(\tau_1, \tau_2)$ denote the distance between τ_1 and τ_2 in the graph G_T.

A *rooted d-tree* is a pair (T, τ), where T is a d-tree and τ is some $(d-1)$-face of T. Let τ be a fixed $(d-1)$-simplex. Given $k \geq 0$ and $\gamma > 0$, we describe a random process that gives rise to a probability space $\mathcal{T}_d(k, \lambda)$ of all d-trees T rooted at τ such that $\text{dist}_T(\tau, \tau') \leq k$ for all $\tau' \in T$. The definition of $\mathcal{T}_d(k, \lambda)$ proceeds by induction on k. $\mathcal{T}_d(0, \gamma)$ is the $(d-1)$-simplex τ. Let $k \geq 0$. A d-tree in $\mathcal{T}_d(k+1, \gamma)$ is generated as follows: First generate a $T \in \mathcal{T}_d(k, \lambda)$ and let $\mathcal{U}$ denote all $\tau' \in T(d-1)$ such that

$dist_T(\tau, \tau') = k$. Then, independently for each $\tau' \in \mathcal{U}$, pick $J = J_{\tau'}$ new vertices $z_1, \ldots, z_J$, where J is Poisson distributed with parameter γ, and add the d-simplices $z_1\tau', \ldots, z_J\tau'$ to T.

We next define the operation of pruning of a rooted d-tree (T, τ). Let $\{\tau_1, \ldots, \tau_\ell\}$ be the set of all free $(d-1)$-faces of T that are distinct from τ, and let σ_i be the unique d-simplex of T that contains τ_i. The d-tree T' obtained from T by removing the simplices $\tau_1, \sigma_1, \ldots, \tau_\ell, \sigma_\ell$ is called the *pruning of T*. Clearly, any $T \in \mathcal{T}_d(k+1, \gamma)$ collapses to its root τ after at most $k+1$ pruning steps. Denote by $\mathcal{C}_d(k+1, \gamma)$ the event that $T \in \mathcal{T}_d(k+1, \gamma)$ collapses to τ after at most k pruning steps, and let $\rho_d(k, \gamma) = \Pr[\mathcal{C}_d(k+1, \gamma)]$. Clearly, $\rho_d(0, \gamma)$ is the probability that $T \in \mathcal{T}_d(1, \gamma)$ consists only of τ, hence

$$\rho_d(0, \gamma) = e^{-\gamma}. \tag{11}$$

Let $\sigma_1, \ldots, \sigma_j$ denote the d-simplices of $T \in \mathcal{T}_d(k+1, \gamma)$ that contain τ and for each $1 \le i \le j$ let $\eta_{i1}, \ldots, \eta_{id}$ be the $(d-1)$-faces of σ_i that are different from τ. Let $T_{i\ell} \in \mathcal{T}_d(k, \lambda)$ denote the subtree of T that grows out of $\eta_{i\ell}$. Clearly, T collapses to τ after at most k pruning steps iff for each $1 \le i \le j$, at least one of the d-trees $T_{i\ell}$ collapses to its root $\eta_{i\ell}$ in at most $k-1$ steps. We therefore obtain the following recursion:

$$
\begin{aligned}
\rho_d(k, \gamma) &= \sum_{j=0}^{\infty} \Pr[J = j](1 - (1 - \rho_d(k-1, \gamma))^d)^j \\
&= \sum_{j=0}^{\infty} \frac{\gamma^j}{j!} e^{-\gamma}(1 - (1 - \rho_d(k-1, \gamma))^d)^j \\
&= \exp(-\gamma(1 - \rho_d(k-1, \gamma))^d).
\end{aligned}
\tag{12}
$$

Equations (11) and (12) imply that the sequence $\{\rho_d(k, \gamma)\}_k$ is non-decreasing and converges to $\rho_d(\gamma) \in (0, 1]$, where $\rho_d(\gamma)$ is the smallest positive solution of the equation

$$u_d(\gamma, x) = \exp(-\gamma(1-x)^d) - x = 0. \tag{13}$$

If $\gamma \ge 0$ is small, then $\rho_d(\gamma) = 1$. Let γ_d denote the infimum of the set of nonnegative γ's for which $\rho_d(\gamma) < 1$. The pair $(\gamma, x) = (\gamma_d, \rho_d(\gamma_d))$ satisfies both $u_d(\gamma_d, \rho_d(\gamma_d)) = 0$ and $\frac{\partial u_d}{\partial x}(\gamma_d, \rho_d(\gamma_d)) = 0$. A straightforward computation shows that $\gamma_d = (dx(1-x)^{d-1})^{-1}$, where $x = \rho_d(\gamma_d)$ is the unique solution of $\exp(-\frac{1-x}{dx}) = x$.

4 The Number of Non-d-Collapsible Complexes

When we discuss d-collapsibility, we only care about the inclusion relation between d-faces and $(d-1)$-faces. Therefore, in this section we can and will simplify matters and consider only the complex that is induced from our (random) choice of d-faces. Namely, for every $i \le d$, a given i-dimensional face belongs to the complex iff it is contained in some of the chosen d-faces.

A complex is a *core* if every $(d-1)$-dimensional face belongs to at least two simplices, so that *not even a single* collapse step is possible.

A core complex is called a *minimal core* complex if none of its proper subcomplexes is a core.

The main goal of this section is to show that with almost certainty there are just two types of minimal core subcomplexes that a sparse random complex can have. It can either be the boundary of a $(d+1)$-simplex, $\partial\Delta_{d+1}$, or it must be very large. Obviously this implies that there are no small non-collapsible subcomplexes which do not contain the boundary of a $(d+1)$-simplex.

Theorem 4.1 *For every $c > 0$ there exists a constant $\delta = \delta(c) > 0$ such that a.a.s. every minimal core subcomplex K of $Y \in Y_d(n, \frac{c}{n})$ with $f_d(K) \leq \delta n^d$, must contain the boundary of a $(d+1)$-simplex.*

Henceforth, we use the convention that *faces* refer to arbitrary dimensions, but unless otherwise specified, the word *simplex* is reserved to mean a d-face.

Our proof uses the first moment method. In the main step of the proof, we obtain an upper bound on $C_d(n, m)$, the number of all minimal core d-dimensional complexes on vertex set $[n] = \{1, 2, \ldots, n\}$, which contain m simplices.

Two simplices are considered *adjacent* if their intersection is a $(d-1)$-face. If $A \cup B$ is a splitting of a minimal core complex, then there is a simplex in A and one in B that are adjacent, otherwise the corresponding subcomplexes are cores as well. Therefore, K can be constructed by successively adding a simplex that is adjacent to an already existing simplex. This consideration easily yields an upper bound of n^{d+m} on $C_d(n, m)$. The point is that if $m = \delta n^d$ for $\delta > 0$ small enough, we get an exponentially smaller (in m) upper bound and this is crucial for our analysis.

Lemma 1 *Let $m = \delta n^d$ and $\delta > 0$ small enough. Then*

$$C_d(n, m) \leq \binom{n^{d-1}}{(d^2 m)^{\frac{d-1}{d}}} n^d n^m (2^{d+1} d^3 \delta^{\frac{1}{d^4}})^m. \qquad (14)$$

Proof Let $b = \left(\frac{d(d+1)\delta}{2}\right)^{\frac{1}{d}}$. A $(d-2)$-face is considered *heavy* or *light* depending on whether it is covered by at least bn $(d-1)$-faces or less. The sets of heavy and light $(d-2)$-faces are denoted by H_{d-2} and L_{d-2}, respectively. We claim that $|H_{d-2}| \leq b^{d-1} n^{d-1}$. To see this note that each simplex contains exactly $d+1$ $(d-1)$-faces, but the complex is a core, so that each $(d-1)$-face is covered at least twice. Consequently, our complex has at most $\frac{m(d+1)}{2}$ $(d-1)$-faces. Likewise, each $(d-1)$-face contains d $(d-2)$-faces. Each heavy $(d-2)$-face is covered at least bn times and the claim follows by the following calculation:

$$|H_{d-2}| \leq \frac{d(d+1)m}{2bn} = \frac{d(d+1)\delta n^d}{2bn} = \frac{d(d+1)\delta n^{d-1}}{2b} = b^{d-1} n^{d-1}.$$

We extend the heavy/light dichotomy to lower dimensions as well. For each $0 \leq i \leq d-3$, an i-face is considered *heavy* if it covered by at least $b \cdot n$ *heavy* $(i+1)$-faces.

Otherwise it is *light*. The sets of heavy/light i-faces are denoted by H_i/L_i, respectively. By counting inclusion relations between heavy faces of consecutive dimensions it is easily seen that $|H_i| \leq \frac{(i+2)|H_{i+1}|}{bn}$ which yields

$$|H_i| \leq \frac{(d-1)!}{(i+1)!}b^{i+1}n^{i+1}.$$

The set of i-dimensional heavy (respectively, light) faces contained in a given face σ is denoted by H_i^{σ} be (respectively, L_i^{σ}).

The bulk of the proof considers a sequence of complexes $C_1, \ldots, C_m = C$, where the complex C_i is obtained from C_{i-1} by adding a single simplex. A $(d-1)$-face σ of C_i can be *saturated* or *unsaturated*. This depends on whether or not every simplex in C_m that contains σ already belongs to C_i. Prior to defining the complexes C_i, we specify the set of heavy $(d-2)$-faces in one of at most $\binom{\binom{n}{d-1}}{b^{d-1}n^{d-1}}$ possible ways. Note that this choice uniquely determines the sets of heavy faces for every dimension $0 \leq i \leq d-3$. We start off with the complex C_1, which has exactly one simplex. Clearly, there are $\binom{n}{d+1}$ possible choices for C_1. We move from C_{i-1} to C_i by adding a single simplex t_i, which covers a *chosen* unsaturated $(d-1)$-face σ_{i-1} of C_{i-1}. Our choices are subject to the condition that every heavy $(d-2)$-face in C_m is one of the heavy $(d-2)$-faces chosen prior to the process. In other words, we must never make choices that create any additional heavy faces in addition to those derived from our preliminary choice. Our goal is to bound the number of choices for this process.

The crux of the argument is a rule for selecting the chosen face. Associated with every face is a vector counting the number of its heavy vertices, its heavy edges, its heavy 2-faces, etc. The chosen face is always lexicographically minimal w.r.t. this vector, breaking ties arbitrarily. A $(d-1)$-face all of whose subfaces are light is called *primary*.

In each step j we expand a $(d-1)$-face σ to a simplex $\sigma \cup y$. Such a step is called a *saving* step if either:

1. The vertex y is heavy.
2. There exists a light $(d-2)$-subface $\tau \subset \sigma$ such that $\tau \cup y$ is contained in a simplex in C_{j-1}.
3. There exists a light subface $\tau \subset \sigma$ such that the face $\tau \cup y$ is heavy.

Note that the number of choices of y in the first case is $\leq |H_0| \leq (d-1)! \cdot b \cdot n$. In the second case, the number of choices for y is at most dbn. In the third case, there are $d-2$ possibilities for the dimension of the light face and for each such dimension i there are at most $\binom{d}{i+1}bn$ choices for y. In all cases, the number of choices for y is at most $\leq d^d bn$. A step that is not saving is considered *wasteful*. For wasteful steps, we bound the number of choices for y by n.

The idea of the proof is that every such a process which produces a minimal core complex must include many saving steps. More specifically, we want to show:

Claim 4.2 *For every d^3 wasteful steps, at least one saving step is carried out.*

Proof The proof of this claim consists of two steps. We show that there is no sequence of $d(d-1)$ consecutive wasteful steps, without the creation of an unsaturated

primary face. Also, the creation of $d + 1$ primary faces necessarily involves a saving step.

If u is a vertex in a $(d-1)$-face σ, let $r_i^\sigma(u)$ be the number of heavy i-faces in σ that contain u. Also, V_i^σ denotes the set of vertices v in σ that are included only in light i-subfaces of σ.

Proposition 4.3 *Let σ and σ' be two consecutively chosen faces, where σ is non-primary and the extension step on σ is wasteful. Then σ' precedes σ in the order of faces and $|V_i^{\sigma'}| \geq |V_i^\sigma| + 1$, where i is the smallest dimension for which $|H_i^\sigma| > 0$.*

Proof Since the extension step on σ is wasteful (and, in particular, not a saving step of type (iii)) and since all j-subfaces of σ are light for $j < i$, every j-face in $\sigma \cup y$ is light. Moreover, every i-subface of $\sigma \cup y$ that contains y is light as well.

We claim that $\sigma' = \sigma \setminus \{u\} \cup \{y\}$, where the vertex u of σ maximizes $r_i^\sigma(v)$ (since $|H_i^\sigma| > 0$, there are vertices v in σ for which $r_i^\sigma(v) > 0$).

Notice that σ has $\binom{d-1}{i} - r_i^\sigma(u)$ more light i-subfaces than does $\tau^u := \sigma \setminus u$. Namely, $|L_i^{\tau^u}| = |L_i^\sigma| - \binom{d-1}{i} + r_i^\sigma(u)$.

Combining the fact that every i-subface of $\sigma \cup y$ that contains y is light we see that in $\tau_y^u := \tau^u \cup y$, $|L_i^{\tau_y^u}| = |L_i^{\tau^u}| + \binom{d-1}{i} = |L_i^\sigma| + r_i^\sigma(u)$. But since $r_i^\sigma(u) > 0$, τ_y^u precedes σ. In this case, τ^u must be a new face that does not belong to the previous complex, or else it would have been preferred over σ. Being a new face, it is necessarily unsaturated. Since u maximizes $r_i^\sigma(u)$ over all vertices in σ, it follows that τ_y^u precedes all other faces created in the expansion. Furthermore, no other face precedes σ or else it would be chosen rather than σ. Thus $\sigma' = \tau_u^y$, as claimed. Notice that $y \in V_i^{\sigma'}$ and also $V_i^\sigma \subseteq V_i^{\sigma'}$ (note that every i-dimensional subfaces of σ' that is not contained in σ is light since it contains y). Thus, $|V_i^{\sigma'}| \geq |V_i^\sigma| + 1$. □

Consider a chosen non-primary face σ and let i be the smallest dimension for which $|H_i^\sigma| > 0$. The previous claim implies that after at most d consecutive wasteful steps the chosen face, θ precedes σ and $|V_i^\theta| = d$. Then $|H_j^\theta| = 0$ for all $j \leq i$ (in particular $|H_i^\theta| = 0$). By repeating this argument $d - 1$ times we conclude that following every series of $d(d-1)$ consecutive wasteful steps, a primary face must be chosen: after at most d consecutive wasteful steps the chosen face can have no heavy vertices. At the end of the next d consecutive wasteful steps, the chosen face has no heavy vertices or heavy edges. Repeating this argument $(d-1)$ times necessarily leads us to a chosen primary face.

Proposition 4.4 *Only saving steps can decrease the number of unsaturated primary faces.*

Proof Let $\sigma = a_1, a_2 \ldots, a_d$ be a primary face and let y be the vertex that expands it. Denote the $(d-1)$-face $\{a_1, a_2 \ldots, a_{i-1}, a_{i+1}, \ldots, a_d\} \cup \{y\}$ by σ^i. Since this is not a saving step of type (i), y is light. It is also not of type (iii) and so $|H_k^{\sigma^i}| = 0$ for every $i = 1, \ldots, d$ and $k = 1, \ldots, d-2$, so that faces σ^i are primary. However this is not a type (ii) saving step, so all the $(d-1)$-faces σ^i must be new. Thus the number of unsaturated primary faces has increased by at least $d - 1$.

The proof of Claim 4.2 is now complete, since at each step at most $d + 1$ faces get covered.

We can turn now to bound the number of minimal core m-simplex complexes C_m. As mentioned, we first specify the heavy $(d-2)$-faces of C_m by specifying a set of $b^{d-1}n^{d-1}$ out of the total of $\binom{n}{d-1}$ $(d-2)$-faces. Then we select the first simplex C_1 and mark all its $(d-1)$-faces as unsaturated. In order to choose the ith step we first decide whether it is a saving or wasteful step, and if it is a saving step, what type it has. There is a total of $d+1$ possible kinds of extensions of the current $(d-1)$-face: A saving step of type (i), (ii), or one of the $d-2$ choices of type (iii) (according to dimension), or a wasteful step. In a saving step, the expanding vertex can be chosen in at most $d^d bn$ ways. The number of possible extension clearly never exceeds n and it is this trivial upper bound that we use for wasteful steps. Finally, we update the labels on the $(d-1)$-faces of a new simplex. We need to decide which of the unsaturated $(d-1)$-faces that are already covered by at least two simplices change their status to saturated. There are at most 2^{d+1} possibilities of such an update. As we saw, at least $\frac{m}{d^3}$ of the steps in such process are saving steps. Consequently, we get the following upper bound on $C_d(n, m)$, the number of minimal core n-vertex d-dimensional complexes with m simplices (in reading the expression below, note that the terms therein correspond in a one-to-one manner to the ingredients that were just listed).

$$\binom{\binom{n}{d-1}}{b^{d-1}n^{d-1}} n^{d+1}(d+1)^{m-1}n^{m-1}(d^d b)^{\frac{m}{d^3}}(2^{d+1})^m$$

$$\leq \binom{n^{d-1}}{(d^2\delta)^{\frac{d-1}{d}}n^{d-1}} n^d n^m ((d+1)2^{d+1}d^{\frac{1}{d^2}}b^{\frac{1}{d^3}})^m$$

$$\leq \binom{n^{d-1}}{(d^2 m)^{\frac{d-1}{d}}} n^d n^m (2^{d+1}d^2(d^2\delta)^{\frac{1}{d^4}})^m$$

$$\leq \binom{n^{d-1}}{(d^2 m)^{\frac{d-1}{d}}} n^d n^m (2^{d+1}d^3\delta^{\frac{1}{d^4}})^m. \qquad \square$$

Proof of Theorem 4.1 We show the assertion with $\delta = \delta(c) = (2^{d+2}d^3 c)^{-d^4}$. Indeed, consider a complex drawn from $Y_d(n, p)$. Let $X_m = X_m(n, p)$ count the number of minimal core subcomplexes with m simplices and which are not copies of $\partial\Delta_{d+1}$. Our argument splits according to whether m is small or large, the dividing line being $m = m_1 = (d^3 \log n)^d$. The theorem speaks only about the range $m \leq m_2 = \delta(c)n^d$. By Lemma 1,

$$\sum_{m=m_1}^{m_2} \mathbb{E}X_m \leq \sum_{m=m_1}^{m_2} C_d(n, m)p^m \leq \sum_{m=m_1}^{m_2} (n^{d-1})^{(d^2 m)^{\frac{d-1}{d}}} n^d \left(2^{d+1}d^3\delta^{\frac{1}{d^4}}c\right)^m$$

$$\leq n^d \sum_{m=m_1}^{m_2} (n^{d-1})^{(d^2 m)^{\frac{d-1}{d}}} 2^{-m} \leq n^d \sum_{m=m_1}^{m_2} (n^{d-1}2^{-\frac{m^{\frac{1}{d}}}{d^2}})^{(d^2 m)^{\frac{d-1}{d}}}$$

$$\leq n^d \sum_{m=m_1}^{m_2} (\tfrac{1}{n})^{(d^2 m)^{\frac{d-1}{d}}} = o(1).$$

It follows that with almost certainty no cores with m simplices occur, where $m_2 \geq m \geq m_1$. We next consider the range $d + 3 \leq m \leq m_1$. Note that a minimal core complex with $m \geq d + 3$ simplices has at most $\frac{d+3}{d+4}m$ vertices. Let $\Delta(u)$ denote the number of simplices that contain the vertex u. Clearly, if $\Delta(u) > 0$ then $\Delta(u) \geq d+1$ (consider a simplex σ that contains u. Every face of the form $\sigma \setminus w$ with $u \neq w \in \sigma$ is covered by a simplex other than σ). It is not hard to verify that if some simplex σ contains two distinct vertices with $\Delta(u) = \Delta(w) = d + 1$ then the complex contains $\partial \Delta_{d+1}$ contrary to the minimality assumption. Let t be the number of vertices u with $\Delta(u) = d + 1$. No simplex contains two such vertices, so that $t \leq \frac{m}{d+1}$. Counting vertices in the complex according to the value of Δ we get

$$(d + 1)t + (d + 2)(v - t) \leq (d + 1)m,$$

where v is the total number of vertices. The conclusion follows.

The expected number of minimal core subcomplexes of $Y_d(n, p)$ that contain $d + 3 \leq m \leq m_1$ simplices satisfies

$$\sum_{m=d+3}^{m_1} \mathbb{E}X_m \leq \sum_{m=d+3}^{m_1} \binom{n}{\frac{d+3}{d+4}m}\left[\left(\frac{d+3}{d+4}m\right)^{d+1}p\right]^m \leq \sum_{m=d+3}^{m_1} n^{\frac{d+3}{d+4}m}\left[m^{d+1}p\right]^m$$

$$\leq \sum_{m=d+3}^{m_1} \left(c^{d+4}\frac{\log^{2d(d+1)(d+4)} n}{n}\right)^{\frac{m}{d+4}} = o(1).$$

Consequently, a.a.s. $Y_d(n, p)$ contains no minimal core subcomplexes of m simplices with $d + 3 \leq m \leq (2^{d+2}d^3c)^{-d^4}n^d$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5 The Threshold for d-Collapsibility

For a complex $Y \subset \Delta_{n-1}^{(d)}$ and a fixed $\tau \in \Delta_{n-1}(d - 1)$, define a sequence of complexes $\{\mathcal{S}_i(Y)\}_{i \geq 0}$ as follows. $\mathcal{S}_0(Y) = \tau$ and for $i \geq 1$ let $\mathcal{S}_i(Y)$ be the union of $\mathcal{S}_{i-1}(Y)$ and the complex generated by all the d-simplices of Y that contain some $\eta \in \mathcal{S}_{i-1}(Y)(d - 1)$. Let $\mathcal{T}_d$ denote the family of all d-trees. Consider the events $A_k, D \subset Y_d(n, p)$ given by

$$A_k = \{\mathcal{S}_k(Y) \in \mathcal{T}_d\}$$

and

$$D = \{\deg_Y(\eta) \leq \log n \text{ for all } \eta \in \Delta_{n-1}(d - 1)\}.$$

Claim 5.1 *Let k and $c > 0$ be fixed and $p = \frac{c}{n}$. Then*

$$\Pr[A_{k+1} \cap D] = 1 - o(1).$$

Proof Fix $\eta \in \Delta_{n-1}(d-1)$. The random variable $\deg_Y(\eta)$ has a binomial distribution $B(n-d, \frac{c}{n})$, hence by the large deviations estimate

$$\Pr[\deg_Y(\eta) > \log n] < n^{-\Omega(\log \log n)}.$$

Therefore, $\Pr[D] = 1 - o(1)$. If $Y \in D$ then $f_0(\mathcal{S}_{k+1}(Y)) = O(\log^{k+1} n)$ and $f_{d-1}(\mathcal{S}_k(Y)) = O(\log^k n)$. Note that $\mathcal{S}_{k+1}(Y)$ is a d-tree iff in its generation process, we never add a simplex of the form ηv such that both $\eta \in \Delta_{n-1}(d-1)$ and $v \in \Delta_{n-1}(0)$ already exist in the complex. Since the number of such pairs is at most $f_0(\mathcal{S}_{k+1}(Y)) f_{d-1}(\mathcal{S}_k(Y))$ it follows that

$$\Pr[A_{k+1} \cap D] \geq \left(1 - \frac{c}{n}\right)^{f_0(\mathcal{S}_{k+1}(Y)) f_{d-1}(\mathcal{S}_k(Y))}$$

$$\geq \left(1 - \frac{c}{n}\right)^{O(\log^{2k+1} n)} = 1 - o(1). \qquad \square$$

For $Y \subset \Delta_{n-1}^{(d)}$ let $r(Y) = f_d(R_\infty(Y))$ be the number of d-simplices remaining in Y after performing all possible d-collapsing steps. For $\tau \in \Delta_{n-1}(d-1)$ let $\Gamma(\tau) = \{\sigma \in \Delta_{n-1}(d) : \sigma \supset \tau\}$.

Claim 5.2 *Let $0 < c < \gamma_d$ be fixed and $p = \frac{c}{n}$. Then for any fixed $\tau \in \Delta_{n-1}(d)$:*

$$\Pr[R_\infty(Y) \cap \Gamma(\tau) \neq \varnothing] = o(1). \qquad (15)$$

Proof Let $\delta > 0$. Since $c < \gamma_d$

$$\lim_{k \to \infty} \rho_d(k, c) = \rho_d(c) = 1.$$

Choose a fixed k such that

$$\rho_d(k, c) > 1 - \frac{\delta}{3}.$$

Claim 5.1 implies that if n is sufficiently large then

$$\Pr[A_{k+1} \cap D] \geq 1 - \frac{\delta}{3}.$$

Next note that if $Y \in A_{k+1}$ then $\mathcal{S}_{k+1} = \mathcal{S}_{k+1}(Y)$ can be generated by the following inductively defined random process: $\mathcal{S}_0 = \tau$. Let $0 \leq i \leq k$. First generate $T = \mathcal{S}_i$ and let $\mathcal{U}$ denote all $\tau' \in T(d-1)$ such that $\text{dist}_T(\tau, \tau') = i$. Then, according to (say) the lexicographic order on $\mathcal{U}$, for each $\tau' \in \mathcal{U}$ pick J new vertices $z_1, \ldots, z_J$ according to the binomial distribution $B(n - n', \frac{c}{n})$, where n' is the number of vertices that appeared up to that point, and add the d-simplices $z_1 \tau', \ldots, z_J \tau'$ to T. Note that the process described above is identical to the d-tree process of Sect. 3, except for the use of the binomial distribution $B(n - n', \frac{c}{n})$ instead of the Poisson distribution $\text{Po}(c)$.

Now if $Y \in A_{k+1} \cap D$ then $n' = O(\log^{k+1} n)$ at all stages of this process. It follows that if n is sufficiently large then the total variation distance between the distributions $\mathcal{S}_{k+1}(Y)$ and $\mathcal{T}_d(k+1, c)$ is less then $\frac{\delta}{3}$. Denote by $\mathcal{C}'_d(k+1, c)$ the event that $\mathcal{S}_{k+1}(Y)$ is in A_{k+1} and collapses to τ in at most k pruning steps. The crucial observation now is that if $Y \in \mathcal{C}'_d(k+1, c)$ then $R_\infty(Y) \cap \Gamma(\tau) = \emptyset$. It follows that

$$\begin{aligned}
\Pr[R_\infty&(Y) \cap \Gamma(\tau) \neq \emptyset] \\
&\leq (1 - \Pr[A_{k+1} \cap D]) + \Pr[Y \notin \mathcal{C}'_d(k+1, c)] \\
&\leq (1 - \Pr[A_{k+1} \cap D]) + d_{TV}(\mathcal{S}_{k+1}(Y), \mathcal{T}_d(k+1, c)) \\
&\quad + (1 - \Pr[\mathcal{C}_d(k+1, c)]) \\
&\leq \tfrac{\delta}{3} + \tfrac{\delta}{3} + \tfrac{\delta}{3} = \delta. \qquad\qquad\qquad \square
\end{aligned}$$

Let

$$\mathcal{G}(Y) = \{\tau \in \Delta_{n-1}(d-1) : R_\infty(Y) \cap \Gamma(\tau) \neq \emptyset\}$$

and let $g(Y) = |\mathcal{G}(Y)|$. For a family $\mathcal{G} \subset \Delta_{n-1}(d-1)$ let $w(\mathcal{G})$ denote the set of all d-simplices $\sigma \in \Delta_{n-1}(d)$ all of whose $(d-1)$-faces are contained in $\mathcal{G}$. Using Claim 5.2 we establish the following

Theorem 5.3 *Let $\delta > 0$ and $0 < c < \gamma_d$ be fixed and let $p = \frac{c}{n}$. Then*

$$\Pr[f_d(R_\infty(Y)) > \delta n^d] = o(1).$$

Proof Let $0 < \varepsilon = \varepsilon(d, c, \delta) < 1$ be a constant whose value will be fixed later. Clearly,

$$\begin{aligned}
\Pr[f_d&(R_\infty(Y)) > \delta n^d] \\
&\leq \Pr[g(Y) > \varepsilon \delta n^d] + \Pr[g(Y) \leq \varepsilon \delta n^d \text{ and } f_d(R_\infty(Y)) > \delta n^d].
\end{aligned}$$

To bound the first summand, note that $E[g] = o(n^d)$ by Claim 5.2. Hence, by Markov's inequality

$$\Pr[g(Y) > \varepsilon \delta n^d] \leq (\varepsilon \delta n^d)^{-1} E[g] = o(1).$$

Next note that

$$\begin{aligned}
\Pr[g(Y) &\leq \varepsilon \delta n^d \text{ and } f_d(R_\infty(Y)) > \delta n^d] \\
&\leq \sum_{\{\mathcal{G} \subset \Delta_{n-1}(d-1) : |\mathcal{G}| = \varepsilon \delta n^d\}} \Pr[|w(\mathcal{G}) \cap Y(d)| > \delta n^d].
\end{aligned}$$

Fix a $\mathcal{G} \subset \Delta_{n-1}(d-1)$ such that $|\mathcal{G}| = \varepsilon \delta n^d$. By the Kruskal-Katona theorem there exists a $C_1 = C_1(d, \delta)$ such that

$$N = |w(\mathcal{G})| \leq C_1 \varepsilon^{\frac{d+1}{d}} n^{d+1}.$$

Applying the large deviation estimate for the binomial distribution $B(N, \frac{c}{n})$ and writing $C_2 = \frac{ecC_1}{\delta}$ we obtain

$$\Pr[|w(\mathcal{G}) \cap Y(d)| > \delta n^d] \le \left(C_2 \varepsilon^{\frac{d+1}{d}} \right)^{\delta n^d}.$$

On the other hand,

$$\binom{\binom{n}{d}}{\varepsilon \delta n^d} \le \left(\frac{e}{\varepsilon \delta} \right)^{\varepsilon \delta n^d}.$$

Choosing ε such that

$$\left(\frac{e}{\varepsilon \delta} \right)^\varepsilon C_2 \varepsilon^{\frac{d+1}{d}} < e^{-1}$$

it follows that

$$\Pr[g(Y) \le \varepsilon \delta n^d \text{ and } f_d(R_\infty(Y)) > \delta n^d] \le \exp(-\delta n^d).$$

Proof of Theorem 1.4 Let $c < \gamma_d$ and $p = \frac{c}{n}$. By Theorem 4.1 there exists a $\delta > 0$ such that a.a.s. any non-d-collapsible subcomplex K of $Y \in Y_d(n, \frac{c}{n})$ such that $f_d(K) \le \delta n^d$ contains the boundary of a $(d+1)$-simplex. It follows that

$$\Pr[Y \text{ non-}d\text{-collapsible}|Y \in \mathcal{F}_{n,d}] = \Pr[f_d(R_\infty(Y)) > 0|Y \in \mathcal{F}_{n,d}]$$
$$\le \Pr[f_d(R_\infty(Y)) > \delta n^d] \cdot \Pr[Y \in \mathcal{F}_{n,d}]^{-1} + \Pr[0 < f_d(R_\infty(Y)) \le \delta n^d|Y \in \mathcal{F}_{n,d}].$$

The first summand is $o(1)$ by Theorem 5.3, and the second summand is $o(1)$ by Theorem 4.1.   □

6 Concluding Remarks

Let us remark that one may show a random process statement slightly stronger than Theorem 4.1 (see [5], where a similar result is shown for the k-core of random graphs). More specifically, let us define the d-dimensional random process $\mathcal{Y}_d = \{Y_d(n, M)\}_{M=0}^{\binom{n}{d+1}}$ as the Markov chain whose stages are simplicial complexes, which starts with the full $(d-1)$-dimensional skeleton of Δ_{n-1} and no d-simplices, and in each stage $Y_d(n, M+1)$ is obtained from $Y_d(n, M)$ by adding to it one d-simplex chosen uniformly at random from all the d-simplices which do not belong to $Y_d(n, M)$. The core of a complex Y is the maximal core subcomplex of Y. Then the following holds.

Theorem 6.1 *There exists a constant $\alpha = \alpha(d) > 0$ such that for almost every d-dimensional random process $\mathcal{Y}_d = \{Y_d(n, M)\}_{M=0}^{\binom{n}{d+1}}$ there exists a stage $\hat{M} = $*

$\hat{M}(\mathcal{Y}_d)$ *such that the core of $Y_d(n, \hat{M})$ is of the size $O(1)$ and consists of boundaries of $(d + 1)$-simplices, while the core of $Y_d(n, \hat{M} + 1)$ contains at least αn^d d-simplices.*

Many questions remain open. The most obvious ones are

- What is the threshold for d-collapsibility of random simplicial complexes in $\mathcal{F}_{n,d}$? We conjecture that it is indeed $p = \gamma_d/n$.
- Find the exact threshold for the nonvanishing of $H_d(Y)$. The first two authors [1] have recently improved the upper bound given in Theorem 1.2 and they conjecture that their new bound is in fact sharp. This in particular would imply that the threshold does not depend on the underlying field.
- Although this question is implicitly included in the above two questions, it is of substantial interest in its own right: Can you show that the two thresholds (for d-collapsibility and for the vanishing of the top homology) are distinct? We conjecture that the two thresholds are, in fact, quite different. In particular, although d-collapsibility is a sufficient condition for the vanishing of H_d, there is only a vanishingly small probability that a random simplicial complex with trivial top homology is d-collapsible.

References

1. Aronshtam, L., Linial, N.: When does the top homology of a random simplicial complex vanish? arXiv:1203.3312
2. Alon, N., Spencer, J.: The Probabilistic Method, 2nd edn. Wiley-Intescience, New York (2000)
3. Cohen, D., Costa, A., Farber, M., Kappeler, T.: Topology of random 2-complexes. Discrete Comput. Geom. **47**, 117–149 (2012)
4. Kozlov, D.: The threshold function for vanishing of the top homology group of random d complexes. Proc. Am. Math. Soc. **138**, 4517–4527 (2010)
5. Łuczak, T.: Size and connectivity of the k-core of a random graph. Discret. Math. **91**, 61–68 (1991)
6. McDiarmid, C.: On the method of bounded differences. In: Siemons, J. (ed.) Surveys in Combinatorics. London Mathematical Society Lecture Note Series, vol. 141, pp. 148–188. Cambridge University Press, Cambridge (1989)
7. Stanley, R.P.: Enumerative combinatorics. In: Cambridge Studies in Advanced Mathematics, vol. 1, 2nd edn. Cambridge University Press, Cambridge (2012)