Phase transition in cohomology groups of non-uniform random simplicial complexes

Oliver Cooley*

Institute of Science and Technology Austria (ISTA)
3400 Klosterneuburg, Austria

oliver.cooley@ist.ac.at

Nicola Del Giudice* Mihyun Kang* Philipp Sprüssel*

Institute of Discrete Mathematics
Graz University of Technology
8010 Graz, Austria

{delgiudice,kang,spruessel}@math.tugraz.at

Submitted: Jul 29, 2021; Accepted: Jun 16, 2022; Published: Jul 29, 2022

Abstract

We consider a generalised model of a random simplicial complex, which arises from a random hypergraph. Our model is generated by taking the downward-closure of a non-uniform binomial random hypergraph, in which for each k, each set of $k+1$ vertices forms an edge with some probability p_k independently. As a special case, this contains an extensively studied model of a (uniform) random simplicial complex, introduced by Meshulam and Wallach [Random Structures & Algorithms 34 (2009), no. 3, pp. 408–417].

We consider a higher-dimensional notion of connectedness on this new model according to the vanishing of cohomology groups over an arbitrary abelian group R. We prove that this notion of connectedness displays a phase transition and determine the threshold. We also prove a hitting time result for a natural process interpretation, in which simplices and their downward-closure are added one by one. In addition, we determine the asymptotic behaviour of cohomology groups inside the critical window around the time of the phase transition.

Mathematics Subject Classifications: 05E45, 05C65, 05C80

*Supported by Austrian Science Fund (FWF): I3747, W1230.

1 Introduction

1.1 Motivation

One of the first and most famous results in the theory of random graphs, due to Erdős and Rényi [20], states that the uniform random graph $G(n, m)$ displays a phase transition threshold for the property of being connected at about $m = \frac{1}{2} n \log n$ edges, where $\log$ denotes the natural logarithm. Almost equivalently, in modern terminology, with high probability the binomial random graph $G(n, p)$ becomes connected around $p = \frac{\log n}{n}$ (see [42]).

The result was subsequently strengthened by Bollobás and Thomason [12] to a *hitting time* result—the random graph process, in which edges are added to an initially empty graph one by one in a uniformly random order, is very likely to become connected at exactly the moment at which the last isolated vertex disappears (i.e. acquires an edge).

More recently, there has been a focus on generalising graphs to higher-dimensional structures. One very well-studied higher-dimensional analogue of graphs is hypergraphs, most often uniform hypergraphs, in which one may consider vertex-connectedness (see e.g. [6–11, 19, 31, 40, 41]) or high-order connectedness (also known as j-tuple-connectedness, e.g. [14–16, 30]), as well as the appearance of spanning structures such as Hamilton cycles (see e.g. [1, 2, 37, 39]).

Simplicial complexes have also seen a great deal of attention as higher-dimensional analogues of graphs. The study of random simplicial complexes was initiated by Linial and Meshulam [32], who studied a model on vertex set $[n]$ in which each 2-simplex is present with probability $p = p(n)$ independently, and all 1-simplices are always present. The notion of connectedness they studied involved the vanishing of the first homology group over $\mathbb{F}_2$ (or equivalently the first cohomology group over $\mathbb{F}_2$), and they proved that this property undergoes a phase transition at threshold $p = \frac{2 \log n}{n}$. This threshold is related to the disappearance of the last isolated 1-simplex (i.e. a 1-simplex that does not lie in any 2-simplex) as was subsequently proved by Kahle and Pittel [30].

Meshulam and Wallach [35] extended the result of [32] to random simplicial k-complexes with full $(k-1)$-skeleton (for any $k \geqslant 2$), proving that the threshold for the vanishing of the $(k-1)$-th (co)homology group over $\mathbb{F}_2$, or indeed over any finite abelian group R, undergoes a phase transition at threshold $p = \frac{k \log n}{n}$. In [13], we proved the corresponding hitting time result for cohomology over $\mathbb{F}_2$, relating cohomological connectedness to the disappearance of the last isolated $(k-1)$-simplex, as a corollary of results about a slightly different model of random simplicial k-complexes generated from a random binomial $(k+1)$-uniform hypergraph by taking the downward-closure (so in particular, the complex does not necessarily have a full $(k-1)$-skeleton). A similar hitting time result in the Linial-Meshulam model and for homology groups over $\mathbb{Z}$ was proved by Łuczak and Peled [34] in the case when $k = 2$ and recently by Newman and Paquette [38] for general $k \geqslant 2$.

Since the work of Linial and Meshulam, many different models of random simplicial complexes have been introduced (see e.g. [17, 22, 26–29, 33]), and several notions of connectedness have been analysed (see e.g. [3, 4, 25, 34, 38]), as well as related concepts such as expansion [24] and bootstrap percolation [23]. In this paper, we consider a model of random simplicial complexes generated from *non-uniform* random hypergraphs, in which edges may have different sizes, and study cohomology groups over an arbitrary (not necessarily finite) abelian group R.

We note that our model includes both the model introduced by Linial and Meshulam, which was extended by Meshulam and Wallach, and the model we introduced in [13] as special cases, and therefore our main result extends and unifies the results of [13], [32], and [35]. We also note that our model is equivalent to the 'upper model' which was recently introduced independently by Farber, Mead, and Nowik [21], although they considered different properties and different ranges of probabilities to the ones we focus on in this paper. A similar model has also been considered by Costa and Farber in [18].

1.2 Model

Throughout the paper let $d \geqslant 2$ be a fixed integer and let R be an abelian group with at least two elements. We use additive notation for the group operation of R and denote the identity element by 0_R. For an integer $k \geqslant 1$, we write $[k] := \{1, \dots, k\}$ and $[k]_0 := \{0, \dots, k\}$. If A is a set with at least k elements, we denote by $\binom{A}{k}$ the family of k-element subsets of A and we call $K \in \binom{A}{k}$ a k-*set* of A.

Definition 1.1. A family $\mathcal{G}$ of non-empty finite subsets of a vertex set V is called a *simplicial complex* on V if it is downward-closed, i.e. if every non-empty set A that is contained in a set $B \in \mathcal{G}$ also lies in $\mathcal{G}$, and if furthermore the singleton $\{v\}$ is in $\mathcal{G}$ for every $v \in V$.

The elements of a simplicial complex $\mathcal{G}$ which have cardinality $i + 1$ are called i-*simplices* of $\mathcal{G}$. If $\mathcal{G}$ has no $(d + 1)$-simplices, then we call it d-*dimensional*, or a d-*complex*.[1] If $\mathcal{G}$ is a d-complex, then for each $j \in [d - 1]_0$ the j-*skeleton* of $\mathcal{G}$ is the j-complex formed by all i-simplices in $\mathcal{G}$ with $i \in [j]_0$.

We define a model of a random d-complex generated from a *non-uniform* random hypergraph, in which sets of vertices have different probabilities of forming an edge depending on their size.

Definition 1.2. For each $k \in [d]$, let $p_k = p_k(n) \in [0, 1] \subset \mathbb{R}$ be given and write $\mathbf{p} := (p_1, \dots, p_d)$. Denote by $G(n, \mathbf{p})$ the (non-uniform) binomial random hypergraph on vertex set $[n]$ in which, for all $k \in [d]$, each element of $\binom{[n]}{k+1}$ forms an edge with probability p_k independently. By $\mathcal{G}(n, \mathbf{p})$, we denote the random d-dimensional simplicial complex on $[n]$ such that

- the 0-simplices of $\mathcal{G}(n, \mathbf{p})$ are the singletons of $[n]$ and
- for each $i \in [d]$, the i-simplices are precisely the $(i+1)$-sets which are contained in edges of $G(n, \mathbf{p})$.

In other words, $\mathcal{G}(n, \mathbf{p})$ is the downward-closure of the set of edges of $G(n, \mathbf{p})$, together with all singletons of $[n]$ (if these are not already in the downward-closure).[2]

[1] Note that we do not require $\mathcal{G}$ to contain any d-simplices in order to be d-dimensional. This is in contrast to the usual terminology, but we adopt this convention for technical convenience.

[2] Note that if $\binom{n}{d+1}p_d$ is small, then it is likely that there are no d-simplices—it is for this reason that we slightly abuse terminology by referring to a d-complex even if there may not be any d-simplices.

Observe that if $\mathbf{p} = (1, \ldots, 1, p)$ for some $p = p(n) \in [0, 1]$, we obtain the model of Meshulam and Wallach, while if $\mathbf{p} = (0, \ldots, 0, p)$, we obtain the model generated from uniform random hypergraphs which we previously introduced in [13]. Indeed, if we were to set $\mathbf{p} = (p, 0, \ldots, 0)$, we would obtain the Erdős-Rényi binomial random graph, although since we assume that $d \geqslant 2$ and later that $p_d \neq 0$, this model will not formally be covered by our results.

We note that while $\mathcal{G}(n, \mathbf{p})$ is generated from $G(n, \mathbf{p})$, which is a multi-parameter, *non-uniform* random hypergraph, it is *homogeneous* in the sense that every vertex behaves identically, and every set of a given size has the same probability of appearing as an edge.

Denote by $H^i(\mathcal{G}; R)$ the i-th cohomology group of a simplicial complex $\mathcal{G}$ with coefficients in R (see (1) in Section 2.1 for a formal definition). It is well-known that $H^0(\mathcal{G}; R) = R$ if and only if $\mathcal{G}$ is connected in the topological sense (see e.g. [36, Theorem 42.1]), which we call *topologically connected* in order to distinguish it from other notions of connectedness. Observe that topological connectedness of $\mathcal{G}$ is equivalent to vertex-connectedness of the underlying hypergraph. For any integer $i \geqslant 1$, the *vanishing* of $H^i(\mathcal{G}; R)$ can be viewed as a 'higher-order connectedness' of $\mathcal{G}$.

Definition 1.3. Given a non-negative integer j, a simplicial complex $\mathcal{G}$ is called *R-cohomologically j-connected* (*j-cohom-connected* for short) if

(a) $H^0(\mathcal{G}; R) = R$;

(b) $H^i(\mathcal{G}; R) = 0$ for all $i \in [j]$.

We note that the analogous definition of connectedness considered by Meshulam and Wallach in [35] was only for the case $j = d - 1$, and only demanded the vanishing of the $(d-1)$-th cohomology group—this was reasonable for their model since with the complete $(d-1)$-dimensional skeleton, the i-th cohomology group must always vanish for all $i \in [d-2]$ (or equal R if $i = 0$).

1.3 Main results

We will consider asymptotic properties of $\mathcal{G}(n, \mathbf{p})$ as the number of vertices n tends to infinity, hence all asymptotics in the paper are with respect to n. In particular, we say that a property or an event holds *with high probability*, abbreviated to *whp*, if the probability tends to 1 as n tends to infinity.

Our first main theorem will relate the j-cohom-connectedness of $\mathcal{G}(n, \mathbf{p})$ to the absence of any *minimal obstructions* to this property. We call these obstructions *copies of* $\hat{M}_{j,k}$ for any k with $j \leqslant k \leqslant d$ (these will be defined later, see Definitions 5.3 and 5.4), and we will see in Section 5 that the presence of any of these configurations in $\mathcal{G}(n, \mathbf{p})$ is a witness for the *non-vanishing* of $H^j(\mathcal{G}(n, \mathbf{p}); R)$ (Corollary 5.10), which is 'minimal' in a natural sense (Lemma 5.11).

In particular, the strongest relation between j-cohom-connectedness and the absence of copies of $\hat{M}_{j,k}$ will be a hitting time result, analogous to the result of Bollobás and Thomason [12] for graphs, for which we will need to turn the random d-complex $\mathcal{G}(n, \mathbf{p})$ into a *process*. We do this by assigning a *birth time* to each k-simplex: more precisely, for each $k \in [d]$

and each $(k+1)$-set $K \in \binom{[n]}{k+1}$ independently, sample a birth time uniformly at random from $[0,1]$. Then $\mathcal{G}(n,\mathbf{p})$ is exactly the complex generated by the $(k+1)$-sets with birth times at most p_k, for all $k \in [d]$, by taking the downward-closure (so some simplices may be present despite having large birth time because they are subsets of a larger simplex which is present).

If we fix a 'direction vector' $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$ of non-negative real numbers (not necessarily less than 1) with $\bar{p}_d \neq 0$, set

$$\mathbf{p} = \tau\bar{\mathbf{p}} \ := \ (\min\{\tau\bar{p}_1, 1\}, \ldots, \min\{\tau\bar{p}_d, 1\}),$$

and gradually increase τ from 0 to

$$\tau_{\max} \ := \ 1/\bar{p}_d,$$

then $\mathcal{G}(n,\mathbf{p})$ becomes a process in which simplices (together with their downward-closure) arrive one by one.[3] We will denote this process by $(\mathcal{G}(n,\tau\bar{\mathbf{p}}))_{\tau\in[0,\tau_{\max}]}$, or sometimes just by $(\mathcal{G}_\tau)$ when the direction vector $\bar{\mathbf{p}}$ is clear from the context. In this way, τ may be thought of as a 'time' parameter. Let us note that if we consider a *snapshot* of the process $(\mathcal{G}_\tau)$ at time $\tau = \tau_0$, then it has the same distribution as $\mathcal{G}(n,\tau_0\bar{\mathbf{p}})$. Therefore we will often give definitions or state and prove results for the random complex $\mathcal{G}_\tau := \mathcal{G}(n,\tau\bar{\mathbf{p}})$ for some appropriate value of τ, and subsequently apply them to the process at that time, meaning in particular that $\mathcal{G}_{\tau_0} \subseteq \mathcal{G}_{\tau_1}$ if $\tau_0 \leqslant \tau_1$, i.e. we have a natural coupling of the random complexes rather than sampling them independently. In other words, for the rest of the paper we take one sample of random birth times uniformly from $[0,1]$ and independently for all simplices, and whenever we refer to $\mathcal{G}_\tau$, we mean the complex generated by the simplices with *scaled birth times* (scaled according to $\bar{\mathbf{p}}$) at most τ (see (2) in Section 3 for the formal definition of scaled birth time).

Note that the evolution of the process $(\mathcal{G}_\tau)$ is unchanged if the direction vector $\bar{\mathbf{p}}$ is scaled by a multiplicative factor. Therefore we would like to scale $\bar{\mathbf{p}}$ so that we expect the last copy of $\hat{M}_{j,k}$ to disappear when τ is close to 1. Indeed, our first main result in particular states that this happens for a specific type of direction vector that we call *j-critical direction vector*, or *j-critical direction* for short, and that will be formally defined in Section 3.1 (Definition 3.3).

Theorem 1.4 (Hitting time). For $j \in [d-1]$ and a j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_d)$ with $\bar{p}_d \neq 0$, let $\tau_{\max} = 1/\bar{p}_d$ and consider the process $(\mathcal{G}_\tau) = (\mathcal{G}(n,\tau\bar{\mathbf{p}}))_{\tau\in[0,\tau_{\max}]}$. Let

$$\tau_j^* \ := \ \sup\{\tau \in \mathbb{R}_{\geqslant 0} \mid \mathcal{G}_\tau \text{ contains a copy of } \hat{M}_{j,k} \text{ for some } k \text{ with } j \leqslant k \leqslant d\}.$$

Then for every function ω of n which tends to infinity as $n \to \infty$, the following statements hold with high probability.

(a) $\tau_j^* = 1 + o\left(\frac{\omega}{\log n}\right)$.

(b) For all $\tau \in [0, \tau_j^*)$, the random d-complex process $(\mathcal{G}_\tau)$ is not R-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_\tau; R) \neq R \quad \text{ or } \quad H^i(\mathcal{G}_\tau; R) \neq 0 \text{ for some } i \in [j].$$

[3]Observe that by time $\tau = \tau_{\max}$, all d-simplices will be present deterministically, and therefore also all simplices of dimension $k \leqslant d$ will be present as part of their downward-closure.

(c) For all $\tau \in [\tau_j^*, \tau_{\max}]$, the random d-complex process $(\mathcal{G}_\tau)$ is R-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_\tau; R) = R \quad \text{and} \quad H^i(\mathcal{G}_\tau; R) = 0 \text{ for all } i \in [j].$$

Observe that in Theorem 1.4 we do not consider j-cohom-connectedness for the case $j = 0$. Indeed, the condition $H^0(\mathcal{G}_\tau; R) = R$ corresponds to topological connectedness of $\mathcal{G}_\tau$, i.e. vertex-connectedness of the underlying (non-uniform) random hypergraph, which has been extensively studied, and for which much stronger results are known (see e.g. [14, 40]). However, topological connectedness is a necessary condition for the j-cohom-connectedness of $\mathcal{G}_\tau$ (see Definition 1.3), therefore in order to make this paper self-contained, this case is treated separately in Lemma 4.2.

Furthermore, we observe that neither j-cohom-connectedness nor the presence of copies of $\hat{M}_{j,k}$ are necessarily monotone properties (as we will see in Example 5.12), which makes the proofs significantly harder. Indeed, it is not immediately clear that j-cohom-connectedness should have a single threshold—in principle, the random d-complex process $(\mathcal{G}_\tau)$ could switch between being j-cohom-connected or not several times. However, Theorem 1.4 implies that with high probability this does not happen and there is indeed a single threshold.

Our second main result gives an asymptotic description of the j-th cohomology group of $\mathcal{G}_\tau$ for values of τ in the *critical window*, i.e. $\tau = 1 + O(1/\log n)$.

Theorem 1.5 (Rank in the critical window). Let $c \in \mathbb{R}$ be a constant and suppose that $(c_n)_{n \geqslant 1}$ is a sequence of real numbers with $c_n \xrightarrow{n\to\infty} c$. Let $j \in [d-1]$, let $\tau = 1 + \frac{c_n}{\log n}$, and consider $\mathbf{p} = \tau\bar{\mathbf{p}}$ for a j-critical direction $\bar{\mathbf{p}}$. Then there exists a constant $\mathcal{E} = \mathcal{E}(c, \bar{\mathbf{p}})$ such that with high probability

$$H^j(\mathcal{G}_\tau; R) \quad = \quad R^Y,$$

where Y is a Poisson random variable with mean $\mathcal{E}$.

The constant $\mathcal{E}$ will be explicitly defined in (5) in Section 3.1.

As a consequence of Theorems 1.4 and 1.5, we derive an explicit expression for the limiting probability of $\mathcal{G}_\tau$ being j-cohom-connected within the critical window.

Corollary 1.6. Let $(c_n)_{n\geqslant 1}$, c, j, τ, $\bar{\mathbf{p}}$, $\mathbf{p}$, and $\mathcal{E}$ be given as in Theorem 1.5. Then

$$\mathbb{P}\left(\mathcal{G}_\tau \text{ is } R\text{-cohomologically } j\text{-connected}\right) \xrightarrow{n\to\infty} \exp(-\mathcal{E}).$$

1.4 Proof techniques

The three statements of the Hitting Time Theorem (Theorem 1.4) follow from auxiliary results presented in Section 4, which in turn are proved throughout the paper.

More precisely, we show in Lemma 4.1 that the choice of a j-critical direction $\bar{\mathbf{p}}$ (Definition 3.3) implies that the last minimal obstruction disappears at around time $\tau = 1$, thus proving statement (a) of Theorem 1.4.

The main ingredient in the proof of Theorem 1.4 (b) will be Lemma 4.4, which states that for every constant $\varepsilon > 0$, whp $H^j(\mathcal{G}_\tau; R) \neq 0$ for every τ in the interval

$$I_j(\varepsilon) \; := \; \left[\frac{\varepsilon}{n}, \tau_j^*\right).$$

To prove this, in Section 8 we split $I_j(\varepsilon)$ into three subintervals and show that whp in each of these there exists a copy of the obstruction $\hat{M}_{j,k}$ for some $j \leqslant k \leqslant d$ (Lemmas 8.1, 8.3 and 8.4), and thus $H^j(\mathcal{G}_\tau; R) \neq 0$ (Corollary 5.10).

In addition, we show that for every $i \in [j-1]$ there exists an appropriate scaling factor τ such that the vector $\tau\bar{\mathbf{p}}$ is an i-critical direction (Lemma 4.5). Thus we can apply Lemma 4.4 with j replaced by i and find intervals $I_i(\varepsilon)$ where whp $H^i(\mathcal{G}_\tau; R) \neq 0$ (Corollary 4.6). We further define an interval $I_0(\varepsilon)$ and show that whp $\mathcal{G}_\tau$ is not topologically connected for every $\tau \in I_0(\varepsilon)$ (Lemma 4.2). In this way we can complete the proof of Theorem 1.4 (b) by showing that we can choose ε such that $[0, \tau_j^*) = \bigcup_{i=0}^{j} I_i(\varepsilon)$ and thus $\mathcal{G}_\tau$ is not j-cohom-connected throughout the subcritical case.

By definition of τ_j^*, whp for any $\tau \geqslant \tau_j^*$ there are no copies of the minimal obstruction $\hat{M}_{j,k}$ for any $k = j, \ldots, d$, thus in order to prove statement (c) of Theorem 1.4 we need to show that whp no other 'larger' obstructions to the vanishing of $H^j(\mathcal{G}_\tau; R)$ appear in the complex. This is given by Lemma 4.7, which we prove in Section 9. We show that the smallest support of any non-zero element of the cohomology group must be *traversable* (Lemma 9.4), a very useful property that allows us to define a search process, with which we can construct such a support. By bounding the number of ways this search process can evolve, we also bound the number of possible supports and the probability that such a non-zero element of the cohomology group exists (Lemmas 9.6 and 9.8).

To prove the Rank Theorem (Theorem 1.5) and Corollary 1.6, in Section 10 we will use the fact that for values of τ 'close' to 1 whp the only obstructions to j-cohom-connectedness are copies of $\hat{M}_{j,k}$ (Corollary 9.14) and that indeed they are a *minimal* set of generators for $H^j(\mathcal{G}_\tau; R)$. We conclude the proof of the Rank Theorem by showing that the number of such obstructions converges in distribution to a Poisson random variable (Lemma 10.1). Finally we prove Corollary 1.6 by applying Theorem 1.5 to determine the probability that the j-th cohomology group vanishes and Theorem 1.4 to show that whp all lower cohomology groups vanish (except the zero-th, which is R).

1.5 Outline of the paper

The paper is structured as follows.

Section 2 gives a short review of some standard concepts of cohomology theory and introduces a structure in a complex whose existence will be our most frequent argument for cohomology groups to *not* vanish. In Section 3, we introduce the parametrisation of a j-critical direction and argue that this parametrisation covers all interesting cases of our main results. Section 4 contains the main auxiliary results that we combine to prove the Hitting Time Theorem (Theorem 1.4). The proofs of the auxiliary results of Section 4 will follow in Sections 5–9. In particular, the results of Section 9 will also lay the foundation of the proof of the Rank Theorem (Theorem 1.5), which is presented in Section 10, together with the proof of Corollary 1.6. In Section 11 we discuss our main results and present some open problems.

Some standard but technical proofs are omitted from the main text, but included in Appendix A for completeness. Finally, in Appendix B we include a glossary of some of the most important terminology and notation used in the paper, for easy reference.

2 Preliminaries from cohomology

2.1 Cohomology

Let us review the standard notions of cohomology groups of a d-dimensional simplicial complex $\mathcal{G}$.

Let $j \in [d]_0 = \{0, 1, \dots, d\}$. To define cohomology groups, one considers *ordered j-simplices*, that is, j-simplices with an ordering of their vertices.[4] We adopt the notation $[v_0, \dots, v_j]$ for a j-simplex whose vertices are ordered $v_0, \dots, v_j$. If $\sigma = [v_0, \dots, v_j]$ is an ordered j-simplex and $i \in [j]_0$, then $[v_0, \dots, \hat{v}_i, \dots, v_j]$ denotes the ordered $(j-1)$-simplex obtained from σ by removing v_i (and preserving the order on the remaining vertices).

Recall that we will be considering cohomology groups over an arbitrary (non-trivial) abelian group R. A function f from the set of ordered j-simplices in $\mathcal{G}$ to R is called a *j-cochain* if $f(\sigma) = -f(\sigma')$ whenever σ' is obtained from σ by exchanging the positions of two vertices in the ordering of the simplex. For a j-cochain f, we define its *support* $\mathrm{supp}(f)$ to be the set of *unordered* simplices σ such that f maps some (and thus every) ordering of σ to a non-zero value.

The set $C^j(\mathcal{G}; R)$ of j-cochains in $\mathcal{G}$ forms a group with respect to pointwise summation, defined by $(f_1+f_2)(\sigma) := f_1(\sigma)+f_2(\sigma)$. For $j \in [d-1]_0$, we define the *coboundary operator* $\delta^j \colon C^j(\mathcal{G}; R) \to C^{j+1}(\mathcal{G}; R)$ by

$$(\delta^j f)([v_0, \dots, v_{j+1}]) := \sum\nolimits_{i=0}^{j+1} (-1)^i f([v_0, \dots, \hat{v}_i, \dots, v_{j+1}]).$$

Clearly, δ^j is a group homomorphism. Furthermore, let δ^{-1} and δ^d denote the unique group homomorphisms $\delta^{-1} \colon \{0\} \to C^0(\mathcal{G}; R)$ and $\delta^d \colon C^d(\mathcal{G}; R) \to \{0\}$. For each $j \in [d]_0$, the j-cochains in $\ker \delta^j$ and in $\operatorname{im} \delta^{j-1}$ are called *j-cocycles* and *j-coboundaries*, respectively.

A straightforward calculation shows that every j-coboundary is also a j-cocycle, i.e. $\operatorname{im} \delta^{j-1} \subseteq \ker \delta^j$. Thus, we can define the *j-th cohomology group of $\mathcal{G}$ with coefficients in R* as the quotient group

$$H^j(\mathcal{G}; R) := \ker \delta^j / \operatorname{im} \delta^{j-1}. \tag{1}$$

2.2 Non-vanishing of cohomology groups

In view of Theorems 1.4 and 1.5, we are particularly interested in when $H^j(\mathcal{G}; R)$ vanishes for $j \in [d-1]$, which happens if and only if every j-cocycle is also a j-coboundary. Hence, we need a criterion for a j-cocycle (or more generally a j-cochain) *not* to be a j-coboundary, which will be provided by Lemma 2.2. To this end, we need the following definition.

Definition 2.1. For any $(j+2)$-set A in a complex $\mathcal{G}$, the collection of all $(j+1)$-sets of A is called a *j-shell* if each of them forms a j-simplex in $\mathcal{G}$.

Observe that in Definition 2.1, the set A does *not* have to form a $(j+1)$-simplex in $\mathcal{G}$. Thus, a j-shell is the boundary of a $(j+1)$-simplex A of the complete complex such that all

[4] When we consider simplices *without* an ordering, we will often simply refer to them as 'simplices' instead of 'unordered simplices'.

simplices in this boundary are simplices of $\mathcal{G}$, but A itself may not be. If the collection of all $(j+1)$-subsets of a $(j+2)$-set A forms a j-shell, with a slight abuse of terminology we also refer to the set A itself as a j-shell.

Lemma 2.2. Let $j \in [d-1]$, let f be a j-cochain in a d-dimensional complex $\mathcal{G}$ on $[n]$ and suppose that there exists $A \in \binom{[n]}{j+2}$ such that

(a) A is a j-shell in $\mathcal{G}$ and

(b) precisely one $(j+1)$-set of A lies in the support of f.

Then f is not a j-coboundary in $\mathcal{G}$.

Proof. Let $\mathcal{G}' := \mathcal{G} \cup \{A\}$ and observe that this is a simplicial complex, because all proper non-empty subsets of A were already simplices in $\mathcal{G}$ by condition (a). Denote the vertices in A by $v_0, \ldots, v_{j+1}$ such that $\{v_1, \ldots, v_{j+1}\} \in \operatorname{supp}(f)$. By (b), this means that

$$(\delta^j f)([v_0, \ldots, v_{j+1}]) = \sum_{i=0}^{j+1} (-1)^i f([v_0, \ldots, \hat{v}_i, \ldots, v_{j+1}]) = f([v_1, \ldots, v_{j+1}]) \neq 0_R.$$

This implies that while f may be a j-cocycle in $\mathcal{G}$, it is certainly not a j-cocycle in $\mathcal{G}'$. Thus in particular f is not a j-coboundary in $\mathcal{G}'$. Since $\mathcal{G}$ and $\mathcal{G}'$ have the same sets of j-simplices and of $(j-1)$-simplices, this means that f is also not a j-coboundary in $\mathcal{G}$. $\square$

3 Parametrisation

3.1 Defining the parametrisation

In this section we define the concept of a j-critical direction, which appears in Theorems 1.4 and 1.5.

Given a direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, let $k \in [d]$ be an index such that $\bar{p}_k \neq 0$, and let K be a $(k+1)$-set with birth time t_K. The *scaled birth time* of K is defined as

$$\tau_K := \frac{t_K}{\bar{p}_k}. \tag{2}$$

If $\bar{p}_k = 0$ we view all $(k+1)$-sets as having infinite scaled birth time. Thus τ_K is distributed uniformly in $[0, 1/\bar{p}_k]$, and $\mathcal{G}_\tau$ consists of all those simplices with *scaled* birth time at most τ, together with their downward-closure[5]. Observe that if a $(k+1)$-set K is contained in an $(l+1)$-set L with $k+1 \leqslant l \leqslant d$ and $\tau_L < \tau_K$, then K forms a k-simplex in $\mathcal{G}_\tau$ for every $\tau \geqslant \tau_L$ and thus exists in $(\mathcal{G}_\tau)$ before its own scaled birth time.

The motivation for the following definitions will become apparent later (see Lemma 5.14 and Section 3.2).

[5]With probability 1 no two simplices have the same scaled birth time, which is important for the process interpretation.

Definition 3.1. Given $j \in [d-1]$, a vector $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$ is called *j-admissible* if for each $1 \leqslant k \leqslant d$ there are real-valued constants $\bar{\alpha}_k$,$\bar{\gamma}_k$, and a function $\bar{\beta}_k = \bar{\beta}_k(n)$ such that

$$\bar{p}_k = \begin{cases} \dfrac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)! & \text{for } j \leqslant k \leqslant d, \\ \dfrac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}} & \text{for } 1 \leqslant k \leqslant j-1, \end{cases} \tag{3}$$

and furthermore

(A1) at least one of $\bar{\alpha}_k, \bar{\gamma}_k$ is zero and neither of them is negative;

(A2) if $\bar{\alpha}_k = 0$, then either $\bar{\beta}_k \equiv 0$ or $\bar{\beta}_k$ is positive and *subpolynomial* in the sense that for every constant $\varepsilon > 0$, we have $\bar{\beta}_k = o(n^{\varepsilon})$, but $\bar{\beta}_k = \omega(n^{-\varepsilon})$;

(A3) if $\bar{\gamma}_k = 0$, then $|\bar{\beta}_k| = o(\log n)$;

(A4) there exists an index $j+1 \leqslant k_0 \leqslant d$ with $\bar{\alpha}_{k_0} > 0$.

The following observation follows immediately from the definition, and will be used implicitly at many points in the paper.

Remark 3.2. If $\bar{\mathbf{p}}$ is j-admissible and $k \geqslant j+1$, then $\bar{p}_k = O\left(\frac{\log n}{n}\right) = o(1)$. In particular, if $\mathbf{p} = \tau \bar{\mathbf{p}}$ for some $\tau = O(1)$, then $p_k = \tau \bar{p}_k \leqslant 1$.

This observation means that, for $k \geqslant j+1$ and for τ not too large, we have that $p_k = \tau \bar{p}_k$ is indeed a probability term and we can use it in calculations without having to replace it by 1. On the other hand, for $k = j$ we often need to be slightly more careful.

Note that some of the properties in Definition 3.1 can be guaranteed simply by scaling $\bar{\mathbf{p}}$ and choosing $\bar{\alpha}_k, \bar{\gamma}_k, \bar{\beta}_k$ appropriately, but that some other properties place restrictions on the direction. However, we will see later (Section 3.2) that it is reasonable to restrict attention to j-admissible vectors $\bar{\mathbf{p}}$. Indeed, by scaling appropriately we can even go further: given a j-admissible vector $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, for every index k with $j \leqslant k \leqslant d$ and $\bar{p}_k \neq 0$ we define the parameters

$$\begin{aligned} \bar{\lambda}_k &:= j+1-\bar{\gamma}_k-(k-j+1)\sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i, \\ \bar{\mu}_k &:= -(k-j+1)\sum\nolimits_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \begin{cases} 0 & \text{if } \bar{p}_k > 1, \\ \log\log n & \text{if } \bar{p}_k \leqslant 1 \text{ and } \bar{\alpha}_k \neq 0, \\ \log(\bar{\beta}_k) & \text{if } \bar{p}_k \leqslant 1 \text{ and } \bar{\alpha}_k = 0, \end{cases} \\ \bar{\nu}_k &:= \begin{cases} -\log((j+1)!) & \text{if } k = j, \\ -\log(j!)-\log(k-j+1)+\log(\bar{\alpha}_k) & \text{if } k \neq j \text{ and } \bar{\alpha}_k \neq 0, \\ -\log(j!)-\log(k-j+1) & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

Note that all $\bar{\lambda}_k$, $\bar{\nu}_k$ are constants, because the $\bar{\alpha}_i, \bar{\gamma}_k$ are constants, while the $\bar{\mu}_k$ are functions of n with $\bar{\mu}_k = o(\log n)$, by Definition 3.1.

Definition 3.3. We say that a j-admissible vector $\bar{\mathbf{p}}$ is a *j-critical direction vector* if

(C1) $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k \leqslant 0$, for all indices k with $j \leqslant k \leqslant d$ and $\bar{p}_k \neq 0$;

(C2) $\bar{\lambda}_{\bar{k}} \log n + \bar{\mu}_{\bar{k}} + \bar{\nu}_{\bar{k}} = 0$, for some $\bar{k}$ with $j \leqslant \bar{k} \leqslant d$.

With slight abuse of terminology, we will omit the word 'vector' and simply talk about *j-critical directions*.

Observe that a j-critical direction is *not* a vector of length 1, but rather it is chosen so that it turns out to be 'critical' for j-cohom-connectedness of $\mathcal{G}(n, \bar{\mathbf{p}})$.

More generally, if we have a vector $\mathbf{p} = (p_1, \ldots, p_d)$ (where we will usually have $\mathbf{p} = \tau\bar{\mathbf{p}}$), we would like to define parameters analogous to those for $\bar{\mathbf{p}}$.

Definition 3.4. Given a vector $\mathbf{p} = (p_1, \ldots, p_d)$, for each $k \in [d]$, define

$$\begin{aligned} \alpha_k &:= \lim_{n\to\infty} \left(\frac{p_k n^{k-j}}{(k-j)! \log n} \right), \\ \gamma_k &:= \sup\{\gamma \in \mathbb{R} \mid p_k n^{k-j+\gamma} = o(1)\}, \\ \beta_k &:= \frac{n^{k-j+\gamma_k} p_k}{(k-j)!} - \alpha_k \log n, \end{aligned}$$

if the limit and the supremum exist.

Furthermore, we define the parameters λ_k, μ_k, and ν_k analogously to (4), with $\bar{\alpha}_k$, $\bar{\gamma}_k$, and $\bar{\beta}_k$ replaced by α_k, γ_k, and β_k, respectively.

The following observation follows directly from the definition.

Remark 3.5. If $\bar{\mathbf{p}}$ is a j-critical direction and $\mathbf{p} = \tau\bar{\mathbf{p}}$ for some $\tau = O(1)$, then the analogue of (A1) also holds for $\mathbf{p}$, i.e. for all $1 \leqslant k \leqslant d$, at least one of α_k, γ_k is zero and neither of them is negative.

In order to prove Theorem 1.5, we will need to take a closer look at how the process behaves within the critical window, which is the range where whp the complex $\mathcal{G}_\tau$ switches from being not j-cohom-connected to being j-cohom-connected. More precisely, we consider $\tau = 1 + O(1/\log n)$ (cf. Theorem 1.5). We also need the following concepts.

Definition 3.6. Given a j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, an index k with $j \leqslant k \leqslant d$ and $\bar{p}_k \neq 0$ is called a *critical dimension* if $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k = O(1)$, i.e. $\bar{\lambda}_k = 0$ and $\bar{\mu}_k = O(1)$ (recall that $\bar{\nu}_k = O(1)$). We denote by $\mathcal{C} = \mathcal{C}(\bar{\mathbf{p}}, j)$ the set of all critical dimensions for the j-critical direction $\bar{\mathbf{p}}$.

It will turn out (Lemma 10.1) that for any $\tau = 1 + O(1/\log n)$, the critical dimensions are precisely those indices k for which there is a positive asymptotic probability of having copies of a reduced version of $\hat{M}_{j,k}$, called $M_{j,k}$ (Definitions 5.5 and 5.6), in $\mathcal{G}_\tau$. Furthermore, if we consider $\tau = 1 + \frac{c_n}{\log n}$ with $c_n \xrightarrow{n\to\infty} c \in \mathbb{R}$, then the constant $\mathcal{E}$ which appeared in Theorem 1.5 is precisely

$$\mathcal{E} := \exp(-c(j+1)) \sum_{k\in\mathcal{C}} \exp(\bar{\mu}_k + \bar{\nu}_k + c\gamma_k), \tag{5}$$

as we will see in the proof of Theorem 1.5 (Section 10). We will also see that for any critical dimension k, the term $\exp(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1))$ is closely related to the number of copies of $M_{j,k}$ (Corollary 6.6).

3.2 Justifying the parametrisation

In this section we discuss the choice of parametrisation of $\bar{\mathbf{p}}$ and the assumptions made for j-admissibility and j-criticality in Definitions 3.1 and 3.3. Note that the arguments in this section are independent of all results and proofs in this paper—rather, they justify why the assumptions made in the main theorems are reasonable and cover all interesting cases.

We first justify the parametrisation of $\bar{\mathbf{p}}$ in terms of the $\bar{\alpha}_k, \bar{\beta}_k, \bar{\gamma}_k$. We note that scaling $\bar{\mathbf{p}}$ by a factor c (which may be a function of n) has no effect on the evolution of the process $(\mathcal{G}_\tau)$, since $\mathcal{G}(n, c\tau\bar{\mathbf{p}}) = \mathcal{G}(n, \tau'\bar{\mathbf{p}})$, where $\tau' = c\tau$. We therefore aim to choose $\bar{\mathbf{p}}$ such that the critical range for j-cohom-connectedness occurs around time $\tau = 1$, i.e. when $\mathbf{p} = \bar{\mathbf{p}}$.

Observe that the probabilities p_i with $i \in [j-1]$ have no influence on the j-th cohomology group $H^j(\mathcal{G}_\tau; R)$. To see this, we note that the j-th cohomology group depends only on the set of $(j-1)$-simplices, the set of j-simplices, and the set of $(j+1)$-simplices of $\mathcal{G}_\tau$. The probabilities p_i with $i \in [j-2]$ have no influence on any of these sets, while p_{j-1} only affects the set of *isolated* $(j-1)$-simplices. Isolated $(j-1)$-simplices, however, have no effect on the set of j-coboundaries, and thus do not influence $H^j(\mathcal{G}_\tau; R)$. Therefore, when we consider whether or not the j-th cohomology group vanishes, we will only take the probabilities $p_j, \ldots, p_d$ into account.

3.2.1 Approximate order: Justifying $\bar{p}_k = O\left(\frac{\log n}{n^{k-j}}\right)$

We first explain why we may assume that $\bar{p}_k = O\left(\frac{\log n}{n^{k-j}}\right)$. In particular, this will imply the assumption $\bar{\gamma}_k \geqslant 0$, once $\bar{\gamma}_k$ is defined (see (8)).

What range of $\mathbf{p}$ do we expect to be critical for j-cohom-connectedness of $\mathcal{G}(n, \mathbf{p})$? Let us first look at a single probability p_k, i.e. consider

$$\mathbf{p} = (0, \ldots, 0, p_k, 0 \ldots, 0).$$

For $R = \mathbb{F}_2$ and $j+1 \leqslant k \leqslant d$, [13, Theorem 1.11] states that the critical range lies around

$$p_k = \frac{(j+1)\log n + \log\log n}{(k-j+1)n^{k-j}}(k-j)!.$$

It is therefore reasonable to expect the critical range for general coefficient group R and general $\mathbf{p} = (p_1, \ldots, p_d)$ to lie around

$$p_k = \frac{\alpha_k \log n + r_k}{n^{k-j}}(k-j)!, \quad \forall j \leqslant k \leqslant d, \tag{6}$$

where each α_k is a non-negative constant, at least one α_k is non-zero, and each $r_k = r_k(n)$ is a function of order $o(\log n)$.

To justify this more precisely, note that if $p_k \geqslant \frac{c_k \log n}{n^{k-j}}(k-j)!$ for some constant $c_k > j+1$, a simple first moment calculation shows that whp $\mathcal{G}_\tau$ has a complete j-skeleton, and therefore if it is j-cohom-connected, adding further k-simplices for $j \leqslant k \leqslant d$ will not change this. Furthermore, it follows from the results of [13] (for $R = \mathbb{F}_2$), and indeed also from Theorem 1.4 (for general R), that the complex will in fact be j-cohom-connected whp if p_k is this large, and therefore it is reasonable to scale the chosen direction $\bar{\mathbf{p}}$ in such a way that $\bar{p}_k = O(\frac{\log n}{n^{k-j}})$ for every $j+1 \leqslant k \leqslant d$.

Thus, let us suppose that for each k with $j \leqslant k \leqslant d$, the limit

$$\bar{\alpha}_k := \lim_{n\to\infty}\left(\frac{\bar{p}_k n^{k-j}}{(k-j)!\log n}\right) \tag{7}$$

exists. This is a reasonable assumption because in terms of phase transitions, we are interested in how the model behaves depending on the asymptotic behaviour of the probabilities p_k, which should not fluctuate between, say, $\frac{\log n}{n^{k-j}}$ and $\frac{2\log n}{n^{k-j}}$. Observe that if $\bar{p}_k = o(\log n/n^{k-j})$, then $\bar{\alpha}_k = 0$. Indeed, we next argue that we may also assume that at least one $\bar{\alpha}_k$ is non-zero.

3.2.2 Existence of k_0: Justifying (A4)

So far we have only guaranteed certain properties of $\bar{\mathbf{p}}$ by scaling appropriately. By rescaling once more if necessary, we can certainly guarantee that $\bar{\alpha}_k \neq 0$ for some $j \leqslant k \leqslant d$, but we would like to ensure that this does not *only* hold for $k = j$, i.e. that there is in fact some $j+1 \leqslant k \leqslant d$ such that $\bar{\alpha}_k \neq 0$. Note that this cannot necessarily be achieved by a simple rescaling without potentially violating the condition that $p_j = O(\log n)$.

Instead, we consider the two cases:

(a) $\bar{p}_j \geqslant 1$ and $\bar{\alpha}_k = 0$ for all $k = j+1, \ldots, d$;

(b) $\bar{p}_j \geqslant 1$ and for some $j+1 \leqslant k \leqslant d$ we have $\bar{\alpha}_k \neq 0$.

We argue that case (a) can be easily reduced to case (b) and therefore we may assume that there exists $k_0 \geqslant j+1$ with $\bar{\alpha}_{k_0} \neq 0$ as stated in (A4).

Indeed, suppose we have the slightly more general case than case (a), that $\bar{p}_j \geqslant 1$ and $\bar{p}_k \leqslant \frac{\log n}{Cn^{k-j}}$ for all $k = j+1, \ldots, d$ and for some sufficiently large C. In this case, a simple second moment argument shows that there exists a constant $c > 0$ such that for $\tau = cn^{-j}$, whp $\mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$ contains an isolated k-simplex for some $0 \leqslant k \leqslant j-1$, which guarantees the existence of an isolated simplex of dimension at most k for any $\tau \leqslant cn^{-j}$, and therefore $(\mathcal{G}_\tau)$ is not j-cohom-connected in the interval $[0, cn^{-j}]$. Furthermore, another second moment argument shows that if C is large enough, whp $\mathcal{G}_1 = \mathcal{G}(n, \bar{\mathbf{p}})$ contains $\Omega(n^{j+\frac{1}{2}})$ isolated j-simplices. Conditioned on its presence in $\mathcal{G}_1$, the probability that an isolated j-simplex was already present in $\mathcal{G}_{cn^{-j}}$ is at least cn^{-j} independently for each such simplex, and therefore with high probability one of these was present throughout the entire range $\tau \in [cn^{-j}, 1]$. In other words, either the presence of isolated k-simplices for some $k \leqslant j-1$ or of isolated j-simplices ensure that whp the process is certainly not j-cohom-connected until the time when it has a complete j-skeleton. Therefore we may increase $\bar{p}_k$ for all $k \geqslant j+1$ by the same factor (equivalent to decreasing $\bar{p}_j$) until $\bar{p}_k = \frac{\log n}{Cn^{k-j}}$ for some k without affecting which appearances of simplices cause the process becomes j-cohom-connected. In other words, we may assume that $\bar{\alpha}_{k_0} > 0$ for some $k_0 \geqslant j+1$.

3.2.3 Lower bound on $\bar{p}_k$: Justifying $\bar{\gamma}_k < \infty$

Furthermore, we may assume that each non-zero probability $\bar{p}_k$ is not 'too small', or in other words that any $\bar{p}_k$ which is very small is in fact 0. More precisely, we have shown the existence of an index k_0 with $\bar{\alpha}_{k_0} \neq 0$, which implies that $\bar{p}_{k_0} = \Theta(\frac{\log n}{n^{k_0-j}})$. Now if $\bar{p}_k \leqslant n^{-(k+k_0-j+1)}$, then a simple first moment calculation shows that whp all k_0-simplices are born (and so in particular the complex is j-cohom-connected) before any k-simplices are born. Thus we may set $\bar{p}_k = 0$ without affecting when the process is j-cohom-connected. Therefore we may assume that

$$\bar{\gamma}_k := \sup\{\gamma \in \mathbb{R} \mid \bar{p}_k n^{k-j+\gamma} = o(1)\} \tag{8}$$

exists for every k with $j \leqslant k \leqslant d$ and $\bar{p}_k \neq 0$. By the existence of the limit in (7), we have $\bar{\gamma}_k \geqslant 0$.

3.2.4 Fine-tuning

Finally, let $\bar{\beta}_k$ be the function of n for which

$$\bar{p}_k = \frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)!. \tag{9}$$

Note that the function $\bar{\beta}_k$ might be negative if $\bar{\alpha}_k \neq 0$.

3.2.5 j-admissibility

So far we have only ensured that properties (A1)–(A3) hold for $j \leqslant k \leqslant d$ (note that (A4) is independent of k). To show that we may assume that these properties also hold for $1 \leqslant k \leqslant j-1$, we use a similar argument to the one in Section 3.2.2: if for some $k \in [j-1]$ we have $\bar{p}_k \geqslant Cn^{j-k}\log n$ for some sufficiently large constant C, then whp $(\mathcal{G}_\tau)$ contains an isolated i-simplex for some $i \in [k]_0$ (and is therefore not j-cohom-connected) until the moment when it has a complete k-skeleton. Therefore we may decrease $\bar{p}_k$ to $Cn^{j-k}\log n$ without changing when the process becomes j-cohom-connected. In other words, we may assume that $\bar{p}_k = O\left(\frac{\log n}{n^{k-j}}\right)$, and thus also that $\bar{\alpha}_k, \bar{\beta}_k, \bar{\gamma}_k$ are well-defined, and properties (A1)–(A3) follow.

3.2.6 j-criticality

It only remains to justify the assumptions of Definition 3.3. These properties can also be guaranteed by appropriate scaling of $\bar{\mathbf{p}}$.

To see this, observe that scaling $\bar{\mathbf{p}}$ by a constant C^* also scales the $\bar{\alpha}_k$ by the same factor C^*, while leaving the $\bar{\gamma}_k$ unchanged. Thus if we let $\bar{\mathbf{p}} = C^*\mathbf{p}$, where $\mathbf{p}$ is a j-admissible direction, since $\alpha_{k_0} > 0$, we have

$$\bar{\lambda}_k = j+1-\bar{\gamma}_k - \Theta(C^*).$$

Thus by choosing C^* large enough, we can ensure that $\bar{\lambda}_k < -1$ for all $k \geqslant j$. Since $\bar{\lambda}_k \log n$ is the main term in $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k$, this would mean that (C1) certainly holds if C^* is large enough. On the other hand, since $\bar{\gamma}_{k_0} = \gamma_{k_0} = 0$, if C^* is small enough we have $\bar{\lambda}_{k_0} \geqslant j+1-\frac{1}{2} > 0$, i.e. at least one of the $\bar{\lambda}_k$ is positive. By continuity, we may choose C^* such that (C1) and (C2) both hold.

4 Hitting Time Theorem: proof of Theorem 1.4

In this section, we provide an outline of the most important auxiliary results of the paper and show how together they prove the Hitting Time Theorem (Theorem 1.4). These auxiliary results are proved throughout the rest of the paper.

4.1 Hitting time and subcritical case

To prove Theorem 1.4 (a), recall that τ_j^* is the birth time of the simplex whose appearance causes the last copy of $\hat{M}_{j,k}$ for any $j \leqslant k \leqslant d$ to disappear, where the definition of $\hat{M}_{j,k}$ appears in Section 5. We want to show that this happens at around time $\tau = 1$. More precisely, we will prove the following.

Lemma 4.1. Let ω be a function of n that tends to infinity as $n \to \infty$. If $\bar{\mathbf{p}}$ is a j-critical direction, then whp

$$1 - \frac{\omega}{\log n} < \tau_j^* < 1 + \frac{\omega}{\log n}.$$

Statement (a) of Theorem 1.4 will follow directly from Lemma 4.1, which is proved in Section 7. Indeed, we will prove a slightly stronger result (Lemma 7.1).

For the subcritical case (i.e. statement (b)) of Theorem 1.4, we first determine the threshold for topological connectedness of $\mathcal{G}(n, \mathbf{p})$, i.e. for when $H^0(\mathcal{G}(n, \mathbf{p}); R) = R$.

Lemma 4.2. There exist positive constants $c^- = c^-(d)$ and $c^+ = c^+(d)$ such that

(a) whp $\mathcal{G}(n, \mathbf{p})$ is not topologically connected if $p_k \leqslant \frac{c^- \log n}{n^k}$ for all $k \in [d]$;

(b) whp $\mathcal{G}(n, \mathbf{p})$ is topologically connected if $p_k \geqslant \frac{c^+ \log n}{n^k}$ for some $k \in [d]$.

In terms of hypergraphs, Lemma 4.2 provides the order of p_k when the non-uniform random hypergraph $G(n, \mathbf{p})$ becomes vertex-connected. Corresponding and far stronger results are well-known for uniform random hypergraphs, but the non-uniform case does not seem to have been considered in the literature. The proof of Lemma 4.2 (a) combines known results about vertex-connectedness of random uniform hypergraphs with a straightforward adaptation of the classical proof of connectedness of random graphs. We include the proof in Appendix A.1 for completeness.

Remark 4.3. In fact, with a slightly more careful extension of the argument, one could strengthen Lemma 4.2 to give the exact threshold. More precisely, if $p_k = \frac{c_k \log n}{n^k}$ for $k \in [d]$, where each c_k may now be a function in n, then $\mathcal{G}(n, \mathbf{p})$ contains isolated vertices whp provided $\sum_{k=1}^{d} \frac{c_k}{k!} = 1 - \omega(\frac{1}{\log n})$, whereas $\mathcal{G}(n, \mathbf{p})$ is topologically connected whp if $\sum_{k=1}^{d} \frac{c_k}{k!} = 1 + \omega(\frac{1}{\log n})$. We omit the details.

In particular, Lemma 4.2 will imply that for every sufficiently small $\varepsilon > 0$, whp the process $(\mathcal{G}_\tau)$ is not topologically connected, and thus also not j-cohom-connected, for every $\tau \in [0, \frac{\varepsilon}{n^j}]$.

In order to cover the whole interval $[0, \tau_j^*)$, the following result, whose proof is in Section 8, will be key.

Lemma 4.4. Let $\varepsilon > 0$ be a constant and define

$$I_j(\varepsilon) := \left[\frac{\varepsilon}{n}, \tau_j^*\right).$$

Then, whp $H^j(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_j(\varepsilon)$.

To prove Lemma 4.4 we will show that whp, for any $\tau \in I_j(\varepsilon)$ there is an index k with $j \leqslant k \leqslant d$ such that a copy of the minimal obstruction $\hat{M}_{j,k}$ exists in $\mathcal{G}_\tau$. In fact, we will show that whp just three minimal obstructions exist in ranges which together cover $I_j(\varepsilon)$ (Lemmas 8.1, 8.3, and 8.4).

For the remaining range of the subcritical interval $[0, \tau_j^*)$, we want to consider the cohomology groups $H^i(\mathcal{G}_\tau; R)$ with $i \in [j-1]$ and determine in which subintervals they do not vanish, i.e. we want to find an analogue of Lemma 4.4 for $H^i(\mathcal{G}_\tau; R)$. To do this, we need to show that starting from a j-critical direction $\bar{\mathbf{p}}$ we can use an appropriate rescaling to obtain an i-critical direction.

Lemma 4.5. If $\bar{\mathbf{p}}$ is a j-critical direction, then for each $i \in [j-1]$ there exist a constant $\eta = \eta_i > 0$ and a function $\epsilon = \epsilon_i(n) = o(1)$ such that the vector $\frac{\eta+\epsilon}{n^{j-i}}\bar{\mathbf{p}}$ is an i-critical direction.

Although Lemma 4.5 is intuitively obvious, the proof requires carefully balancing the values of η and ϵ so that (C1) and (C2) in Definition 3.3 will be satisfied with j replaced by i, which involves rather tedious computations. We therefore delay the proof until Appendix A.2.

Using Lemma 4.5, for general $i \in [j]$ we can consider the hitting time τ_i^* for the disappearance of the last minimal obstruction $\hat{M}_{i,k}$. More precisely, for the j-critical direction $\bar{\mathbf{p}}$ as in Theorem 1.4, consider the process $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_{\tau\in[0,\tau_{\max}]}$ and for each $i \in [j]$ let

$$\tau_i^* \ := \ \sup\{\tau \in \mathbb{R}_{\geqslant 0} \mid \mathcal{G}_\tau \text{ contains a copy of } \hat{M}_{i,k} \text{ for some } k \text{ with } i \leqslant k \leqslant d\}. \tag{10}$$

Observe that for $i = j$, this matches the definition of τ_j^* in Theorem 1.4. We derive the following result from Lemmas 4.1, 4.4, and 4.5.

Corollary 4.6. Let $\varepsilon > 0$ be a constant and $i \in [j]$. Define

$$I_i(\varepsilon) \ := \ \left[\frac{\varepsilon}{n^{j-i+1}}, \tau_i^*\right).$$

Then, whp

(a) $H^i(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_i(\varepsilon)$;

(b) there exists a positive constant $\eta = \eta_i$ such that $\tau_i^* = \frac{\eta+o(1)}{n^{j-i}}$.

Indeed, Lemma 4.1 implies that $\eta = \eta_j = 1$ for $i = j$ in statement (b) of Corollary 4.6.

Proof. For any $i \in [j-1]$, by Lemma 4.5 we can appropriately scale $\bar{\mathbf{p}}$ to obtain an i-critical direction $\frac{\eta+\epsilon}{n^{j-i}}\bar{\mathbf{p}}$. Thus, by Lemmas 4.4 and 4.1 applied with j replaced by i, we obtain (a) and (b), respectively.

For $i = j$, Lemmas 4.4 and 4.1 apply directly to $\bar{\mathbf{p}}$. ☐

4.2 Supercritical case

In Theorem 1.4 (c) we consider $\tau \in [\tau_j^*, \tau_{\max}]$. By definition of τ_j^*, we know that whp in this range there is no copy of the minimal obstruction $\hat{M}_{j,k}$ to j-cohom-connectedness for any $j \leqslant k \leqslant d$, but we also have to exclude other types of obstructions. Indeed, in Section 9 we prove the following.

Lemma 4.7. Whp for every $\tau \in [\tau_j^*, \tau_{\max}]$, we have $H^j(\mathcal{G}_\tau; R) = 0$.

Observe that since the choice of j was arbitrary, Lemma 4.7 also holds when j is replaced by any $i \in [j-1]$.

4.3 Proof of Theorem 1.4

We now apply the auxiliary results of Sections 4.1 and 4.2 to prove the Hitting Time Theorem (Theorem 1.4).

(a) Fix a function ω of n which tends to infinity as $n \to \infty$. To show that whp $\tau_j^* = 1 + o(\omega/\log n)$, it suffices to apply Lemma 4.1 with any function ω' which tends to infinity but satisfies $\omega' = o(\omega)$, e.g. picking $\omega' = \sqrt{\omega}$ will suffice.

(b) For $\varepsilon > 0$, define

$$I_0(\varepsilon) \;:=\; \left[0, \frac{\varepsilon}{n^j}\right].$$

By definition of j-admissibility (Definition 3.1), for every $k \in [d]$ we have $\bar{p}_k = O(\log n/n^{k-j})$. Thus, by Lemma 4.2 we can choose ε small enough such that whp

$$\text{for every } \tau \in I_0(\varepsilon), \qquad H^0(\mathcal{G}_\tau; R) \neq R. \tag{11}$$

Now consider the intervals $I_i(\varepsilon)$. By Corollary 4.6 (b), we can choose ε small enough (namely $\varepsilon < \eta_i$ for every $i \in [j-1]$) such that whp for each such i

$$I_i(\varepsilon) \cap I_{i+1}(\varepsilon) = \left[\frac{\varepsilon}{n^{j-i}}, \tau_i^*\right) \neq \emptyset,$$

and thus

$$\bigcup_{i=0}^{j} I_i(\varepsilon) = [0, \tau_j^*). \tag{12}$$

By Corollary 4.6 (a), for any $\varepsilon > 0$ whp $H^i(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_i(\varepsilon)$. Thus, by choosing ε such that conditions (11) and (12) hold simultaneously, whp the process $(\mathcal{G}_\tau)$ is not j-cohom-connected for all $\tau \in [0, \tau_j^*)$, as required.

(c) Recalling that $\bar{p}_k = O(\log n/n^{k-j})$ for any $k \in [d]$ by Definition 3.1, Lemma 4.2 implies that we can find a positive constant ϑ such that whp $H^0(\mathcal{G}_\tau; R) = R$ for every $\tau \in [\frac{\vartheta}{n^j}, \tau_{\max}]$, which whp contains the interval $[\tau_j^*, \tau_{\max}]$ by Lemma 4.1.

Furthermore, by Lemma 4.7 applied for every $i \in [j]$, whp $H^i(\mathcal{G}_\tau; R) = 0$ for every $\tau \in [\tau_i^*, \tau_{\max}]$, which contains $[\tau_j^*, \tau_{\max}]$ by Corollary 4.6 (b). Thus, whp the process $(\mathcal{G}_\tau)$ is j-cohom-connected for every $\tau \in [\tau_j^*, \tau_{\max}]$, as required. ☐

5 Minimal obstructions

In this section we define copies of $\hat{M}_{j,k}$ (Definitions 5.3 and 5.4) and we explain why these objects can be interpreted as *minimal obstrucions* to j-cohom-connectedness.

For the rest of the paper, let $j \in [d-1]$ be fixed. We first introduce the following necessary concepts.

Definition 5.1. Given $k \geqslant j$ and a $(k+1)$-set K, a $(j+1)$-set $J \subseteq K$ is K*-localised* if every simplex σ with $J \subseteq \sigma$ is such that $\sigma \subseteq K$.

Note that we do not demand that J is a j-simplex—if it is not, then it is trivially K-localised for any $K \supseteq J$ since there is no simplex $\sigma \supseteq J$.

Definition 5.2. Let k be an integer with $j \leqslant k \leqslant d$. Given a k-simplex K in a d-dimensional simplicial complex $\mathcal{G}$, we say that a collection $\mathcal{F} = \{P_0, \ldots, P_{k-j}\}$ of j-simplices forms a j*-flower in* K (see Figure 1) if

(F1) $K = \bigcup_{i=0}^{k-j} P_i$;

(F2) there exists C with $|C| = j$ that is contained in P_i for every $i \in [k-j]_0$.

We call the j-simplices P_i the *petals* and the set C the *centre* of the j-flower $\mathcal{F}$. When j is clear from the context, we often refer to the j-flower $\mathcal{F}$ simply as a *flower*.

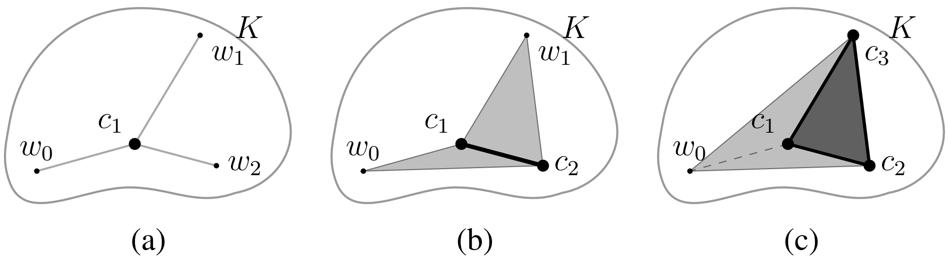

Figure 1: Examples of j-flowers in a k-simplex K, for $k = 3$ and $j = 1, 2, 3$.
(a) The 1-flower in K with centre $C = \{c_1\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1, 2$ (grey).
(b) The 2-flower in K with centre $C = \{c_1, c_2\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1$ (grey).
(c) The 3-flower in K with centre $C = \{c_1, c_2, c_3\}$ (bold black) and the unique petal $P_0 = C \cup \{w_0\} = K$ (grey).

Observe that for each k-simplex K and each $(j-1)$-simplex $C \subset K$, there is a unique j-flower in K with centre C, namely

$$\mathcal{F}(K, C) \ := \ \{C \cup \{w\} \mid w \in K \setminus C\}. \tag{13}$$

The flower $\mathcal{F}(K, C)$ could equivalently be described as the join of the $(j-1)$-simplex C with the $k+j-1$ vertices of $K \setminus C$. Note that if $k = j$, then any choice of a centre $C \subset K$ produces the same flower $\mathcal{F}(K, C) = \{K\}$.

Definition 5.3. Let k be an integer with $j+1 \leqslant k \leqslant d$. We say that a 4-tuple (K, C, w, a) forms a *copy of* $\hat{M}_{j,k}$ (see Figure 2) in a simplicial complex $\mathcal{G}$ if

(M1) K is a k-simplex in $\mathcal{G}$;

(M2) C is a $(j-1)$-simplex in K such that every petal of the flower $\mathcal{F} = \mathcal{F}(K, C)$ is K-localised;

(M3) $w \in K \setminus C$ and $a \in [n] \setminus K$ are such that $C \cup \{w\} \cup \{a\}$ is a j-shell in $\mathcal{G}$.

We call the j-simplex $C \cup \{w\}$ the *base* and a the *apex vertex* of the j-shell $C \cup \{w\} \cup \{a\}$. Every other j-simplex in the j-shell $C \cup \{w\} \cup \{a\}$ is called a *side* of the j-shell.

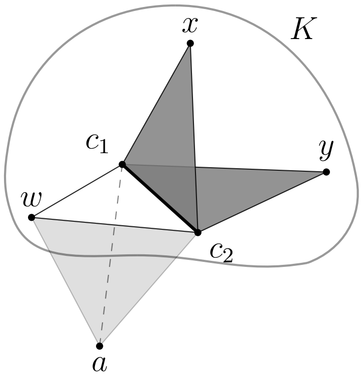

Figure 2: A copy of $\hat{M}_{j,k}$, for $k = 4$ and $j = 2$.
(a) The k-simplex K contains the flower $\mathcal{F}(K, C)$ with centre $C = \{c_1, c_2\}$, whose petals $C \cup \{w\}$, $C \cup \{x\}$, and $C \cup \{y\}$ are not present in any simplex which is not contained in K.
(b) The $(j+2)$-set $C \cup \{w\} \cup \{a\}$ is a j-shell with apex vertex $a \notin K$ and whose base is the petal $C \cup \{w\}$.

We aim to give an analogous definition for the case $k = j$ and it will be convenient to use unified terminology. However, as observed before, if K is a $(j+1)$-set, then a j-flower $\mathcal{F}(K, C)$ is always equal to K itself, independently of the choice of the centre C in K. In particular, in this case condition (M2) simply says that K is an *isolated j-simplex* in $\mathcal{G}$, i.e. a j-simplex that is not contained in any other simplex of $\mathcal{G}$. This means that given K, the sets that would be required to be simplices or not in $\mathcal{G}$ do not change for different choices of the centre C, and therefore we do not want to consider two copies of $\hat{M}_{j,j}$ to be distinct if they share the same j-simplex but have different centres. For this reason, to define $\hat{M}_{j,j}$ we will use the following 'canonical' choice for the centre.

Definition 5.4. We say that a 4-tuple (K, C, w, a) forms a *copy of* $\hat{M}_{j,j}$ in $\mathcal{G}$ if

- K is an isolated j-simplex in $\mathcal{G}$;
- $a \in [n] \setminus K$ is such that $K \cup \{a\}$ is a j-shell in $\mathcal{G}$;

- C consists of the first j vertices of K in the increasing order on $[n]$, and w is the last vertex of K in this order.

The notions of *base*, *apex vertex*, and *side* are analogous to Definition 5.3.

It is easy to see that a copy of $\hat{M}_{j,j}$ in Definition 5.4 satisfies conditions (M1)–(M3) of Definition 5.3.

Let us now define a 'reduced' version of $\hat{M}_{j,k}$, denoted by $M_{j,k}$, by omitting the condition (M3) on the j-shell $C \cup \{w\} \cup \{a\}$ in Definitions 5.3 and 5.4.

Definition 5.5. Let k be an integer with $j+1 \leqslant k \leqslant d$. A pair (K, C) is called a *copy of* $M_{j,k}$ if it satisfies the following conditions.

(M1) K is a k-simplex in $\mathcal{G}$;

(M2) C is a $(j-1)$-simplex in K such that every petal of the flower $\mathcal{F} = \mathcal{F}(K, C)$ is K-localised.

Similarly to $\hat{M}_{j,k}$, we also define an analogous concept for the case $k = j$.

Definition 5.6. A pair (K, C) is called a *copy of* $M_{j,j}$ if

- K is an isolated j-simplex;
- C consists of the first j vertices in K in the increasing order on $[n]$.

We will see later (Corollary 6.4) that the shell required for (M3) in Definition 5.3 (and the analogous condition in Definition 5.4) is very likely to exist if τ is 'large enough', which will be the case well before the critical range for the disappearance of $M_{j,k}$. Thus the presence of $\hat{M}_{j,k}$ and of $M_{j,k}$ are essentially equivalent events for sufficiently large τ, allowing us to switch our focus to the simpler $M_{j,k}$.

We also define the following random variables, which we will later use to count the number of minimal obstructions in the complex (e.g. Lemma 5.14).

Definition 5.7. For $j \leqslant k \leqslant d$, let

$$X_{j,k} = X_{j,k}(\tau) \ := \ |\{ \text{ copies of } M_{j,k} \text{ in } \mathcal{G}_\tau \ \}|$$

and

$$\hat{X}_{j,k} = \hat{X}_{j,k}(\tau) \ := \ |\{ \text{ copies of } \hat{M}_{j,k} \text{ in } \mathcal{G}_\tau \ \}|.$$

We now justify our interpretation of $\hat{M}_{j,k}$ as a minimal obstruction to j-cohom-connectedness, first observing that it is certainly an obstruction (Corollary 5.10). To show this, we define a j-cocycle which is not a j-coboundary—the function we choose will depend only on the underlying copy (K, C) of $M_{j,k}$.

Definition 5.8. Let $M = (K, C)$ be a copy of $M_{j,k}$ in a simplicial complex.

(a) We denote by $\mathrm{Ord}(K, C)$ the (unique) ordering $v_0, \ldots, v_{n-1}$ of all vertices in $[n]$ such that $C = \{v_0, \ldots, v_{j-1}\}$, $K = \{v_0, \ldots, v_k\}$, and furthermore the vertices within C, within $K \setminus C$, and within $[n] \setminus K$ are ordered according to the increasing order in $[n]$.

(b) Given $\mathrm{Ord}(K, C)$, for any $r \in R$ we define the following j-cochain $f_{M,r}$. For every ordered j-simplex $\sigma = [v_{i_0}, \ldots, v_{i_j}]$ with $i_0 < \cdots < i_j$, we set

$$f_{M,r}(\sigma) := \begin{cases} r & \text{if } \sigma \in \mathcal{F}(K, C), \text{ i.e. } i_s = s \text{ for } 0 \leqslant s \leqslant j-1 \text{ and } j \leqslant i_j \leqslant k, \\ 0_R & \text{otherwise,} \end{cases}$$

and we extend this function to all j-simplices with different orderings so as to obtain a j-cochain.

Proposition 5.9. Let $M = (K, C)$ be a copy of $M_{j,k}$ in a simplicial complex $\mathcal{G}$ and let f be a j-cochain whose support is contained within the flower $\mathcal{F}(K, C)$. Then the following hold.

(a) The j-cochain f is a j-cocycle if and only if $f = f_{M,r}$ for some $r \in R$.

(b) Suppose that there exist $w \in K \setminus C$ and $a \in [n] \setminus K$ such that (K, C, w, a) is a copy of $\hat{M}_{j,k}$ in $\mathcal{G}$. Then f is a j-cocycle but not a j-coboundary if and only if $f = f_{M,r}$ for some $r \in R \setminus \{0_R\}$.

Proof. (a) First observe that if $k = j$ then $\mathcal{F}(K, C) = \{K\}$ and K is an isolated j-simplex by Definition 5.6. Hence a j-cochain with support contained in K is necessarily of the form $f_{M,r}$, where $r \in R$ is the value it assigns to (the appropriate ordering of) K, and is a j-cocycle since no $(j+1)$-simplex contains K.

Now consider k with $k \geqslant j+1$. Let $K = \{v_0, \ldots, v_k\}$ and $C = \{v_0, \ldots, v_{j-1}\}$ according to $\mathrm{Ord}(K, C)$, and let ρ be a $(j+1)$-simplex. By (M2), $(\delta^j f)(\rho) = 0_R$ follows immediately unless $C \subset \rho \subseteq K$. We may therefore assume that $\rho = \{v_0, \ldots, v_{j-1}, v_{i_1}, v_{i_2}\}$, with $j \leqslant i_1 \leqslant i_2 \leqslant k$. Then we have

$$(\delta^j f)(\rho) = (-1)^j f([v_0, \ldots, v_{j-1}, v_{i_2}]) + (-1)^{j+1} f([v_0, \ldots, v_{j-1}, v_{i_1}]).$$

This implies that f is a j-cocycle if and only if on each petal $[v_0, \ldots, v_{j-1}, v_i]$ with $j \leqslant i \leqslant k$, it takes the same value $r \in R$, i.e. $f = f_{M,r}$.
(b) If f is a j-cocycle and not a j-coboundary, by (a) we already know that there exists $r \in R$ such that $f = f_{M,r}$. If $r = 0_R$, then $f \equiv 0_R$ and thus f is a j-coboundary, a contradiction.

Conversely, if $f = f_{M,r}$ for some $r \in R \setminus \{0_R\}$, then f is a j-cocycle by (a). Furthermore, property (M3) in Definition 5.3 implies that $C \cup \{w\} \cup \{a\}$ is a j-shell that meets the support of f in precisely one petal, namely in $C \cup \{w\}$. Thus Lemma 2.2 yields that f is not a j-coboundary. □

Corollary 5.10. Suppose that in a simplicial complex $\mathcal{G}$ the 4-tuple $M = (K, C, w, a)$ forms a copy of $\hat{M}_{j,k}$. Then $H^j(\mathcal{G}; R) \neq 0$.

Proof. For any $r \in R \setminus \{0_R\}$, the function $f_{M,r}$ defined in Definition 5.8 is a j-cocyle but not a j-coboundary by Proposition 5.9 (b), i.e. the cohomology class of $f_{M,r}$ is a non-zero element of $H^j(\mathcal{G}; R)$. $\square$

The next lemma shows that copies of $\hat{M}_{j,k}$ are also (in a natural sense) *minimal* obstructions. Given a k-simplex K and a collection $\mathcal{S}$ of j-simplices, define $\mathcal{S}_K$ to be the set of j-simplices of $\mathcal{S}$ contained in K.

Lemma 5.11. Let $\mathcal{S}$ be the support of a j-cocycle f in a d-complex $\mathcal{G}$. Then for each k with $j+1 \leqslant k \leqslant d$ and each k-simplex K,

(a) either $\mathcal{S}_K = \emptyset$ or both $|\mathcal{S}_K| \geqslant k-j+1$ and $\bigcup_{\sigma \in \mathcal{S}_K} \sigma = K$;

(b) if $|\mathcal{S}_K| = k-j+1$, then $\mathcal{S}_K$ forms a j-flower in K.

Note in particular that Lemma 5.11 implies that the support $\mathcal{S}$ of any non-trivial j-cocycle satisfies at least one of the following three properties:

- $\mathcal{S}_K$ is empty for every k-simplex K;
- $|\mathcal{S}| \geqslant k-j+2$;
- $|\mathcal{S}| = k-j+1$ and $\mathcal{S}$ forms a j-flower in some k-simplex K.

Since in the third case a j-shell containing a petal, with its single additional apex vertex, is the simplest (though by no means the only) way of ensuring that the corresponding j-cocycle is not a j-coboundary, this justifies why a copy of $\hat{M}_{j,k}$ may be considered a minimal obstruction to the vanishing of the j-th cohomology group.

Proof of Lemma 5.11. (a) Suppose $\mathcal{S}_K \neq \emptyset$ and let $\sigma_0 \in \mathcal{S}_K$. Denote the vertices of σ_0 and of $K \setminus \sigma_0$ by $u_0, \dots, u_j$ and by $v_1, \dots, v_{k-j}$, respectively. For each $i \in [k-j]$, the ordered $(j+1)$-simplex $[u_0, \dots, u_j, v_i]$ has to be mapped to 0_R by $\delta^j f$ and thus the underlying unordered simplex $\sigma_0 \cup \{v_i\}$ contains some j-simplex $\sigma_i \in \mathcal{S}_K \setminus \{\sigma_0\}$, which therefore contains v_i. The simplices $\sigma_0, \dots, \sigma_{k-j}$ are distinct, because each v_i lies in σ_i but in no other $\sigma_{i'}$. Therefore $|\mathcal{S}_K| \geqslant k-j+1$ and

$$K \supseteq \bigcup_{i=0}^{k-j} \sigma_i \supseteq \sigma_0 \cup \{v_1, \dots, v_{k-j}\} = K.$$

(b) Suppose now that $\mathcal{S}_K = \{\sigma_0, \dots, \sigma_{k-j}\}$ with $\sigma_0, \dots, \sigma_{k-j}$ defined as above. For $2 \leqslant i \leqslant k-j$ (if such indices exist), the $(j+1)$-simplex $\sigma := \sigma_1 \cup \{v_i\}$ contains σ_1, but no $\sigma_{i'}$ with $i' \notin \{1, i\}$. By the choice of f as a j-cocycle, $\delta^j f$ maps each ordering of σ to 0_R and thus σ has to contain at least two elements of $\mathcal{S}_K$, implying that $\sigma_i \subset \sigma$. This means that

$$\sigma_1 \cap \sigma_i = \sigma \setminus \{v_1, v_i\} = \sigma_0 \cap \sigma_1.$$

As this holds for all i, $\mathcal{S}_K$ forms a flower in K with centre $C = \sigma_0 \cap \sigma_1$. $\square$

The proofs of our main results (Theorems 1.4 and 1.5) are significantly more difficult than might naively be expected due to the fact that both the presence of copies of $\hat{M}_{j,k}$ and j-cohom-connectedness in $\mathcal{G}_\tau$ are *not* monotone properties. Indeed, we observe that in Definition 5.3 while (M1) and (M3) are monotone increasing properties, property (M2) is monotone decreasing. Thus, in principle the random process $(\mathcal{G}_\tau)$ could oscillate between being j-cohom-connected or not, as the following example shows.

Example 5.12. We consider the case $j = 1$. Let $\mathcal{G}$ be the simplicial complex on vertex set $\{1, 2, 3, 4\}$ generated by the hypergraph with edges $\{1, 2\}$, $\{2, 3\}$, and $\{3, 4\}$, as in Figure 3. It is easy to see that $\mathcal{G}$ is 1-cohom-connected and thus contains no copies of $\hat{M}_{1,k}$ for any $k \geqslant 1$. If we add the 2-simplex $\{1, 3, 4\}$ (and its downward-closure) to $\mathcal{G}$, the 4-tuple $(\{1, 3, 4\}, \{3\}, 1, 2)$ creates a copy of $\hat{M}_{1,2}$ and thus we obtain a complex $\mathcal{G}'$ which is not 1-cohom-connected. Adding the 2-simplex $\{1, 2, 3\}$ to $\mathcal{G}'$ yields the complex $\mathcal{G}''$ which is again 1-cohom-connected and thus contains no copies of $\hat{M}_{1,k}$ for any $k \geqslant 1$.

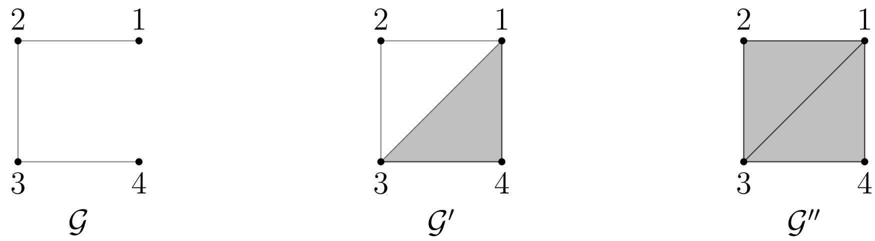

Figure 3: Adding simplices might create new copies of $\hat{M}_{j,k}$ or destroy existing ones.

In order to determine the critical range for the disappearance of copies of $M_{j,k}$, in Lemma 5.14 we will calculate the expectation of $X_{j,k}$, i.e. the number of copies of $M_{j,k}$ (Definition 5.7). We first estimate the probability of (M2). Define

$$\bar{q} = \bar{q}(\bar{\mathbf{p}}, n, j) \ := \ \prod_{k=j+1}^{d} (1 - \bar{p}_k)^{\binom{n-j-1}{k-j}}. \tag{14}$$

Observe that $\bar{q}$ is the probability that a given set of $j + 1$ vertices (which may or may not form a j-simplex) is not in any k-simplex of $\mathcal{G}(n, \bar{\mathbf{p}})$ (i.e. at time $\tau = 1$) for any $k \geqslant j + 1$. Moreover if $\bar{\mathbf{p}}$ is a j-admissible direction, by Definition 3.1 we have

$$\bar{q} = (1 + o(1)) \exp\left(-\sum_{k=j+1}^{d} \left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right), \tag{15}$$

because by (A1) at least one of $\bar{\alpha}_k, \bar{\gamma}_k$ is zero and thus $\frac{\bar{\alpha}_k}{n^{\bar{\gamma}_k}} = \bar{\alpha}_k$.

The next lemma implies that for any $\tau = O(1)$, the probability of (M2) in $\mathcal{G}_\tau$ is approximately $\bar{q}^{\,\tau(k-j+1)}$—we state the lemma in a slightly more general setting, since we will need to apply it in different situations (for example when calculating the second moment of $X_{j,k}$).

Proposition 5.13. Let $\tau = O(1)$ and let $\mathbf{p} = (p_1, \dots, p_d) = \tau\bar{\mathbf{p}}$ for a j-admissible direction $\bar{\mathbf{p}}$. Let $\mathcal{J}$ be a collection of $O(1)$ many $(j + 1)$-sets in $[n]$ and let $\mathcal{S}$ be a collection of $O(1)$ many sets of vertices of size between $j+2$ and $d+1$. Let A be the event that no $(j+1)$-set of $\mathcal{J}$ lies in any k-simplex K of $\mathcal{G}_\tau$ with $j+1 \leqslant k \leqslant d$ and $K \notin \mathcal{S}$. Then

$$\mathbb{P}(A) = (1+o(1))\bar{q}^{\,\tau|\mathcal{J}|}.$$

The proof of this proposition is a straightforward asymptotic computation and appears in Appendix A.3. We now apply Proposition 5.13 to calculate the expectation of $X_{j,k}$, for $\tau = O(1)$.

Suppose first that $k \geqslant j+1$. There are $\binom{n}{k+1}\binom{k+1}{j} = (1+o(1))\frac{n^{k+1}}{j!(k-j+1)!}$ ways to choose a pair (K, C) that might form a copy of $M_{j,k}$. The $(k+1)$-set K forms a k-simplex in $\mathcal{G}_\tau$ with probability p_k (recall that $p_k \leqslant 1$ by Remark 3.2). By Proposition 5.13 applied with $\mathcal{J} = \mathcal{F}(K, C)$ being the set of petals and with $\mathcal{S} = \{\sigma \subseteq K : |\sigma| \geqslant j+2\}$, the probability that (M2) holds is $(1+o(1))\bar{q}^{\,\tau(k-j+1)}$. Therefore,

$$\mathbb{E}(X_{j,k}) = (1+o(1))\frac{n^{k+1}p_k}{j!(k-j+1)!}\bar{q}^{\,\tau(k-j+1)}. \tag{16}$$

The case $k = j$ is very similar, but for a pair (K, C) to form a copy of $M_{j,j}$ we only require that the set K is an isolated $(j+1)$-simplex, since the centre C is uniquely defined (see Definition 5.6). On the other hand, we need to be careful if $p_j > 1$, since then p_j must be replaced by 1 in any probability calculations. We have

$$\mathbb{E}(X_{j,j}) = (1+o(1))\frac{n^{j+1}\min\{p_j, 1\}}{(j+1)!}\bar{q}^{\,\tau}. \tag{17}$$

In the next lemma we use (16) and (17) to obtain an explicit expression for $\log\left(\mathbb{E}(X_{j,k})\right)$, which we will need in Section 6. Recall that given a vector $\mathbf{p}$, the parameters λ_k, μ_k, and ν_k are as defined in Definition 3.4.

Lemma 5.14. Let $\mathbf{p} = (p_1, \ldots, p_d) = \tau\bar{\mathbf{p}}$ for a j-admissible direction $\bar{\mathbf{p}}$, where $\tau = O(1)$, but $\tau = \Omega(n^{-c})$ for some positive constant c. Then the number $X_{j,k}$ of copies of $M_{j,k}$ in $\mathcal{G}_\tau$ satisfies

$$\log\left(\mathbb{E}(X_{j,k})\right) = \lambda_k \log n + \mu_k + \nu_k + o(1)$$

for all $j \leqslant k \leqslant d$ with $\bar{p}_k \neq 0$.

Proof. Observe that in Definition 3.4, α_k and γ_k exist, because $\tau = O(1)$ and $\tau = \Omega(n^{-c})$, respectively. Thus, λ_k, μ_k, and ν_k can be defined analogously to (4). Moreover, since $\bar{\alpha}_i/n^{\bar{\gamma}_i} = \bar{\alpha}_i$ for each $j+1 \leqslant i \leqslant d$ by (A1) in Definition 3.1, we have

$$\begin{aligned}
\bar{q}^{\,\tau} &\overset{(15)}{=} (1+o(1))\exp\left(-\tau\sum\nolimits_{i=j+1}^{d}\left(\bar{\alpha}_i\log n + \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}}\right)\right)\\
&\overset{(3)}{=} (1+o(1))\exp\left(-\sum\nolimits_{i=j+1}^{d}\tau\bar{p}_i\frac{n^{i-j}}{(i-j)!}\right)\\
&= (1+o(1))\exp\left(-\sum\nolimits_{i=j+1}^{d}p_i\frac{n^{i-j}}{(i-j)!}\right)\\
&= (1+o(1))\exp\left(-\sum\nolimits_{i=j+1}^{d}\left(\alpha_i\log n + \frac{\beta_i}{n^{\gamma_i}}\right)\right).
\end{aligned} \tag{18}$$

Suppose first that $k \geqslant j+1$. Recall that in this case $p_k \leqslant 1$ by Remark 3.2. Furthermore, we observe that

$$\frac{n^{k+1}p_k}{j!(k-j+1)!} = \frac{n^{j+1-\gamma_k}(\alpha_k \log n + \beta_k)}{j!(k-j+1)}, \tag{19}$$

by Definition 3.4. We thus have

$$\begin{aligned}
\log\left(\mathbb{E}(X_{j,k})\right) &\overset{(16)}{=} \log\left((1+o(1))\frac{n^{k+1}p_k}{j!(k-j+1)!}\bar{q}^{\,\tau(k-j+1)}\right) \\
&\overset{(19)}{=} (j+1-\gamma_k)\log n + \log(\alpha_k \log n + \beta_k) - \log(j!) - \log(k-j+1) \\
&\qquad + (k-j+1)\log\left(\bar{q}^{\,\tau}\right) + o(1) \\
&\overset{(18)}{=} \left(j+1-\gamma_k-(k-j+1)\sum\nolimits_{i=j+1}^{d}\alpha_i\right)\log n \\
&\qquad - (k-j+1)\sum\nolimits_{i=j+1}^{d}\frac{\beta_i}{n^{\gamma_i}} + \log(\alpha_k\log n + \beta_k) \\
&\qquad - \log(j!) - \log(k-j+1) + o(1).
\end{aligned}$$

Now (4), together with the fact that $p_k \leqslant 1$, implies that

$$\log\left(\mathbb{E}(X_{j,k})\right) = \lambda_k \log n + \mu_k + \nu_k + o(1),$$

as required.

If $k = j$, substituting (18) into (17) we have

$$\mathbb{E}(X_{j,j}) = (1+o(1))\frac{n^{j+1}\min\{p_j, 1\}}{(j+1)!}\exp\left(-\sum\nolimits_{i=j+1}^{d}\left(\alpha_i \log n + \frac{\beta_i}{n^{\gamma_i}}\right)\right).$$

Therefore we obtain

$$\begin{aligned}
\log(\mathbb{E}(X_{j,j})) = \left(j+1-\sum\nolimits_{i=j+1}^{d}\alpha_i\right)\log n + \sum\nolimits_{i=j+1}^{d}\frac{\beta_i}{n^{\gamma_i}} - \log((j+1)!) \\
+ \log\left(\min\{p_j, 1\}\right) + o(1).
\end{aligned} \tag{20}$$

We first consider the case when $p_j \leqslant 1$, in which case by Definition 3.4 we have $\alpha_j = 0$, and so

$$\log\left(\min\{p_j,1\}\right) = \log p_j = \log\left(\frac{\beta_j}{n^{\gamma_j}}\right) = -\gamma_j \log n + \log \beta_j.$$

Substituting this into (20), we obtain

$$\begin{aligned}
\log(\mathbb{E}(X_{j,j})) &= \left(j+1-\gamma_j-\sum\nolimits_{i=j+1}^{d}\alpha_i\right)\log n + \sum\nolimits_{i=j+1}^{d}\frac{\beta_i}{n^{\gamma_i}} + \log\beta_j \\
&\qquad\qquad - \log((j+1)!) + o(1) \\
&= \lambda_j \log n + \mu_j + \nu_j + o(1)
\end{aligned}$$

as required.

On the other hand, if $p_j > 1$, then we must have $\gamma_j = 0$. Furthermore,

$$\log\left(\min\{p_j, 1\}\right) = \log 1 = 0,$$

and therefore (20) gives

$$\begin{aligned}\log(\mathbb{E}(X_{j,j})) &= \Big(j+1-\sum\nolimits_{i=j+1}^{d}\alpha_i\Big)\log n + \sum\nolimits_{i=j+1}^{d}\frac{\beta_i}{n^{\gamma_i}} - \log((j+1)!) + o(1)\\ &= \lambda_j \log n + \mu_j + \nu_j + o(1)\end{aligned}$$

since we are in the case when $k = j$ and $p_j > 1$. □

Recall that in our main Theorems 1.4 and 1.5 we consider a j-critical direction $\bar{\mathbf{p}}$ which in particular is a j-admissible direction (cf. Definitions 3.1 and 3.3). Thus Proposition 5.13 and Lemma 5.14 are applicable. By Lemma 5.14, heuristically the critical range for the disappearance of copies of $M_{j,k}$ is when $\lambda_k \log n + \mu_k = O(1)$, that is, $\lambda_k = 0$ and $\mu_k = O(1)$. This justifies the conditions (C1) and (C2) in Definition 3.3, which together with Lemma 5.14 yield that for $\mathbf{p} = \bar{\mathbf{p}}$,

$$\mathbb{E}(X_{j,\bar{k}}) = 1 + o(1) \quad \text{and} \quad \mathbb{E}(X_{j,k}) \leqslant 1 + o(1) \text{ for all other indices } k. \tag{21}$$

In other words, heuristically $\bar{\mathbf{p}}$ is in a critical range for the disappearance of copies of $M_{j,\bar{k}}$, while for all other k, $\bar{\mathbf{p}}$ is either in or already beyond the critical range for the disappearance of copies of $M_{j,k}$. We will see later (Corollary 6.4) that in this range, whp all copies (K, C) of $M_{j,k}$ can be extended to copies (K, C, w, a) of $\hat{M}_{j,k}$. Thus $\bar{\mathbf{p}}$ is also in the critical range for the disappearance of minimal obstructions.

Recall that in Theorem 1.4, we consider

$$\tau_j^* \ := \ \sup\{\tau \in \mathbb{R}_{\geqslant 0} \mid \mathcal{G}_\tau \text{ contains a copy of } \hat{M}_{j,k} \text{ for some } k \text{ with } j \leqslant k \leqslant d\}.$$

In other words, τ_j^* is the scaled birth time of a simplex whose appearance causes the last minimal obstruction to disappear. We denote the dimension of this obstruction by ℓ, i.e. let ℓ be the index such that this obstruction is a copy of $\hat{M}_{j,\ell}$. For future reference, we collect the definitions of the special indices $\bar{k}, k_0, \ell$ which we have fixed so far.

Definition 5.15. Let $\bar{k}, k_0, \ell$ be integers such that

(a) $j \leqslant \bar{k} \leqslant d$ and $\bar{\lambda}_{\bar{k}} \log n + \bar{\mu}_{\bar{k}} + \bar{\nu}_{\bar{k}} = 0$ (see (C2));

(b) $j + 1 \leqslant k_0 \leqslant d$ and $\bar{\alpha}_{k_0} \neq 0$ (see (A4));

(c) at time τ_j^*, a (last) copy of $\hat{M}_{j,\ell}$ vanishes.

6 Finding minimal obstructions

To prove Lemma 4.4, the strategy is to show that whp a copy of $\hat{M}_{j,k}$ (for some $j \leqslant k \leqslant d$) exists in $\mathcal{G}_\tau$ for every $\tau \in I_j(\varepsilon) = [\varepsilon/n, \tau_j^*)$. Therefore in this section we study the behaviour of the minimal obstructions $\hat{M}_{j,k}$.

We start by showing that at the beginning of the interval $I_j(\varepsilon)$ we will already have a growing number of copies of $\hat{M}_{j,k_0}$, where k_0 is as in Definition 5.15 (b).

Lemma 6.1. Let $\varepsilon > 0$ be constant. If $\tau = \frac{\varepsilon}{n}$, then whp $\mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$ contains $\Theta((\log n)^{j+2})$ copies of $\hat{M}_{j,k_0}$ whose associated copies of M_{j,k_0} are all distinct.

The proof of Lemma 6.1 is a standard but slightly technical application of the second moment method and similar ideas were used in [13, Proof of Lemma 4.4]. We therefore postpone the proof to Appendix A.4.

In Lemma 6.3 we will show that in a range closer to criticality (i.e. for τ closer to 1) j-shells are very likely to exist. In order to formulate the statement, we first define the operation of 'adding a simplex'.

Definition 6.2. Given a complex $\mathcal{G}$ on vertex set V and a non-empty set $B \subseteq V$, we define $\mathcal{G} + B$ to be the complex obtained by adding the set B and its downward-closure to $\mathcal{G}$, i.e.

$$\mathcal{G} + B \ := \ \mathcal{G} \cup \{2^B \setminus \{\emptyset\}\}.$$

Observe that if B is already a simplex of $\mathcal{G}$, then $\mathcal{G} + B = \mathcal{G}$.

Lemma 6.3. For every $\varepsilon > 0$ there exists a constant $\zeta > 0$ such that if $\tau \geqslant \frac{\varepsilon}{\log n}$, then whp for every $(j+1)$-set B, the complex $\mathcal{G}_\tau + B$ contains at least ζn many j-shells that contain B.

Proof. Let $L_1, L_2, \ldots, L_{j+1}$ be the j-sets contained in B. We are interested in the vertices $a \in [n] \setminus B$ such that $B \cup \{a\}$ forms a j-shell in $\mathcal{G}_\tau + B$, i.e. such that $L_i \cup \{a\}$ is a j-simplex in $\mathcal{G}_\tau$ for every $i \in [j+1]$. We only consider a certain type of such j-shells, obtaining a lower bound on their total number.

Let $A, D \subset [n]$ be disjoint sets, both of size $\lceil n/3 \rceil$ and disjoint from B. Recall (Definition 5.15 (b)) that k_0 is an index with $j+1 \leqslant k_0 \leqslant d$ such that $\bar{\alpha}_{k_0} = 0$. We consider (potential) j-shells $B \cup \{a\}$ formed in the following way:

- the (apex) vertex a is in A;
- for each $i \in [j+1]$ there exists a set $R_i \subset D$, with $|R_i| = k_0 - j$, such that $L_i \cup \{a\}$ forms a j-simplex in $\mathcal{G}_\tau$ (and thus also in $\mathcal{G}_\tau + B$) as a subset of the k_0-simplex $R_i' := L_i \cup \{a\} \cup R_i$ with scaled birth time at most τ (i.e. with birth time at most $p_{k_0} = \tau \bar{p}_{k_0}$).

Since a different choice of the triple (L_i, a, R_i) never gives the same simplex R_i', we have independence in the following calculations.

For fixed L_i and a, the probability that no such R_i exists is

$$\begin{aligned}(1 - p_{k_0})^{\binom{|D|}{k_0 - j}} &\leqslant \exp\left(-\left(p_{k_0} \frac{n^{k_0 - j}}{4^{k_0 - j}(k_0 - j)!}\right)\right) \\ &\leqslant (1 + o(1)) \exp\left(-\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0 - j}}\right) \leqslant \exp\left(-\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0}}\right), \qquad (22)\end{aligned}$$

where we used that by (A3) we have $\bar{\beta}_{k_0} = o(\log n)$ since $\bar{\gamma}_{k_0} = 0$.

For any $a \in A$, let E_a be the event that $B \cup \{a\}$ is a j-shell in $\mathcal{G}_\tau + B$. Using (22), we obtain

$$\mathbb{P}(E_a) \geqslant \left(1 - \exp\left(\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0}}\right)\right)^{j+1} =: c > 0.$$

Since the events E_a are independent, the number of such j-shells dominates Bi$(\lceil n/3 \rceil, c)$. By Chernoff's bound, we have

$$\mathbb{P}\Big(\mathrm{Bi}\left(\lceil n/3 \rceil, c\right) < \zeta n\Big) \leqslant \exp\left(-\frac{\left(\frac{cn}{3} - \zeta n\right)^2}{\frac{2cn}{3}}\right) = \exp\left(-\frac{n}{6c}\left(c - 3\zeta\right)^2\right).$$

Choosing $0 < \zeta < c/3$ and by taking a union bound over all possible choices for the set B, we obtain that the probability that there are less than ζn many j-shells containing B is bounded above by

$$\binom{n}{j+1} \exp\left(-\frac{n}{6c}\left(c - 3\zeta\right)^2\right) = o(1),$$

as required. □

As an immediate corollary, we obtain that in this range, whp every copy of $M_{j,k}$ can be extended to a copy of $\hat{M}_{j,k}$, allowing us to consider just copies of $M_{j,k}$ as obstructions to j-cohom-connectedness.

Corollary 6.4. Let $\varepsilon > 0$ be constant and consider the process $(\mathcal{G}_\tau)$. Then whp any copy (K, C) of $M_{j,k}$ which exists in $\mathcal{G}_\tau$ for any $\tau \geqslant \frac{\varepsilon}{\log n}$ can be extended to a copy (K, C, w, a) of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$.

Proof. Let (K, C) be any pair of sets that *could* form a copy of $M_{j,k}$, i.e. K is a $(k+1)$-subset of $[n]$ and C is a j-subset of K (and, if $k = j$, then C consists of the first j vertices of K in the increasing order on $[n]$). Fix any vertex $w = w_{K,C} \in K \setminus C$. Lemma 6.3 implies that at time $\tau = \frac{\varepsilon}{\log n}$, whp for *all* such pairs (K, C), there are linearly many j-shells in $\mathcal{G}_\tau + (C \cup \{w\})$ that contain $C \cup \{w\}$. For each (K, C), only $O(1)$ many of these j-shells can be subsets of K. Therefore there exist vertices $a = a_{K,C} \in [n] \setminus K$ and $w = w_{K,C} \in K \setminus C$ such that whp, for every pair (K, C), the sides of the j-shell $C \cup \{w\} \cup \{a\}$ are all present as j-simplices in $\mathcal{G}_{\frac{\varepsilon}{\log n}}$, and therefore for any $\tau \geqslant \frac{\varepsilon}{\log n}$ such that the pair (K, C) forms a copy of $M_{j,k}$ in $\mathcal{G}_\tau$, also (K, C, w, a) forms a copy of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$. □

The following proposition describes more precisely the parameters in Definition 3.4 for $\mathbf{p} = \tau\bar{\mathbf{p}}$ and τ 'close' to 1, in terms of the analogous parameters defined in (3) and (4) for $\bar{\mathbf{p}}$.

Proposition 6.5. Let $\tau = 1 + \xi$ with $\xi = \xi(n) = o(1)$ and let $\mathbf{p} = \tau\bar{\mathbf{p}}$. Then for all $j \leqslant k \leqslant d$ with $\bar{p}_k \neq 0$,

$$\begin{aligned} \alpha_k &= \bar{\alpha}_k, & \beta_k &= (1+\xi)\,\bar{\beta}_k + \bar{\alpha}_k \xi \log n, & \gamma_k &= \bar{\gamma}_k, \\ \lambda_k &= \bar{\lambda}_k, & \mu_k &= \bar{\mu}_k - (1 + o(1))(k - j + 1)\xi \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n, & \nu_k &= \bar{\nu}_k. \end{aligned}$$

The proof of Proposition 6.5 consists of straightforward, but tedious computations, and is therefore postponed to Appendix A.5.

We derive the following corollary about the expected number of copies of $M_{j,k}$ in the critical window, which will be crucial for the proof of the Rank Theorem (Theorem 1.5).

Corollary 6.6. Let $c \in \mathbb{R}$ be a constant and suppose $(c_n)_{n\geqslant 1}$ is a sequence of real numbers such that $c_n \xrightarrow{n\to\infty} c$. Let $\tau = \left(1 + \frac{c_n}{\log n}\right)$ and $\mathbf{p} = \tau\bar{\mathbf{p}}$. Then for any k with $j \leqslant k \leqslant d$,

$$\mathbb{E}(X_{j,k}) = \begin{cases} (1+o(1))\exp(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)) & \text{if } k \text{ is a critical dimension,} \\ o(1) & \text{otherwise.} \end{cases}$$

Proof. For any $j \leqslant k \leqslant d$, Proposition 6.5 applied with $\xi = \frac{c_n}{\log n}$ tells us that

$$\lambda_k = \bar{\lambda}_k, \qquad \mu_k = \bar{\mu}_k - (1+o(1))(k-j+1)c_n \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i, \qquad \nu_k = \bar{\nu}_k. \tag{23}$$

If k is not a critical dimension, by Definition 3.3 and Lemma 5.14 we have

$$\log\left(\mathbb{E}(\bar{X}_{j,k})\right) = \bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k + o(1) \to -\infty,$$

where $\bar{X}_{j,k}$ denotes the number of copies of $M_{j,k}$ in $\mathcal{G}(n, \bar{\mathbf{p}})$ (i.e. for $\tau = 1$) and thus $\mathbb{E}(\bar{X}_{j,k}) = o(1)$. Hence, by applying Lemma 5.14 at time τ we have

$$\mathbb{E}(X_{j,k}) \overset{(23)}{=} (1+o(1))\mathbb{E}(\bar{X}_{j,k}) = o(1),$$

as required.

If k is a critical dimension, we have $\bar{\lambda}_k = 0$ and $\bar{\mu}_k = O(1)$ (see Definition 3.6). Thus by Lemma 5.14 we have

$$\begin{aligned} \mathbb{E}(X_{j,k}) &= \exp(\lambda_k \log n + \mu_k + \nu_k + o(1)) \\ &\overset{(23)}{=} (1+o(1)) \frac{\exp(\bar{\mu}_k + \bar{\nu}_k)}{\exp\left((k-j+1)c_n \sum_{i=j+1}^{d} \bar{\alpha}_i\right)} \\ &= (1+o(1))\exp(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)), \end{aligned}$$

where we are using that $0 = \bar{\lambda}_k = j + 1 - \bar{\gamma}_k - (k-j+1)\sum_{i=j+1}^{d} \bar{\alpha}_i$. ☐

As the last result of this section, we show that for τ slightly less than 1, whp we have many copies of $M_{j,k}$.

Lemma 6.7. Let $\omega_0 = \omega_0(n) = o(\log n)$ be a function that tends to infinity and let $\tau = \left(1 - \frac{1}{\omega_0}\right)$. Then there exists a constant $c > 0$ such that for any critical dimension $k \geqslant j$, whp there are at least $\exp(\frac{c \log n}{\omega_0})$ many copies of $M_{j,k}$ in $\mathcal{G}_\tau$.

The proof is similar to the proof of Lemma 6.1, although the second moment calculation is significantly simpler without the j-shell of $\hat{M}_{j,k}$, and can be found in Appendix A.6.

7 Determining the hitting time: proof of Lemma 4.1

In this section we consider the hitting time τ_j^* for the disappearance of the last minimal obstruction, i.e.

$$\tau_j^* \ := \ \sup\{\tau \in \mathbb{R}_{\geqslant 0} \mid \mathcal{G}_\tau \text{ contains a copy of } \hat{M}_{j,k} \text{ for some } k \text{ with } j \leqslant k \leqslant d\},$$

as defined in Theorem 1.4. We will show that whp this happens at around the claimed threshold $\tau = 1$ (Lemma 4.1).

Consider the time

$$\tau' \ := \ 1 - \frac{\log\log n}{10d\log n},$$

and let τ'' be the first scaled birth time larger than τ' such that there are no copies of $\hat{M}_{j,k}$ in $\mathcal{G}_{\tau''}$. Lemmas 6.3 and 6.7 tell us that whp $\mathcal{G}_{\tau'}$ contains a growing number of copies of $\hat{M}_{j,k}$, thus by definition of τ_j^* we have $\tau'' \leqslant \tau_j^*$. The following main result of this section says that they are in fact equal whp, and indeed whp both are close to 1.

Lemma 7.1. Whp $\tau_j^* = \tau''$. Furthermore, suppose ω is a function of n that tends to infinity as $n \to \infty$. Then, whp

$$1 - \frac{\omega}{\log n} < \tau_j^* < 1 + \frac{\omega}{\log n}.$$

Observe that Lemma 4.1 is an immediate corollary of Lemma 7.1. To prove Lemma 7.1, we will need one further concept and some auxiliary results.

Definition 7.2. Given an integer k with $j \leqslant k \leqslant d$, a k-simplex K is called a *local j-obstacle* if it contains at least $k - j + 1$ many j-simplices that are K-localised.

In particular a j-simplex is a local j-obstacle if and only if it is isolated. More generally, any copy of $M_{j,k}$ for $j \leqslant k \leqslant d$ is certainly a local j-obstacle, although a local j-obstacle is not necessarily an obstruction to j-cohom-connectedness.

Lemma 7.3. Whp, for all $\tau \geqslant \tau'$, every local j-obstacle in $\mathcal{G}_\tau$ also exists in $\mathcal{G}_{\tau'}$.

Proof. We will prove the statement for local obstacles of size $k + 1$, for some $j \leqslant k \leqslant d$. The lemma then follows by applying a union bound over all k.

We first note that by Remark 3.2, $\bar{p}_k < 1$ if $k \geqslant j + 1$. On the other hand, if $k = j$ and $\bar{p}_j \geqslant \frac{1}{\tau'} = 1 + o(1)$, every j-simplex is present in $\mathcal{G}_{\tau'}$ deterministically, and therefore the statement of the lemma trivially holds.

Thus in the following we may assume that

$$\text{either} \quad k \geqslant j + 1 \qquad \text{or} \quad k = j \text{ and } \bar{p}_j < \frac{1}{\tau'} < 1 + \frac{\log\log n}{9d\log n}. \tag{24}$$

Although in the second case we indeed have $\bar{p}_j < \frac{1}{\tau'}$, we would incur some technical difficulties if the probability $\bar{p}_j$ is very 'close' to $\frac{1}{\tau'}$. Hence, in the following calculations we need to replace τ' by a slightly smaller value. More precisely, we consider

$$\tau^- \ := \ 1 - \frac{\log\log n}{5d\log n}.$$

We will show that whp for any $\tau \geqslant \tau^-$, a local j-obstacle in $\mathcal{G}_\tau$ also exists in $\mathcal{G}_{\tau^-}$, thus obtaining the statement for any $\tau \geqslant \tau' > \tau^-$ as well. In particular, observe that

$$\tau^- \bar{p}_k < 1 \qquad \text{for every } k \geqslant j,$$

by Remark 3.2 and (24).

Fix $k \geqslant j$, let K be a $(k+1)$-set, and recall that $\tau_K \in [0, 1/\bar{p}_k]$ denotes its scaled birth time, i.e. if t_K is the birth time of K as a k-simplex in $\mathcal{G}_\tau$, then $\tau_K = t_K/\bar{p}_k$ (see (2)). In order to *become* a local j-obstacle in $\mathcal{G}_\tau$ for some $\tau > \tau^-$, K must contain a collection $\mathcal{J}$ of $k-j+1$ many $(j+1)$-sets such that the following conditions are satisfied:

(L1) $\tau_K > \tau^-$;

(L2) every $J \in \mathcal{J}$ is K-localised in $\mathcal{G}_{\tau^-}$;

(L3) K is born as a k-simplex before any other simplex I that contains some $J \in \mathcal{J}$, but which is not contained in K, i.e. $\tau_K < \tau_I$ for all such I.

Fix the $(k+1)$-set K and the collection $\mathcal{J}$ of $(j+1)$-sets in K. For this choice of K and $\mathcal{J}$, we denote by L_1, L_2, and L_3 the events that conditions (L1), (L2), and (L3) hold, respectively.

By definition of our model, we have that

$$\mathbb{P}(L_1) = (1 - \tau^- \bar{p}_k). \tag{25}$$

In order to compute $\mathbb{P}(L_2 \mid L_1)$, first observe that L_2 is independent of L_1. By Proposition 5.13 applied with $\mathbf{p} = \tau^- \bar{\mathbf{p}}$, we have

$$\begin{aligned} \mathbb{P}(L_2 \mid L_1) &= \mathbb{P}(L_2) \\ &= (1+o(1))\bar{q}^{\,\tau^-(k-j+1)} \\ &\overset{(16),(17)}{=} \left(\frac{\Theta(1)\cdot \mathbb{E}(\bar{X}_{j,k})}{n^{k+1}\min\{\bar{p}_k, 1\}}\right)^{\tau^-} \overset{(\bar{p}_k < 1+o(1))}{=} \left(\frac{\Theta(1)\cdot \mathbb{E}(\bar{X}_{j,k})}{n^{k+1}\bar{p}_k}\right)^{\tau^-}, \end{aligned} \tag{26}$$

where $\bar{X}_{j,k}$ denotes the number of copies of $M_{j,k}$ in $\mathcal{G}_1 = \mathcal{G}(n, \bar{\mathbf{p}})$ (i.e. $\tau = 1$), and thus $\mathbb{E}(\bar{X}_{j,k}) \leqslant 1 + o(1)$ (see (21)).

We now want to bound $\mathbb{P}\big(L_3 \mid (L_1 \wedge L_2)\big)$. For any i such that $j+1 \leqslant i \leqslant d$ and for any $J \in \mathcal{J}$, there are $\binom{n-k-1}{i-j}$ many $(i+1)$-sets which contain J and whose remaining vertices are outside K. In order for L_3 to hold, all these $(i+1)$-sets (among others) must be born as simplices after K and observe that all of these $(i+1)$-sets are distinct for different choices of J. It will be convenient to pick $i = k_0$, recalling from Definition 5.15 (b) that $k_0 \geqslant j+1$ is such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$. Thus we have a family $\mathcal{Z}$ of

$$z := |\mathcal{Z}| = (k-j+1)\binom{n-k-1}{k_0 - j} = \Theta\left(\frac{\log n}{\bar{p}_{k_0}}\right) \tag{27}$$

many *bad* (k_0+1)-sets whose scaled birth times are uniformly distributed in the interval $[\tau^-, \frac{1}{\bar{p}_{k_0}}]$ (since the corresponding simplices are not present in $\mathcal{G}_{\tau^-}$ by L_2), but must all be larger than τ_K, in order for L_3 to hold. Similarly, conditioned on L_1, the scaled birth time τ_K is uniformly distributed in $[\tau^-, \frac{1}{\bar{p}_k}]$. This allows us to prove the following.

Claim 7.4. Let L_3' be the event that K is born as a k-simplex before any of the bad (k_0+1)-sets in $\mathcal{Z}$. Then

$$\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2)\big) = (1+o(1))\frac{\bar{p}_k}{z\bar{p}_{k_0}(1-\tau^-\bar{p}_k)} \overset{(27)}{=} \Theta\left(\frac{\bar{p}_k}{(1-\tau^-\bar{p}_k)\log n}\right).$$

Before proving Claim 7.4, let us first show how it implies Lemma 7.3. Note that $L_3 \subseteq L_3'$, thus Claim 7.4 in particular implies that

$$\mathbb{P}\big(L_3 \mid (L_1 \wedge L_2)\big) \leqslant \mathbb{P}\big(L_3' \mid (L_1 \wedge L_2)\big) = \Theta\left(\frac{\bar{p}_k}{(1-\tau^-\bar{p}_k)\log n}\right). \tag{28}$$

Putting (25), (26), and (28) together we have

$$\begin{aligned}\mathbb{P}(L_1 \wedge L_2 \wedge L_3) &= \mathbb{P}(L_1)\mathbb{P}(L_2 \mid L_1)\mathbb{P}\big(L_3 \mid (L_1 \wedge L_2)\big)\\ &= O\left((1-\tau^-\bar{p}_k)\left(\frac{\mathbb{E}(\bar{X}_{j,k})}{n^{k+1}\bar{p}_k}\right)^{\tau^-}\frac{\bar{p}_k}{(1-\tau^-\bar{p}_k)\log n}\right).\end{aligned}$$

Recalling that $\mathbb{E}(\bar{X}_{j,k}) \leqslant 1+o(1)$ by (21), we thus deduce that

$$\mathbb{P}(L_1 \wedge L_2 \wedge L_3) = O\left(\frac{\bar{p}_k^{1-\tau^-}}{n^{\tau^-(k+1)}\log n}\right). \tag{29}$$

There are $\Theta(n^{k+1})$ choices for the $(k+1)$-set K and, once K is fixed, there are $\Theta(1)$ choices for the collection $\mathcal{J}$ of $k-j+1$ many $(j+1)$-sets in K. Since $\bar{\mathbf{p}}$ is a j-admissible direction (cf. Definition 3.1) we know that $\bar{p}_k = O\left(\frac{\log n}{n^{k-j}}\right)$, hence the expected numbers of pairs $(K,\mathcal{J})$ satisfying (L1), (L2), and (L3) is

$$\begin{aligned}\Theta(n^{k+1})\mathbb{P}(L_1 \wedge L_2 \wedge L_3) &\overset{(29)}{=} O\left(\frac{\left(n^{k+1}\bar{p}_k\right)^{1-\tau^-}}{\log n}\right)\\ &= O\left(\frac{n^{(j+1)(1-\tau^-)}}{(\log n)^{\tau^-}}\right)\\ &= O\left(\exp\left(\left(\frac{j+1}{5d}-(1-o(1))\right)\log\log n\right)\right)\\ &= O\left(\exp\left(\frac{-\log\log n}{2}\right)\right) \xrightarrow{n\to\infty} 0.\end{aligned}$$

Therefore by Markov's inequality, whp there are no such pairs $(K,\mathcal{J})$, as required.

All that is left in order to prove Lemma 7.3 is the proof of Claim 7.4. We split the proof into two cases, according to which of $\bar{p}_k$ and $\bar{p}_{k_0}$ is larger. In both cases, we will use the fact that, since $k_0 \geqslant j+1$, by Remark 3.2 we have $\frac{1}{\bar{p}_{k_0}} = \omega(1)$, and thus

$$\frac{1}{\bar{p}_{k_0}} - \tau^- = (1+o(1))\frac{1}{\bar{p}_{k_0}}. \tag{30}$$

Case 1: $\frac{1}{\bar{p}_k} \geqslant \frac{1}{\bar{p}_{k_0}}$. Let S be the event that $\tau_K \leqslant \frac{1}{\bar{p}_{k_0}}$. Note that $L_3' \subseteq S$, hence

$$\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2)\big) = \mathbb{P}\big(S \mid (L_1 \wedge L_2)\big)\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2 \wedge S)\big). \tag{31}$$

Recall that conditioned on L_1, the birth time τ_K is uniformly distributed in $[\tau^-, \frac{1}{\bar{p}_k}]$. Therefore, since S is independent of L_2, we have

$$\mathbb{P}\big(S \mid (L_1 \wedge L_2)\big) = \mathbb{P}(S \mid L_1) = \frac{1/\bar{p}_{k_0} - \tau^-}{1/\bar{p}_k - \tau^-} \overset{(30)}{=} (1+o(1))\frac{\bar{p}_k}{\bar{p_{k_0}}(1-\tau^-\bar{p}_k)}. \tag{32}$$

Moreover, conditioned on S, the set K has the same birth time distribution as the z many bad (k_0+1)-sets in $\mathcal{Z}$ and all these birth times are independent, thus

$$\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2 \wedge S)\big) = \frac{1}{1+z} = (1+o(1))\frac{1}{z}\,, \tag{33}$$

because $z = \omega(1)$ by (27) and the fact that $p_{k_0} = o(1)$.

Putting (31), (32), and (33) together, we have

$$\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2)\big) = (1+o(1))\frac{\bar{p}_k}{z\bar{p}_{k_0}(1-\tau^-\bar{p}_k)},$$

as claimed.

Case 2: $\frac{1}{\bar{p}_k} < \frac{1}{\bar{p}_{k_0}}$. First observe that if $k = j$, by (24) we have $\bar{p}_j < 1 + \frac{\log\log n}{9d\log n}$, while if $k \geqslant j+1$, then $\bar{p}_k < 1$ by Remark 3.2. Thus by the definition of τ^-, for any $k \geqslant j$ we have

$$1 - \tau^-\bar{p}_k > \frac{\log\log n}{12d\log n}. \tag{34}$$

Let $\mathcal{Z}_{\bar{p}_k}$ be the set of bad (k_0+1)-sets in $\mathcal{Z}$ with birth times in the interval $[\tau^-, \frac{1}{\bar{p}_k}]$ and let $\zeta_k := |\mathcal{Z}_{\bar{p}_k}|$. Since the birth times of the sets in $\mathcal{Z}$ are uniformly distributed in $[\tau^-, \frac{1}{\bar{p}_{k_0}}]$, the random variable ζ_k has binomial distribution $\mathrm{Bi}\left(z, \frac{1/\bar{p}_k - \tau^-}{1/\bar{p}_{k_0} - \tau^-}\right)$ and observe that

$$\mathbb{E}(\zeta_k) = z \cdot \frac{1/\bar{p}_k - \tau^-}{1/\bar{p}_{k_0} - \tau^-} \overset{(30)}{=} (1+o(1))\frac{z\bar{p}_{k_0}}{\bar{p}_k}(1-\tau^-\bar{p}_k). \tag{35}$$

Since $z\bar{p}_{k_0} = \Theta(\log n)$ by (27) and $\bar{p}_k < 1 + \frac{\log\log n}{9d\log n}$, we obtain

$$\mathbb{E}(\zeta_k) \overset{(34)}{=} \Omega\left(\log n \cdot \frac{\log\log n}{12d\log n}\right) = \Omega\left(\log\log n\right) \to \infty.$$

By the Chernoff bound, the probability that ζ_k is not within a multiplicative factor $1 \pm \frac{1}{\mathbb{E}(\zeta_k)^{1/4}}$ of the mean is $\exp(-\Omega(\mathbb{E}(\zeta_k)^{1/2}))$.

Furthermore, conditioned on the value of ζ_k (and the events L_1, L_2), the probability of L_3' is $\frac{1}{1+\zeta_k}$, because the birth times of K and of the bad sets in $\mathcal{Z}_{\bar{p}_k}$ all have the same (conditional) distribution. Thus, since $\mathbb{E}(\zeta_k) \to \infty$, we have

$$\begin{aligned}\mathbb{P}\big(L_3' \mid (L_1 \wedge L_2)\big) &= \frac{1+o(1)}{1+\Big(1 \pm \frac{1}{\mathbb{E}(\zeta_k)^{1/4}}\Big)\mathbb{E}(\zeta_k)} + \exp(-\Omega(\mathbb{E}(\zeta_k)^{1/2})) \\ &= \frac{1+o(1)}{\mathbb{E}(\zeta_k)} \overset{(35)}{=} (1+o(1))\frac{\bar{p}_k}{z\bar{p}_{k_0}(1-\tau^{-}\bar{p}_k)},\end{aligned}$$

as claimed. □

We are now ready to prove the main result of this section.

Proof of Lemma 7.1. Observe that in particular a copy of $M_{j,k}$ (or, more precisely, the associated k-simplex) is a local j-obstacle. Lemma 7.3 shows that if a local j-obstacle is present in $\mathcal{G}_\tau$ for some $\tau \geqslant \tau'$, then whp it already existed in $\mathcal{G}_{\tau'}$. Therefore, if $\tau'' < \tau_j^*$ then a copy of $\hat{M}_{j,k}$ appears in between these two times, but whp the associated copy of $M_{j,k}$ would already exist at time τ'' and thus whp would form a copy of $\hat{M}_{j,k}$ at that time too, by Lemma 6.3. This cannot happen by definition of τ'', which gives $\tau'' = \tau_j^*$ whp, as required.

To prove the second statement, observe that by Lemmas 6.3 and 6.7, whp we have $\tau_j^* > 1 - \frac{\omega}{\log n}$, proving the lower bound.

In the proof of the upper bound it will be convenient to assume that $\omega = o(\log n)$—this assumption is permissible since the statement becomes stronger for smaller ω.

For $\tau = 1 + \frac{\omega}{\log n}$ and any $j \leqslant k \leqslant d$, applying (16) (for $k \geqslant j+1$) or (17) (for $k = j$) to $X_{j,k}$ and to the number $\bar{X}_{j,k}$ of copies of $M_{j,k}$ at time 1, we deduce that

$$\begin{aligned}\mathbb{E}(X_{j,k}) \;&=\; (1+o(1))\mathbb{E}(\bar{X}_{j,k}) \cdot \bar{q}^{(\tau-1)(k-j+1)} \\ &\overset{(15)}{=} (1+o(1))\mathbb{E}(\bar{X}_{j,k}) \cdot \exp\left(-\frac{\omega}{\log n}(k-j+1)\sum\nolimits_{k=j+1}^{d}\left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right) \\ &\leqslant\; (1+o(1))\mathbb{E}(\bar{X}_{j,k})\exp(-\bar{\alpha}_{k_0}\omega) = o(1),\end{aligned}$$

where we are using that $\mathbb{E}(\bar{X}_{j,k}) \leqslant 1 + o(1)$ by (21) and that the index k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$.

Hence, by Markov's inequality whp there are no copies of $M_{j,k}$ and thus also no copies of $\hat{M}_{j,k}$. This means that whp $\tau'' < 1 + \frac{\omega}{\log n}$, and we have already shown that whp $\tau_j^* = \tau''$. □

8 Subcritical case: proof of Lemma 4.4

In this section we first derive some auxiliary results and combine them to prove Lemma 4.4, which plays a crucial role in the proof of the subcritical case (i.e. statement (b)) of Theorem 1.4.

Given a constant $\varepsilon > 0$, in order to show that whp $H^j(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_j(\varepsilon) = [\varepsilon/n, \tau_j^*)$, we split this range into three separate intervals,

$$\left[\frac{\varepsilon}{n}, \tau_j^*\right) = \left[\frac{\varepsilon}{n}, \frac{\delta}{\log n}\right] \cup \left[\frac{\delta}{\log n}, 1 - \frac{1}{(\log n)^{1/3}}\right] \cup \left[1 - \frac{1}{(\log n)^{1/3}}, \tau_j^*\right)$$

for some constant $\delta > 0$, and prove that for each of these ranges there is some k and one copy of $\hat{M}_{j,k}$ which exists throughout the subinterval (Lemmas 8.1, 8.3, and 8.4).

Lemma 8.1. For every constant $\varepsilon > 0$, there exists a constant $\delta > 0$ such that whp there is at least one copy of $\hat{M}_{j,k_0}$ that is present in the process $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_\tau$ for all values

$$\tau \in \left[\frac{\varepsilon}{n}, \frac{\delta}{\log n}\right].$$

Remark 8.2. Indeed, Lemma 8.1 would also hold with k_0 replaced by any index k with $j+1 \leqslant k \leqslant d$ such that $\bar{\alpha}_k \neq 0$.

Proof of Lemma 8.1. By Lemma 6.1, there exist constants $0 < c_1 < c_2$ (depending on ε) such that whp the number $\hat{X}_{j,k_0}$ of copies of $\hat{M}_{j,k_0}$ in $\mathcal{G}_{\frac{\varepsilon}{n}}$ satisfies

$$c_1(\log n)^{j+2} \leqslant \hat{X}_{j,k_0} \leqslant c_2(\log n)^{j+2},$$

and all these copies of $\hat{M}_{j,k_0}$ originate from distinct copies of M_{j,k_0}. We will show that whp at least one of these copies survives (i.e. remains a copy of $\hat{M}_{j,k_0}$) until time $\tau = \frac{\delta}{\log n}$, for a suitable constant $\delta > 0$.

For each index k with $j+1 \leqslant k \leqslant d$, call a $(k+1)$-set *dangerous* if it is *not* a k-simplex in $\mathcal{G}_{\frac{\varepsilon}{n}}$ and contains a petal of at least one copy of $\hat{M}_{j,k_0}$. Since there are at most $c_2(k-j+1)(\log n)^{j+2}$ petals, and each is contained in at most $\binom{n-j-1}{k-j} \leqslant \frac{n^{k-j}}{(k-j)!}$ many $(k+1)$-sets, setting $c_3 = \max_k \frac{(k-j+1)c_2}{(k-j)!}$, for all k the number of dangerous $(k+1)$-sets is at most

$$c_3(\log n)^{j+2} n^{k-j}.$$

For each dangerous $(k+1)$-set, the probability that it becomes a simplex by time $\tau = \delta/(\log n)$ is the probability that its scaled birth time is at most $\delta/(\log n)$ conditioned on the event that it is at least ε/n, which is

$$\frac{\left(\frac{\delta}{\log n} - \frac{\varepsilon}{n}\right)\bar{p}_k}{1 - \frac{\varepsilon}{n}\bar{p}_k} \leqslant \frac{\delta \bar{p}_k}{\log n} \leqslant \frac{(k-j)!(\bar{\alpha}_k + 1)\delta}{n^{k-j}}.$$

Setting $c_4 := \max_k (k-j)!(\bar{\alpha}_k + 1)$, the number of dangerous $(k+1)$-sets that turn into k-simplices in the time interval we are considering is dominated by

$$\mathrm{Bi}\left(c_3(\log n)^{j+2} n^{k-j}, \frac{c_4\delta}{n^{k-j}}\right),$$

so by a Chernoff bound, we deduce that the number of dangerous sets of *any* size that turn into simplices by time $\tau = \delta/(\log n)$ is whp smaller than

$$2(d-j)c_3c_4\delta(\log n)^{j+2}.$$

Note that each $(k+1)$-set can contain at most $\binom{k+1}{j+1} \leqslant \binom{d+1}{j+1} =: c_5$ petals, and therefore each of these dangerous sets makes at most c_5 copies of $\hat{M}_{j,k_0}$ disappear by becoming a simplex. If we choose $\delta < \frac{c_1}{2(d-j)c_3c_4c_5}$, then whp the number of copies of $\hat{M}_{j,k_0}$ that disappear by time $\tau = \delta/(\log n)$ is at most

$$2(d-j)c_3c_4c_5\delta(\log n)^{j+2} < c_1(\log n)^{j+2} \leqslant \hat{X}_{j,k_0}.$$

In other words, at least one copy of $\hat{M}_{j,k_0}$ that exists at the beginning of the interval survives until the end of the interval. ☐

Lemma 8.3. For every constant $\delta > 0$ and every critical dimension k with $j \leqslant k \leqslant d$, whp there is a copy of $\hat{M}_{j,k}$ that is present in the process $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_\tau$ for all values

$$\tau \in \left[\frac{\delta}{\log n}, 1 - \frac{1}{(\log n)^{1/3}}\right].$$

Proof. By Lemma 6.7 with $\omega_0 = (\log n)^{1/3}$, whp there are more than $\exp(\sqrt{\log n})$ many copies of $M_{j,k}$ in $\mathcal{G}_\tau$ at the upper end $\tau = 1 - \frac{1}{(\log n)^{1/3}}$ of the interval. Observe that only $\Theta(1)$ such copies can share the same k-simplex K, thus we have $\Omega(\exp(\sqrt{\log n}))$ many copies with distinct k-simplices. For each such copy (K, C), the scaled birth time of K is uniformly distributed within

$$\left[0, \min\left\{1 - \frac{1}{(\log n)^{1/3}}, \frac{1}{\bar{p}_k}\right\}\right],$$

meaning that (K, C) formed a copy of $M_{j,k}$ at time $\tau = \frac{\delta}{\log n}$ with probability

$$\frac{\delta/(\log n)}{\min\left\{1 - \frac{1}{(\log n)^{1/3}}, \frac{1}{\bar{p}_k}\right\}} \geqslant \frac{\delta}{\log n}.$$

The birth times of the simplices K are independent, thus the probability that at least one of them was present at time $\tau = \frac{\delta}{\log n}$ is at least

$$1 - \left(1 - \frac{\delta}{\log n}\right)^{\Theta\left(\exp(\sqrt{\log n})\right)} \geqslant 1 - \exp\left(-\Theta\left(\frac{\exp(\sqrt{\log n})}{\log n}\right)\right) = 1 - o(1).$$

In other words, whp some copy (K, C) of $M_{j,k}$ that exists at time $\tau = 1 - \frac{1}{(\log n)^{1/3}}$ already existed at time $\tau = \frac{\delta}{\log n}$. By Corollary 6.4 applied at time $\tau = \frac{\delta}{\log n}$, whp there exist $w \in K \setminus C$ and $a \in [n] \setminus K$ such that (K, C, w, a) is a copy of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$ and therefore throughout the interval $\left[\frac{\delta}{\log n}, 1 - \frac{1}{(\log n)^{1/3}}\right]$, as claimed. ☐

Lemma 8.4. Whp the minimal obstruction which vanishes at time τ_j^* (defined in Theorem 1.4) was already present in $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_\tau$ for all values τ with

$$\tau \in \left[1 - \frac{1}{(\log n)^{1/3}}, \tau_j^*\right).$$

Proof. Recall that by Definition 5.15, the last minimal obstruction to vanish is a copy (K, C, w, a) of $\hat{M}_{j,\ell}$. Similar to the proof of Lemma 8.3, the birth time of K is uniformly distributed within $\big[0, \min\{\tau_j^*, 1/\bar{p}_\ell\}\big)$ and by Lemma 7.1, whp $\tau_j^* \leqslant 1 + \frac{1}{(\log n)^{1/3}}$. Conditioned on this high probability event, the probability that K already existed as a simplex in $\mathcal{G}_\tau$ at time $\tau = 1 - \frac{1}{(\log n)^{1/3}}$ is at least

$$\frac{1 - \frac{1}{(\log n)^{1/3}}}{1 + \frac{1}{(\log n)^{1/3}}} = 1 - o(1),$$

i.e. whp (K, C) formed a copy of $M_{j,k}$ already at time $\tau = 1 - \frac{1}{(\log n)^{1/3}}$. By Corollary 6.4, this means that whp there exist $\tilde{w}, \tilde{a}$ such that $(K, C, \tilde{w}, \tilde{a})$ forms a copy of $\hat{M}_{j,\ell}$ throughout the interval. ☐

Proof of Lemma 4.4. Lemmas 8.1, 8.3, and 8.4 imply that whp for any $\tau \in [\varepsilon/n, \tau_j^*)$ a copy of $\hat{M}_{j,k}$ (for some $j \leqslant k \leqslant d$) exists in $\mathcal{G}_\tau$. Therefore, for any τ in this range $H^j(\mathcal{G}_\tau; R) \neq 0$ by Corollary 5.10. ☐

9 Critical and supercritical cases: proof of Lemma 4.7

In this section we present some auxiliary results and prove Lemma 4.7, which we used in Section 4 to show that whp the process $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_\tau$ is j-cohom-connected for all $\tau \geqslant \tau_j^*$ (Theorem 1.4 (c)). Furthermore, the results of this section will be fundamental for the proof of Theorem 1.5 (Section 10).

Recall that in order to have $H^j(\mathcal{G}_\tau; R)$ not vanishing, $\mathcal{G}_\tau$ would have to admit a *bad function*, i.e. a j-cocycle that is not a j-coboundary. We aim to show that no bad function exists by considering what such a function with smallest possible support might look like, if it exists. We show that the support must be *traversable* (Definition 9.3, Lemma 9.4), and then use this property to show that whp the support cannot be small (Lemma 9.6). Subsequently, we use traversability and a result of Meshulam and Wallach [35] to show that whp the support cannot be large (Lemma 9.8), which is a contradiction.

However, so far this only proves that for any $\tau \geqslant \tau_j^*$, whp $H^j(\mathcal{G}_\tau; R) = 0$. We need to know that whp, for any $\tau \geqslant \tau_j^*$ the group $H^j(\mathcal{G}_\tau; R)$ vanishes (i.e. with a different order of quantifiers). We achieve this by observing that at time τ_j^* the j-th cohomology group is zero, and proving that whp no new bad functions can appear (Lemma 9.15).

Slightly more generally than described above, we will actually prove that for τ large enough, but slightly smaller than τ_j^*, the only bad functions that exist are the result of copies of $M_{j,k}$ existing.

Definition 9.1. Let (K, C) be a copy of $M_{j,k}$ in a d-complex $\mathcal{G}$. We say that a j-cochain $f \in C^j(\mathcal{G})$ *arises from* (K, C) if its support $\mathcal{S} = \mathrm{supp}(f)$ is such that

$$\mathcal{S} = \mathcal{F}(K, C).$$

Note that we have no restrictions (apart from not being 0) on the values from the group R that such a j-cochain f takes on its support $\mathcal{S}$.

We say that a j-cocycle f (i.e. $f \in \ker(\delta^j) \subseteq C^j(\mathcal{G})$) is *generated by copies of* $M_{j,k}$ if it belongs to the same cohomology class as some $f_1 + f_2 + \ldots + f_m$, where each f_i is a j-cocycle that arises from a copy of M_{j,k_i}. We denote by $\mathcal{N}_\mathcal{G}$ the set of j-cocycles in $\mathcal{G}$ that are *not* generated by copies of $M_{j,k}$. If $\mathcal{G} = \mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$, we will ease notation by defining $\mathcal{N}_\tau := \mathcal{N}_{\mathcal{G}_\tau}$.

The goal is to prove that whp for all $\tau \geqslant \tau_j^*$ we have $\mathcal{N}_\tau = \emptyset$. Since in this range there are no copies of $\hat{M}_{j,k}$ and thus by Corollary 6.4 whp also no copies of $M_{j,k}$, this will imply that whp $H^j(\mathcal{G}_\tau; R) = 0$. To this end, we need the following notation.

Definition 9.2. For every τ, we denote by f_τ a function in $\mathcal{N}_\tau$ with smallest support $\mathcal{S}_\tau$, if such a function exists.

In order to bound the number of possible such supports $\mathcal{S}_\tau$, we first show (Lemma 9.4) that $\mathcal{S}_\tau$ must satisfy the following concept of *traversability*.

Definition 9.3. Let $(\mathcal{S}, \mathcal{T})$ be a pair where $\mathcal{S}$ is a collection of j-simplices in $\mathcal{G}_\tau$ and $\mathcal{T}$ is a collection of simplices in $\mathcal{G}_\tau$ of dimensions $j+1, \ldots, d$.

We say that $\mathcal{S}$ is $\mathcal{T}$*-traversable* if it *cannot* be partitioned into two non-empty subsets such that every simplex of $\mathcal{T}$ contains elements of $\mathcal{S}$ in at most one of the two subsets. Equivalently, $\mathcal{S}$ is $\mathcal{T}$-traversable if for every $J, J' \in \mathcal{S}$ there exists a sequence $J = J_0, J_1, \ldots, J_m = J'$ of j-simplices in $\mathcal{S}$ and a sequence $K_1, \ldots, K_m$ of simplices in $\mathcal{T}$ (not necessarily all of the same dimension) such that $(J_{i-1} \cup J_i) \subseteq K_i$ for all $i \in [m]$. We may think of these sequences of simplices as a generalisation of a path between J and J', and thus traversability may be considered a form of connectedness.

We say that $\mathcal{S}$ is *traversable* (in $\mathcal{G}_\tau$) if it is $\mathcal{T}$-traversable with $\mathcal{T}$ consisting of *all* k-simplices of $\mathcal{G}_\tau$ for every $j+1 \leqslant k \leqslant d$.

Lemma 9.4. For every τ, the support $\mathcal{S}_\tau$, if it exists, is traversable.

Proof. Suppose $\mathcal{S}_\tau$ is not traversable and let $\mathcal{S}_\tau = \mathcal{S}_{(1)} \dot\cup \mathcal{S}_{(2)}$ be a partition into non-empty parts such that (in particular) each $(j+1)$-simplex of $\mathcal{G}_\tau$ contains elements of $\mathcal{S}_\tau$ in at most one of the two parts.

For $i = 1, 2$, let $f_{(i)}$ be the j-cochain defined by

$$f_{(i)}(\sigma) = \begin{cases} f_\tau(\sigma) & \text{if } \sigma \in \mathcal{S}_{(i)}, \\ 0_R & \text{otherwise.} \end{cases}$$

Suppose that ρ is a $(j+1)$-simplex that contains j-simplices from only one $\mathcal{S}_{(i)}$, without loss of generality from $\mathcal{S}_{(1)}$ and not $\mathcal{S}_{(2)}$. Then trivially $(\delta^j f_{(2)})(\rho) = 0$ and $(\delta^j f_{(1)})(\rho) = (\delta^j f_\tau)(\rho) =$

0, because $f_\tau \in \ker \delta^j$. Thus both functions $f_{(i)}$ are j-cocycles, and neither of them lies in $\mathcal{N}_\tau$ by the minimality of $\mathcal{S}_\tau$. Hence $f_\tau = f_{(1)} + f_{(2)}$ is generated by copies of $M_{j,k}$, since this property is closed under summation, a contradiction to $f_\tau \in \mathcal{N}_\tau$. □

It is clear that given a traversable $\mathcal{S}$ in $\mathcal{G}_\tau$, there exists a minimal collection $\mathcal{T}$ of simplices of $\mathcal{G}_\tau$ such that $\mathcal{S}$ is $\mathcal{T}$-traversable and every $\sigma \in \mathcal{T}$ has scaled birth time at most τ. We fix some such minimal collection and denote it by $\mathcal{T}(\mathcal{S})$. We also define the sequence $\mathbf{t}(\mathcal{S}) = (t_{j+1}, \dots, t_d)$ where $t_k \geqslant 0$ is the number of k-simplices in $\mathcal{T}(\mathcal{S})$, for every $j+1 \leqslant k \leqslant d$.

Remark 9.5. We note that if $\mathcal{S}$ is traversable, $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ can be explored in a natural way using a breadth-first search process: start from some j-simplex in $\mathcal{S}$ and reveal all k-simplices of $\mathcal{T}(\mathcal{S})$ containing it, for every $k \in \{j+1, \dots, d\}$. We thus 'discover' any further j-simplices of $\mathcal{S}$ contained in these simplices of $\mathcal{T}(\mathcal{S})$, and from each of these j-simplices in turn we repeat the process. The $\mathcal{T}(\mathcal{S})$-traversability of $\mathcal{S}$ implies that all j-simplices of $\mathcal{S}$ (and also all simplices of $\mathcal{T}(\mathcal{S})$) are discovered in this process.

This viewpoint allows us to observe some important properties. First, note that by the minimality of $\mathcal{T}(\mathcal{S})$, every simplex of $\mathcal{T}(\mathcal{S})$ must contain a previously undiscovered j-simplex of $\mathcal{S}$, and therefore

$$|\mathcal{T}(\mathcal{S})| = \sum_{k=j+1}^{d} t_k \leqslant |\mathcal{S}|. \tag{36}$$

On the other hand, each k-simplex of $\mathcal{T}(\mathcal{S})$ is discovered from a j-simplex it contains, and therefore contains at most $k-j$ previously undiscovered vertices. Thus if v is the number of vertices that are contained in some j-simplex of $\mathcal{S}$, we have

$$v \leqslant (j+1) + \sum_{k=j+1}^{d} (k-j) t_k. \tag{37}$$

In the next lemma we show that at around $\tau = 1$, while we may have copies of $M_{j,k}$, whp there are no 'small' traversable supports of j-cocycles other than those arising from these $M_{j,k}$.

Lemma 9.6. Let $\tau = 1 + o(1)$ and let $h \in \mathbb{R}^+$ be a constant. Then whp there is no j-cocycle in $\mathcal{G}_\tau$ with traversable support of size $s \leqslant h$, apart from those arising from copies of $M_{j,k}$.

In particular, whp $|\mathcal{S}_\tau| > h$, if it exists.

Proof. We want to bound the expected number of pairs $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$, where $\mathcal{S}$ is a traversable support of a j-cocycle not arising from a copy of $M_{j,k}$ and with size $s \leqslant h$.

Recall that v is the number of vertices that are contained in some j-simplex of $\mathcal{S}$ and $\mathbf{t}(\mathcal{S}) = (t_{j+1}, \dots, t_d)$ is such that $t_k \geqslant 0$ indicates the number of k-simplices in $\mathcal{T}(\mathcal{S})$. By (36) we have

$$\sum_{k=j+1}^{d} t_k \leqslant s \leqslant h. \tag{38}$$

Since $\mathcal{S}$ is the support of a j-cocycle, by Lemma 5.11 (a) if a k-simplex contains an element in $\mathcal{S}$ then all its $k+1$ vertices are contained in some j-simplex of $\mathcal{S}$. This means that the $s\binom{n-v}{k-j}$ many $(k+1)$-sets containing a j-simplex in $\mathcal{S}$ and $k-j$ vertices not in any j-simplex of $\mathcal{S}$ are not allowed to be simplices. We thus obtain that the probability that a fixed pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ has all the necessary properties (in terms of which simplices exist and which do not) is bounded from above by

$$\begin{aligned}
&\prod_{k=j+1}^{d} p_k^{t_k}(1-p_k)^{s\binom{n-v}{k-j}} \\
&\qquad = \prod_{k=j+1}^{d} \left((1+o(1))\bar{p}_k\right)^{t_k} \exp\left(-(1+o(1))\bar{p}_k\left(s\frac{n^{k-j}}{(k-j)!}\right)\right) \\
&\qquad = O\left(\prod_{k=j+1}^{d} \left(n^{-(k-j+\bar{\gamma}_k)+o(1)}\right)^{t_k} \exp\left(-(1+o(1))s\left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right)\right) \\
&\qquad = O\left(n^{-\sum_{k=j+1}^{d}\left((k-j+\bar{\gamma}_k)t_k + s\bar{\alpha}_k\right)+o(1)}\right),
\end{aligned}$$

where we used the observation that $\bar{\beta}_k$ can only be negative if $\bar{\alpha}_k \neq 0$, in which case $\bar{\beta}_k = o(\log n)$ and $\bar{\gamma}_k = 0$ (see (A2)).

Let $\mathbf{t} = (t_{j+1}, \dots, t_d)$ and denote by $E_{s,v,\mathbf{t}}$ the event that a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with $\mathcal{S}$ a traversable support of size s on v vertices and $\mathbf{t}(\mathcal{S}) = \mathbf{t}$ exists. Equations (37) and (38) together imply that $v \leqslant j+1+(d-j)h = O(1)$, and therefore there are $O(n^v)$ ways of choosing such a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$, meaning that

$$\mathbb{P}\left(E_{s,v,\mathbf{t}}\right) = O\left(n^{v-\sum_k (k-j)t_k - \sum_k(\bar{\gamma}_k t_k + s\bar{\alpha}_k)+o(1)}\right).$$

By (37), we have

$$v - \sum_{k=j+1}^{d}(k-j)t_k \leqslant j+1.$$

Moreover, for an index i such that $t_i \geqslant 1$ (such an index exists, because otherwise the support would be empty), by Lemma 5.11 and since the considered j-cocycle does not arise from a copy of $M_{j,k}$, it holds that $s \geqslant i-j+2$. Recalling that

$$\bar{\lambda}_i = j+1-\bar{\gamma}_i-(i-j+1)\sum_{k=j+1}^{d}\bar{\alpha}_k \overset{\text{(C1),(C2)}}{\leqslant} 0$$

and that $\sum_{k=j+1}^{d}\bar{\alpha}_k > 0$, we have

$$\sum_{k=j+1}^{d}(\bar{\gamma}_k t_k + s\bar{\alpha}_k) \geqslant \bar{\gamma}_i + (i-j+2)\sum_{k=j+1}^{d}\bar{\alpha}_k = j+1-\bar{\lambda}_i + \sum_{k=j+1}^{d}\bar{\alpha}_k > j+1.$$

Thus,

$$\mathbb{P}\left(E_{s,v,\mathbf{t}}\right) = o(1).$$

Since by (37) and (38) there are only constantly many choices for the values s, v, and $\mathbf{t}$, the probability that any such pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ exists is $o(1)$. □

For supports of larger sizes, we will need a lower bound on the number of $(j+2)$-sets that are not allowed to be $(j+1)$-simplices in $\mathcal{G}_\tau$. Such a bound is given by Meshulam and Wallach [35, Proposition 3.1], where it was stated for the case when the cohomology groups considered are over any *finite* abelian group R. We observe however, that the proof given in [35] still works without the additional condition that R is finite. We include this proof in Appendix A.7 for completeness.

Proposition 9.7 ([35, Proposition 3.1]). Let $n \geqslant j+2$ and let Δ be the downward-closure of the $(n-1)$-simplex on $[n]$. Let $f \in C^j(\Delta; R)$ have support $\mathcal{S}$ and suppose that any other j-cochain of the form $f + \delta^{j-1}g$, where $g \in C^{j-1}(\Delta; R)$, has support of size at least $|\mathcal{S}|$. Denote by $\mathcal{D}(f)$ the support of $\delta^j f$, i.e. all $(j+1)$-simplices in Δ such that for some ordering $[v_0, \ldots, v_j]$ (and thus for all orderings) it holds that $\sum_{i=0}^{j}(-1)^i f[v_0, \ldots, \hat{v}_i, \ldots, v_j] \neq 0_R$. Then

$$|\mathcal{D}(f)| \geqslant \frac{n}{j+2}|\mathcal{S}|.$$

The following lemma shows that whp in the supercritical case a smallest support in $\mathcal{N}_\tau$ cannot be 'large'.

Lemma 9.8. There exists a positive constant $\tilde{h} \in \mathbb{R}^+$ such that whp for all $\tau \geqslant \tau_0 := 1 - \frac{1}{(\log n)^{1/3}}$ we have $|\mathcal{S}_\tau| < \tilde{h}$ (if $\mathcal{S}_\tau$ exists).

We will prove this lemma with the help of several auxiliary results (Claims 9.9 to 9.13). To give a better overall picture of the proof, and since the claims are mostly either elementary or technical in nature, we will first complete the proof of Lemma 9.8 with the help of these claims and postpone their proofs to Appendices A.8 to A.12.

Proof of Lemma 9.8. We first note that, if τ is large enough to imply that there exists a dimension $k \in \{j+1, \ldots, d\}$ with $p_k = \tau\bar{p}_k \geqslant 1$, then the result is trivial: all k-simplices are present and there is no bad function, so $\mathcal{S}_\tau$ does not exist. We will therefore assume for the remainder of the proof that $\tau_0 \leqslant \tau \leqslant \min_{k\in\{j+1,\ldots,d\}} 1/\bar{p}_k$, and in particular $p_k \leqslant 1$ for each $k \geqslant j+1$.

By Lemma 9.4 if $\mathcal{S}_\tau$ exists it is traversable. Consider a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ where $\mathcal{S}$ is a traversable support of size s, and thus can be found via the search process described in Remark 9.5. For such a pair, we define the *exploration matrix* $B = (b_{i,k})$ for $i \in [s]$ and $k \in \{j+1, \ldots, d\}$, where $b_{i,k} \geqslant 0$ is the number of k-simplices of $\mathcal{T}(\mathcal{S})$ we discover from the i-th j-simplex of $\mathcal{S}$ in the search process.

For a fixed exploration matrix B, we can bound the number of ways the exploration process can proceed, thus bounding the number of pairs $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with exploration matrix B.

Claim 9.9. The number of pairs $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ in which $\mathcal{S}$ is traversable and has exploration matrix B is at most

$$n^{j+1}\frac{\prod_k \left(\binom{n}{k-j}2^{\binom{k+1}{j+1}}\right)^{t_k}}{\prod_{i,k} b_{i,k}!}.$$

Claim 9.9 provides an upper bound for the number of sets that are candidates for $\mathcal{S}_\tau$. Our next aim is to bound the probability that a fixed function f with traversable support $\mathcal{S}$ turns out to be f_τ. To this end, we would like to know that for $k \in \{j+1, \ldots, d\}$, the number of $(k+1)$-sets that are not allowed to be k-simplices in $\mathcal{G}_\tau$ in order for f_τ to be a j-cocycle is 'large'. We will prove this by applying Proposition 9.7 to f_τ (if it exists).

In order to see that f_τ (if it exists) satisfies the hypothesis of Proposition 9.7, we will need the following auxiliary result, which we prove by a simple first moment argument.

Claim 9.10. Whp every $\mathcal{G}_\tau$ with $\tau \geqslant \tau_0$ has full $(j-1)$-skeleton.

For the rest of the proof, we condition on the high probability event in Claim 9.10. This means that, with Δ as in Proposition 9.7, we have $C^{j-1}(\Delta; R) = C^{j-1}(\mathcal{G}_\tau; R)$. Now let g be a $(j-1)$-cochain in Δ (and thus also in $\mathcal{G}_\tau$). Consider $f_\tau + \delta^{j-1} g$ as a j-cochain f in $\mathcal{G}_\tau$ and as a j-cochain f_Δ in Δ. Clearly, the support of f_Δ contains the support of f, which in turn has size at least $|\mathcal{S}_\tau|$, because $\mathcal{S}_\tau$ is minimal. In particular, f_τ satisfies the hypothesis of Proposition 9.7. This enables us to prove that there are 'many' $(k+1)$-sets that are not allowed to be k-simplices in $\mathcal{G}_\tau$ in order for f_τ to be a j-cocycle.

Claim 9.11. For $j+1 \leqslant k \leqslant d$ and a j-cocycle f, denote by $\mathcal{D}_k(f)$ the set of $(k+1)$-sets in $[n]$ that contain at least one element of $\mathcal{D}(f)$. There exists a positive constant h_0 such that, if f_τ exists, then for all k with $j+1 \leqslant k \leqslant d$,

$$|\mathcal{D}_k(f_\tau)| \geqslant h_0 |\mathcal{S}_\tau| n^{k-j}.$$

The elements in $\mathcal{D}_k(f_\tau)$ are not allowed to be k-simplices in $\mathcal{G}_\tau$ and thus all must have birth time larger than τ. We can use this fact, together with the upper bound for the number of 'candidates' for $\mathcal{S}_\tau$, to prove that $\mathcal{S}_\tau$ existing and having a fixed exploration matrix is unlikely.

Claim 9.12. There exist positive constants $h_1, \tilde{h}_1$ such that the following holds for every $s \geqslant \tilde{h}_1$. For every $(s \times (d-j))$-matrix B, the probability r_B that f_τ exists and $\mathcal{S}_\tau$ has size s and exploration matrix B satisfies

$$r_B \leqslant \frac{n^{-h_1 s}}{\prod_{i,k} b_{i,k}!}.$$

We determine a lower bound for the denominator $\prod_{i,k} b_{i,k}!$ and sum over all possible exploration matrices B, so as to deduce that whp $\mathcal{S}_\tau$ cannot be larger than a given large constant.

Claim 9.13. There exists a positive constant h_2 such that for every $\tilde{h} \geqslant \tilde{h}_1$, the probability that $\mathcal{S}_\tau$ exists with $|\mathcal{S}_\tau| \geqslant \tilde{h}$ is at most

$$n^{-h_2 \tilde{h}/2}.$$

Claim 9.13 implies that for *a fixed time* $\tau \geqslant \tau_0$, whp $|\mathcal{S}_\tau| < \tilde{h}$ (if it exists). Our aim, however, is to prove that whp $|\mathcal{S}_\tau| < \tilde{h}$ *simultaneously for all* $\tau \geqslant \tau_0$. To this end, observe that there are $O(n^{d+1})$ simplices σ with scaled birth times $\tau_\sigma \geqslant \tau_0$. Taking a union bound over all these scaled birth times, we deduce from Claim 9.13 that

$$\mathbb{P}(\exists \tau \geqslant \tau_0 \text{ such that } \mathcal{S}_\tau \text{ exists and } |\mathcal{S}_\tau| \geqslant \tilde{h}) = O(n^{d+1-h_2\tilde{h}/2}).$$

Thus, Lemma 9.8 holds for $\tilde{h} > \max\left\{\tilde{h}_1, \frac{2(d+1)}{h_2}\right\}$. □

Lemma 9.8 implies that whp traversable supports of j-cocycles of 'large' size do not exist in the whole supercritical range. For supports of constant size, this is given by Lemma 9.6 only for $\tau = 1 + o(1)$. We therefore derive the following result, stating that for τ 'close' to 1 whp every j-cocycle is generated by copies of $M_{j,k}$.

Corollary 9.14. For every $\tau = 1 + O\left(\frac{1}{\log n}\right)$, we have $\mathcal{N}_\tau = \emptyset$ whp.

Proof. For every such τ Lemma 9.8 holds and thus there exists a constant $\tilde{h}$ such that either $\mathcal{S}_\tau$ does not exist or $|\mathcal{S}_\tau| \leqslant \tilde{h}$. But by Lemma 9.6 (with $h = \tilde{h}$), whp if $\mathcal{S}_\tau$ exists then $|\mathcal{S}_\tau| > \tilde{h}$. Thus whp $\mathcal{S}_\tau$ does not exist, i.e. $\mathcal{N}_\tau = \emptyset$. $\square$

To exclude the existence of 'small' supports throughout the entire supercritical case, we show that if a new obstruction appears, then the simplex whose addition to the complex creates the obstruction in fact forms a local j-obstacle, which whp does not exist in this range by Lemma 7.3.

Lemma 9.15. Let K be the simplex with smallest scaled birth time $\tau_K \geqslant \tau_j^*$ such that $\mathcal{N}_{\tau_K} \neq \emptyset$ (if it exists). Then whp K forms a local j-obstacle in $\mathcal{G}_{\tau_K}$.

Proof. First observe that by Lemma 7.1 and Corollary 9.14, whp $\mathcal{N}_{\tau_j^*} = \emptyset$, and thus whp $\tau_K > \tau_j^*$. For the rest of this proof, we condition on this high probability event.

Suppose now that $|K| = k+1$ and let $\tau \geqslant \tau_j^*$ be such that $\mathcal{G}_{\tau_K} = \mathcal{G}_\tau + K$. If $\mathcal{S}_{\tau_K} \cap \mathcal{G}_\tau \neq \emptyset$, let $\mathcal{S}$ be a maximal subset of $\mathcal{S}_{\tau_K}$ which is traversable in $\mathcal{G}_\tau$ and let f be the j-cochain in $\mathcal{G}_\tau$ defined by

$$f(\sigma) = \begin{cases} f_{\tau_K}(\sigma) & \text{if } \sigma \in \mathcal{S}; \\ 0_R & \text{otherwise.} \end{cases}$$

Then f is a j-cocycle in $\mathcal{G}_\tau$ because every i-simplex of $\mathcal{G}_\tau$, for $i = j+1, \ldots, d$, containing some element of $\mathcal{S}$ cannot contain other j-simplices in $\mathcal{S}_{\tau_k} \setminus \mathcal{S}$ by the maximality of $\mathcal{S}$, and because f_{τ_K} is a j-cocycle.

Moreover, by Lemma 9.8 there exists a positive constant $\tilde{h}$ such that whp $|\mathcal{S}| \leqslant |\mathcal{S}_{\tau_K}| < \tilde{h}$. Lemma 6.3 implies that whp each j-simplex of $\mathcal{S}$ lies in a linear number of j-shells in $\mathcal{G}_\tau$ and at most $|\mathcal{S}| - 1$ many of them can contain other elements of $\mathcal{S}$. This means that whp there are j-shells in $\mathcal{G}_\tau$ that meet $\mathcal{S}$ in a single j-simplex, and thus f is a bad function in $\mathcal{G}_\tau$. Since f cannot be generated by copies of $M_{j,k}$, because $\tau \geqslant \tau_j^*$ and thus no copies of $M_{j,k}$ exist, this yields $\mathcal{N}_\tau \neq \emptyset$, a contradiction to the choice of K.

Hence, whp the j-simplices of $\mathcal{S}_\tau$ are all contained in K and are not in other simplices of $\mathcal{G}_{\tau_K}$. Then whp K forms a local j-obstacle in $\mathcal{G}_{\tau_K}$, because $|\mathcal{S}_{\tau_K}| \geqslant k - j + 1$ by Lemma 5.11. $\square$

Corollary 9.16. Whp, for all $\tau \geqslant \tau_j^*$ we have $\mathcal{N}_\tau = \emptyset$.

Proof. By Lemma 7.1 and Corollary 9.14 whp $\mathcal{N}_{\tau_j^*}$ is empty. If there is $\tau > \tau_j^*$ such that $\mathcal{N}_\tau \neq \emptyset$, then Lemma 9.15 tells us that the simplex whose birth creates a j-cocycle that is not generated by copies of $M_{j,k}$ would create a local j-obstacle. But by Lemma 7.1 whp $\tau > \tau_j^* \geqslant \tau' = 1 - \frac{\log \log n}{10 d \log n}$, thus by Lemma 7.3 whp no new local j-obstacle can appear in $\mathcal{G}_\tau$. $\square$

We can now use Corollary 9.16 to prove Lemma 4.7

Proof of Lemma 4.7. By the definition of τ_j^*, there are no copies of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$ for any $\tau \geqslant \tau_j^*$, so in order for the j-th cohomology group *not* to vanish, $\mathcal{N}_\tau$ would have to be non-empty, but this is excluded by Corollary 9.16. $\square$

10 Rank in the critical window: proofs of Theorem 1.5 and Corollary 1.6

In order to prove the Rank Theorem (Theorem 1.5), we first want to describe the asymptotic joint distribution of the number of copies of $M_{j,k}$ within the critical window. To this end, we will make use of Lemma 10.1, for which we need the following notation. Given a sequence $(X_k)_{k\in I}$ of random variables (for some finite, ordered index set I) we denote by $\mathcal{L}(X_k)$ the probability distribution of X_k and by $\mathcal{L}\left((X_k)_{k\in I}\right)$ the joint probability distribution of the sequence $(X_k)_{k\in I}$. If $X_k = X_k(n)$ for every $k \in I$, we say that the sequence $(X_k)_{k\in I}$ *converges in distribution* to the sequence of random variables $(Y_k)_{k\in I}$ if $\mathbb{P}\big((X_k)_{k\in I} = (x_k)_{k\in I}\big) \xrightarrow{n\to\infty} \mathbb{P}\big((Y_k)_{k\in I} = (x_k)_{k\in I}\big)$ for every sequence of values $(x_k)_{k\in I}$, and we write $(X_k)_{k\in I} \xrightarrow{\mathrm{d}} (Y_k)_{k\in I}$.

We adopt the convention that $\mathrm{Po}(0) \equiv 0$ and recall the definition of a critical dimension from Definition 3.6. The following lemma provides the main ingredient to prove Theorem 1.5.

Lemma 10.1. Let $c \in \mathbb{R}$ be a constant and $(c_n)_{n\geqslant 1}$ be a sequence of real numbers such that $c_n \xrightarrow{n\to\infty} c$. For any $j \leqslant k \leqslant d$ define

$$\mathcal{E}_k := \begin{cases} \exp(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)) & \text{if } k \text{ is a critical dimension,} \\ 0 & \text{otherwise,} \end{cases}$$

and let $\tau = 1 + \frac{c_n}{\log n}$. Then, setting $\mathbf{X} := (X_{j,j}, X_{j,j+1}, \ldots, X_{j,d})$, we have

$$\mathcal{L}(\mathbf{X}) \xrightarrow{\mathrm{d}} (\mathrm{Po}(\mathcal{E}_j), \ldots, \mathrm{Po}(\mathcal{E}_d)).$$

Before proving Lemma 10.1, let us show how it implies the Rank Theorem (Theorem 1.5).

Proof of Theorem 1.5. Consider $\mathcal{X} = \sum_{k=j}^{d} X_{j,k}$ and define

$$\mathcal{E} := \exp(-c(j+1)) \sum_{k\in\mathcal{C}} \exp(\bar{\mu}_k + \bar{\nu}_k + c\bar{\gamma}_k),$$

where $\mathcal{C} = \mathcal{C}(\bar{\mathbf{p}}, j)$ is the set of critical dimensions for the j-critical direction $\bar{\mathbf{p}}$. By Lemma 10.1, we have

$$\mathcal{L}(\mathcal{X}) \xrightarrow{\mathrm{d}} \sum_{k=j}^{d} \mathrm{Po}(\mathcal{E}_k) = \mathrm{Po}\left(\sum_{k\in\mathcal{C}} \exp\left(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)\right)\right) = \mathrm{Po}(\mathcal{E}). \tag{39}$$

We first show that $H^j(\mathcal{G}_\tau; R) = R^{\mathcal{X}}$ whp. Let $M_1, M_2, \ldots, M_{\mathcal{X}}$ denote the copies of $M_{j,k}$, for every $j \leqslant k \leqslant d$ that are present in $\mathcal{G}_\tau$. By Corollary 9.14 we have $\mathcal{N}_\tau = \emptyset$ whp and by Proposition 5.9 (a) we know that the only j-cocycles arising from M_i are of the form f_{M_i,r_i} with $r_i \in R$. Thus whp each cohomology class contains an element of the form $\sum_{i=1}^{\mathcal{X}} f_{M_i,r_i}$ with $r_i \in R$, i.e. whp the set of cohomology classes of those elements generates $H^j(\mathcal{G}_\tau; R)$.

We now need to show that if we take two tuples $(r_1, \ldots, r_{\mathcal{X}}) \neq (r'_1, \ldots, r'_{\mathcal{X}})$ with $r_i, r'_i \in R$ for every i, then the cohomology classes of $\sum_{i=1}^{\mathcal{X}} f_{M_i,r_i}$ and of $\sum_{i=1}^{\mathcal{X}} f_{M_i,r'_i}$ are distinct. Note that this is equivalent to showing that if $(r_1, \ldots, r_{\mathcal{X}})$ is not the 0_R-vector, then $f = \sum_{i=1}^{\mathcal{X}} f_{M_i,r_i}$ is not in the same cohomology class as the 0_R-function, i.e. the j-cochain f is not a j-coboundary.

We first observe that by Markov's inequality $\mathcal{X} = \sum_{k=j}^{d} X_{j,k} = o(n)$, because $\mathbb{E}(\mathcal{X}) = O(1)$ by Corollary 6.6. We further claim that for each $j \leqslant k \leqslant d$ whp no two copies of $M_{j,k}$ share the same k-simplex. Indeed, for $k = j$ by Definition 5.6 all copies of $M_{j,j}$ come from different j-simplices. If $k \geqslant j+1$, for two copies of $M_{j,k}$ sharing the same k-simplex there are $\binom{n}{k+1}$ ways to choose the common k-simplex and $O\left(\binom{k+1}{j}^2\right)$ ways to choose the centres of the two flowers. Moreover, these two copies are present in $\mathcal{G}_\tau$ with probability $O(p_k \bar{q}^{\,\tau(2k-2j+1)})$, because the common $(k+1)$-set is a k-simplex in $\mathcal{G}_\tau$ with probability p_k and the two flowers can share at most one petal, thus in total there are at least $(k-j+1)+(k-j) = 2k-2j+1$ many j-simplices that are petals, and these satisfy (M2) with probability at most $(1+o(1))\bar{q}^{\,\tau(2k-2j+1)}$ by Proposition 5.13.

Therefore, for $k \geqslant j+1$, the expected number of pairs of copies of $M_{j,k}$ with the same k-simplex is

$$O\left(\binom{n}{k+1}\binom{k+1}{j}^2 p_k \bar{q}^{\,\tau(2k-2j+1)}\right) \overset{(16)}{=} O\left(\mathbb{E}(X_{j,k})\bar{q}^{\,\tau(k-j)}\right) = O(\bar{q}^{\,\tau(k-j)}),$$

because $\mathbb{E}(X_{j,k}) = O(1)$. Furthermore, we have that

$$\bar{q} \overset{(15)}{=} O\left(\exp\left(-\sum_{i=j+1}^{d}\left(\bar{\alpha}_i \log n + \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}}\right)\right)\right) = O(n^{-\bar{\alpha}_{k_0}/2}) = o(1), \tag{40}$$

where we are using that k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$ (see (A1) and (A4) in Definition 3.1). Since $k \geqslant j+1$ and $\tau = 1+o(1)$ we also have that $\bar{q}^{\,\tau(k-j)} = o(1)$, and thus by Markov's inequality, whp there exist no such pairs of $M_{j,k}$.

Hence, by condition (M2) in Definition 5.5, whp the f_{M_i,r_i} have pairwise disjoint supports, and in particular, for our choice of the r_i, the support S of f is not empty. Pick a j-simplex $L \in S$. Lemma 6.3 yields that in the range of τ we are considering, whp L is contained in $\Theta(n)$ many j-shells which meet only in L, and therefore at most $|S| \leqslant \sum_{k=j}^{d}(k-j+1)X_{j,k} \leqslant (d-j+1)\mathcal{X} = o(n)$ of them can contain another j-simplex in S. Thus whp there exists a j-shell that meets the support of f only in L, i.e. f is not a j-coboundary by Lemma 2.2.

We therefore have $H^j(\mathcal{G}_\tau; R) = R^{\mathcal{X}}$ whp. Since $\mathcal{L}(\mathcal{X}) \xrightarrow{\mathrm{d}} \mathrm{Po}(\mathcal{E})$ by (39), there exists a coupling $Y \sim \mathrm{Po}(\mathcal{E})$ such that $\mathcal{X} = Y$ whp. Thus, whp

$$H^j(\mathcal{G}_\tau; R) = R^Y,$$

as required. $\square$

We now combine Theorems 1.4 and 1.5 to prove Corollary 1.6

Proof of Corollary 1.6. As in the proof of Theorem 1.4, by applying Lemma 4.7 and Corollary 4.6 (b) we deduce that for $\tau = 1 + \frac{c_n}{\log n}$, whp $H^i(\mathcal{G}_\tau; R) = 0$ for every $i \in [j-1]$. Furthermore, whp $H^0(\mathcal{G}_\tau; R) = R$ by Lemma 4.2. This in particular implies that

$$\mathbb{P}\left(\mathcal{G}_\tau \text{ is } j\text{-cohom-connected}\right) = \mathbb{P}\left(H^j(\mathcal{G}_\tau; R) = 0\right) + o(1).$$

Moreover, by Theorem 1.5 whp $H^j(\mathcal{G}_\tau; R) = R^Y$ with $Y \sim \mathrm{Po}(\mathcal{E})$, hence

$$\mathbb{P}\left(\mathcal{G}_\tau \text{ is } j\text{-cohom-connected}\right) = (1+o(1))\mathbb{P}\big(\mathrm{Po}(\mathcal{E}) = 0\big) = (1+o(1))\exp\left(-\mathcal{E}\right),$$

as required. □

All that is left to complete the proof of Theorem 1.5 is to prove Lemma 10.1. To do this, we use the Poisson approximation method, specifically a result from [5] that requires the following notion. Given a discrete set H, the *total variation distance* between the distributions of two H-valued random variables Y and Z is defined by

$$d_{TV}\left(\mathcal{L}(Y), \mathcal{L}(Z)\right) := \frac{1}{2}\sum\nolimits_{h\in H} |\mathbb{P}(Y=h) - \mathbb{P}(Z=h)|.$$

Lemma 10.2 ([5, Theorem 10.J]). Given a set Γ with a partition $\Gamma = \dot\cup_{k=1}^{r}\Gamma_k$ and a collection $(I_a)_{a\in\Gamma}$ of indicator random variables defined on a common probability space, let

- $\pi_a := \mathbb{P}(I_a = 1)$, for every $a \in \Gamma$;
- $W_k := \sum_{a\in\Gamma_k} I_a$, for $k \in [r]$;
- $\mathbf{W} := (W_1, \ldots, W_r)$;
- $m_k := \mathbb{E}(W_k) = \sum_{a\in\Gamma_k} \pi_a$, for $k \in [r]$;
- $\mathbf{m} := (m_1, \ldots, m_r)$.

Suppose that for each $a \in \Gamma$ there exist random variables $(J_{ba})_{b\in\Gamma}$ defined on the same probability space as $(I_b)_{b\in\Gamma}$ with

$$\mathcal{L}\left((J_{ba})_{b\in\Gamma}\right) = \mathcal{L}\left((I_b)_{b\in\Gamma}|I_a = 1\right).$$

Then

$$d_{TV}\Big(\mathcal{L}\big(\mathbf{W}\big), \mathrm{Po}(\mathbf{m})\Big) \leqslant \sum\nolimits_{a\in\Gamma} \pi_a \left(\pi_a + \sum\nolimits_{b\neq a} \mathbb{E}|J_{ba} - I_b|\right),$$

where $\mathrm{Po}(\mathbf{m})$ denotes the joint Poisson distribution $(\mathrm{Po}(m_1)\ldots, \mathrm{Po}(m_r))$.

It is easy to see that if there exists $\tilde{\mathbf{m}} := (\tilde{m}_1, \ldots, \tilde{m}_r)$ such that for every $k \in [r]$, $\tilde{m}_k \in \mathbb{R}$ and $m_k = m_k(n) \xrightarrow{n\to\infty} \tilde{m}_k$, then

$$\mathcal{L}\big(\mathbf{W}\big) \xrightarrow{\mathrm{d}} \mathrm{Po}(\tilde{\mathbf{m}})$$

if and only if

$$d_{TV}\left(\mathcal{L}\big(\mathbf{W}\big), \mathrm{Po}(\mathbf{m})\right) \xrightarrow{n\to\infty} 0.$$

Proof of Lemma 10.1. We will first show that we can apply Lemma 10.2 with $W_k = X_{j,k}$ and $m_k = \mathbb{E}(X_{j,k})$ for $k = j, \ldots, d$. Subsequently, we show the bound on the total variation distance is indeed $o(1)$ and that $\lim_{n\to\infty}\mathbb{E}(X_{j,k}) = \mathcal{E}_k$.

We want to define the set Γ of potential copies of $M_{j,k}$ in $\mathcal{G}_\tau$ for each $j \leqslant k \leqslant d$. As in the proof of Lemma 6.7, we consider the sets

$$\mathcal{T}_k = \left\{(K, C) : K \in \binom{[n]}{k+1}, C \in \binom{K}{j}\right\},$$

for each $j + 1 \leqslant k \leqslant d$. Furthermore we define the set $\mathcal{T}_j$ analogously but with the additional condition that given a $(j+1)$-set K, then the set C consists of the first j vertices of K according to the increasing order of $[n]$ (cf. Definition 5.6). Following the notation of Lemma 10.2, we set $\Gamma := \dot{\bigcup}_{k=j}^{d} \mathcal{T}_k$ and we use $a = (K_a, C_a)$ to denote an element of Γ.

For any $a \in \mathcal{T}_k \subseteq \Gamma$, we define the following quantities:

- $k_a := k$;
- I_a is the indicator random variable of the event that a forms a copy of $M_{j,k}$;
- $\pi_a := \mathbb{P}(I_a = 1) = \mathbb{E}(I_a)$;
- $\mathcal{B}_a$ is the collection of *forbidden sets* for a, i.e.

$$\mathcal{B}_a = \{B \subset [n] : |B| \leqslant d + 1, B \not\subseteq K_a, B \supset P \text{ for some } P \in \mathcal{F}(K_a, C_a)\},$$

  where $\mathcal{F}(K_a, C_a) = \{C_a \cup \{w\} \mid w \in K_a \setminus C_a\}$ is the j-flower in K_a with centre C_a (see Definition 5.2 and (13)). In other words, $\mathcal{B}_a$ is the collection of subsets of $[n]$ that are not allowed to be simplices in $\mathcal{G}_\tau$ in order for a to form a copy of $M_{j,k}$ (cf. (M2) in Definition 5.5).

Observe that if $a = (K_a, C_a) \in \mathcal{T}_j$, then $\mathcal{F}(K_a, C_a) = \{K_a\}$, therefore the set $\mathcal{B}_a$ consists of all subsets of $[n]$ (of cardinality at most $d + 1$) that contain K_a, except for K_a itself.

Given a family $\mathcal{D}$ of sets of vertices, we say that the indicator random variable of an event E *depends only on* $\mathcal{D}$ if E only depends on whether the sets in $\mathcal{D}$ are simplices in $\mathcal{G}_\tau$ or not. Observe that by Definitions 5.5 and 5.6 we can write

$$I_a = \mathbb{1}\left\{\{K_a \in \mathcal{G}_\tau\} \wedge \{B \notin \mathcal{G}_\tau, \forall B \in \mathcal{B}_a\}\right\},$$

therefore the random variable I_a depends only on the family of sets

$$\mathcal{D}_a := \{K_a\} \cup \mathcal{B}_a.$$

We now aim to define the random variables $(J_{ba})_{b\in\Gamma}$ needed to apply Lemma 10.2. Given $a, b \in \Gamma$, we define the events

- $E^1_{ba} = \{K_b \in \mathcal{G}_\tau\} \vee \{K_b = K_a\}$,
- $E^2_{ba} = \{B \notin \mathcal{G}_\tau, \forall B \in \mathcal{B}_b \setminus \mathcal{B}_a\}$,
- $E^3_{ba} = \{K_b \notin \mathcal{B}_a\} \wedge \{K_a \notin \mathcal{B}_b\}$,

and the indicator random variable

$$J_{ba} = \mathbb{1}\ \{E^1_{ba} \wedge E^2_{ba} \wedge E^3_{ba}\}. \tag{41}$$

We claim that

$$\mathcal{L}\left((J_{ba})_{b\in\Gamma}\right) = \mathcal{L}\left((J_{ba})_{b\in\Gamma}|I_a = 1\right). \tag{42}$$

To see this, let $\mathcal{D}_{ba}$ be the family of sets of vertices which J_{ba} depends only on. If $K_b \in \mathcal{B}_a$ or $K_a \in \mathcal{B}_b$, by (41) and the definition of E^3_{ba} we have $J_{ba} = 0$ deterministically and we set $\mathcal{D}_{ba} := \emptyset$. Otherwise, by (41) we have

$$\mathcal{D}_{ba} := (\{K_b\} \setminus \{K_a\}) \cup (\mathcal{B}_b \setminus \mathcal{B}_a)$$

and if $K_b = K_a$ then K_a is not in $\mathcal{B}_b$ because the event E^3_{ba} holds, hence we have $\mathcal{D}_{ba} = \mathcal{D}_b \setminus \mathcal{D}_a$. In particular, this implies that $\mathcal{D}_{ba}$ and $\mathcal{D}_a$ are always disjoint and this holds for every $b \in \Gamma$, thus the joint distribution of $(J_{ba})_{b\in\Gamma}$ does not change if we condition on $I_a = 1$, yielding (42).

We further claim that for every $b \in \Gamma$

$$\big((I_a = 1) \wedge (J_{ba} = 1)\big) \quad \Longleftrightarrow \quad \big((I_a = 1) \wedge (I_b = 1)\big). \tag{43}$$

Suppose $I_a = J_{ba} = 1$. We have $K_b \in \mathcal{G}_\tau$ by E^1_{ba} and the fact that $K_a \in \mathcal{G}_\tau$, since $I_a = 1$. Moreover, $I_a = 1$ yields that none of the sets in $\mathcal{B}_a$ is in $\mathcal{G}_\tau$ and by definition of E^2_{ba} also none of the sets in $\mathcal{B}_b \setminus \mathcal{B}_a$ is in $\mathcal{G}_\tau$, therefore in particular every set in $\mathcal{B}_b$ is not in $\mathcal{G}_\tau$. Thus, by definition of I_b we have that $I_b = 1$.

Vice versa, suppose that $I_a = I_b = 1$. By definition of I_b, clearly the events E^1_{ba} and E^2_{ba} hold. Moreover, I_a and I_b can only both be equal to 1 if K_b is not forbidden for a and K_a is not forbidden for b, i.e. the event E^3_{ba} must hold. Thus, it follows that $J_{ba} = 1$. This proves (43).

Hence, conditioned on $I_a = 1$, for every $b \in \Gamma$ (43) yields that J_{ba} and I_b are the same random variable, and thus in particular

$$\mathcal{L}\big((J_{ba})_{b\in\Gamma}|I_a = 1\big) = \mathcal{L}\big((I_b)_{b\in\Gamma}|I_a = 1\big). \tag{44}$$

In total, we have

$$\mathcal{L}\big((J_{ba})_{b\in\Gamma}\big) \overset{(42)}{=} \mathcal{L}\big((J_{ba})_{b\in\Gamma}|I_a = 1\big) \overset{(44)}{=} \mathcal{L}\big((I_b)_{b\in\Gamma}|I_a = 1\big). \tag{45}$$

Since $X_{j,k} = \sum_{a\in\mathcal{T}_k} I_a$ for any $j \leqslant k \leqslant d$, we can therefore apply Lemma 10.2. Setting $Z_k := \mathrm{Po}(\mathbb{E}(X_{j,k}))$ independently for each k, we obtain

$$d_{TV}\left(\mathcal{L}(\mathbf{X}), \big(Z_j, \dots, Z_d\big)\right) \leqslant \sum_{a\in\Gamma} \pi_a \left(\pi_a + \sum_{b\neq a} \mathbb{E}|J_{ba} - I_b|\right). \tag{46}$$

We want to show that the right-hand side of (46) is $o(1)$. Recall that for every $b \in \Gamma$ by (14) and Proposition 5.13 we have

$$\mathbb{E}(I_b) = \pi_b = (1 + o(1))p_{k_b}\bar{q}^{\,\tau(k_b-j+1)}, \tag{47}$$

and therefore (cf. (16))

$$\mathbb{E}(X_{j,k}) = \sum\nolimits_{b \in \mathcal{T}_k} \mathbb{E}(I_b) = \Theta(n^{k+1} p_k \bar{q}^{\,\tau(k-j+1)}) = O(1), \tag{48}$$

where the last equality holds because we are considering $\mathcal{G}_\tau$ within the critical window. Furthermore, by (45) we have

$$\mathbb{E}(J_{ba}) = \mathbb{P}(I_b = 1 | I_a = 1). \tag{49}$$

We now fix $a \in \Gamma$ and estimate the sum $\sum_{b \neq a} \mathbb{E}|J_{ba} - I_a|$, by distinguishing some cases.

Case 1: $K_b = K_a$. First observe that since $b \neq a$, this case can only be possible if $k_b = k_a \geqslant j+1$ and $C_b \neq C_a$. Moreover, conditioned on $I_a = 1$, i.e. a forming a copy of M_{j,k_a}, there are $\binom{k_a+1}{j} - 1 = O(1)$ ways to choose b such that $K_b = K_a$ and $C_b \neq C_a$. Furthermore, b forms a copy of $M_{j,k_b} = M_{j,k_a}$ with probability $O(\bar{q}^{\,\tau(k_a-j)})$, because $K_b = K_a$ already exists in $\mathcal{G}_\tau$ as simplex (and so there is no $p_{k_b} = p_{k_a}$ term) and because the flower $\mathcal{F}(K_b, C_b)$ can share at most one petal with the flower $\mathcal{F}(K_a, C_a)$ (and so we lose at most one factor $\bar{q}^{\,\tau}$). Thus if we set

$$B_1 = B_1(a) \ := \ \{b \in \Gamma : b \neq a, K_b = K_a\}$$

we have

$$\begin{aligned}
\sum\nolimits_{b \in B_1} \mathbb{E}|J_{ba} - I_b| \ &\leqslant \ \sum\nolimits_{b \in B_1} \big(\mathbb{E}(J_{ba}) + \mathbb{E}(I_b)\big) \\
&\overset{(49)}{=} \sum\nolimits_{b \in B_1} \big(\mathbb{P}(I_b = 1 | I_a = 1) + \pi_b\big) \\
&\overset{(47)}{=} O\left(1\right) \cdot \big(O(\bar{q}^{\,\tau(k_a-j)}) + O(p_{k_a} \bar{q}^{\,\tau(k_a-j+1)})\big) \\
&= O(\bar{q}^{\,\tau(k_a-j)}) = o(1),
\end{aligned}$$

where the last equality follows from the facts that $\tau(k_a - j) > 1 + o(1)$ and $\bar{q} = o(1)$ (cf. (40)).

Case 2: $K_b \neq K_a$, but $K_b \in \mathcal{B}_a$ or $K_a \in \mathcal{B}_b$. This means that the event E^3_{ba} does not happen, thus $J_{ba} = 0$ deterministically by (41).

Case 2.1: $K_b \in \mathcal{B}_a$. Given K_a, the set K_b must contain at least $j+1$ vertices of K_a in order to be forbidden for K_a, because K_b contains at least one petal (i.e. $(j+1)$-set) of the flower $\mathcal{F}(K_a, C_a)$. Hence, there are $O(n^{k_b - j})$ possible choices for b, and thus if we set

$$B_{2.1} = B_{2.1}(a) \ := \ \{b \in \Gamma : b \neq a, K_b \neq K_a, K_b \in \mathcal{B}_a\},$$

we have

$$\begin{aligned}
\sum\nolimits_{b \in B_{2.1}} \mathbb{E}|J_{ba} - I_b| \ &\leqslant \ \sum\nolimits_{b \in B_{2.1}} \mathbb{E}(I_b) \\
&\overset{(47)}{=} \sum\nolimits_{k=j}^{d} O\left(n^{k-j} p_k \bar{q}^{\,\tau(k-j+1)}\right) \\
&\overset{(48)}{=} \sum\nolimits_{k=j}^{d} O\left(\frac{\mathbb{E}(X_{j,k})}{n^{j+1}}\right) = o(1).
\end{aligned}$$

Case 2.2: $K_a \in \mathcal{B}_b$. Set

$$B_{2.2} = B_{2.2}(a) \;:=\; \{b \in \Gamma : b \neq a, K_b \neq K_a, K_a \in \mathcal{B}_b\}.$$

By exchanging the roles of K_a and K_b in Case 2.1, with the same argument we have

$$\sum\nolimits_{b \in B_{2.2}} \mathbb{E}|J_{ba} - I_b| = o(1).$$

Case 3: $K_b \neq K_a$, $K_b \notin \mathcal{B}_a$, and $K_a \notin \mathcal{B}_b$. This case contains almost all the summands of $\sum_{b \neq a} \mathbb{E}|J_{ba} - I_a|$, thus we need the main terms in the sum to cancel. The event E^3_{ba} holds, yielding that if $I_b = 1$ then also $J_{ba} = 1$, that is $J_{ba} \geqslant I_b$ deterministically and therefore

$$\mathbb{E}|J_{ba} - I_b| = \mathbb{E}(J_{ba}) - \mathbb{E}(I_b). \tag{50}$$

There are $(k_b - j + 1)$ (potential) petals in b each contained in $\binom{n-j-1}{k-j}$ many $(k+1)$-sets that must not form k-simplices in $\mathcal{G}_\tau$ in order for b to form a copy of $M_{j,k}$, for each $j+1 \leqslant k \leqslant d$. However some of these forbidden $(k+1)$-sets might be double-counted because they contain more than one petal in b, and additionally some of these forbidden $(k+1)$-sets might be forbidden for both a and b, and therefore we already know that they are not simplices if we condition on $I_a = 1$. In either case, any of these $(k+1)$-sets contains at least two petals and so at least $j+2$ vertices are already fixed, thus there are $O\left(\binom{n-(j+2)}{k+1-(j+2)}\right) = O(n^{k-j-1})$ many $(k+1)$-sets that we have to exclude when counting the sets of size $k+1$ that are forbidden for b. In other words, the number of $(k+1)$-sets that must not be simplices is $(k_b - j + 1)\binom{n-j-1}{k-j} - O(n^{k-j-1})$, yielding

$$\begin{aligned}
\mathbb{E}(J_{ba}) &\overset{(49)}{=} \mathbb{P}(I_b = 1 | I_a = 1) \\
&= p_{k_b} \prod\nolimits_{k=j+1}^{d} (1 - \tau \bar{p}_k)^{(k_b - j+1)\binom{n-j-1}{k-j} - O(n^{k-j-1})} \\
&\overset{(14)}{=} p_{k_b} \bar{q}^{\,\tau(k_b - j+1)} \prod\nolimits_{k=j+1}^{d} (1 - \tau \bar{p}_k)^{-O(n^{k-j-1})} \\
&= p_{k_b} \bar{q}^{\,\tau(k_b - j+1)} \exp\left(\sum\nolimits_{k=j+1}^{d} O\left(\frac{\log n}{n^{k-j}} n^{k-j-1} \right) \right) \\
&= (1 + o(1)) p_{k_b} \bar{q}^{\,\tau(k_b - j+1)},
\end{aligned}$$

and thus

$$\mathbb{E}|J_{ba} - I_b| \overset{(50)}{=} \mathbb{E}(J_{ba}) - \mathbb{E}(I_b) = \mathbb{E}(J_{ba}) - \pi_b \overset{(47)}{=} o(p_{k_b} \bar{q}^{\,\tau(k_b - j+1)}). \tag{51}$$

Given a, the number of $b \in \mathcal{T}_k$ satisfying the conditions of Case 3 is $O(n^{k+1})$, hence if we set

$$B_3 = B_3(a) \;:=\; \{b \in \Gamma : b \neq a, K_b \neq K_a, K_b \notin \mathcal{B}_a, K_a \notin \mathcal{B}_b\},$$

we have

$$\begin{aligned}
\sum\nolimits_{b \in B_3} \mathbb{E}|J_{ba} - I_b| &\overset{(51)}{=} \sum\nolimits_{k=j}^{d} O(n^{k+1}) \cdot o(p_k \bar{q}^{\,\tau(k-j+1)}) \\
&\overset{(48)}{=} o(1) \cdot \sum\nolimits_{k=j}^{d} O\left(\mathbb{E}(X_{j,k})\right) = o(1).
\end{aligned}$$

Since $\{b \in \Gamma : b \neq a\} = B_1 \cup B_{2.1} \cup B_{2.2} \cup B_3$, putting all the cases together we have that $\sum_{b\neq a} \mathbb{E}|J_{ba} - I_b| = o(1)$ for any fixed $a \in \Gamma$, as required.

Observe that for symmetry reasons, the quantity $\sum_{b\neq a} \mathbb{E}|J_{ba} - I_b| = o(1)$ remains the same if the sum is over $b \neq a'$ with $k_{a'} = k_a$. Thus we have

$$\begin{aligned}\sum\nolimits_{a\in\Gamma} \pi_a \sum\nolimits_{b\neq a} \mathbb{E}|J_{ba} - I_b| &= \sum\nolimits_{k=j}^{d} o(1) \cdot \sum\nolimits_{a\in\Gamma_k} \pi_a \\ &= \sum\nolimits_{k=j}^{d} o(1) \cdot \mathbb{E}(X_{j,k}) \\ &\overset{(48)}{=} \sum\nolimits_{k=j}^{d} o(1) \cdot O(1) = o(1). \end{aligned} \tag{52}$$

The right-hand side of (46) is therefore

$$\begin{aligned}\sum\nolimits_{a\in\Gamma} \pi_a \left(\pi_a + \sum\nolimits_{b\neq a} \mathbb{E}|J_{ba} - I_b|\right) &\overset{(52)}{=} \left(\sum\nolimits_{a\in\Gamma} \pi_a^2\right) + o(1) \\ &\leqslant \left(\max_{a\in\Gamma} \pi_a\right) \left(\sum\nolimits_{a\in\Gamma} \pi_a\right) + o(1) \\ &= \left(\max_{a\in\Gamma} \pi_a\right) \left(\sum\nolimits_{k=j}^{d} \mathbb{E}(X_{j,k})\right) + o(1) \\ &\overset{(48)}{=} \left(\max_{a\in\Gamma} \pi_a\right) \cdot O(1) + o(1),\end{aligned}$$

and

$$\begin{aligned}\max_{a\in\Gamma} \pi_a &\overset{(47)}{=} \max_{j\leqslant k\leqslant d} \left((1 + o(1))p_k \bar{q}^{\,\tau(k-j+1)}\right) \\ &\overset{(48)}{=} \max_{j\leqslant k\leqslant d} \frac{\mathbb{E}(X_{j,k})}{\Theta(n^{k+1})} = O\left(\frac{1}{n^{j+1}}\right) = o(1).\end{aligned}$$

In conclusion, we have

$$d_{TV}\Big(\mathcal{L}(\mathbf{X}), \big(Z_j, \dots, Z_d\big)\Big) \xrightarrow{n\to\infty} 0.$$

Since by Corollary 6.6, $\lim_{n\to\infty} \mathbb{E}(X_{j,k}) = \mathcal{E}_k$ for every $j \leqslant k \leqslant d$, we have

$$\mathcal{L}(\mathbf{X}) \xrightarrow{\mathrm{d}} \left(\mathrm{Po}\big(\mathcal{E}_j\big), \dots, \mathrm{Po}\big(\mathcal{E}_d\big)\right),$$

as required. □

11 Concluding remarks

11.1 Non-triviality of cohomology groups

To prove Theorem 1.4 (b), our strategy was to show that for every $\varepsilon > 0$ and for each $i \in [j]$, whp $H^i(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_i(\varepsilon) = [\varepsilon/n^{j-i+1}, \tau_i^*)$, because of the existence of copies of

$\hat{M}_{i,k}$ for some $i \leqslant k \leqslant d$ throughout the interval $I_i(\varepsilon)$ (Corollary 4.6). However, it is likely that $H^i(\mathcal{G}_\tau; R)$ would already be non-trivial for even smaller τ. In particular, it would be interesting to precisely determine from which point on $H^i(\mathcal{G}_\tau; R) \neq 0$ whp and in this case to describe its rank, analogously to Theorem 1.5.

11.2 Dimension of the last minimal obstruction

In Theorem 1.5 we obtain an asymptotic description of the j-th cohomology group in the critical window. More strongly, in this regime Lemma 10.1 yields the asymptotic (joint) distribution of the number of copies of $M_{j,k}$, for every index k with $j \leqslant k \leqslant d$. This leads to the natural question: what is (the asymptotic probability distribution of) the dimension of the last copy of $M_{j,k}$ that vanishes?

11.3 Determining the critical dimensions

For a given j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, it is interesting to determine which indices k with $j \leqslant k \leqslant d$ represent critical dimensions for $\bar{\mathbf{p}}$. Informally, for each dimension there is an 'obstruction surface' representing when the last minimal obstructions of that dimension disappear as a function of the direction of the process. The critical dimensions are those whose corresponding obstruction surface is furthest from the origin along the appropriate direction, and there may be more than one critical dimension at the intersection of two or more obstruction surfaces. We may also consider the 'critical surface' as consisting of those pieces of all obstruction surfaces which lie furthest from the origin in the appropriate direction.

More formally, recall from Definition 3.6 that k is a critical dimension if $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k = O(1)$. The main term of this expression is $\bar{\lambda}_k \log n$; all other terms are $o(\log n)$. In particular, we have $\bar{\lambda}_k \leqslant 0$ for all k and $\bar{\lambda}_k = 0$ for all critical dimensions. Recall that by (4), $\bar{\lambda}_k$ is a linear combination of the parameters $\bar{\gamma}_k$, $\bar{\alpha}_{j+1}$, …, $\bar{\alpha}_d$, namely

$$\bar{\lambda}_k = j + 1 - \bar{\gamma}_k - (k - j + 1) \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i.$$

Plotting the ranges where $\bar{\lambda}_k = 0$ as a function of those parameters, for $j \leqslant k \leqslant d$, is thus a way to visualise the parameters of a j-critical direction $\bar{\mathbf{p}}$, as well as what dimensions are critical for $\bar{\mathbf{p}}$. For example, if $\bar{\gamma}_j < \frac{j+1}{2}$, then $k = j$ is the only critical dimension and the parameters $\bar{\alpha}_{j+1}, \ldots, \bar{\alpha}_d$ lie in the $(d-j-1)$-dimensional hyperplane defined by $\sum_{i=j+1}^{k} \bar{\alpha}_i = j + 1 - \bar{\gamma}_j$. For $\bar{\gamma}_j = \frac{j+1}{2}$ and $\bar{\gamma}_{j+1} = 0$ however, both $k = j$ and $k = j + 1$ are critical dimensions, and the hyperplane for $\bar{\alpha}_{j+1}, \ldots, \bar{\alpha}_d$ is given by $\sum_{i=j+1}^{k} \bar{\alpha}_i = \frac{j+1}{2}$.

We present some examples of these plots in Figure 4.

11.4 Integer homology

Recently, Newman and Paquette [38] proved a hitting time result for the $(d-1)$-th homology group over $\mathbb{Z}$ in the Linial-Meshulam model (the case $d = 2$ was previously proved by Łuczak and Peled [34]); this is a stronger result than for the corresponding cohomology group. It would be interesting to know whether the analogous result also holds in $\mathcal{G}_\tau$, i.e. does the j-th

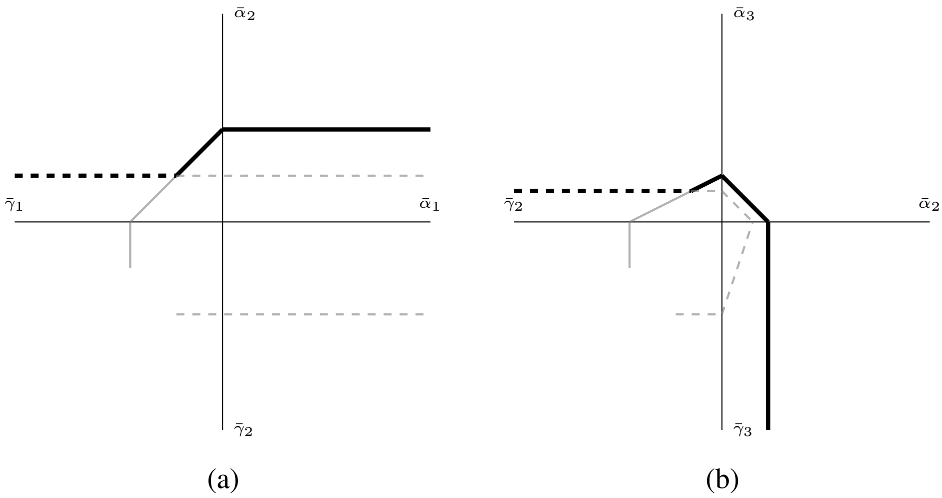

Figure 4: Plots of obstruction surfaces given by the equations $\bar{\lambda}_k = 0$ with $k = 1, 2$ in (a) and $k = 2, 3$ in (b), respectively. For each k, we use the same axis for $\bar{\alpha}_k$ and $\bar{\gamma}_k$, because property (A1) in Definition 3.1 states that only one of $\bar{\alpha}_k, \bar{\gamma}_k$ is non-zero.
The bold sections denote the critical surfaces, that is, the ranges for which $\bar{\mathbf{p}}$ is a j-critical direction and the corresponding k is a critical dimension. The striped portions in the lower left corners indicate the regions in which whp $\mathcal{G}_\tau$ has no simplices of positive dimensions.
(a) The obstruction surfaces $\bar{\lambda}_1 = 0$ (plain) and $\bar{\lambda}_2 = 0$ (dashed), for $d = 2$, $j = 1$. Recall that the equation $\bar{\lambda}_1 = 0$ refers to copies of $M_{1,1}$, i.e. isolated 1-simplices in $\mathcal{G}_\tau$.
(b) The obstruction surfaces $\bar{\lambda}_2 = 0$ (plain) and $\bar{\lambda}_3 = 0$ (dashed), for $d = 3$, $j = 1$. In this case, the plots are under the condition that $\bar{p}_1 = 0$.

homology group with integer coefficients vanish at the same time as the last copy of $\hat{M}_{j,k}$ for any k disappears?

References

[1] P. Allen, J. Böttcher, Y. Kohayakawa, and Y. Person. Tight Hamilton cycles in random hypergraphs. *Random Structures Algorithms* 46(3):446–465, 2015.

[2] P. Allen, C. Koch, O. Parczyk, and Y. Person. Finding tight Hamilton cycles in random hypergraphs faster. *LATIN 2018: Theoretical informatics*, 28–36, Lecture Notes in Comput. Sci., 10807, *Springer, Cham*, 2018.

[3] L. Aronshtam and N. Linial. When does the top homology of a random simplicial complex vanish? *Random Structures Algorithms*, 46(1):26–35, 2015.

[4] L. Aronshtam, N. Linial, T. Łuczak, and R. Meshulam. Collapsibility and vanishing of top homology in random simplicial complexes. *Discrete Comput. Geom.*, 49(2):317–334, 2013.

[5] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.

[6] M. Behrisch, A. Coja-Oghlan, and M. Kang. The order of the giant component of random hypergraphs. *Random Structures Algorithms*, 36(2):149–184, 2010.

[7] M. Behrisch, A. Coja-Oghlan, and M. Kang. Local limit theorems for the giant component of random hypergraphs. *Combin. Probab. Comput.*, 23(3):331–366, 2014.

[8] B. Bollobás and O. Riordan. Asymptotic normality of the size of the giant component in a random hypergraph. *Random Structures Algorithms*, 41(4):441–450, 2012.

[9] B. Bollobás and O. Riordan. Counting connected hypergraphs via the probabilistic method. *Combin. Probab. Comput.* 25(1):1, 21–75, 2016.

[10] B. Bollobás and O. Riordan. Exploring hypergraphs with martingales. *Random Structures Algorithms*, 50(3):325–352, 2017.

[11] B. Bollobás and O. Riordan. Counting dense connected hypergraphs via the probabilistic method. *Random Structures Algorithms*, 53(2):185–220, 2018.

[12] B. Bollobás and A. Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[13] O. Cooley, N. Del Giudice, M. Kang, and P. Sprüssel. Vanishing of cohomology groups of random simplicial complexes. *Random Structures Algorithms*, 56:461–500, 2020.

[14] O. Cooley, M. Kang, and C. Koch. Threshold and hitting time for high-order connectedness in random hypergraphs. *Electron. J. Combin.*, 23(2):#P2.48, 14, 2016.

[15] O. Cooley, M. Kang, and C. Koch. The size of the giant high-order component in random hypergraphs. *Random Structures Algorithms*, 53(2):238–288, 2018.

[16] O. Cooley, M. Kang, and Y. Person. Largest components in random hypergraphs. *Combin. Probab. Comput.*, pages 1–22, 2018.

[17] A. Costa, M. Farber, and D. Horak. Fundamental groups of clique complexes of random graphs. *Trans. London Math. Soc.*, 2(1):1–32, 2015.

[18] A. Costa and M. Farber. Random simplicial complexes. In *Configuration Spaces: Geometry, Topology and Representation Theory*, Springer INdAM Series, volume 14, pages 129–153. Springer, Cham, 2016.

[19] R. W. R. Darling and J. R. Norris. Structure of large random hypergraphs. *Ann. Appl. Probab.*, 15(1A):125–152, 2005.

[20] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[21] M. Farber, L. Mead, and T. Nowik. Random simplicial complexes, duality and the critical dimension. to appear in *J. Topol. Anal.* (2020).

[22] N. Fountoulakis, T. Iyer, C. Mailler, and H. Sulzbach. Dynamical Models for Random Simplicial Complexes. arXiv:1910.12715.

[23] N. Fountoulakis and M. Przykucki. High-dimensional bootstrap processes in evolving simplicial complexes. arXiv:1910.10139.

[24] N. Fountoulakis and M. Przykucki. Algebraic and combinatorial expansion in random simplicial complexes. *Random Structures Algorithms* 60(3):339–366, 2022.

[25] C. Hoffman, M. Kahle, and E. Paquette. The threshold for integer homology in random d-complexes. *Discrete Comput. Geom.*, 57(4):810–823, 2017.

[26] M. Kahle. The neighborhood complex of a random graph. *J. Combin. Theory Ser. A*, 114(2):380–387, 2007.

[27] M. Kahle. Topology of random clique complexes. *Discrete Math.*, 309(6):1658–1671, 2009.

[28] M. Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. *Ann. of Math.*, 179(3):1085–1107, 2014.

[29] M. Kahle. Topology of random simplicial complexes: a survey. In *Algebraic topology: applications and new directions*, volume 620 of *Contemp. Math.*, pages 201–221. Amer. Math. Soc., Providence, RI, 2014.

[30] M. Kahle and B. Pittel. Inside the critical window for cohomology of random k-complexes. *Random Structures Algorithms*, 48(1):102–124, 2016.

[31] M. Karoński and T. Łuczak. Random hypergraphs. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 283–293. János Bolyai Math. Soc., Budapest, 1996.

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

[33] N. Linial and Y. Peled. On the phase transition in random simplicial complexes. *Ann. of Math.*, 184(3):745–773, 2016.

[34] T. Łuczak and Y. Peled. Integral homology of random simplicial complexes. *Discrete Comput. Geom.*, 59(1):131–142, 2018.

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

[36] J. Munkres. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA, 1984.

[37] R. Nenadov and N. Škorić. Powers of Hamilton cycles in random graphs and tight Hamilton cycles in random hypergraphs. *Random Structures Algorithms* 54(1):187–208, 2019.

[38] A. Newman and E. Paquette. The integer homology threshold in $Y_d(n, p)$. arXiv:1808.10647.

[39] O. Parczyk and Y. Person. Spanning structures and universality in sparse hypergraphs. *Random Structures Algorithms*, 49(4):819–844, 2016.

[40] D. Poole. On the strength of connectedness of a random hypergraph. *Electron. J. Combin.*, 22(1):#P1.69, 16, 2015.

[41] J. Schmidt-Pruzan and E. Shamir. Component structure in the evolution of random hypergraphs. *Combinatorica*, 5(1):81–94, 1985.

[42] V. E. Stepanov. On the probability of the connectedness of a random graph $\mathcal{G}_m(t)$. *Theory Probab. Appl.*, 15:55–67, 1970.

A Proofs of auxiliary results

A.1 Proof of Lemma 4.2

(a) Since topological connectedness is a monotone property, it is enough to prove the statement in the case when $p_k = \frac{c^- \log n}{n^k}$ for all $k \in [d]$, which will be convenient in the proof.

Let U be the number of isolated vertices in $\mathcal{G}(n, \mathbf{p})$. Every vertex is contained in $\binom{n-1}{k}$ many $(k+1)$-sets, each of which does not form an i-simplex with probability $(1 - p_k)$, all independently. Hence each vertex is isolated with probability $\prod_{k=1}^{d}(1-p_k)^{\binom{n-1}{k}}$, and therefore we have

$$\begin{aligned}\mathbb{E}(U) = n\prod_{k=1}^{d}(1-p_k)^{\binom{n-1}{k}} &\geqslant n \cdot \exp\left(-\sum_{k=1}^{d}\frac{n^k}{k!}\left(p_k + O(p_k^2)\right)\right)\\ &\geqslant n \cdot \exp\left(-\sum_{k=1}^{d}\frac{n^k}{k!}\cdot\frac{c^-\log n}{n^k} + o(1)\right)\\ &= (1+o(1))n^{1-\tilde{d}c^-},\end{aligned}$$

where $\tilde{d} := \sum_{k=1}^{d} 1/k!$. Moreover, the probability that two fixed distinct vertices are both isolated is

$$\begin{aligned}\prod_{k=1}^{d}(1-p_k)^{2\binom{n-1}{k}-\binom{n-2}{k-1}} &\leqslant \exp\left(-\sum_{k=1}^{d}p_k\binom{n-1}{k}\left(2-\frac{k}{n-1}\right)\right)\\ &\leqslant \exp\left(-2\sum_{k=1}^{d}\frac{c^-\log n}{k!}\left(1+O\left(\frac{1}{n}\right)\right)\right)\\ &\leqslant (1+o(1))n^{-2\tilde{d}c^-}.\end{aligned}$$

By choosing c^- such that $c^- < 1/\tilde{d}$ (so in particular $\mathbb{E}(U) \to \infty$), we obtain

$$\begin{aligned}\mathbb{E}(U^2) &\leqslant \mathbb{E}(U) + n(n-1)(1+o(1))n^{-2\tilde{d}c^-}\\ &= \mathbb{E}(U) + (1+o(1))n^{2(1-\tilde{d}c^-)} = (1+o(1))\mathbb{E}(U)^2,\end{aligned}$$

so by Chebyshev's inequality whp there are isolated vertices, implying that whp $\mathcal{G}(n, \mathbf{p})$ is not topologically connected.

(b) Consider $\tilde{\mathbf{p}}$ obtained from $\mathbf{p}$ by replacing all probabilities except p_k by zero, where $k \in [d]$ is an index such that $p_k \geqslant \frac{c^+ \log n}{n^k}$. If $k = 1$, then (b) follows from the corresponding results for graphs. For $k \geqslant 2$, (b) holds because of the fact that we can choose c^+ such that whp $\mathcal{G}(n, \tilde{\mathbf{p}})$ is topologically connected by [13, Lemma 4.1]. □

We note that the proof idea of Lemma 4.2 is a standard generalisation of the very well-known hitting time result for graphs: whp the random graph process becomes connected at exactly the moment its last isolated vertex disappears. Indeed, Theorem 1.4 is also a generalisation of this result, albeit a far more complex one.

The vertex-connectedness threshold for uniform random hypergraphs, which we quoted from [13] for the proof of (b), also follows as a special case of earlier and much stronger results from [14] and from [40]. The proof in [13] has the advantage that it is a simple and elementary extension of the standard graph argument.

A.2 Proof of Lemma 4.5

We prove the statement for $i = j-1$; for general $i \in [j-1]$ it suffices to iterate the procedure $j-i$ times.

We thus need to show that we can choose a positive constant $\eta = \eta_{j-1}$ and a function $\epsilon = \epsilon_{j-1}(n) = o(1)$ such that

$$\bar{\mathbf{p}}' = \bar{\mathbf{p}}'(\eta,\epsilon) \;:=\; \frac{\eta+\epsilon}{n}\bar{\mathbf{p}}$$

is a $(j-1)$-critical direction (Definition 3.3).

Recall that the j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1,\ldots,\bar{p}_d)$ is in particular a j-admissible direction (Definition 3.1), i.e. for every $k \in [d]$

$$\bar{p}_k = \begin{cases} \dfrac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)! & \text{for } j \leqslant k \leqslant d, \\ \dfrac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}} & \text{for } 1 \leqslant k \leqslant j-1, \end{cases}$$

with $\bar{\alpha}_k$, $\bar{\beta}_k = \bar{\beta}_k(n)$, and $\bar{\gamma}_k$ satisfying conditions (A1)–(A4).

This implies that $\bar{\mathbf{p}}'$ is a $(j-1)$-admissible direction: for every $k \in [d]$, we have

$$\bar{p}'_k = \frac{\eta+\epsilon}{n}\bar{p}_k = \begin{cases} \dfrac{\bar{\alpha}'_k \log n + \bar{\beta}'_k}{n^{k-(j-1)+\bar{\gamma}'_k}}(k-(j-1))! & \text{for } j-1 \leqslant k \leqslant d, \\ \dfrac{\bar{\alpha}'_k \log n + \bar{\beta}'_k}{n^{k-(j-1)+\bar{\gamma}'_k}} & \text{for } 1 \leqslant k \leqslant j-2, \end{cases}$$

where

$$\begin{aligned} \bar{\alpha}'_k &:= \begin{cases} \frac{\eta\bar{\alpha}_k}{k-j+1} & \text{for } j \leqslant k \leqslant d, \\ \eta\bar{\alpha}_k & \text{for } 1 \leqslant k \leqslant j-1, \end{cases} \\ \bar{\beta}'_k &:= \begin{cases} \frac{\eta\bar{\beta}_k + \epsilon\bar{\alpha}_k \log n + \epsilon\bar{\beta}_k}{k-j+1} & \text{for } j \leqslant k \leqslant d, \\ \eta\bar{\beta}_k + \epsilon\bar{\alpha}_k \log n + \epsilon\bar{\beta}_k & \text{for } 1 \leqslant k \leqslant j-1, \end{cases} \\ \bar{\gamma}'_k &:= \bar{\gamma}_k, \end{aligned} \tag{53}$$

and it is easy to check that for any choices of η and ϵ, the parameters $\bar{\alpha}'_k$, $\bar{\beta}'_k$, and $\bar{\gamma}'_k$ satisfy conditions (A1)–(A4) in Definition 3.1.

We now want to prove that, for the appropriate choices of η and ϵ, the vector $\bar{\mathbf{p}}'$ is also a $(j-1)$-critical direction (Definition 3.3), i.e.

(C1′) $\bar{\lambda}'_k \log n + \bar{\mu}'_k + \bar{\nu}'_k \leqslant 0, \qquad$ for all indices k with $j-1 \leqslant k \leqslant d$ and $\bar{p}'_k \neq 0$;

(C2′) $\bar{\lambda}'_{\bar{k}} \log n + \bar{\mu}'_{\bar{k}} + \bar{\nu}'_{\bar{k}} = 0, \qquad$ for some $\bar{k}$ with $j-1 \leqslant \bar{k} \leqslant d$,

where the parameters $\bar{\lambda}'_k$, $\bar{\mu}'_k$, and $\bar{\nu}'_k$ are defined as in (4), but with j replaced by $j-1$.

Recall that $\bar{\lambda}'_k$ and $\bar{\nu}'_k$ are constants, while each $\bar{\mu}'_k$ is a function with $\bar{\mu}'_k = o(\log n)$. Thus, (C1′) and (C2′) will both hold if and only if

$$\begin{array}{ll} \bar{\lambda}'_k \leqslant 0 & \text{for every } j-1 \leqslant k \leqslant d \text{ with } \bar{p}'_k \neq 0; \\ \bar{\lambda}'_{\bar{k}} = 0 & \text{for some } j-1 \leqslant \bar{k} \leqslant d; \end{array} \tag{54}$$

and

$$\begin{array}{ll} \bar{\mu}'_k \leqslant -\bar{\nu}'_k & \text{for every } j-1 \leqslant k \leqslant d \text{ such that } \bar{\lambda}'_k = 0; \\ \bar{\mu}'_{\bar{k}} = -\bar{\nu}'_{\bar{k}} & \text{for some } \bar{k} \text{ such that } \bar{\lambda}'_{\bar{k}} = 0. \end{array} \tag{55}$$

Let us consider (54) first. For every k with $\bar{p}'_k \neq 0$ (which is equivalent to $\bar{p}_k \neq 0$), we have

$$\begin{aligned} \bar{\lambda}'_k &= j - \bar{\gamma}'_k - (k-j+2) \sum\nolimits_{i=j}^{d} \bar{\alpha}'_i \\ &\overset{(53)}{=} j - \bar{\gamma}_k - \eta(k-j+2) \sum\nolimits_{i=j}^{d} \frac{\bar{\alpha}_i}{i-j+1}. \end{aligned}$$

Recall that k_0 is an index with $j+1 \leqslant k_0 \leqslant d$ such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$ (cf. (A1) and (A4) for $\bar{\mathbf{p}}$). We observe that

$$\lim_{\eta \to 0} \bar{\lambda}'_{k_0} = j > 0 \qquad \text{and} \qquad \lim_{\eta \to \infty} \bar{\lambda}'_k = -\infty \quad \text{for every } j-1 \leqslant k \leqslant d \text{ with } \bar{p}'_k \neq 0.$$

Hence, since each $\bar{\lambda}'_k$ is a continuous function of η, we can choose $\eta > 0$ such that (54) holds.

For the rest of the proof, let this value of η be fixed. We will now show that we can choose the function ϵ so that (55) is satisfied as well. To simplify notation, whenever we consider $\bar{\mu}'_k$ or $\bar{\nu}'_k$ in the following, we will assume that $\bar{\lambda}'_k = 0$.

By (4), with j replaced by $j-1$, we have

$$\bar{\nu}'_k = \begin{cases} -\log(j!) & \text{if } k = j-1, \\ -\log((j-1)!) - \log(k-j+2) + \log(\bar{\alpha}'_k) & \text{if } k \neq j-1 \text{ and } \bar{\alpha}'_k \neq 0, \\ -\log((j-1)!) - \log(k-j+2) & \text{otherwise,} \end{cases}$$

which by (53) is independent from ϵ. Furthermore,

$$\bar{\mu}'_k = -(k-j+2) \sum\nolimits_{i=j}^{d} \frac{\bar{\beta}'_i}{n^{\bar{\gamma}'_i}} + \begin{cases} 0 & \text{if } \bar{p}'_k > 1, \\ \log\log n & \text{if } \bar{p}'_k \leqslant 1 \text{ and } \bar{\alpha}'_k \neq 0, \\ \log(\bar{\beta}'_k) & \text{if } \bar{p}'_k \leqslant 1 \text{ and } \bar{\alpha}'_k = 0. \end{cases} \tag{56}$$

By (53),

$$\sum\nolimits_{i=j}^{d} \frac{\bar{\beta}'_i}{n^{\bar{\gamma}'_i}} = \sum\nolimits_{i=j}^{d} \frac{\eta\bar{\beta}_i + \epsilon\bar{\alpha}_i \log n + \epsilon\bar{\beta}_i}{(i-j+1)n^{\bar{\gamma}_i}}.$$

Now (A1)–(A4) imply that $c := \sum_i \frac{\bar{\alpha}_i}{(i-j+1)n^{\bar{\gamma}_i}}$ is a positive constant, while $\sum_i \frac{\bar{\beta}_i}{(i-j+1)n^{\bar{\gamma}_i}} = o(\log n)$. Moreover, all possible summands in the case distinction of (56) are $o(\log n)$ as well. Therefore, there exists a positive function $f(n) = o(\log n)$, which does not depend on the choice of ϵ, such that

$$|\bar{\mu}'_k + c(k-j+2)\epsilon \log n| \leqslant f(n)$$

for all $j-1 \leqslant k \leqslant d$ with $\bar{p}'_k \neq 0$. Fix a function ω of n with $\omega \to \infty$, but $\omega = o\left(\frac{\log n}{f(n)}\right)$. Then for all indices k as above,

$$\bar{\mu}'_k \xrightarrow{n\to\infty} \begin{cases} \infty & \text{for } \epsilon = -\frac{1}{\omega}, \\ -\infty & \text{for } \epsilon = \frac{1}{\omega}. \end{cases}$$

Therefore, by continuity we can choose $\epsilon = o(1)$ such that (55) holds. Since we have now found η and ϵ such that (54) and (55) simultaneously hold, (C1′) and (C2′) are both satisfied, i.e. $\bar{\mathbf{p}}'$ is a $(j-1)$-critical direction, as required. ☐

A.3 Proof of Proposition 5.13

We first observe that for $j+1 \leqslant k \leqslant d$, the number of $(k+1)$-sets which contain at least two distinct $(j+1)$-sets of $\mathcal{J}$ is at most $\binom{|\mathcal{J}|}{2}\binom{n}{k-j-1} = O(n^{k-j-1})$, and therefore the number of $(k+1)$-sets that must *not* be k-simplices in order for A to hold is

$$|\mathcal{J}|\binom{n-j-1}{k-j} - O(n^{k-j-1}) - |\mathcal{S}| = |\mathcal{J}|\binom{n-j-1}{k-j} - O(n^{k-j-1}).$$

Thus, since $p_k n^{k-j-1} = o(1)$, we have

$$\begin{aligned} \mathbb{P}(A) &= \prod_{k=j+1}^{d} (1-p_k)^{|\mathcal{J}|\binom{n-j-1}{k-j} - O(n^{k-j-1})} \\ &= (1+o(1)) \prod_{k=j+1}^{d} (1-\tau\bar{p}_k)^{|\mathcal{J}|\binom{n-j-1}{k-j}} \\ &= (1+o(1)) \exp\left(-|\mathcal{J}|\tau \sum_{k=j+1}^{d} \binom{n-j-1}{k-j}\bar{p}_k\right) \\ &\overset{(14)}{=} (1+o(1))\bar{q}^{\,\tau|\mathcal{J}|}, \end{aligned}$$

as claimed. ☐

A.4 Proof of Lemma 6.1

Let $\hat{\mathcal{T}}$ be the set of all 4-tuples (K, C, w, a) that might form a copy of $\hat{M}_{j,k_0}$ in $\mathcal{G}_\tau$ (i.e. all sizes and containment relations are correct, but we make no assumptions about which simplices are present or absent), and let $\hat{T} = (K, C, w, a) \in \hat{\mathcal{T}}$. Property (M1) holds with probability

$$p_{k_0} = \frac{\varepsilon}{n}\bar{p}_{k_0} \overset{(3)}{=} \Theta\left(\frac{\log n}{n^{k_0-j+1}}\right), \tag{57}$$

since the choice of k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ (see Definition 5.15 (b)). By Proposition 5.13 the probability that (M2) holds is $(1+o(1))\bar{q}^{\,\tau(k-j+1)}$, where since $\tau = \varepsilon/n$ we have

$$\bar{q}^{\,\tau} \overset{(15)}{=} (1+o(1)) \exp\left(-\frac{\varepsilon}{n} \cdot \sum\nolimits_{k=j+1}^{d} \left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right) = 1 + o(1),$$

and therefore (M2) holds—independently of whether (M1) holds—with probability $1 + o(1)$.

In order to calculate the probability that (M3) also holds, first observe that if (M2) holds, then no simplex can contain more than one side of the (potential) j-shell $C \cup \{w\} \cup \{a\}$. Thus, conditioned on the event that (M1) and (M2) hold, each of the $j+1$ sides of $C \cup \{w\} \cup \{a\}$ forms a j-simplex independently with probability

$$1 - \prod\nolimits_{k=j}^{d} (1-p_k)^{\binom{n-j-1}{k-j} + O(n^{k-j-1})} = (1+o(1))r,$$

where

$$r \ := \ \sum\nolimits_{k=j}^{d} \frac{p_k n^{k-j}}{(k-j)!} = \Theta\left(\frac{\log n}{n}\right). \tag{58}$$

Combining all the probabilities, we obtain

$$\mathbb{P}(\hat{T} \text{ forms a copy of } \hat{M}_{j,k_0}) = (1+o(1)) p_{k_0} r^{j+1}. \tag{59}$$

Recall that $\hat{X}_{j,k_0}$ denotes the number of copies of $\hat{M}_{j,k_0}$ in $\mathcal{G}_\tau$. Now (59) implies that

$$\begin{aligned} \mathbb{E}(\hat{X}_{j,k_0}) &= (1+o(1)) \binom{n}{k_0+1} \binom{k_0+1}{j} (k_0 - j + 1)(n - k_0 - 1) p_{k_0} r^{j+1} \\ &= (1+o(1)) \frac{p_{k_0} r^{j+1} n^{k_0+2}}{j!(k_0-j)!} \overset{(57),(58)}{=} \Theta((\log n)^{j+2}). \end{aligned} \tag{60}$$

We now aim to calculate the corresponding second moment $\mathbb{E}((\hat{X}_{j,k_0})^2)$. Given two 4-tuples $\hat{T}_1 = (K_1, C_1, w_1, a_1)$ and $\hat{T}_2 = (K_2, C_2, w_2, a_2)$, we define

- $I = I(\hat{T}_1, \hat{T}_2) \ := \ (K_1 \cup \{a_1\}) \cap (K_2 \cup \{a_2\})$ and $i \ := \ |I|$;
- $s = s(\hat{T}_1, \hat{T}_2) \ := \ \begin{cases} 1 & \text{if } K_1 = K_2, \\ 2 & \text{otherwise;} \end{cases}$
- $\mathcal{J}_x$ to be the set of all $(j+1)$-subsets of $\{C_x \cup \{a_x\} \cup \{w_x\}\}$ for $x = 1, 2$ and

  $$t = t(\hat{T}_1, \hat{T}_2) \ := \ |(\mathcal{J}_1 \cup \mathcal{J}_2) \setminus \{C_1 \cup \{w_1\}, C_2 \cup \{w_2\}\}|,$$

  i.e. the number of $(j+1)$-sets that are sides of the (potential) j-shells of $\hat{T}_1$ and $\hat{T}_2$, but not a base of either j-shell.

If $s = 2$ and the intersection of the two simplices contains a petal, then $\hat{T}_1$ and $\hat{T}_2$ cannot both form a copy of $\hat{M}_{j,k_0}$, because (M2) would be violated. In the following, we therefore assume that this is not the case.

Clearly, (M1) holds for both $\hat{T}_1$ and $\hat{T}_2$ simultaneously with probability $(p_{k_0})^s$, while conditioned on (M1), by Proposition 5.13, the probability that (M2) holds for both $\hat{T}_1$ and $\hat{T}_2$ simultaneously is (at least) $(1+o(1))\bar{q}^{\,\tau 2(k_0-j+1)} = 1+o(1)$. Conditioned on (M1) and (M2) holding, observe that each of the t sides of the (potential) j-shells lies in some k-simplex (and hence forms a j-simplex) with probability r. Moreover, no simplex in $\mathcal{G}_\tau$ can contain more than two of those sides (at most one from each potential shell since otherwise it would contain a petal, which is ruled out by the conditioning on (M2)). Furthermore, the probability of a side lying in any k-simplex that contains two distinct sides is

$$1 - \prod\nolimits_{k=j+1}^{d} (1-p_k)^{O(n^{k-j-1})} = O\left(\frac{\log n}{n^2}\right) \overset{(58)}{=} o(r^2).$$

Therefore, the probability that all t sides form j-simplices is $(1+o(1))r^t$ and thus

$$\mathbb{P}(\hat{T}_1, \hat{T}_2 \text{ both form copies of } \hat{M}_{j,k_0}) = (1+o(1))(p_{k_0})^s r^t. \tag{61}$$

Define $\hat{\mathcal{T}}^2(i,s,t)$ to be the set of pairs $(\hat{T}_1,\hat{T}_2) \in \hat{\mathcal{T}} \times \hat{\mathcal{T}}$ with parameters i, s and t. With this notation, (61) implies that

$$\mathbb{E}((\hat{X}_{j,k_0})^2) = (1+o(1)) \sum_{(i,s,t)} \sum_{(\hat{T}_1,\hat{T}_2)\in\hat{\mathcal{T}}^2(i,s,t)} (p_{k_0})^s r^t.$$

Observe that $|\hat{\mathcal{T}}^2(i,s,t)| = O(n^{2k_0+4-i})$.We can now estimate the contributions of all the summands, distinguishing according to the possible values of s and i.

Case 1: s=1. This means that $K_1 = K_2$, and thus $k_0 + 1 \leqslant i \leqslant k_0 + 2$.

- $i = k_0 + 1$. In this case $a_1 \neq a_2$ and thus the sets of sides of the two j-shells would be disjoint, i.e. $t = 2j + 2$. Therefore we get a contribution of order

  $$O\left(p_{k_0} r^{2j+2} n^{2k_0+4-(k_0+1)}\right) \overset{(60)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{p_{k_0} n^{k_0+1}}\right) \overset{(57)}{=} o\left((\mathbb{E}(\hat{X}_{j,k_0})^2\right).$$

  Indeed, in order to prove the final property of the lemma, that the associated copies of M_{j,k_0} are distinct, we observe something even stronger: we have

  $$p_{k_0} r^{2j+2} n^{2k_0+4-(k_0+1)} = \Theta\left(\frac{(\log n)^{2j+3}}{n^j}\right) \overset{(57),(58)}{=} o(1).$$

  Thus by Markov's inequality, whp there are no two copies of $\hat{M}_{j,k_0}$ that share the same k_0-simplex but have distinct apex vertices.

- $i = k_0 + 2$. The two j-shells have the same apex vertex and thus the j-shells coincide if and only if they have the same base. This means that $t \geqslant j+1$, which gives a contribution of order

  $$O(p_{k_0} r^{j+1} n^{2k_0+4-(k_0+2)}) \overset{(60)}{=} O\left(\mathbb{E}(\hat{X}_{j,k_0})\right) = o\left(\mathbb{E}(\hat{X}_{j,k_0})^2\right).$$

Case 2: s=2.

- $i = 0$. We show that this case represents the dominant contribution to $\mathbb{E}((\hat{X}_{j,k_0})^2)$. The two j-shells are disjoint, hence $t = 2j + 2$. Observe that we have

$$\binom{n}{k_0+1}\binom{k_0+1}{j}(k_0-j+1)(n-k_0-1) = (1+o(1))\frac{n^{k_0+2}}{j!(k_0-j)!}$$

choices for $\hat{T}_1$. For any fixed $\hat{T}_1$, the number of choices for $\hat{T}_2$ that yield $i = 0$ is

$$\binom{n-k_0-2}{k_0+1}\binom{k_0+1}{j}(k_0-j+1)(n-2k_0-3) = (1+o(1))\frac{n^{k_0+2}}{j!(k_0-j)!}.$$

Thus, the contribution of all such pairs is

$$(1+o(1))\frac{p_{k_0}^2 r^{2j+2} n^{2k_0+4}}{(j!(k_0-j)!)^2} \overset{(60)}{=} (1+o(1))\mathbb{E}(\hat{X}_{j,k_0})^2.$$

- $1 \leqslant i \leqslant j$. In this case $\hat{T}_1$ and $\hat{T}_2$ cannot share a j-simplex of their shells, i.e. $t = 2j + 2$. Therefore the contribution is

$$O\left(p_{k_0}^2 r^{2j+2} n^{2k_0+4-i}\right) \overset{(60)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{n^i}\right) = o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

- $i = j + 1$. Here, $\hat{T}_1$ and $\hat{T}_2$ can share at most one j-simplex of their shells, which means $t \geqslant 2j + 1$ and we have a contribution of order

$$O\left(p_{k_0}^2 r^{2j+1} n^{2k_0+4-(j+1)}\right) \overset{(60)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{rn^{j+1}}\right) \overset{(58)}{=} o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

- $j + 2 \leqslant i \leqslant k_0 + 2$. In this case $t \geqslant j$, because $\hat{T}_1$ and $\hat{T}_2$ may share their j-shells, meaning that at least $j + 2$ many j-simplices must be present, but have different bases, i.e. up to two sides of the (potential) j-shells may be automatically present as j-simplices because of K_1 and K_2. Therefore the contribution is

$$O\left(p_{k_0}^2 r^{j} n^{2k_0+4-i}\right) \overset{(60)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{r^{j+2}n^i}\right) \overset{(58)}{=} o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

Summing over all cases shows that $\mathbb{E}((\hat{X}_{j,k_0})^2) = (1 + o(1))\mathbb{E}(\hat{X}_{j,k_0})^2$, as desired. Thus, Chebyshev's inequality implies that $\hat{X}_{j,k_0} = (1 + o(1))\mathbb{E}(\hat{X}_{j,k_0})$ whp.

Finally, recall that in the case $s = 1$, $i = k_0 + 1$, we observed that whp there are no two copies of $\hat{M}_{j,k_0}$ that contain a common M_{j,k_0}, in which case all copies of $\hat{M}_{j,k_0}$ must have distinct associated copies of M_{j,k_0}, as claimed. $\square$

A.5 Proof of Proposition 6.5

Observe that for each $k \geqslant j$ we have

$$\begin{aligned} p_k = \frac{\alpha_k \log n + \beta_k}{n^{k-j+\gamma_k}}(k-j)! &= (1+\xi)\,\frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)! \\ &= \frac{\bar{\alpha}_k \log n + (1+\xi)\,\bar{\beta}_k + \bar{\alpha}_k \xi \log n}{n^{k-j+\bar{\gamma}_k}}(k-j)!, \end{aligned}$$

and therefore the first three statements follow directly. Furthermore, since λ_k and ν_k are dependent only on α_k and γ_k, and not on β_k, the fourth and sixth statements also follow.

For $k \geqslant j+1$, recall that $p_k \leqslant 1$ by Remark 3.2 and therefore we have

$$\mu_k = -(k-j+1)\sum\nolimits_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} + \begin{cases} \log\log n & \text{if } \alpha_k \neq 0, \\ \log(\beta_k) & \text{if } \alpha_k = 0. \end{cases}$$

We have that

$$\begin{aligned} \sum_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} &= \sum_{i=j+1}^{d} \frac{(1+\xi)\,\bar{\beta}_i + \bar{\alpha}_i \xi \log n}{n^{\bar{\gamma}_i}} \\ &\overset{\text{(A1)}}{=} \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \xi \cdot \left(\sum_{i=j+1}^{d} \bar{\alpha}_i \log n + \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} \right) \\ &\overset{\text{(A2)-(A3)}}{=} \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \xi \cdot \left(\sum_{i=j+1}^{d} \bar{\alpha}_i \log n + \sum_{i:\bar{\gamma}_i \neq 0} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + o(\log n) \right) \\ &= \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + (1+o(1))\xi \sum_{i=j+1}^{d} \bar{\alpha}_i \log n. \end{aligned} \tag{62}$$

Observe that (62) does not depend on k. For all k with $\alpha_k = 0$, note that μ_k contains the additional term

$$\log(\beta_k) = \log((1+\xi)\bar{\beta}_k) = \log(\bar{\beta}_k) + O(\xi),$$

while for all k with $\alpha_k \neq 0$, we have the same additional term $\log\log n$ in both μ_k and $\bar{\mu}_k$. Thus in total we have

$$\mu_k = \bar{\mu}_k - (1+o(1))(k-j+1)\xi \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n.$$

On the other hand, for $k = j$, observe that (A1) implies that if $p_j \leqslant 1$, then $\alpha_j = \bar{\alpha}_j = 0$. Furthermore,

$$\text{if } p_j \leqslant 1 \leqslant \bar{p}_j, \text{ then } \bar{\alpha}_j = \bar{\gamma}_j = 0 \text{ and } \bar{p}_j = \bar{\beta}_j = 1 + O(\xi). \tag{63}$$

Therefore, we have

$$\begin{aligned}
\mu_j &= -\sum\nolimits_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} + \begin{cases} 0 & \text{if } p_j > 1, \\ \log(\beta_j) & \text{if } p_j \leqslant 1 \end{cases} \\
&= -\sum\nolimits_{i=j+1}^{d} \frac{(1+\xi)\,\bar{\beta}_i + \bar{\alpha}_i \xi \log n}{n^{\bar{\gamma}_i}} + \begin{cases} 0 & \text{if } p_j > 1, \\ \log(\bar{\beta}_j) + \log(1+\xi) & \text{if } p_j \leqslant 1 \end{cases} \\
&= \bar{\mu}_j - \xi \sum\nolimits_{i=j+1}^{d} \frac{\bar{\beta}_i + \bar{\alpha}_i \log n}{n^{\bar{\gamma}_i}} + O(\xi) + \begin{cases} \log(\bar{\beta}_j) & \text{if } p_j \leqslant 1 \leqslant \bar{p}_j, \\ 0 & \text{otherwise} \end{cases} \\
&\overset{(62),(63)}{=} \bar{\mu}_j - (1+o(1))\,\xi \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n,
\end{aligned}$$

as required. $\square$

A.6 Proof of Lemma 6.7

We will prove the lemma with $c = \bar{\alpha}_{k_0}/4$, where k_0 is as defined in Definition 5.15 (b).

First observe that by Proposition 6.5 applied with $\xi = -1/\omega_0$, for any $j \leqslant k \leqslant d$ we have $\lambda_k = \bar{\lambda}_k$, $\nu_k = \bar{\nu}_k$, and

$$\mu_k = \bar{\mu}_k + (1+o(1))(k-j+1)\frac{1}{\omega_0}\sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n \geqslant \bar{\mu}_k + \frac{\bar{\alpha}_{k_0} \log n}{2\omega_0},$$

where we are using that $k - j + 1 \geqslant 1$. Thus we have

$$\lambda_k \log n + \mu_k + \nu_k \geqslant \bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k + \frac{\bar{\alpha}_{k_0} \log n}{2\omega_0},$$

therefore Lemma 5.14 and the fact that $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k = O(1)$ (because k is a critical dimension) imply that

$$\mathbb{E}(X_{j,k}) \geqslant \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{2\omega_0} + O(1)\right) \geqslant \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{3\omega_0}\right), \tag{64}$$

and thus in particular $\mathbb{E}(X_{j,k}) = \omega(1)$.

In order to apply a second moment argument, we will show that

$$\mathbb{E}((X_{j,k})^2) = (1+o(1))\mathbb{E}(X_{j,k})^2,$$

implying that whp $X_{j,k}$ is concentrated around its expectation. We first consider the case when $k \geqslant j+1$.

Let $\mathcal{T}_k$ denote the family of pairs $T = (K, C)$, where $K \subset [n]$ with $|K| = k+1$ and C is a j-subset of K. Each of these pairs may form a copy of $M_{j,k}$ with K as k-simplex and C as centre of the flower $\mathcal{F}(K, C)$.

Given two pairs $T_1 = (K_1, C_1)$ and $T_2 = (K_2, C_2)$, we define

- $s = s(T_1, T_2) \;:=\; \begin{cases} 1 & \text{if } K_1 = K_2, \\ 2 & \text{otherwise;} \end{cases}$
- $\mathcal{F}_h \;:=\; \mathcal{F}(K_h, C_h)$ for $h = 1, 2$;
- $t = t(T_1, T_2) \;:=\; |\mathcal{F}_1 \cup \mathcal{F}_2|$, i.e. the total number of (potential) petals.

By Proposition 5.13, the probability that two pairs in $\mathcal{T}_k$ both form a copy of $M_{j,k}$ is $(1 + o(1))p_k^s \bar{q}^{\,\tau t}$. With this observation, we can determine the contribution to $\mathbb{E}((X_{j,k})^2)$ made by those pairs with a fixed value of s.

- $s = 1$. Petals can be shared, but certainly $t \geqslant k - j + 1$ and the contribution is at most of order
$$O\left(n^{k+1} p_k \bar{q}^{\,\tau(k-j+1)}\right) \overset{(16)}{=} O(\mathbb{E}(X_{j,k})) = o(\mathbb{E}(X_{j,k})^2).$$
- $s = 2$. By definition, a petal cannot lie in any other k-simplex and thus only the pairs with $t = 2(k - j + 1)$ have a positive probability of both forming a copy of $M_{j,k}$. The number of such pairs is
$$(1 + o(1)) \binom{n}{k+1}^2 \binom{k+1}{j}^2$$
and thus these pairs provide a contribution of
$$(1 + o(1)) \binom{n}{k+1}^2 \binom{k+1}{j}^2 p_k^2 \bar{q}^{\,\tau 2(k-j+1)} \overset{(16)}{=} (1 + o(1))\mathbb{E}(X_{j,k})^2.$$

In total, we therefore have $\mathbb{E}((X_{j,k})^2) = (1 + o(1))\mathbb{E}(X_{j,k})^2$.

We now consider the case $k = j$. The proof is similar but simpler, since for a pair (K, C) to form a copy of $M_{j,j}$ we only require K to be an isolated j-simplex, and C to be the canonical choice (see Definition 5.6). On the other hand, we need to be careful if $p_j > 1$, since then p_j must be replaced by 1 in any probability calculations.

Recall that since j is a critical dimension we have that $\bar{\lambda}_j \log n + \bar{\mu}_j + \bar{\nu}_j = O(1)$ (see Definition 3.6). For the second moment of $X_{j,j}$, we count pairs of isolated j-simplices according to the size of their intersection i. Applying Proposition 5.13, we obtain
$$\begin{aligned}
\mathbb{E}((X_{j,j})^2) \;&=\; \mathbb{E}(X_{j,j}) + \sum_{i=0}^{j} \binom{n}{j+1}\binom{j+1}{i}\binom{n-j-1}{j+1-i} \min\{p_j, 1\}^2 \cdot (1 + o(1))\bar{q}^{\,2\tau} \\
&\overset{(17)}{\leqslant} \mathbb{E}(X_{j,j})^2 \left(1 + o(1) + \sum_{i=1}^{j} O(n^{-i})\right) \\
&=\; \mathbb{E}(X_{j,j})^2 \left(1 + o(1)\right).
\end{aligned}$$

Thus in both cases we have $\mathbb{E}((X_{j,k})^2) = (1 + o(1))\mathbb{E}(X_{j,k})^2$ and so by Chebyshev's inequality whp
$$X_{j,k} = (1 + o(1))\mathbb{E}(X_{j,k}) \overset{(64)}{>} \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{4\omega_0}\right),$$
as required. ∎

A.7 Proof of Proposition 9.7

Given an ordered m-simplex $\Phi = [v_0, \dots, v_m]$ and a vertex $u \notin \Phi$, define the ordered $(m+1)$-simplex

$$[u, \Phi] \; := \; [u, v_0, \dots, v_m],$$

for any $m \in [n-2]_0$.

Let $v \in [n]$ and consider a j-cochain f as in the statement. We define the $(j-1)$-cochain f_v that maps every $(j-1)$-simplex ρ to the value

$$f_v(\rho) = \begin{cases} f([v,\rho]) & \text{if } v \notin \rho; \\ 0_R & \text{otherwise.} \end{cases}$$

For any ordered j-simplex $\sigma = [v_0, \dots, v_j]$ we have

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) = f(\sigma) - \sum_{i=0}^{j} (-1)^i f_v\left([v_0, \dots, \hat{v}_i, \dots, v_j]\right). \tag{65}$$

If $v \notin \sigma$, then

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) \overset{(65)}{=} f(\sigma) - \sum_{i=0}^{j} (-1)^i f\left([v, v_0, \dots, \hat{v}_i, \dots, v_j]\right) = (\delta^j f)([v,\sigma]),$$

by definition of the operator δ^j.

If $v \in \sigma$ then $v = v_l$ for some $l \in [j]_0$, implying that $f_v([v_0, \dots, \hat{v}_i, \dots, v_j]) = 0$ for every $i \neq l$ and $f_v([v_0, \dots, \hat{v}_l, \dots, v_j]) = (-1)^l f(\sigma)$. Thus

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) \overset{(65)}{=} f(\sigma) - (-1)^{2l} f(\sigma) = 0_R.$$

Putting everything together

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) = \begin{cases} (\delta^j f)([v,\sigma]) & \text{if } v \notin \sigma; \\ 0_R & \text{otherwise.} \end{cases} \tag{66}$$

Recalling that every j-cochain of the form $f + g$ with g a j-coboundary has support of size at least $|\mathcal{S}|$, we have

$$n|\mathcal{S}| \leqslant \sum_{v \in [n]} |\text{supp}(f - \delta^{j-1} f_v)| = |\{(v,\sigma) : v \in [n],\ \sigma \in \text{supp}(f - \delta^{j-1} f_v)\}|. \tag{67}$$

For a pair (v,σ), by (66) it holds that σ is in the support of $f - \delta^{j-1} f_v$ if and only if $v \notin \sigma$ and the $(j+1)$-simplex $[v,\sigma]$ is in the support of $(\delta^j f)(u\sigma)$. Hence

$$\begin{aligned} n|\mathcal{S}| &\overset{(67)}{\leqslant} |\{(v,\rho) : v \in \rho,\ \rho \in \text{supp}(\delta^j f)\}| \\ &= (j+2)|\text{supp}(\delta^j f)| \\ &= (j+2)|\mathcal{D}(f)|, \end{aligned}$$

as required. ☐

A.8 Proof of Claim 9.9

Consider the j-simplex in step i of the exploration process described in Remark 9.5: there are at most $\binom{n}{k-j}$ many $(k+1)$-sets which we could potentially discover from this j-simplex, and of these we must choose $b_{i,k}$. From each of the chosen k-simplices, we find at most $\binom{k+1}{j+1} - 1 < \binom{k+1}{j+1}$ undiscovered j-simplices of $\mathcal{S}$. This holds for every $k \in \{j+1, \ldots, d\}$, thus this can happen in at most $\prod_{k=j+1}^{d} \binom{\binom{n}{k-j}}{b_{i,k}} 2^{\binom{k+1}{j+1} b_{i,k}}$ different ways. Hence, considering the choices for the initial j-simplex, the number of pairs $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with $\mathcal{S}$ traversable and with exploration matrix B is bounded from above by

$$\binom{n}{j+1} \prod_{i,k} \binom{\binom{n}{k-j}}{b_{i,k}} 2^{\binom{k+1}{j+1} b_{i,k}} \leqslant n^{j+1} \frac{\prod_k \left(\binom{n}{k-j} 2^{\binom{k+1}{j+1}} \right)^{t_k}}{\prod_{i,k} b_{i,k}!},$$

using that $\sum_{i=1}^{s} b_{i,k} = t_k$, for each $k \in \{j+1, \ldots, d\}$. □

A.9 Proof of Claim 9.10

The expected number of j-sets that do not form a $(j-1)$-simplex in $\mathcal{G}_\tau$ is bounded from above by

$$\begin{aligned} \binom{n}{j} (1 - p_{k_0})^{\binom{n-j}{k_0-j+1}} &\leqslant \binom{n}{j} (1 - \tau_0 \bar{p}_{k_0})^{\binom{n-j}{k_0-j+1}} \\ &\leqslant (1 + o(1)) n^j \exp\left(-\tau_0 n \cdot \frac{\bar{\alpha}_{k_0} \log n + \bar{\beta}_{k_0}}{k_0 - j + 1} \right) = o(1), \end{aligned}$$

thus by Markov's inequality whp $\mathcal{G}_\tau$ has a complete $(j-1)$-dimensional skeleton. □

A.10 Proof of Claim 9.11

Recall that we condition on the high probability event in Claim 9.10 and that this implies that Proposition 9.7 can be applied to f_τ (if it exists). Suppose that f_τ exists and write $s := |\mathcal{S}_\tau|$. Then $\mathcal{D}(f_\tau)$ comprises at least $\frac{sn}{j+2}$ many $(j+2)$-sets by Proposition 9.7. Each $A \in \mathcal{D}(f_\tau)$ is not allowed to be part of k-simplices of $\mathcal{G}_\tau$, for every $k \in \{j+1, \ldots, d\}$. There are $\binom{n-j-2}{k-j-1}$ many $(k+1)$-sets in $[n]$ that contain A, each of which contains $\binom{k+1}{j+2}$ many $(j+2)$-sets. Thus we have

$$|\mathcal{D}_k(f_\tau)| \geqslant \frac{sn\binom{n-j-2}{k-j-1}}{(j+2)\binom{k+1}{j+2}} \geqslant h_0 s n^{k-j}$$

for some positive constant h_0 and for every $k = j+1, \ldots, d$. □

A.11 Proof of Claim 9.12

Let a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with $\mathbf{t}(\mathcal{S}) = (t_{j+1}, \ldots, t_d)$ be given and recall that $s \geqslant \sum_{k=j+1}^{d} t_k =: Y_{\mathbf{t}}$ by (38). By Claim 9.11, the probability that $\mathcal{S}_\tau = \mathcal{S}$ is at most

$$\prod_{k=j+1}^{d} p_k^{t_k}(1-p_k)^{h_0 n^{k-j} s} \leqslant \tau^{Y_{\mathbf{t}}} \prod_{k=j+1}^{d} \bar{p}_k^{t_k}(1-\tau\bar{p}_k)^{h_0 n^{k-j} Y_{\mathbf{t}}} =: x(\tau). \tag{68}$$

The function $x(\tau)$ is positive and its derivative (with respect to τ) is

$$\frac{dx}{d\tau} = \frac{x(\tau)Y_{\mathbf{t}}}{\tau}\left(1 - \tau h_0 \sum_{k=j+1}^{d} \frac{n^{k-j}\bar{p}_k}{1-\tau\bar{p}_k}\right).$$

Recalling that the index k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$, we deduce that

$$\frac{dx}{d\tau} \leqslant \frac{x(\tau)Y_{\mathbf{t}}}{\tau}\left(1 - \tau h_0 \frac{\bar{\alpha}_{k_0}\log n + \bar{\beta}_{k_0}}{1-\tau\bar{p}_{k_0}}\right) < 0$$

for $\tau = \omega(1/\log n)$. Thus, since the derivative of $x(\tau)$ is negative throughout the whole range $\tau \geqslant \tau_0 = 1 - o(1)$, we have $x(\tau) \leqslant x(\tau_0)$ for all $\tau \geqslant \tau_0$, and therefore in the following calculations we may substitute τ_0 for τ.

Now Claim 9.9 implies that

$$\begin{aligned}
r_B \prod_{i,k} b_{i,k}! &\leqslant x(\tau_0) n^{j+1} \prod_{k=j+1}^{d} \left(\binom{n}{k-j} 2^{\binom{k+1}{j+1}}\right)^{t_k} \\
&\overset{(68)}{\leqslant} n^{j+1} \prod_{k=j+1}^{d} \left(\binom{n}{k-j} 2^{\binom{k+1}{j+1}} \tau_0 \bar{p}_k\right)^{t_k} (1-\tau_0\bar{p}_k)^{h_0 n^{k-j} Y_{\mathbf{t}}} \\
&\overset{(3)}{\leqslant} n^{j+1} \prod_{k} \left(\left(\Theta(1)\frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)^{t_k}\right. \\
&\qquad\qquad \left.\cdot \exp\left(-(1+o(1))h_0 Y_{\mathbf{t}}\left(\frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{\bar{\gamma}_k}}(k-j)!\right)\right)\right) \\
&\leqslant n^{j+1}\left(O(\log n)\right)^{Y_{\mathbf{t}}} n^{-(1+o(1))h_0\bar{\alpha}_{k_0}(k_0-j)! Y_{\mathbf{t}}} \\
&\leqslant n^{j+1} n^{-\hat{h}_0 Y_{\mathbf{t}}},
\end{aligned}$$

where $\hat{h}_0 := \frac{h_0\bar{\alpha}_{k_0}(k_0-j)!}{2}$. Now suppose that $s \geqslant \tilde{h}_2 \geqslant \frac{2(j+1)\sum_k \binom{k+1}{j+1}}{\hat{h}_0}$. Since $Y_{\mathbf{t}} \geqslant \frac{s}{\sum_k \binom{k+1}{j+1}}$, we can find another positive constant h_1 such that

$$r_B \leqslant \frac{n^{-h_1 s}}{\prod_{i,k} b_{i,k}!},$$

as desired. $\square$

A.12 Proof of Claim 9.13

For any exploration matrix $B = (b_{i,k})$, define $\mathbf{u}(B) = (u_{j+1}, \dots, u_d)$, where

$$u_k \;:=\; |\{i : b_{i,k} \geqslant n^{h_1/(d-j+1)}\}|.$$

Conversely, given $\mathbf{u} = (u_{j+1}, \dots, u_d)$ with $u_k \geqslant 0$, let $\mathcal{B}_\mathbf{u}$ be the set of all matrices B such that $\mathbf{u}(B) = \mathbf{u}$. Observe that each $B \in \mathcal{B}_\mathbf{u}$ is an $(s \times (d-j))$-matrix. There are $\prod_k \binom{s}{u_k}$ choices for which entries are large (i.e. which contribute to u_k), at most $n^{h_1/(d-j+1)}$ possibilities for each of the small entries and, since the sum of all the entries is $\sum_k t_k \leqslant s$, at most s possibilities for each of the large entries. Thus we obtain the (rather crude) upper bound

$$\begin{aligned} |\mathcal{B}_\mathbf{u}| &\leqslant \left(\prod\nolimits_k \binom{s}{u_k}\right) \left(n^{h_1/(d-j+1)}\right)^{s(d-j)-\sum_k u_k} s^{\sum_k u_k} \\ &\leqslant s^{2\sum_k u_k} \left(n^{h_1/(d-j+1)}\right)^{s(d-j)-\sum_k u_k}. \end{aligned} \tag{69}$$

Moreover, for $B \in \mathcal{B}_\mathbf{u}$

$$\prod\nolimits_{i,k} b_{i,k}! \geqslant \left(\left(n^{\frac{h_1}{d-j+1}}\right)!\right)^{\sum_k u_k} \geqslant n^{n^{\frac{h_1}{d-j+2}} \sum_k u_k}. \tag{70}$$

Putting everything together, the probability r_s that $\mathcal{S}_\tau$ of fixed size $s \geqslant \tilde{h}_1$ exists (together with the collection $\mathcal{T}(\mathcal{S}_\tau)$ of simplices) satisfies

$$\begin{aligned} r_s &\leqslant \sum\nolimits_\mathbf{u} \sum\nolimits_{B \in \mathcal{B}_\mathbf{u}} r_B \overset{\text{(Cl. 9.12)}}{\leqslant} \sum\nolimits_\mathbf{u} |\mathcal{B}_\mathbf{u}| \frac{n^{-h_1 s}}{\prod_{i,k} b_{i,k}!} \\ &\overset{(69),(70)}{\leqslant} \sum\nolimits_\mathbf{u} \left[\prod\nolimits_k \left(\frac{s^2}{n^{\frac{h_1}{d-j+1}} n^{n^{\frac{h_1}{d-j+2}}}}\right)^{u_k}\right] \frac{n^{\frac{h_1}{d-j+1} s(d-j)}}{n^{h_1 s}} \\ &\leqslant \sum\nolimits_\mathbf{u} 1 \cdot n^{-\frac{h_1}{d-j+1} s} \leqslant (s+1)^{d-j} \cdot n^{-\frac{h_1}{d-j+1} s} \leqslant n^{-h_2 s}, \end{aligned}$$

for some positive constant h_2.

Thus, the probability that $\mathcal{S}_\tau$ exists with $|\mathcal{S}_\tau| \geqslant \tilde{h} \geqslant \tilde{h}_1$ is at most

$$\sum\nolimits_{s \geqslant \tilde{h}} r_s \leqslant \sum\nolimits_{s \geqslant \tilde{h}} n^{-h_2 s} \leqslant \frac{n^{-h_2 \tilde{h}}}{1 - n^{-h_2}} \leqslant n^{-h_2 \tilde{h}/2},$$

which concludes the proof of Claim 9.13. $\square$

B Glossary

For the reader's convenience, we include a glossary of some of the most important terminology and notation defined in the paper.

B.1 Combinatorial terminology

Term	**Informal description**	**First defined**
j-shell	all j-simplices on $j+2$ vertices	Def. 2.1
K-localised set J	all simplices containing J are contained in K	Def. 5.1
j-flower	j-simplices within a k-simplex containing a common centre	Def. 5.2
copy of $\hat{M}_{j,k}$	K-localised j-flower with j-shell containing one petal	Def. 5.3
copy of $M_{j,k}$	K-localised j-flower	Def. 5.5
local j-obstacle	$(k+1)$-set K containing $k-j+1$ many K-localised j-simplices	Def. 7.2
traversability	notion of connectedness on j-simplices	Def. 9.3

B.2 Cohomology terminology

Term	**Informal description**	**First defined**
j-cochain	function on ordered j-simplices	p. 8
$C^j(\mathcal{G}; R)$	group of j-cochains	p. 8
coboundary operator δ^j	generates $(j+1)$-cochain from j-cochain	p. 8
j-cocycle	j-cochain in $\ker \delta^j$, i.e. all $(j+1)$-simplices have zero boundary	p. 8
j-coboundary	j-cochain in $\operatorname{im} \delta^{j-1}$, i.e. generated from $(j-1)$-cochain	p. 8
$H^j(\mathcal{G}; R)$	j-th cohomology group over R: $\ker \delta^j / \operatorname{im} \delta^{j-1}$	p. 8

B.3 Probabilities and birth times

Symbol	Informal description	First defined
$\bar{\mathrm{p}}$	vector defining process direction	p. 5
$\tau_{\max}$	τ at end of process: $1/\bar{p}_d$	p. 5
τ_j^*	time when last $\hat{M}_{j,k}$ disappears	p. 5
τ'	$1 - \frac{\log\log n}{10d\log n}$	p. 30
τ''	first time after τ' that no $M_{j,k}$ exists	p. 30

B.4 Random variables

Variable	Informal description	First defined
$X_{j,k}$	number of copies of $M_{j,k}$ in $\mathcal{G}_\tau$	Def. 5.7
$\hat{X}_{j,k}$	number of copies of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$	Def. 5.7

B.5 Parameters

Parameter	Informal description	First defined
$\bar{\alpha}_k$	logarithmic term in $\bar{p}_k$	p. 10
$\bar{\beta}_k$	sublogarithmic term in $\bar{p}_k$	p. 10
$\bar{\gamma}_k$	polynomial correction exponent in $\bar{p}_k$	p. 10
$\bar{p}_k$	probability of k-simplex existing at $\tau = 1$: $\dfrac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)!$	p. 10
$\bar{\lambda}_k$	$j + 1 - \bar{\gamma}_k - (k - j + 1)\sum_{i=j+1}^{d} \bar{\alpha}_i$	p. 10
$\bar{\mu}_k$	$-(k-j+1)\sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \begin{cases} \log\log n & \text{if } \bar{\alpha}_k \neq 0, \\ \log(\bar{\beta}_k) & \text{if } \bar{\alpha}_k = 0 \end{cases}$	p. 10
$\bar{\nu}_k$	$\begin{cases} -\log((j+1)!) & \text{if } k = j, \\ -\log(j!) - \log(k-j+1) + \log(\bar{\alpha}_k) & \text{if } \bar{\alpha}_k \neq 0, \\ -\log(j!) - \log(k-j+1) & \text{otherwise} \end{cases}$	p. 10
$\bar{k}$	index $j \leqslant \bar{k} \leqslant d$ such that $\bar{\lambda}_{\bar{k}} \log n + \bar{\mu}_{\bar{k}} + \bar{\nu}_{\bar{k}} = 0$	p. 11, (see also Def. 5.15, p. 26)
k_0	index $j + 1 \leqslant k_0 \leqslant d$ such that $\bar{\alpha}_{k_0} \neq 0$	p. 10, (see also Def. 5.15, p. 26)
ℓ	index $j \leqslant \ell \leqslant d$ such that at time τ_j^*, a copy of $\hat{M}_{j,\ell}$ vanishes	p. 26