Nicola Del Giudice, Dott. Mag.

Random hypergraphs and random simplicial complexes

DOCTORAL THESIS

to achieve the university degree of

Doktor der Naturwissenschaften

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Mihyun Kang, Ph.D.

Institute of Discrete Mathematics

Graz, April 2020

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Contents

1 Introduction and main results

1.1 Introduction

1.1.1 Motivation

A graph can be viewed as a mathematical tool to model a *network*, i.e. a collection of objects (called *vertices*) linked by pairwise relations (called *edges*). As real-life examples of graphs, we can consider social networks, where each pair of users can be linked if they accept each other as friends, or road maps, where two different points are correlated if there exists a feasible route between the two locations that they represent. In studying such networks, a typical question that may arise is: starting from a given vertex, is it possible to reach any other vertex using a sequence of the edges between pairs of vertices? If the answer is affirmative, we say that the graph representing the network is *connected*. When the graph is not connected, can we describe the "substructures" that satisfy this property, i.e. the *connected components*?

Indeed, when dealing with real-life networks we usually have to consider a huge number of vertices and often some "randomness" is involved: we can think of a navigation system where there is a certain probability that a known itinerary to our final destination becomes unusable due to roadworks. This can be mathematically formulated as a *random graph* model, where given a set of vertices the edges are present (or not) according to some probability distribution. Investigating properties such as connectedness in random graphs translates into probabilistic questions, for example: how can we choose the edge probability (as a function of the number of vertices) in such a way that the graph is very likely to be connected if the number of vertices is considerably large? How does the structure of the connected components evolve when increasing the edge probability?

Finally, in modelling real-life networks the concept of classical graphs might not be sufficient. For instance, the vertices do not necessarily have to be connected by pairwise relations, but might "cluster" in groups of various sizes. In other words, a network can be defined by a set of vertices and a collection of relations between subsets of vertices of arbitrary cardinality, thus leading to "higher-dimensional" notions of graphs. Furthermore, for these more general networks, more than one natural concept of higher-dimensional connectedness can be defined.

The major aim of this thesis is to analyse random discrete structures that represent higher-dimensional analogues of graphs such as *random hypergraphs* and *random simplicial complexes*, for which we investigate several notions of *high-order connectedness*.

1.1.2 Phase transitions in random graphs

With their series of seminal papers [28–31], Erdős and Rényi can be considered the founders of the theory of random graphs. Although their results are stated for the *uniform random graph* model $G(n, m)$ (a graph chosen uniformly at random from among all graphs on vertex set $[n] := \{1, 2, \ldots, n\}$ with m edges), currently it is more common to study a different model, namely the *binomial random graph* $G(n, p)$. This model is defined as a graph on vertex set $[n]$, where each pair of vertices forms an edge with probability p independently. Indeed, it is well known that $G(n, p)$ is essentially equivalent to $G(n, m)$ if the edge probability p is approximately $m/\binom{n}{2}$ (see e.g. [39, Section 1.4]). In this way, all the classical results by Erdős and Rényi can be transferred to the $G(n, p)$ model. Thus, in this thesis we restrict our attention to $G(n, p)$.

When dealing with $G(n, p)$, we consider the edge probability p as a function of the number of vertices, i.e. $p = p(n)$, and investigate the behaviour of $G(n, p)$ for very large n. In particular, Erdős and Rényi studied structural properties of $G(n, p)$ that occur *with high probability* (*whp* for short), meaning that the property holds with probability tending to 1 as n tends to infinity. Remarkably, for many of these properties even a small change in the probability p can induce a drastic difference in the behaviour of $G(n, p)$; in other words, $G(n, p)$ undergoes a *phase transition* with respect to these properties.

One of the most famous results by Erdős and Rényi [28] deals with the property of $G(n, p)$ being connected: there exists a sharp *threshold*, namely $\frac{\log n}{n}$, such that if p is "below" this value then whp $G(n, p)$ is not connected, while for p "above" the threshold then the situation dramatically changes and whp $G(n, p)$ is connected. They also investigated connectedness around the threshold, in the so-called *critical window*, by determining an explicit asymptotic expression for the probability of $G(n, p)$ being connected.

Theorem 1.1.1 ([28]). *Let $(c_n)_{n\geq 1}$ be a sequence of real numbers and let $p = \frac{\log n + c_n}{n}$. Then*

$$\Pr\left(G(n,p) \textit{ is connected}\right) \xrightarrow{n\to\infty} \begin{cases} 0 & \textit{if } \ c_n \to -\infty, \\ e^{-e^{-c}} & \textit{if } \ c_n \to c \in \mathbb{R}, \\ 1 & \textit{if } \ c_n \to +\infty. \end{cases}$$

The value $\frac{\log n}{n}$ is closely related to the presence of isolated vertices, which can be viewed as the "smallest" obstacles to the connectedness of $G(n, p)$. Indeed, a first moment calculation shows that around this value the expected number of isolated vertices becomes very small. This suggests that these minimal obstructions are also the *critical* ones, in the sense that we expect $G(n, p)$ to become connected whp around the same time when all isolated vertices disappear.

This intuition was made formal by Bollobás and Thomason [15], who studied the *random graph process*, in which starting from the empty graph on vertex set $[n]$, at any following time-step one edge is chosen uniformly at random from among all edges not already selected and added to the graph. At time m the process is distributed as

$G(n,m)$, and therefore also asymptotically equivalent to $G(n,p)$, for the appropriate choice of p. Thus we may view $G(n,p)$ as a process, where if we slowly increase p from 0 to 1, the edges appear one by one. Bollobás and Thomason proved the following.

Theorem 1.1.2 ([15]). *With high probability the random graph process becomes connected at the time when the last isolated vertex disappears.*

A similar phenomenon to Theorem 1.1.1 also characterises $G(n,p)$ in an earlier stage of its evolution, i.e. before the critical threshold for connectedness. In [29], Erdős and Rényi studied the asymptotic *order* the connected components, i.e. the number of vertices they contain, again highlighting a surprising behaviour. If the probability p is smaller than $\frac{1}{n}$, whp $G(n,p)$ consists of a collection of "small" components of order at most logarithmic, whereas as soon as p becomes larger than the critical value $\frac{1}{n}$, whp the largest component reaches linear order and all the other components are of at most logarithmic order. In other words, $\frac{1}{n}$ is the threshold for what is commonly known as the emergence of a unique *giant component*. Interestingly, the component structure also changes dramatically when p is below or above this threshold. Indeed, in the subcritical regime whp the largest components are *acyclic*, and thus trees. In the supercritical regime, the largest component, which suddenly reaches linear size, undergoes a drastic change in its structure: whp the newly-born giant component is *complex*, i.e. contains at least two cycles, while the other components are trees or unicyclic.

Later on, this phenomenon was studied in more depth and the classical results by Erdős and Rényi were refined in several respects. In particular, Bollobás [9] and Łuczak [50] examined the evolution of the largest components considering $p = \frac{1\pm\varepsilon}{n}$, where $\varepsilon = \varepsilon(n)$ tends to zero "slowly enough", and obtained more precise bounds on the component orders. We summarise these results as follows.

Theorem 1.1.3 ([9, 29, 50]). *Let $\varepsilon = \varepsilon(n)$ be such that $\varepsilon \to 0$ and $\bar{\lambda} := \varepsilon^3 n \to \infty$. For every $i \in \mathbb{N}$, let $C_i = C_i(G(n,p))$ denote the order (i.e. number of vertices) of the i-th largest connected component in $G(n,p)$. Then with high probability the following holds.*

(i) If $p = \frac{1-\varepsilon}{n}$, for every constant $i \geq 1$ the i-th largest component is a tree. Furthermore, for any function ω of n such that $\omega \to \infty$,

$$\left| C_1 - \alpha^{-1}\left(\log\bar{\lambda} - \frac{5}{2}\log\log\bar{\lambda}\right)\right| \leq \omega\varepsilon^{-2},$$

where $\alpha := -\varepsilon - \log(1-\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$, and $C_i = (2+o(1))\varepsilon^{-2}\log\bar{\lambda}$ for $i \geq 2$.

(ii) If $p = \frac{1+\varepsilon}{n}$, the largest component is complex and

$$C_1 = (1+o(1))2\varepsilon n.$$

Moreover, for every constant $i \geq 2$ the i-th largest component is a tree with $C_i = (2+o(1))(\varepsilon^{-2}\log\bar{\lambda})$ and in particular, for any function ω of n such that

$\omega \to \infty$,

$$\left| C_2 - \beta^{-1}\left(\log \bar{\lambda} - \frac{5}{2}\log\log \bar{\lambda}\right)\right| \le \omega\varepsilon^{-2},$$

where $\beta := \varepsilon - \log(1+\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$.

Observe that Theorem 1.1.3 states that in the supercritical case the largest component is substantially larger than the second largest, thus showing that there exists a unique giant component. Moreover, the evolution of the components displays a *symmetry* property for $G(n,p)$: the supercritical random graph with the giant component removed behaves as the subcritical random graph.

We also note that the coefficient $-\frac{5}{2}$ before the $\log\log\bar{\lambda}$ factor arises from the asymptotic number of various families of labelled trees, as we will explain in more detail in Section 1.2.1.

1.1.3 Random hypergraphs

Inspired by these classical results on $G(n,p)$, many different random models and concepts of connectedness have been investigated. In this thesis, we are particularly interested in *higher-dimensional* analogues of random graphs, for which several different notions of connectedness can be defined at various complexity levels.

Among all the possible generalisations of a graph, one of the most natural is *hypergraphs*. Just as a graph consists of a set of vertices and a set of pairs of vertices (called edges), a *k-uniform hypergraph* consists of a set of vertices and a set of *hyperedges*, where each hyperedge is a set of k vertices (also called *k-set*), for $k \ge 2$. Thus, a graph can also be viewed as a 2-uniform hypergraph and for $k \ge 3$ we obtain a higher-dimensional structure. Similarly, we can generalise the random graph model $G(n,p)$: given integers $n, k \ge 2$, the *binomial random k-uniform hypergraph* $\mathcal{H}_p = \mathcal{H}(k;n,p)$ is the k-uniform hypergraph defined on vertex set $[n]$, where each k-set of vertices forms a hyperedge with probability $p = p(n)$ independently.

An obvious notion of connectedness in a k-uniform hypergraph is so-called *vertex-connectedness*: we say that two vertices v and w in a k-uniform hypergraph are connected if there exists a sequence $e_1, \ldots, e_\ell$ of hyperedges such that $v \in e_1$, $w \in e_\ell$ and for each $i = 1, \ldots, \ell-1$ we have $e_i \cap e_{i+1} \neq \emptyset$. It is easy to see that vertex-connectedness defines an equivalence relation on the set of vertices, whose equivalence classes are the connected components.

Vertex-connectedness has been extensively studied and analogues of the classical results on random graphs have been obtained for the random hypergraph. In many cases, the proof ideas of the graph case naturally extend to this higher-dimensional context. For example, similarly as in Theorems 1.1.1 and 1.1.2, the threshold for vertex-connectedness in $\mathcal{H}_p$ corresponds to the threshold for disappearance of the last isolated vertex, that occurs at around $p = \frac{\log n}{n^{k-1}}(k-1)!$, as follows from both [19] and [57] as a special case. As for the evolution of the vertex-connected components, Schmidt-Pruzan and Shamir [59] first underlined how the random hypergraph shows a similar behaviour to the random graph. They proved that $\mathcal{H}_p$ undergoes a phase transition at around $p = \frac{1}{(k-1)\binom{n}{k-1}}$, after which the largest component suddenly passes from being of logarithmic order to linear order, a result that was later refined (see e.g. [7,8,13,47,58]), thus obtaining an analogue of Theorem 1.1.3.

Nonetheless, vertex-connectedness is not the only natural notion of connectedness in k-uniform hypergraphs. Indeed, this is already clear from the definition: two vertices are considered connected if they belong to the endpoints of a common sequence of intersecting hyperedges, but there are several different ways of defining this sequence. For instance, we may distinguish different notions of connecting sequences according to the size of the intersection between two consecutive hyperedges. This leads to the following definition of *high-order* connectedness.

Definition 1.1.4. Given $j \in [k-1]$, a pair $\{J_a, J_b\}$ of j-sets of vertices in a k-uniform hypergraph is called *j-tuple-connected* (or *j-connected*, for short) if there exists a sequence of hyperedges $e_1, \dots, e_\ell$ such that $J_a \subset e_1$, $J_b \subset e_\ell$, and $|e_i \cap e_{i+1}| \geq j$ for each $i \in [\ell-1]$. A *j-(tuple-)component* is a maximal collection of pairwise j-connected j-sets. We say that a k-uniform hypergraph on vertex set $[n]$ is *j-(tuple-)connected* if all the j-sets of vertices lie in the same j-component.

Observe that the case $j = 1$ corresponds to vertex-connectedness, which has been widely analysed, as mentioned before. In contrast, the case $j \geq 2$ is not yet well-understood. Indeed, in this case the proof ideas from the graph case do not extend so easily, mainly because j-components consist of j-sets rather than vertices and thus it is possible that two different j-components share common vertices (as long as they do not share j-sets). For this reason, the behaviour of the order of the j-components, i.e. the number of j-sets they contain, is strongly dependent on their structure (i.e. how the j-sets intersect with each other), making the analysis more complicated.

Concerning the existence of a threshold for the j-connectedness of $\mathcal{H}_p$, the case $j = 1$ follows from the result on vertex-connectedness by Schimdt-Pruzan and Shamir [59], while the case $j = k-1$ was proved by Kahle and Pittel [44]. More generally, Cooley, Kang, and Koch [19] determined the threshold for the j-connectedness of $\mathcal{H}_p$ for any $j \in [k-1]$. They proved that the property of being j-connected undergoes a sharp phase transition at $p = \frac{j \log n}{\binom{n-j}{k-j}}$. Even more strongly, they proved a hitting time result analogous to Theorem 1.1.2, by showing that this value corresponds to the threshold for the disappearance of the last isolated j-set, i.e. a j-set which is not contained in any hyperedge.

Cooley, Kang, and Person [22] first studied the evolution of the j-components before the threshold for j-connectedness, for any $j \in [k-1]$. Similarly as the case $j = 1$, a phase transition occurs at around probability $p = \frac{1}{(\binom{k}{j}-1)\binom{n-j}{k-j}}$: whp before this critical value the largest j-component has at most logarithmic order, whereas once above the critical value the largest j-component is approximately of order $\Theta(n^j)$. Observe that we should not expect the largest j-component to reach only linear size because, as noted before, it consists of j-sets rather than vertices and thus should contain a constant fraction of all the $\binom{n}{j} = \Theta(n^j)$ many j-sets. Cooley, Kang, and Koch [20] later improved this result, by giving a more precise bound on the asymptotic order of the largest j-component in the supercritical regime and proving that the second largest component has much smaller order, thus showing that whp a unique giant j-component of order $\Theta(n^j)$ emerges.

Theorem 1.1.5 ([20, 22]). *Given integers $k \geq 2$ and $j \in [k-1]$, let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \to 0$, $\varepsilon^3 n^j \to \infty$ and $\varepsilon^2 n^{1-2\delta} \to \infty$, for some constant $\delta > 0$. Then with high probability the following holds.*

(i) If $p = \frac{1-\varepsilon}{(\binom{k}{j}-1)\binom{n-j}{k-j}}$, all j-components of the random hypergraph $\mathcal{H}(k; n, p)$ have order $O(\varepsilon^{-2}\log n)$.

(ii) If $p = \frac{1+\varepsilon}{(\binom{k}{j}-1)\binom{n-j}{k-j}}$, the order of the largest j-component of the random hypergraph $\mathcal{H}(k; n, p)$ is $(1 \pm o(1))\frac{2\varepsilon}{\binom{k}{j}-1}\binom{n}{j}$, while all other j-components have order $o(\varepsilon n^j)$.

Comparing this result with Theorem 1.1.3, we notice that the analogy with the graph case is not complete. In particular, the subcritical case was not fully analysed: what is the precise asymptotic order of the largest j-component in the subcritical random hypergraph? And what can we say about the structure of the largest j-component in this regime? Can it be described as a "higher-dimensional tree" as the graph case would suggest, or does it give rise to a more complex structure?

In Chapter 2, we address these questions and derive an analogous result to Theorem 1.1.3. In particular, we study the precise asymptotic *size* (i.e. number of hyperedges) of the largest j-components in the subcritical hypergraph and obtain a structural description of these components, showing that whp they are *hypertrees* (Theorem 1.2.1).

1.1.4 Random simplicial complexes

An alternative approach to find a higher-dimensional extension of a graph is given by simplicial complexes. A *simplicial complex* (or *complex* for short) $\mathcal{C}$ is defined by a vertex set V and a collection of non-empty subsets of V, called *simplices*, with the additional properties that this collection is *downward-closed*, i.e. every non-empty subset of a simplex also lies in $\mathcal{C}$, and furthermore the singleton $\{v\}$ is in $\mathcal{C}$ for every $v \in V$. For any non-negative integer k, the elements of $\mathcal{C}$ of cardinality $k+1$ are called *k-simplices* and we say that $\mathcal{C}$ is a *k-dimensional complex* (or simply *k-complex*) if $\mathcal{C}$ has no $(k+1)$-simplices. If $\mathcal{C}$ is k-dimensional, for each $0 \leq j \leq k-1$ the *j-skeleton* of $\mathcal{C}$ is the j-dimensional (sub)complex formed by all i-simplices in $\mathcal{C}$ with $0 \leq i \leq j$. Observe that a graph can be viewed as a 1-complex, where the vertices are the 0-simplices and the edges are the 1-simplices.

Simplicial complexes are typical objects studied in algebraic topology and in particular *homology theory* can be used to define a higher-dimensional notion of connectedness. Informally speaking, homology and cohomology groups can be viewed as algebraic objects which describe "holes" of various dimensions in simplicial complexes (see Section 3.2.3 for the formal definition and an overview about cohomology notation). For example, the vertex-connected components of a simplicial complex are related to the zero-th cohomology group, in the sense that asking this group to be "as small as possible" is equivalent to requiring that the complex consists of a single component, i.e. there are no zero-dimensional holes. Similarly, the dimension of higher cohomology groups gives information about the presence of higher-dimensional holes. This motivates the choice of defining a notion of higher-dimensional connectedness by the *vanishing* of the cohomology groups for every (reasonable) dimension, as we will see more formally later (Section 1.2.2).

Linial and Meshulam [48] first introduced a model of random 2-dimensional simplicial complexes, which was later generalised by Meshulam and Wallach [53] for general dimension $k \geq 2$. The random k-complex $\mathcal{Y}_p = \mathcal{Y}(k; n, p)$ is a k-dimensional complex on vertex set $[n]$ with full $(k-1)$-dimensional skeleton, where every $(k+1)$-set forms a k-simplex with probability p independently. The presence of the full $(k-1)$-skeleton excludes the presence of i-dimensional holes for every $i \in [k-2]$. For example, in $\mathcal{Y}(2; n, p)$ all the possible 1-simplices (i.e. edges) formed by pair of vertices in $[n]$ are included, yielding a 2-complex which is automatically vertex-connected, and thus does not contain any zero-dimensional holes. Hence, in order to have a connected k-complex we only have to preclude the existence of $(k-1)$-dimensional holes. In particular, Linial and Meshulam (for $k = 2$), and subsequently Meshulam and Wallach (for $k \geq 2$), investigated the vanishing of the $(k-1)$-th cohomology group $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)$ with coefficients in the two-element field $\mathbb{F}_2$ and showed that it has a sharp threshold at $p = \frac{k \log n}{n}$.

Theorem 1.1.6 ([48, 53])**.** *Let $k \geq 2$ and let ω be any function of n which tends to infinity as $n \to \infty$. Then with high probability the following holds.*

(i) If $p = \frac{k \log n - \omega}{n}$, then $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) \neq 0$;

(ii) If $p = \frac{k \log n + \omega}{n}$, then $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) = 0$.

Indeed, Meshulam and Wallach [53] showed that the same result holds with the coefficient group $\mathbb{F}_2$ replaced by any finite abelian group.

Similarly as we discussed for $G(n, p)$, the complex $\mathcal{Y}_p$ can also be naturally interpreted as a process, where k-simplices appear one by one. Kahle and Pittel [44] further proved a hitting time result (analogous to Theorem 1.1.2) but only for the case $k = 2$, relating the threshold for cohomological connectedness to the disappearance of the last *minimal obstruction*. More precisely, they showed that whp $H^1(\mathcal{Y}(2; n, p); \mathbb{F}_2)$ vanishes at the exact moment when the last isolated 1-simplex—i.e. a 1-simplex which is not contained in any 2-simplex of $\mathcal{Y}(2; n, p)$—disappears. Furthermore, to achieve a full analogy with Theorem 1.1.1, they also looked at the behaviour of $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)$ for general $k \geq 2$ inside the critical window given by the threshold, i.e. for $p = \frac{k \log n + O(1)}{n}$. They showed that the dimension of the $(k-1)$-th cohomology group tends in distribution to a Poisson random variable and thus obtained an asymptotic expression for the probability of this group vanishing.

Theorem 1.1.7 ([44])**.** *Let $k \geq 2$ and $(c_n)_{n \geq 1}$ be a sequence of real numbers such that $c_n \to c \in \mathbb{R}$. If $p = \frac{k \log n + c_n}{n}$, then $\dim\left(H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)\right)$ converges in distribution to a*

Poisson random variable with expectation $e^{-c}/k!$. *In particular, we have*

$$\Pr\left(H^{k-1}(\mathcal{Y}_p;\mathbb{F}_2)=0\right) \xrightarrow{n\to\infty} e^{-e^{-c}/k!}.$$

Since then, many other concepts of connectedness as well as different models of random simplicial complexes have been investigated (see e.g. [43] for an overview). In this thesis, we try to bridge the gap between hypergraphs and simplicial complexes, by considering random complexes which in some sense originate from a random hypergraph. In contrast with $\mathcal{Y}_p$, where the *full* $(k-1)$-skeleton is added, we define the random k-complex $\mathcal{G}_p = \mathcal{G}(k;n,p)$ where, once the k-simplices are chosen with probability p independently, we include only those simplices that are truly *necessary* to have a simplicial complex (see Definition 1.2.3). In other words, $\mathcal{G}_p$ can be viewed as the k-complex generated from a random $(k+1)$-uniform hypergraph by taking the downward-closure of the hyperedges.

Since in the model $\mathcal{G}_p$ the $(k-1)$-dimensional skeleton is not automatically complete, we need to consider all cohomology groups of dimension up to $k-1$. For this reason, for every $j\in[k-1]$ we introduce $\mathbb{F}_2$-*cohomological* j-*connectedness* of $\mathcal{G}_p$ as the vanishing of the cohomology groups $H^i(\mathcal{G}_p;\mathbb{F}_2)$ for all $i\in[j]$ and the zero-th cohomology group being isomorphic to $\mathbb{F}_2$ (Definition 1.2.4).

In Chapter 3, we study this property and prove results analogous to Theorems 1.1.6 and 1.1.7. More specifically, we determine the *sharp threshold* for $\mathbb{F}_2$-cohomological j-connectedness in $\mathcal{G}_p$ and prove a *hitting time* result, by relating this threshold to the disappearance of the last *minimal obstruction* (Theorem 1.2.5). Furthermore, we analyse the asymptotic distribution of the dimension of the j-th cohomology group, within the *critical window* associated with this threshold (Theorem 1.2.7).

In Chapter 4 we will also consider a generalisation of both $\mathcal{Y}_p$ and $\mathcal{G}_p$, i.e. a model of simplicial complex originated from the downward-closure of a *non-uniform* hypergraph (Definition 1.2.8), for which we examine the vanishing of the cohomology groups over an *arbitrary* abelian group. Similarly to the uniform case, we show a hitting time result for this property (Theorem 1.2.9) and study the behaviour of the cohomology groups in the critical window (Theorem 1.2.10).

1.1.5 Overview

This thesis is structured as follows.

In Section 1.2 we present the main results of the thesis. More precisely, in Section 1.2.1 we study the *subcritical* regime of the phase transition for the appearance of the giant j-component in the binomial random k-uniform hypergraph (Theorem 1.1.5). We derive a more precise asymptotic description of the largest j-components in this range, determining their size, order, and structure (Theorem 1.2.1 and Corollary 1.2.2). In Sections 1.2.2 and 1.2.3, we consider random simplicial complexes obtained from the downward-closure of the *uniform* and *non-uniform* random hypergraphs (Definitions 1.2.3 and 1.2.8, respectively). For each of these models, we define a notion of high-order connectedness according to the vanishing of the cohomology groups with coefficients over some suitable abelian group. We determine the sharp threshold for this property and prove a hitting time result for a natural process interpretation of these random complex models (Theorems 1.2.5 and 1.2.9). Furthermore, we investigate the asymptotic behaviour of the cohomology groups inside the critical window (Theorems 1.2.7 and 1.2.10).

Section 1.3 is devoted to the key techniques that are used to prove the main results of the thesis. First, in Section 1.3.1 we present a two-type branching process and the upper coupling argument that we use to analyse the structure of the j-components in the subcritical random hypergraph. In Section 1.3.2 we investigate the vanishing of cohomology groups of random simplicial complexes and underline the relation between the threshold for this property and the absence of minimal obstructions. Indeed, showing the existence of this threshold turns out to be particularly challenging, because this notion of cohomological connectedness is *not* a monotone property (see e.g. Examples 3.3.2 and 4.4.11). Thus, we need to develop sophisticated tools in order to overcome the difficulties arising from the non-monotonicity. In Section 1.3.3 we present how we can suitably subdivide the subcritical regime into intervals and ensure the presence of obstructions throughout each of the intervals. In Section 1.3.4 we define a breadth-first search process, which we use in the supercritical case to carefully bound the number of "obstacles" to cohomological connectedness and thus prove that whp they do not exist. We take a closer look at non-uniform simplicial complexes in Section 1.3.5, describing how we can determine the critical direction for the vanishing of the cohomology groups. In Section 1.4 we discuss open problems that could lead to future research.

The thesis is based on the following research papers.

(1) *Subcritical random hypergrhaps, high-order components, and hypertrees*
O. Cooley, W. Fang, N. Del Giudice, and M. Kang, accepted for publication in SIAM Journal on Discrete Mathematics (2019), 1–30.

An extended abstract appeared in the Proceedings of Analytic Algorithmics and Combinatorics (ANALCO19), pp. 111–118.

(2) *Vanishing of cohomology groups of random simplicial complexes*
O. Cooley, N. Del Giudice, M. Kang, and P. Sprüssel, Random Structures & Algorithms 56 (2020), 461–500.

An extended abstract appeared in the Proceedings of the 29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018), pp. 7:1–7:14.

(3) *Cohomology groups of non-uniform random simplicial complexes*
O. Cooley, N. Del Giudice, M. Kang, and P. Sprüssel, in preparation (2020), 1–56.

An extended abstract appeared in the Proceedings of the European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2019), Acta Math. Univ. Comenianae Vol. LXXXVIII, 3 (2019), pp. 553–560.

In this thesis, these papers correspond to Chapters 2–4, respectively.

1.2 Main results

In this section we describe the main results of this thesis.

In Section 1.2.1 we first discuss the notion of j-tuple-connectedness (Definition 1.1.4) in the binomial random k-uniform hypergraph $\mathcal{H}_p$. We consider the phase transition for the appearance of the giant j-component (Theorem 1.1.5) and focus on the subcritical regime. We determine the asymptotic *size* (i.e. number of hyperedges) of the largest j-components in the subcritical $\mathcal{H}_p$, and we give a description of their structure, showing that whp they are *hypertrees*, a higher-dimensional counterpart of trees in graphs (Theorem 1.2.1), analogously to Theorem 1.1.3.

Sections 1.2.2 and 1.2.3 deal with random simplicial complexes generated by the downward-closure of random hypergraphs. We define a notion of higher-order connectedness as the vanishing of the cohomology groups over a suitable coefficient group, for which we prove results analogous to Theorems 1.1.1 and 1.1.2 (among other results).

More specifically, in Section 1.2.2 we consider the random complex obtained as the downward-closure of the *uniform* random hypergraph and study $\mathbb{F}_2$-cohomological j-connectedness. We find the sharp threshold for this property and show that whp this corresponds to the time when the last minimal obstruction disappears (Theorem 1.2.5). Furthermore, we obtain a precise asymptotic expression for the probability of the complex being $\mathbb{F}_2$-cohomologically j-connected within the critical window (Theorem 1.2.7).

In Section 1.2.3 we define a more general model of simplicial complex generated by *non-uniform* hypergraphs, where sets of vertices of any size have different probabilities of forming hyperedges. For this model, we investigate the vanishing of cohomology groups over an *arbitrary abelian group* R and obtain a hitting time result which relates the threshold for this property to the absence of minimal obstruction (Theorem 1.2.9). Finally, we obtain an asymptotic description of the j-th cohomology group over R inside the critical window associated with the threshold (Theorem 1.2.10).

1.2.1 Subcritical random hypergraphs

We consider the binomial random k-uniform hypergraph $\mathcal{H}_p = \mathcal{H}(k; n, p)$ and the notion of j-tuple-connectedness (Definition 1.1.4) for $j \in [k-1]$. Theorem 1.1.5 states that $\mathcal{H}_p$ undergoes a phase transition at

$$p_g := \frac{1}{\left(\binom{k}{j} - 1\right)\binom{n-j}{k-j}}, \tag{1.1}$$

where a unique giant j-component of order (i.e. number of j-sets it contains) $\Theta(n^j)$ emerges. We take a closer look at the subcritical regime (i.e. statement (i) of Theorem 1.1.5), where the largest j-component has at most logarithmic order. In the main result of this section (Theorem 1.2.1) we precisely compute the asymptotic *size* (i.e. number of hyperedges) of the largest j-components and determine their structure. Indeed, we show that whp the largest j-components are *hypertrees*, meaning that their structure naturally extends the notion of tree in a graph to this higher-dimensional context. This will allow us to also determine the precise asymptotic *order* of the largest j-components (Corollary 1.2.2).

Note that in Definition 1.1.4 we defined j-connectedness between a pair of j-sets through the existence of a sequence of hyperedges linking the two j-sets, where every two consecutive hyperedges intersect in at least j vertices. Hence, when considering a j-component $\mathcal{J}$, although consisting of a collection of j-sets it is naturally equipped with the set of hyperedges that are used for the connecting sequences, i.e. the set $\mathcal{K}_{\mathcal{J}}$ of hyperedges containing any of the j-sets in $\mathcal{J}$. With a slight abuse of terminology, we say that a j-component $\mathcal{J}$ also contains the set of hyperedges $\mathcal{K}_{\mathcal{J}}$. Thus in the following, rather than investigating the order of the largest j-components we focus on their *size*. In the probability range of our interest the size and the order of the largest j-components will only differ approximately by a constant factor, thus we also derive a result on the order (Corollary 1.2.2).

In view of a generalisation of Theorem 1.1.3, we are interested in describing the structure of the largest j-components. To this end, we define a higher-dimensional notion of a tree. A *hypertree j-component* (or simply *hypertree*) is a j-component containing as many j-sets as possible given its size, i.e. if we denote by s the size and by t the order of the j-component, then $t = 1 + \left(\binom{k}{j} - 1\right) s$. Indeed, the case $k = 2$ and $j = 1$ gives the classical concept of tree in a graph, where given the number s of edges, we have $t = 1 + s$ many vertices.

We show that the largest subcritical j-components are whp hypertrees. We also determine their asymptotic size.

Theorem 1.2.1. *Let $k \geq 2$ be an integer, $j \in [k-1]$, and $\varepsilon = \varepsilon(n)$ with $0 < \varepsilon < 1$ such that $\frac{\varepsilon^4 n}{(\log n)^2} \to \infty$. For $i \in \mathbb{N}$, let $\mathcal{L}_i = \mathcal{L}_i(\mathcal{H}_p)$ be the i-th largest j-component of $\mathcal{H}_p = \mathcal{H}_p(k; n, p)$, and let L_i be the size (i.e. number of hyperedges) of $\mathcal{L}_i$. Let $m \in \mathbb{N}$ be fixed.*

If $p = (1-\varepsilon)p_g$, with p_g defined in (1.1), then with high probability for any $1 \leq i \leq m$, the component $\mathcal{L}_i$ is a hypertree and has size

$$L_i = \delta^{-1}\left(\log \lambda - \frac{5}{2}\log\log \lambda + O_p(1)\right),$$

where $\delta := -\varepsilon - \log(1-\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$ and $\lambda := \varepsilon^3\binom{n}{j}$.

We observe that the tree-like structure of the components is a key ingredient of our proof. This is reflected in the coefficient $-\frac{5}{2}$ before the $\log\log\lambda$ factor in Theorem 1.2.1, which is the same as Theorem 1.1.3. Indeed, the asymptotic number of trees on t vertices is of the form $c \cdot t!\gamma^t t^{-5/2}$, with c and γ depending on the properties of the considered family of trees. The critical exponent $-\frac{5}{2}$ is thus transferred to the coefficient $-\frac{5}{2}$ of the $\log\log\lambda$ factor.

Theorem 1.2.1 tells us that whp the largest components are hypertrees. In this case, we exactly know the relation between size and order of the j-components, therefore we obtain the following result.

Corollary 1.2.2. *Let $k, j, \varepsilon, \delta, \lambda, \mathcal{L}_i, p$ be as in Theorem 1.2.1. For $i \in \mathbb{N}$, let $M_i = M_i(\mathcal{H}_p)$ be the order (i.e. number of j-sets) of $\mathcal{L}_i$. Let $m \in \mathbb{N}$ be fixed. Then with high probability for any $1 \leq i \leq m$ we have*

$$M_i = \left(\binom{k}{j} - 1\right) L_i + 1 = \left(\binom{k}{j} - 1\right) \delta^{-1} \left(\log \lambda - \frac{5}{2} \log \log \lambda + O_p(1)\right).$$

1.2.2 Random simplicial complexes: uniform case

We consider the random k-complex $\mathcal{G}_p$ generated from the binomial random $(k+1)$-uniform hypergraph by taking the downward-closure of the hyperedges (Definition 1.2.3) and study $\mathbb{F}_2$-cohomological j-connectedness for any $j \in [k-1]$, defined as the vanishing of the i-th cohomology group with coefficients in $\mathbb{F}_2$ for any $i \in [j]$ and the zero-th cohomology group being isomorphic to $\mathbb{F}_2$ (Definition 1.2.4). This notion of higher-dimensional connectedness proves to be not-necessarily monotone (see Example 3.3.2), thus the existence of a single threshold for this property is not guaranteed. Nevertheless, we determine a single sharp threshold for $\mathbb{F}_2$-cohomological j-connectedness in $\mathcal{G}_p$ and in addition prove a hitting time result, relating this threshold to the disappearance of the last minimal obstruction (Theorem 1.2.9). This will also imply an analogous hitting time result for the model $\mathcal{Y}_p$ introduced in [48, 53], which Kahle and Pittel [44] proved only in the case when $\mathcal{Y}_p$ is a 2-complex (i.e. $k = 2$). Furthermore, we study the dimension of the j-th cohomology group inside the critical window given by the threshold, thus obtaining a precise asymptotic expression for the probability of $\mathcal{G}_p$ being j-cohom-connected (Theorem 1.2.7), a result analogous to Theorem 1.1.7.

We consider the following model of random k-complex.

Definition 1.2.3. We denote by $\mathcal{G}_p = \mathcal{G}(k; n, p)$ the random k-dimensional simplicial complex on vertex set $[n]$ such that

- the 0-simplices are the singletons of $[n]$;
- the k-simplices are the hyperedges of the binomial random $(k+1)$-uniform hypergraph $\mathcal{H}(k+1; n, p)$;
- for each $j \in [k-1]$, the j-simplices are exactly the $(j+1)$-subsets of hyperedges of $\mathcal{H}(k+1; n, p)$.

Denote by $H^i(\mathcal{G}_p; \mathbb{F}_2)$ the i-th cohomology group of the complex $\mathcal{G}_p$ with coefficients in $\mathbb{F}_2$ (see Section 3.2.3 for the definition). As mentioned before, $H^0(\mathcal{G}_p; \mathbb{F}_2)$ being isomorphic to $\mathbb{F}_2$ is equivalent to *topological connectedness* of $\mathcal{G}_p$, i.e. the vertex-connectedness of the underlying random $(k+1)$-uniform hypergraph (see e.g. [55, Theorem 42.1]). Hence, we obtain a stronger notion of connectedness by also requiring the higher cohomology groups to be "as small as possible".

Definition 1.2.4. Given a positive integer j, the random simplicial complex $\mathcal{G}_p$ is $\mathbb{F}_2$-*cohomologically j-connected* (*j-cohom-connected* for short) if

- $H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2$;
- $H^i(\mathcal{G}_p; \mathbb{F}_2) = 0$ for all $i \in [j]$.

In terms of proving analogues of Theorems 1.1.1 and 1.1.6, the analysis becomes challenging, because in principle $\mathcal{G}_p$ might oscillate between being j-cohom-connected and not being j-cohom-connected, and thus the existence of a single threshold for this property is not guaranteed. Nonetheless, we show that

$$p_j := \frac{(j+1)\log n + \log\log n}{(k-j+1)n^{k-j}}(k-j)! \tag{1.2}$$

is a sharp threshold for j-cohom-connectedness of $\mathcal{G}_p$ (Theorem 1.2.5).

Indeed, we prove a stronger result analogous to Theorem 1.1.2, by relating this sharp threshold to the disappearance of the last *minimal obstruction* to j-cohom-connectedness, which we call M_j (Definition 3.1.10). We will later see that a copy of M_j is a k-simplex in $\mathcal{G}_p$ equipped with a specific configuration, and is a witness for the j-th cohomology group $H^j(\mathcal{G}_p; \mathbb{F}_2)$ *not* vanishing. Thus, a copy of M_j represents an obstruction to j-cohom-connectedness and furthermore this obstruction is "minimal" in a natural way.

The complex $\mathcal{G}_p$ can also be interpreted as a process. We assign a birth time chosen uniformly at random in $[0,1]$ to each $(k+1)$-set in $[n]$ and, by gradually increasing p from 0 to 1, we see the k-simplices (i.e. the $(k+1)$-sets and their downward-closure) appearing in sequence, ordered by their birth times. With this interpretation, we can define p_{M_j} as the birth time of the k-simplex which causes the last copy of the minimal obstruction M_j to disappear. The following result states that whp p_{M_j} is the hitting time for the j-cohom-connectedness of $\mathcal{G}_p$ and is indeed "close" to p_j defined in (1.2).

Theorem 1.2.5. *Let $k \geq 2$ be an integer and let ω be any function of n which tends to infinity as $n \to \infty$. For each $j \in [k-1]$, with high probability the following holds.*

(i) p_{M_j} satisfies

$$\frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)! < p_{M_j} < \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!.$$

(ii) For all $p < p_{M_j}$, $\mathcal{G}_p$ is not $\mathbb{F}_2$-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_p; \mathbb{F}_2) \neq \mathbb{F}_2 \quad \textit{or} \quad H^i(\mathcal{G}_p; \mathbb{F}_2) \neq 0 \textit{ for some } i \in [j].$$

(iii) For all $p \geq p_{M_j}$, $\mathcal{G}_p$ is $\mathbb{F}_2$-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2 \quad \textit{and} \quad H^i(\mathcal{G}_p; \mathbb{F}_2) = 0 \textit{ for all } i \in [j].$$

As a corollary, we generalise the hitting time result proved by Kahle and Pittel [44] for $\mathcal{Y}_p$. They related the vanishing of the first cohomology group of $\mathcal{Y}_p(2; n, p)$ to the disappearance of the last minimal obstruction, namely an isolated 1-simplex. We extend this result to the process associated with $\mathcal{Y}_p$ for general $k \geq 2$, considering an isolated $(k-1)$-simplex—i.e. a $(k-1)$-simplex which is not contained in any k-simplex of $\mathcal{Y}_p$—as minimal obstruction.

Corollary 1.2.6. *Let $k \geq 2$ be an integer. Then with high probability $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)$ vanishes at exactly the moment when the last isolated $(k-1)$-simplex disappears from $\mathcal{Y}_p$.*

Finally, we study the asymptotic distribution of the dimension of $H^j(\mathcal{G}_p; \mathbb{F}_2)$ inside the critical window given by the threshold p_j, thus also obtaining an expression for the probability of $\mathcal{G}_p$ being j-cohom-connected (analogously to Theorem 1.1.7).

Theorem 1.2.7. *Let $k \geq 2$ be an integer, $j \in [k-1]$, and $(c_n)_{n\geq 1}$ be a sequence of real numbers such that $c_n \to c \in \mathbb{R}$. If*

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

then $\dim\left(H^j(\mathcal{G}_p; \mathbb{F}_2)\right)$ converges in distribution to a Poisson random variable with expectation $\lambda_j := \frac{(j+1)e^{-c}}{(k-j+1)^2 j!}$, while with high probability $H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2$ and $H^i(\mathcal{G}_p; \mathbb{F}_2)$ vanishes for every $i \in [j-1]$. In particular,

$$\Pr\left(\mathcal{G}_p \textit{ is } j\textit{-cohom-connected}\right) \xrightarrow{n\to\infty} e^{-\lambda_j}.$$

1.2.3 Random simplicial complexes: non-uniform case

In this section, we define a more general model of random simplicial complex arising from a *non-uniform* random hypergraph. Given a d-tuple $\mathbf{p} = (p_1, \ldots, p_d)$ of probabilities, the d-complex $\mathcal{G}(n, \mathbf{p})$ is obtained from the downward-closure of the random hypergraph on vertex set $[n]$ where each set of $k+1$ vertices forms a hyperedge with probability p_k independently, for any $k \in [d]$ (Definition 1.2.8). For this model, we consider a stronger notion of connectedness, namely the vanishing of cohomology groups over an *arbitrary* abelian group R (not only over $\mathbb{F}_2$). Analogously to the uniform case, we overcome the difficulties arising from the non-monotonicity and show that this notion of R-cohomological connectedness displays a phase transition at around the time when all the minimal obstructions disappear (Theorem 1.2.9). We also investigate the behaviour of the cohomology groups within the critical window (Theorem 1.2.10).

We now give the formal definition of our model $\mathcal{G}(n, \mathbf{p})$.

Definition 1.2.8. Let $d \geq 2$ be an integer. For each $k \in [d]$, let $p_k = p_k(n) \in [0, 1] \subset \mathbb{R}$ be given and write $\mathbf{p} := (p_1, \ldots, p_d)$. Denote by $\mathcal{H}(n, \mathbf{p})$ the binomial (non-uniform) random hypergraph on vertex set $[n]$ in which, for all $k \in [d]$, each element of $\binom{[n]}{k+1}$ forms a hyperedge 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 $[n]$;
- for each $k \in [d]$, the k-simplices are precisely the $(k+1)$-sets which are contained in hyperedges of $\mathcal{H}(n, \mathbf{p})$.

Given a positive integer j, for this model we study a stronger notion of j-cohom-connectedness by considering cohomology groups over an arbitrary abelian group R. More precisely, we say that $\mathcal{G}(n,\mathbf{p})$ is *R-cohomologically j-connected* if the cohomology groups $H^i(\mathcal{G}(n,\mathbf{p});R)=0$ for all $i\in[j]$ and $H^0(\mathcal{G}(n,\mathbf{p});R)=R$.

In order to prove a hitting time result (analogous to Theorem 1.1.2), we are interested in determining what the minimal obstruction to R-cohomological j-connectedness is. As it turns out, a copy of M_j defined for the uniform case remains the minimal obstruction to the vanishing of the j-th cohomology group over any abelian group R. However, in $\mathcal{G}(n,\mathbf{p})$ any k-simplex with $j\le k\le d$ can potentially give rise to a minimal obstruction. Therefore we have to consider obstructions of different sizes and thus define a slightly more general structure which we call $\hat{M}_{j,k}$ (Definitions 4.4.2 and 4.4.3).

Although R-cohomological j-connectedness is not a monotone property (see Example 4.4.11), we will show the existence of a single threshold for this property, which is related to the absence of any minimal obstruction $\hat{M}_{j,k}$. To do this, we turn the d-complex $\mathcal{G}(n,\mathbf{p})$ into a process by assigning a birth time to each k-simplex for any $k\in[d]$, as discussed for the uniform model $\mathcal{G}_p$. In particular, if we fix a "direction" $\bar{\mathbf{p}}=(\bar{p}_1,\dots,\bar{p}_d)$ of non-negative real numbers (with $\bar{p}_d\neq 0$), set $\mathbf{p}=\tau\bar{\mathbf{p}}:=(\min\{\tau\bar{p}_1,1\},\dots,\min\{\tau\bar{p}_d,1\})$ and gradually increase the parameter τ from 0 to $\tau_{\max}:=1/\bar{p}_d$, then $\mathcal{G}(n,\mathbf{p})$ can be viewed as a process where simplices and their downward-closure arrive one by one. Thus, given a direction $\bar{\mathbf{p}}$, we can use τ as a time parameter and study the evolution of the process $\mathcal{G}_\tau=\mathcal{G}(n,\tau\bar{\mathbf{p}})$ for $\tau\in[0,\tau_{\max}]$.

Moreover, we note that the evolution of $\mathcal{G}_\tau$ is unchanged if $\bar{\mathbf{p}}$ is scaled by a multiplicative factor and therefore we scale $\bar{\mathbf{p}}$ in such a way that the last copy of $\hat{M}_{j,k}$ (for some k) is likely to disappear when τ is close to 1. Indeed, this happens whp when choosing $\bar{\mathbf{p}}$ as a direction that we call *j-critical* and we formally define later (Definition 4.2.3).

Theorem 1.2.9. *Let $d\ge 2$ be an integer. For $j\in[d-1]$ and a j-critical direction $\bar{\mathbf{p}}=(\bar{p}_1,\bar{p}_2,\dots,\bar{p}_d)$ with $\bar{p}_d\neq 0$, let $\mathcal{G}_\tau=\mathcal{G}(n,\tau\bar{\mathbf{p}})$, let $\tau_{\max}=1/\bar{p}_d$, and let*

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

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

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

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

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

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

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

The final main result deals with τ in the critical window, i.e. $\tau=1+O(1/\log n)$. In this range, we provide an asymptotic description of the j-th cohomology group of $\mathcal{G}_\tau$ with coefficients in R.

Theorem 1.2.10. *Let $(c_n)_{n\geq 1}$ be a sequence of real numbers with $c_n \to c \in \mathbb{R}$. Let $j \in [d-1]$, $\tau = 1 + \frac{c_n}{\log n}$ 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*

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

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

1.3 Key techniques

To prove the main results of this thesis, we make use of several (standard) probabilistic and enumerative techniques, including first and second moment methods, method of moments, multivariate Poisson approximation, and generating functions. In this section, we instead provide an overview of the key non-standard methods and concepts that are introduced in the thesis.

In Section 1.3.1 we consider j-tuple-connectedness in random hypergraphs and describe the main techniques used in Chapter 2. We will show that we can naturally explore the j-components via a breadth-first search process. Nonetheless, to efficiently obtain bounds on the components' size and order, we need to define a *two-type branching process*, that we use as an *upper coupling* of the component search-process.

In Sections 1.3.2–1.3.4, we present the essential ideas and techniques needed in Chapters 3 and 4 for the study of random simplicial complexes. In particular, in Section 1.3.2 we introduce the *minimal obstruction* to the vanishing of the cohomology groups and determine the probability range where we expect the last minimal obstruction to disappear.

Showing that p_j is the threshold for j-cohom-connectedness proves to be quite sophisticated, because this notion of connectedness is *not* intrinsically a monotone property. In order to overcome the challenges caused by non-monotonicity, we make use of various techniques. In Section 1.3.3, we show how we can partition the subcritical regime in suitable *subintervals* and prove that whp *throughout* each of these subintervals a minimal obstruction exists. This enables us to prove that whp for *any* p before the threshold $\mathcal{G}_p$ is not j-cohom-connected.

For the analysis of the supercritical case, in Section 1.3.4 we introduce auxiliary structures, called *local obstacles*, whose presence in $\mathcal{G}_p$ behaves in a monotone way. Moreover, we use the concept of *traversability* to develop a *search process* that allows us to carefully bound the number of remaining obstructions.

Many of these techniques can be extended to the model $\mathcal{G}(n, \mathbf{p})$ generated by the non-uniform random hypergraph, which is studied in Chapter 4. Nevertheless in this case, we have to show that a phase transition happens if we make the process associated with the complex evolve along a *j-critical direction*, which we describe in Section 1.3.5.

1.3.1 Coupling with a two-type branching process

In Theorem 1.2.1, which is proved in Chapter 2, we consider j-connectedness (Definition 1.1.4) in the random k-uniform hypergraph $\mathcal{H}_p$. We investigate the asymptotic size and structure of the largest j-components, in the subcritical regime with respect to p_g defined in (1.1).

One natural approach to analyse a j-component is to define a component search process via a breadth-first search algorithm which explores the j-sets and the hyperedges of the component. However, to obtain precise bounds on the size and order of the largest j-components we upper couple this component search process with a two-type branching process (Lemma 1.3.1), whose evolution is easier to analyse.

For the *component search process*, we consider j-sets and k-sets of $[n]$, which we label *neutral*, *active*, or *explored* and use a queue to stock active j-sets and k-sets. Initially all j-sets and k-sets are neutral. The algorithm is then defined as follows.

- We start by choosing a j-set J_0, which we label as active and put in the queue.
- If the queue is not empty, an element pops out from the queue.
  - If the popped element is a j-set J_*, then we query every k-set K containing J_* and if K is neutral and forms a hyperedge in $\mathcal{H}_p$, then we label it as active and add it to the queue. Once we have queried every K, we label J^* as explored.
  - If the popped element is a k-set K_*, then we consider every j-set $J \subset K_*$ and if J is neutral, then we label it active and add it to the queue, whereas K^* is labelled as explored.
- If the queue is empty, the algorithm stops and we have found the whole j-component containing J_0.

Nonetheless, in terms of finding an upper bound on the size of the j-components, the described search algorithm is difficult to analyse. We therefore introduce the *two-type branching process* $\mathcal{T}$, where each vertex is either of type j or of type k, representing a j-set or k-set of $[n]$, respectively. The branching process $\mathcal{T}$ starts with a vertex J_0 of type j as the root and evolves as follows.

- For each vertex J of type j (i.e. for every j-set J), we consider the collection $\mathcal{K}_J$ of all k-sets containing J to be the set of possible children of J. For each $K \in \mathcal{K}_J$, we generate a new vertex of type k independently with probability p.
- For each such new vertex K of type k, we then attach $\binom{k}{j} - 1$ many new vertices of type j, corresponding to the j-sets in $\binom{K}{j} \setminus \{J\}$.

Thus, an instance of $\mathcal{T}$ constructs a *rooted two-type tree*, where the number of children of a vertex of type j is distributed as $\mathrm{Bi}\left(\binom{n-j}{k-j}, p\right)$, whereas each vertex of type k has $\binom{k}{j} - 1$ children deterministically. This enables us to accurately bound the number of possible instances of $\mathcal{T}$ of fixed size (i.e. number of vertices of type k) via enumerative techniques (Lemma 2.4.1).

The key observation is that we can use the two-type branching process $\mathcal{T}$ to control the evolution of the component search process. More precisely, in Section 2.3 we prove the following.

Lemma 1.3.1. *There exists an upper coupling of the component search process on j-components of $\mathcal{H}_p$ with $\binom{n}{j}$ copies of the two-type branching process $\mathcal{T}$. In particular, for any given $s \in \mathbb{N}$, the expected number of j-components of $\mathcal{H}_p$ of size at least s is not larger than the expected number of instances of $\mathcal{T}$ of size at least s.*

With this coupling argument, we can use the bounds on the number of possible instances of $\mathcal{T}$ to derive an upper bound on the expected number of j-components with size larger than a fixed value. By applying Markov's inequality, we can then show that if we run (at most) $\binom{n}{j}$ component search processes (each one starting from a different j-set), whp no j-component of size larger than the value claimed in Theorem 1.2.1 will be discovered, thus obtaining an upper bound on the size of the largest j-components (Lemma 2.2.1).

For a matching lower bound, we introduce *wheels* (see Section 2.5), which can be viewed as a higher-dimensional analogue of cycles in graphs, and thus they are obstacles for a j-component to be a hypertree. By bounding the number of wheels from above (Lemma 2.5.1), in Section 2.5 we are able to show that whp all j-components of sufficiently large size are hypertrees and to find the desired lower bound on the size of the largest j-component by a second moment argument (Lemma 2.2.2).

1.3.2 Determining the threshold via minimal obstructions

In order to determine a threshold for j-cohom-connectedness of the random k-complex $\mathcal{G}_p$ (Theorem 1.2.5), we need to identify the *minimal obstruction* to the vanishing of the cohomology groups of $\mathcal{G}_p$ and prove that this is also the *critical* obstruction, meaning that whp the disappearance of all the minimal obstructions from $\mathcal{G}_p$ corresponds to the time at which $\mathcal{G}_p$ becomes j-cohom-connected.

As we will see more formally in Chapter 3, for any $j \in [k-1]$ a non-trivial element of the cohomology group $H^j(\mathcal{G}_p; \mathbb{F}_2)$ is an equivalence class represented by an $\mathbb{F}_2$-valued function defined on the j-simplices of $\mathcal{G}_p$, which satisfies specific properties (see Definition 3.2.4). Hence, the existence of such a function—which we call a *bad function*—yields the non-vanishing of $H^j(\mathcal{G}_p; \mathbb{F}_2)$. We are interested in determining under which conditions we are able to define a bad function whose *support*, i.e. the set of j-simplices mapped to 1, is *minimal*.

Observe that in the model $\mathcal{Y}_p$ defined in [48, 53], which includes the full $(k-1)$-dimensional skeleton, the minimal obstruction to the vanishing of $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)$ is an isolated $(k-1)$-simplex, i.e. a $(k-1)$-simplex which is not contained in any k-simplex of $\mathcal{Y}_p$. We cannot expect the same type of minimal obstruction in our model $\mathcal{G}_p$, because by definition isolated $(k-1)$-simplices do not exist in $\mathcal{G}_p$. We therefore define M_j (Definition 3.1.10), a specific configuration of simplices that is a witness for the existence of a minimal bad function, if present in $\mathcal{G}_p$ (see Section 3.3.1).

Instead of providing a formal argument, here we locate the candidate threshold for j-cohom-connectedness by considering a simplified version of the minimal obstruction. A pair (K, C) is said to form a *copy of M_j^-* (see Figure 1.1) in $\mathcal{G}_p$ if K is a k-simplex and $C \subseteq K$ is a $(j-1)$-simplex such that every j-simplex of the set

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

is contained in no other simplex of $\mathcal{G}_p$.

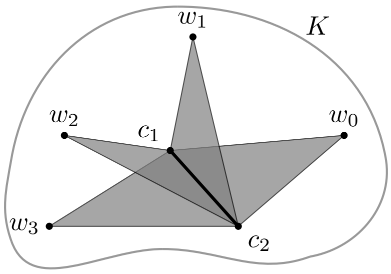


Figure 1.1: A copy of M_j^- for $k = 5$ and $j = 2$, where $K = \{c_1, c_2, w_0, w_1, w_2, w_3\}$ is a k-simplex and $C = \{c_1, c_2\}$ is a $(j-1)$-simplex contained in K. Each j-simplex $C \cup \{w_i\}$ for $i = 0, 1, 2, 3$ (grey) is contained in no other k-simplex except K.

A copy of the minimal obstruction M_j can be described as a copy of M_j^- with the additional property that one of the j-simplices in $\mathcal{F}(K, C)$ is part of a j-*cycle*, i.e. a collection of j-simplices such that every $(j-1)$-simplex is contained in an even number of j-simplices of this collection (see Section 3.2.3 for an overview on cohomology notation). Indeed, we will show that in the probability range of our interest the required j-cycle is very likely to exist (Lemma 3.4.6), thus the presence of M_j^- and of M_j are essentially equivalent events. Hence, in order to heuristically guess the threshold for our studied property, we compute the expected number of copies of M_j^- and determine when this expectation becomes constant.

Consider a pair (K, C), where K is a $(k+1)$-set and $C \subseteq K$ is a j-set. By definition of $\mathcal{G}_p$, the set K forms a k-simplex with probability p. Moreover, each of the $(j+1)$-sets in $\mathcal{F}(K, C)$ is not present in any other k-simplex with approximately probability $(1-p)^{\binom{n}{k-j}} \approx \exp\left(-\frac{n^{k-j}}{(k-j)!}p\right)$, and in total we have $k - j + 1$ many such j-sets. Since there are $\Theta(n^{k+1})$ ways of choosing such a pair (K, C), we can approximate the expected number of copies of M_j^- by $\Theta\left(n^{k+1} p \exp\left(-\frac{(k-j+1)n^{k-j}}{(k-j)!}p\right)\right)$. We are interested in the value of p such that $n^{k+1} p \exp\left(-\frac{(k-j+1)n^{k-j}}{(k-j)!}p\right) = 1$. With easy computations, we derive that this holds approximately when

$$p = \frac{(j+1)\log n + \log\log n + O(1)}{(k-j+1)n^{k-j}}(k-j)!,$$

which corresponds to the value p_j defined in (1.2).

In Theorem 1.2.5 (i), we in fact formally prove that the value

$$p_{M_j} := \sup\{p \in [0,1] \mid \mathcal{G}_p \text{ contains a copy of } M_j\}, \tag{1.3}$$

(i.e. the birth time of the k-simplex which causes the last minimal obstruction M_j to disappear) is whp close to the heuristic value p_j.

1.3.3 Subcritical case: splitting the range and covering the intervals

In order to prove the subcritical case (i.e. statement (ii)) of Theorem 1.2.5, a standard application of the second moment method does not suffice, because of the non-monotonicity of j-cohom-connectedness. We thus have to find different techniques to circumvent the challenges produced by non-monotonicity.

More precisely, it is not sufficient to show the existence whp of some bad function slightly before the hitting time p_{M_j} (defined in (1.3)), because it is not guaranteed that this obstruction was present throughout the subcritical regime. Rather we have to show that whp $\mathcal{G}_p$ is not j-cohom-connected for *any* p up to p_{M_j}, i.e. for any such p there exists $0 \leq i \leq j$ such that $H^i(\mathcal{G}_p; \mathbb{F}_2)$ is "non-trivial". We achieve this by splitting the subcritical probability range $[0, p_{M_j})$ in $j+1$ subintervals and by showing that whp minimal obstructions to the vanishing of some of the cohomology groups exist throughout all of these subintervals.

We first consider the zero-th cohomology group. The threshold p_{M_0} for the vertex-connectedness of the uniform hypergraph which generates the random complex $\mathcal{G}_p$ is known (see e.g. [19, 56, 57]), thus we have that whp $H^0(\mathcal{G}_p; \mathbb{F}_2) \neq \mathbb{F}_2$ for every $p \in [0, p_{M_0})$.

To complete the subcritical case, our strategy is to consider for $j \in [k-1]$ the interval $[p_{M_{j-1}}, p_{M_j})$ and to show that whp for any p in this range the j-th cohomology group of $\mathcal{G}_p$ does not vanish, due to the presence of minimal obstructions. To this end, we prove the following.

Lemma 1.3.2. *Let $j \in [k-1]$. With high probability, there exist three copies $\mathcal{M}^\ell$, with $\ell = 1, 2, 3$, of the minimal obstruction M_j such that for all $p \in [p_{M_{j-1}}, p_{M_j})$, the copy $\mathcal{M}^\ell$ is present in $\mathcal{G}_p$ for some ℓ. In particular, whp $H^j(\mathcal{G}_p; \mathbb{F}_2) \neq 0$ for all $p \in [p_{M_{j-1}}, p_{M_j})$.*

For any $i \in [j]$, we consider the minimal obstruction M_i to the vanishing of the i-th cohomology group and the related hitting time p_{M_i}. We then apply Lemma 1.3.2 with j replaced by any $i \in [j]$ and obtain that whp for any p in the interval $[p_{M_{i-1}}, p_{M_i})$ we have $H^i(\mathcal{G}_p; \mathbb{F}_2) \neq 0$. In this way, we can show that whp $\mathcal{G}_p$ is not j-cohom-connected in

$$[0, p_{M_0}) \cup \bigcup_{i=1}^{j} [p_{M_{i-1}}, p_{M_i}) \quad \overset{\text{(whp)}}{=} \quad [0, p_{M_j}),$$

thus proving Theorem 1.2.5 (ii).

To prove Lemma 1.3.2, we subdivide $[p_{M_{j-1}}, p_{M_j})$ into three subintervals

$$[p_{M_{j-1}}, p_{M_j}) \ = \ I_1 \cup I_2 \cup I_3,$$

and show that whp a copy of M_j exists throughout each subinterval I_ℓ with $\ell = 1, 2, 3$.

Observe that we cannot expect to cover the interval just by one copy of M_j. Indeed, we know that p_{M_j} is the birth time of the k-simplex which causes the last copy of M_j to disappear, but the k-simplex which formed this copy of M_j is very unlikely to already exist at around time $p_{M_{j-1}}$, because $p_{M_{j-1}} \ll p_{M_j}$ by Theorem 1.2.5 (i).

In order to cover the three subintervals, we use a combination of a second moment method and concentration arguments (such as Chernoff bounds), depending on the specific range of probability. In particular, we use the following approach.

- At the beginning of I_1, whp $\mathcal{G}_p$ contains "many" copies of M_j, at least one of which survives throughout the interval I_1.
- At the end of I_2, whp there are "many" copies of M_j, one of which already existed at the beginning of this interval.
- The last copy of M_j, which disappears at time p_{M_j}, whp already existed at the beginning of the interval I_3.

1.3.4 Supercrtitical case: local obstacles and search process

To prove Theorem 1.2.5 (iii), we will show that whp for every $p \geq p_{M_j}$ no bad functions can appear in $\mathcal{G}_p$. By definition, there are no copies of M_j after p_{M_j}, but we also have to exclude other "larger" obstructions.

We therefore want to bound the number of possible bad functions that can arise in the supercritical case, and thus also bound the probability that they exist. However, non-monotonicity of j-cohom-connectedness imposes particular challenges. Indeed, we need to prove that the probability of $\mathcal{G}_p$ *not* being j-cohom-connected is small enough that we can apply a union bound over *all* $p \geq p_{M_j}$.

We first consider the notion of a *local obstacle* (Definition 3.4.8), which is a k-simplex that contains at least $k - j - 1$ many j-simplices which are not present in any other k-simplex of the complex. We show that after the threshold p_{M_j} no new local obstacle can appear, because, in contrast to j-cohom-connectedness, the presence of a local obstacle is a monotone decreasing property in $\mathcal{G}_p$ (Lemma 3.4.9). We then prove that if the support of a bad function appears in $\mathcal{G}_p$ for some $p > p_{M_j}$ and has constant size, then it would create a local obstacle, whose presence was excluded in this regime.

For supports of larger size, we need to adopt a different approach. We consider the cohomology classes of bad functions and we show that the smallest support of elements in any of these classes satisfies the following property of *traversability* (Lemma 3.5.4).

Definition 1.3.3. Let S be a collection of j-simplices in $\mathcal{G}_p$. For $\sigma_1, \sigma_2 \in S$, we set

$$\sigma_1 \sim \sigma_2 \quad \text{if} \quad \sigma_1 \text{ and } \sigma_2 \text{ lie in a common } k\text{-simplex.}$$

We say that the set S is *traversable* if the transitive closure of $\sim$ is $S \times S$.

This in particular means that such a support can be discovered via a breadth-first search process: start from any j-simplex in the support and query all $(k+1)$-sets containing it. For any of these sets which forms a k-simplex, we discover all other j-simplices within this k-simplex which belong to the support and continue the process from them according to some pre-determined order, but we explore only $(k+1)$-sets which would give us some previously undiscovered j-simplex. The traversability of the support guarantees that we discover all of its j-simplices in this process.

Thus we can use this search process to accurately count the number of large "bad" supports and in turn obtain a suitable bound on the probability that any such support exists in $\mathcal{G}_p$ for some p. In this way, we can prove that $\mathcal{G}_p$ is j-cohom-connected for all $p \geq p_{M_j}$ *simultaneously.*

1.3.5 Critical direction

In Chapter 4 we study the random d-complex $\mathcal{G}(n, \mathbf{p})$ arising from the random non-uniform hypergraph (Definition 1.2.8), where given the d-tuple of probabilities $\mathbf{p} = (p_1, \ldots, p_d)$, every $(k+1)$-set (together with its downward-closure) is present as a k-simplex with probability p_k, for every $k \in [d]$. We consider R-cohomological j-connectedness defined as the vanishing of the cohomology groups over an arbitrary abelian group R and show that this property undergoes a phase transition (Theorem 1.2.9).

We also prove that whp this phase transition occurs around the time when the last minimal obstruction $\hat{M}_{j,k}$ (Definitions 4.4.2 and 4.4.3) disappears. We will see that a copy $\hat{M}_{j,k}$ represents a generalisation of the structure M_j defined for $\mathcal{G}_p$, which in the model $\mathcal{G}(n, \mathbf{p})$ can be formed by a k-simplex for any $j \leq k \leq d$. In this way, we are able to extend the main ideas used in the uniform model $\mathcal{G}_p$ to this more general case. In particular, similarly as in the uniform case, we partition the subcritical regime and show the existence of minimal obstructions throughout any of the considered "subregions". For the analysis of the supercritical case, we generalise the concept of local obstacles and the definition of the search process for traversable supports, to obtain suitable bounds on the probability that obstructions exist in this regime.

In order to adapt these techniques to the non-uniform case, we will interpret $\mathcal{G}(n, \mathbf{p})$ as a process. As explained in Section 1.2.3, we fix a direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$ and set $\mathbf{p} = \tau\bar{\mathbf{p}}$, so that we can study the evolution of the process $\mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$ in terms of the time parameter τ. In Theorem 1.2.9 we consider a particular direction $\bar{\mathbf{p}}$, which we call *j-critical*, such that the last copy of $\hat{M}_{j,k}$ is likely to disappear at time τ "close" to 1, i.e. when $\mathbf{p}$ is roughly $\bar{\mathbf{p}}$.

We now give an idea of how we can determine such a j-critical direction. First of all, for a given tuple $\mathbf{p} = (p_1, \ldots, p_d)$ we have that the probabilities p_k with $k \in [j-1]$ have no influence on the j-th cohomology group $H^j(\mathcal{G}_\tau; R)$, so we only have to consider the probabilities p_k for $j \leq k \leq d$. If we look at a single probability p_k, i.e. consider

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

when $R = \mathbb{F}_2$ and $j+1 \leq k \leq d$, Theorem 1.2.5 yields that the critical range for j-cohom-connectedness lies around

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

Therefore, even when considering general coefficient group R, 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 $j+1 \leq k \leq d$.

More precisely, we will see that we can restrict our attention to a *j-admissible* direction $\bar{\mathbf{p}}$ (Definition 4.2.1), for which there are real-valued constants $\bar{\alpha}_k$, $\bar{\gamma}_k$, and a function $\bar{\beta}_k = \bar{\beta}_k(n)$ (satisfying specific properties) such that for each $k \in [d]$ we have

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

Furthermore, for any k we find an asymptotic expression for the expected number of obstructions $\hat{M}_{j,k}$, in terms of the parameters $\bar{\alpha}_k$, $\bar{\beta}_k$, and $\bar{\gamma}_k$ (Lemma 4.4.13). In this way, a j-critical direction can be defined as a j-admissible vector such that there exists a dimension k with $j \leq k \leq d$ which is "critical" for the disappearance of the last copy of $\hat{M}_{j,k}$, while whp all the other minimal obstructions have already disappeared. We discuss this parametrisation in more detail in Sections 4.2.1 and 4.10.

1.4 Discussion

In this section we summarise the main results of this thesis and discuss future research work.

In Section 1.4.1 we consider our results on the subcritical random hypergraph and examine the conditions we impose to prove Theorem 1.2.1. Moreover, we address the symmetry phenomenon that the random graph displays at the moment of the threshold for the appearance of the giant component, aiming to describe a similar phenomenon for j-tuple connectedness in the random hypergraph.

In Section 1.4.2 we discuss alternative models of random simplicial complexes, as well as other notions of higher-dimensional connectedness that can be defined in $\mathcal{G}_p$ in order to better understand the structural properties of our model.

1.4.1 Weakening conditions and symmetry phenomenon

The main contribution of Chapter 2 is Theorem 1.2.1, where we determine the asymptotic size of the largest j-components of the binomial random hypergraph $\mathcal{H}_p$, in the regime before the appearance of the giant j-component. More precisely, our result holds under the condition $(\varepsilon^4 n)/(\log n)^2 \to \infty$, which is required to prove the lower bound on the size. Although the lower bound might still hold for smaller ε, in this case we no longer expect the largest components to be hypertrees, thus new proof techniques would be needed. On the other hand for the upper bound, the aforementioned condition can be replaced by the two weaker conditions $\varepsilon^3 n^j \to \infty$ and $(\varepsilon^2 n^{k-j})/\log n \to \infty$. In particular, the condition $\varepsilon^3 n^j \to \infty$ was conjectured in [20] to be the critical window, whereas further study is required to determine if the condition $(\varepsilon^2 n^{k-j})/\log n \to \infty$ is solely needed for our specific proof strategy or represents a necessary restriction (thus disproving the conjectured critical window).

As mentioned in Section 1.1.2, the random graph $G(n,p)$ exhibits a symmetry phenomenon in the range of the phase transition that occurs around $p = 1/n$ (Theorem 1.1.3): the supercritical random graph with the giant component removed is approximately distributed as the subcritical random graph, meaning that the connected components have very similar order and structure in these two cases. It would be interesting to show the occurrence of analogous symmetric behaviour for $\mathcal{H}_p$ at around the threshold for the appearance of the giant j-component. However, this case is more challenging: although the asymptotic order of the giant j-component was determined in [20], the analysis of the structure of the remaining j-components is made complicated by the fact that they consist of j-sets rather than vertices. Thus, the distribution of the supercritical hypergraph with the giant j-component removed strongly depends on how the j-sets in the giant j-component are distributed across the vertex set.

1.4.2 Alternative random complexes and other topological aspects

The higher-dimensional analogue of graphs that we study in Chapters 3 and 4 is provided by random simplicial complexes that arise from the downward-closure of random hypergraphs. Indeed, this is not the only possibility for defining random complexes. For instance, one can naturally generate a simplicial complex starting from a random graph. One example is the *flag complex* or *clique complex* on vertex set $[n]$, where every non-empty subset $S \subset [n]$ forms a simplex if and only if S forms a clique in the binomial random graph $G(n,p)$. Another alternative is given by the random *neighbourhood complex*, whose simplices correspond to all non-empty sets of vertices with a common neighbour in $G(n,p)$. Topological properties of the flag complex and the neighbourhood complex have been analysed in [24, 41, 42] and [40], respectively, and further study might lead to a richer understanding of these structures.

On the other hand, in Theorem 1.2.5 we determine the (single) sharp threshold for j-cohom-connectedness in $\mathcal{G}_p$, defined as the vanishing of cohomology groups over $\mathbb{F}_2$, but there are other possible concepts of higher-dimensional connectedness that stem from algebraic topology. Famously, the *simple connectivity* of a simplicial complex, defined as the vanishing of its first homotopy group, is another interesting topological aspect of such structures. Indeed, Babson, Hoffman, and Kahle [5] determined the threshold for the simple connectivity of the random 2-complex $\mathcal{Y}_p$ defined in [48]. It would be interesting to investigate such a property in our model $\mathcal{G}_p$ as well, in the probability range where the two models do not coincide. Furthermore, requiring *all* the homotopy groups $\pi_1(\mathcal{G}_p), \ldots, \pi_j(\mathcal{G}_p)$ to vanish would yield an even stronger notion of connectedness.

In Theorem 1.2.7 we determine the asymptotic distribution of the dimension of $H^j(\mathcal{G}_p; \mathbb{F}_2)$ within the critical window of the threshold for j-cohom-connectedness. However, one might also be interested in characterising this dimension before the point of the phase transition, for instance by exactly determining the regime where whp the j-th cohomology group first becomes non-trivial.

2 Subcritical random hypergraphs, high-order components, and hypertrees

2.1 Introduction

2.1.1 Motivation

One of the most prominent results on random graphs is the so-called *phase transition* in the order of the largest components, first discussed by Erdős and Rényi in their seminal work [29]: a small change in the edge density around the critical value drastically alters the structure and order of the largest components. Their result was improved for example by Bollobás [9] and Łuczak [50] and is often stated for the binomial random graph $\mathcal{G}(n,p)$, a graph with vertex set $[n] := \{1,\dots,n\}$ in which each pair of vertices is present as an edge with probability p independently. Throughout the paper, log denotes the natural logarithm and we say that an event holds *with high probability* (*whp* for short) if the probability that it holds tends to 1 as $n \to \infty$.

Theorem 2.1.1 ([9, 10, 29, 50])**.** *Let $0 < \varepsilon < 1$ be a constant or a function in n satisfying $\varepsilon \to 0$ and $\bar{\lambda} := \varepsilon^3 n \to \infty$. For each $i \in \mathbb{N}$, let $C_i = C_i(\mathcal{G}(n,p))$ denote the number of vertices in the i-th largest component in $\mathcal{G}(n,p)$.*

(1) If $p = (1-\varepsilon)n^{-1}$, then whp every component is either a tree or unicyclic, and for every constant $i \geq 1$, the i-th largest component is a tree. Furthermore, for any function $\omega = \omega(n) \to \infty$, whp

$$\left| C_1 - \alpha^{-1}\left(\log \bar{\lambda} - \frac{5}{2}\log\log\bar{\lambda}\right)\right| \leq \omega\varepsilon^{-2},$$

where $\alpha = -\varepsilon - \log(1-\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$ and $C_i = (2+o(1))\varepsilon^{-2}\log(\bar{\lambda})$ for $i \geq 2$.

(2) If $p = (1+\varepsilon)n^{-1}$, then whp the largest component contains at least two cycles and

$$C_1 = (1+o(1))2\varepsilon n.$$

Furthermore, for every $i \geq 2$, whp the i-th largest component is a tree with $C_i = (2+o(1))(\varepsilon^{-2}\log\bar{\lambda})$ and in particular, for any function $\omega = \omega(n) \to \infty$, whp

$$\left| C_2 - \beta^{-1}\left(\log \bar{\lambda} - \frac{5}{2}\log\log\bar{\lambda}\right)\right| \leq \omega\varepsilon^{-2},$$

where $\beta = \varepsilon - \log(1+\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$.

Note that in the supercritical random graph (i.e. when $p = (1+\varepsilon)n^{-1}$), the largest component is substantially larger than the second largest component and therefore it is also called the *giant* component. Theorem 2.1.1 displays a *symmetry* between the structure of the subcritical random graph and the supercritical random graph with the giant component removed.

Since this result, various models of random graphs have been introduced and analysed for similar phenomena. For instance, considerable attention has been paid to random regular graphs (see e.g. [10, 62] for an overview). As for $\mathcal{G}(n,p)$, this random graph model is "homogeneous", meaning that all vertices are equivalent, and whp all vertex degrees lie within a small range. In contrast, many random models for "real-world" graphs show highly inhomogeneous behaviour with a large spread of degrees, whose distributions often follow a power law. Thus, also the phase transition in inhomogeneous random graphs such as scale-free graphs [12] and distance graphs [1] has been investigated [11], with a wide analysis of the subcritical case (see e.g. [37, 61]).

Higher-dimensional analogues of random graphs and their phase transitions have also drawn particular attention. The most commonly studied higher-dimensional analogue of $\mathcal{G}(n,p)$ is the binomial random k-uniform hypergraph $\mathcal{H}^k(n,p)$ defined below. Amongst other properties, vertex-connectedness [2, 7, 8, 13, 14, 45, 46, 57–59] and high-order connectedness (also known as j-tuple-connectedness) [20–22] of $\mathcal{H}^k(n,p)$ have been extensively studied and will be discussed in more detail in Section 2.1.3. In parallel, Linial and Meshulam instigated research into random simplicial complexes, which have since also been extensively studied [17, 44, 48, 49, 53].

Before stating our results, we introduce the necessary concepts. Let $k \geq 2$ and $1 \leq j \leq k-1$ be integers. A *k-uniform hypergraph* H is a pair $H = (V,E)$, where V is the set of *vertices* and $E \subseteq \binom{V}{k}$, a collection of k-element subsets of V, is the set of *hyperedges*. An ℓ-element subset of V is called an *ℓ-set of V* (or *ℓ-set* for short). A pair $\{J_1, J_2\}$ of j-sets are called *j-tuple-connected* (*j-connected* for short) in H if there is a sequence of hyperedges $K_1, \ldots, K_m$ such that $J_1 \subset K_1$, $J_2 \subset K_m$ and $|K_i \cap K_{i+1}| \geq j$ for all $1 \leq i \leq m-1$. Additionally, any j-set is always j-connected to itself. The *j-connected components* (*j-components* for short) of H are equivalence classes of the relation $\sim_j$ defined by $J_1 \sim_j J_2$ if and only if J_1 and J_2 are j-connected. In other words, a j-component is a maximal collection of j-sets that are pairwise j-connected. Given a j-component $\mathcal{J}$, it also comes naturally with a set of hyperedges $\mathcal{K}_{\mathcal{J}}$, which are the hyperedges containing any of the j-sets in $\mathcal{J}$. In a slight abuse of terminology, we say that the j-component $\mathcal{J}$ also contains the hyperedges $\mathcal{K}_{\mathcal{J}}$. The *order* of a j-component denotes the number of j-sets it contains, and the *size* of a j-component denotes the number of hyperedges it contains. A *hypertree j-component* (a *hypertree* for short) is a j-component that contains as many j-sets as possible *given its size*, i.e. if it has size s and order t, then $t = 1 + \left(\binom{k}{j} - 1\right)s$. The case $k=2$ and $j=1$ corresponds to the classical concepts in a graph.

We denote by $\mathcal{H}^k(n,p)$ the *random k-uniform hypergraph* with vertex set $[n]$, in which each hyperedge is present with probability p independently. When parameters are clear from the context, we use $\mathcal{H}$ as a shorthand for $\mathcal{H}^k(n,p)$. The following higher-dimensional analogue of the random graph phase transition for $\mathcal{H}$ and j-connectedness was obtained in [20–22].

Theorem 2.1.2 ([20–22])**.** *Given integers $k \geq 2$ and $1 \leq j \leq k-1$, let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \to 0$, $\varepsilon^3 n^j \to \infty$ and $\varepsilon^2 n^{1-2\delta} \to \infty$, as $n \to \infty$, for some constant $\delta > 0$. Let*

$$\bar{p}_0 := \left(\binom{k}{j} - 1\right)^{-1} \binom{n}{k-j}^{-1}.$$

(1) If $p = (1-\varepsilon)\bar{p}_0$, then whp all j-components of $\mathcal{H}^k(n,p)$ have order $O(\varepsilon^{-2}\log n)$.

(2) If $p = (1+\varepsilon)\bar{p}_0$, then whp the order of the largest j-component of $\mathcal{H}^k(n,p)$ is $(1 \pm o(1))\frac{2\varepsilon}{\binom{k}{j}-1}\binom{n}{j}$, while all other j-components have order $o(\varepsilon n^j)$.

The aim of this paper is to strengthen Theorem 2.1.2 in view of Theorem 2.1.1 by taking a closer look at the *subcritical* case and addressing the following natural questions.

(a) What are the precise asymptotic order and size of the largest j-component of $\mathcal{H}^k(n,p)$?

(b) What does the largest j-component look like? Is it whp a hypertree or some other more complex structure?

2.1.2 Main result

In Theorems 2.1.1 and 2.1.2, the order of a j-component was studied. In this paper, we concentrate on the size of a j-component, i.e. the number of hyperedges it contains. Therefore, whenever we talk about the *i-th largest j-component*, the ranking is determined by the size rather than the order (i.e. the number of j-sets it contains). Observe that in the range of the hyperedge probability in our study, whp the order and size of the largest j-component only differ roughly by a multiplicative constant $c_0 = \binom{k}{j} - 1$. As a consequence, we also derive the order of the largest j-components.

Before stating our main result, we introduce the following notation. Let $(X_n)_{n\geq 0}$ be a sequence of random variables and $(a_n)_{n\geq 0}$ a sequence of positive real numbers. We say that $X_n = O_p(a_n)$ if for every $\gamma > 0$ there exist C_γ and $n_0 \in \mathbb{N}$ such that $\Pr(|X_n| \leq C_\gamma a_n) > 1 - \gamma$ for every $n \geq n_0$. It is easy to see that $X_n = O_p(a_n)$ if and only if $\Pr(|X_n| \leq \omega a_n) \to 1$ for every function $\omega = \omega(n) \to \infty$.

Theorem 2.1.3. *Given integers $k \geq 2$ and $1 \leq j \leq k-1$, and $\varepsilon = \varepsilon(n)$ with $0 < \varepsilon < 1$ such that*

$$\frac{\varepsilon^4 n}{(\log n)^2} \to \infty, \tag{2.1}$$

let

$$c_0 := \binom{k}{j} - 1, \quad p_0 := c_0^{-1}\binom{n-j}{k-j}^{-1}, \quad \delta := -\varepsilon - \log(1-\varepsilon) \quad \textit{and} \quad \lambda := \varepsilon^3\binom{n}{j}.$$

For $i \in \mathbb{N}$, let $\mathcal{L}_i = \mathcal{L}_i(\mathcal{H}^k(n,p))$ be the i-th largest j-component of $\mathcal{H}^k(n,p)$, and let L_i be the size of $\mathcal{L}_i$. Let $m \in \mathbb{N}$ be fixed. If $p = (1-\varepsilon)p_0$, then whp for any $1 \le i \le m$, the component $\mathcal{L}_i$ is a hypertree, and has size

$$L_i = \delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda + O_p(1)\right).$$

Furthermore, even if we replace condition (2.1) by the weaker conditions

$$\varepsilon^3 n^j \to \infty, \tag{2.2}$$

$$\frac{\varepsilon^2 n^{k-j}}{\log n} \to \infty, \tag{2.3}$$

we still have the same upper bound

$$L_1 \le \delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda + O_p(1)\right).$$

We note that the critical probability p_0 differs from $\bar{p}_0$ (defined in Theorem 2.1.2) by a factor of $1 + O(n^{-1})$. This is because we analyse a range closer to criticality, which requires a more precise value of p_0.

We also note that the coefficient $-5/2$ before the $\log\log\lambda$ factor in Theorem 2.1.3 is the same as that in Theorem 2.1.1 for graphs, and arises from the universal asymptotic behaviour of various families of labelled trees, i.e. connected acyclic graphs. More precisely, the asymptotic number of trees on t vertices in such a family has the form $c \cdot t!\gamma^t t^{-5/2}$, with c and γ depending on the precise nature of the family. The proofs of both Theorem 2.1.1 and Theorem 2.1.3 involve asymptotic counting of such families of trees, and the coefficient $-5/2$ comes from the common polynomial factor $t^{-5/2}$. In the case when the trees are rooted, which is more commonly considered, the exponent $-5/2$ would become $-3/2$ (see [33, Section VII.3]) – the extra factor of t comes from the choice of the root.

Since Theorem 2.1.3 states that, as long as ε satisfies condition (2.1), whp the largest j-components are *hypertrees* (and therefore for each such component its size s and order t satisfy $s = c_0 t + 1$), as a corollary we can determine their order and obtain the following result.

Corollary 2.1.4. *Let $k, j, \varepsilon, c_0, \delta, \lambda, \mathcal{L}_i, p$ be as in Theorem 2.1.3, where ε satisfies condition (2.1). For $i \in \mathbb{N}$, let $M_i = M_i\left(\mathcal{H}^k(n,p)\right)$ be the order of $\mathcal{L}_i$. Let $m \in \mathbb{N}$ be fixed. Then whp for any $1 \le i \le m$ we have*

$$M_i = c_0 L_i + 1 = c_0\delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda + O_p(1)\right).$$

Note that when $k = 2$ and $j = 1$ (i.e. the graph case) Corollary 2.1.4 gives exactly $M_1 = \delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda + O_p(1)\right)$, as stated in Theorem 2.1.1-(1).

2.1.3 Related work

The case $j = 1$ of j-tuple-connectedness corresponds to vertex-connectedness of the random hypergraph $\mathcal{H}^k(n,p)$ and is the most studied among the higher-dimensional analogues of the phase transition in $\mathcal{G}(n,p)$. Enumeration results for the asymptotic number of 1-connected k-uniform hypergraphs were obtained by Karoński and Łuczak [46], later improved by Andriamampianina and Ravelomanana [2] via enumerative techniques. The threshold for the emergence of the giant 1-component, i.e. $p = (k-1)^{-1}\binom{n-1}{k-1}^{-1}$, was first determined by Schmidt-Pruzan and Shamir [59]. Subsequently, Karoński and Łuczak [47] studied the distribution of the order of the largest component in the early supercritical regime. The studied range was extended by Ravelomanana and Rijamamy [58], although they only computed the expected order of the largest component and not its distribution. Behrisch, Coja-Oghlan, and Kang [7, 8] provided central and local limit theorems for $p(k-1)\binom{n-1}{k-1} > 1+\varepsilon$, with $\varepsilon > 0$ arbitrarily small but fixed, while more recently Bollobás and Riordan [13] showed that the distribution of the order of the largest component tends to a normal distribution for every $\varepsilon = \omega(n^{-1/3})$.

Vertex-connectedness was also studied for a model of non-uniform random hypergraphs, in which the probability for a hyperedge of size t to belong to the hypergraph depends on a parameter ω_t. In particular, de Panafieu [26] determined the critical value at which the first *complex* component (i.e. connected component with more than one cycle) appears.

In contrast, the case $j \geq 2$ is not yet well-understood. Cooley, Kang, and Person [22] determined the threshold $\bar{p}_0$ for the appearance of the giant j-component and subsequently Cooley, Kang, and Koch [20, 21] refined this result by determining the asymptotic order of the largest j-component after the threshold and by showing that the second largest component has much smaller order (see Theorem 2.1.2). However, the subcritical regime was not analysed, which is the aim of this paper.

Moreover, the analysis of the vertex-connectedness case shows a symmetry property: the supercritical hypergraph with the giant component removed behaves as a subcritical hypergraph (with slightly modified parameters). It is not immediately clear that such a behaviour should also hold in the high-order case ($j \geq 2$) and we believe that our study will help to obtain results in this direction.

2.2 Proof of Theorem 2.1.3

In order to prove Theorem 2.1.3, we bound the size of the largest j-component from above and below: we prove that whp there is no j-component of size larger than the claimed value (Lemma 2.2.1) and that for any size smaller than the claimed value, there is at least one j-component (and indeed a hypertree component) with larger size (Lemma 2.2.2).

In Section 2.3, we define the *component search process* which explores a j-component starting from a j-set, and a *two-type branching process* which gives an upper coupling on the search process (Lemma 2.3.1). We will make use of these processes to obtain the following bound on sizes of j-components:

Lemma 2.2.1. *Let $k, j, \varepsilon, \delta, \lambda, p$ be given as in Theorem 2.1.3, where ε satisfies the weaker conditions (2.2) and (2.3) but not necessarily the stronger condition (2.1), and let $K(n) \to \infty$. Whp, $\mathcal{H}^k(n,p)$ contains no j-component of size larger than*

$$\delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda + K(n)\right).$$

Lemma 2.2.1 will be proved in Section 2.4. To this end, we bound from above (and below) the number of possible instances of the two-type branching process of a certain size (Lemma 2.4.1), which are the so-called *rooted assigned two-type trees*. Using the coupling argument of Lemma 2.3.1, we obtain an upper bound on the expected number of j-components whose size is larger than a fixed value. We conclude by applying Markov's inequality to prove that if we run (at most) $\binom{n}{j}$ component search processes (each one starting from a different j-set), whp no component of size larger than the claimed value will be discovered.

In Section 2.5, we obtain the lower bound on the sizes of the largest j-components and show that they are indeed hypertrees.

Lemma 2.2.2. *Let $k, j, \varepsilon, \delta, \lambda, p$ be given as in Theorem 2.1.3, where ε satisfies (2.1), and let $K(n) \to \infty$. For any constant $m \in \mathbb{N}$, whp the m largest components of $\mathcal{H}^k(n,p)$ are hypertree components, each of size at least*

$$\delta^{-1}\left(\log\lambda - \frac{5}{2}\log\log\lambda - K(n)\right).$$

To prove Lemma 2.2.2, in Section 2.5 we introduce *wheels*, which can be viewed as a higher-dimensional analogue of cycles in graphs, and bound their number from above (Lemma 2.5.1). This will allow us to prove that for a certain range of s, most of the instances of our two-type branching process of size s correspond to *hypertree components* (Lemma 2.5.2). Thus, we derive the desired lower bound by estimating the expected number of j-sets in large hypertree components (Lemma 2.5.3) and by a second moment argument. Finally, we show that whp in the considered range all components of sufficiently large size are hypertrees (Lemma 2.5.4).

Proof of Theorem 2.1.3. In the case $\frac{\varepsilon^4 n}{(\log n)^2} \to \infty$, i.e. when (2.1) holds, Theorem 2.1.3 follows directly from Lemmas 2.2.1 and 2.2.2.

In the case when (2.1) is replaced by the weaker conditions (2.2) and (2.3), we observe that Lemma 2.2.1 still holds and gives the required upper bound. □

2.3 Search process and branching process

Throughout the paper, we let k, j, ε, c_0, p_0, δ, λ, p be as given in Theorem 2.1.3. For positive integers $n \geq k$, we write $n_{(k)}$ for the *falling factorial* $n_{(k)} := n(n-1)(n-2)\cdots(n-k+1) = n!/(n-k)!$.

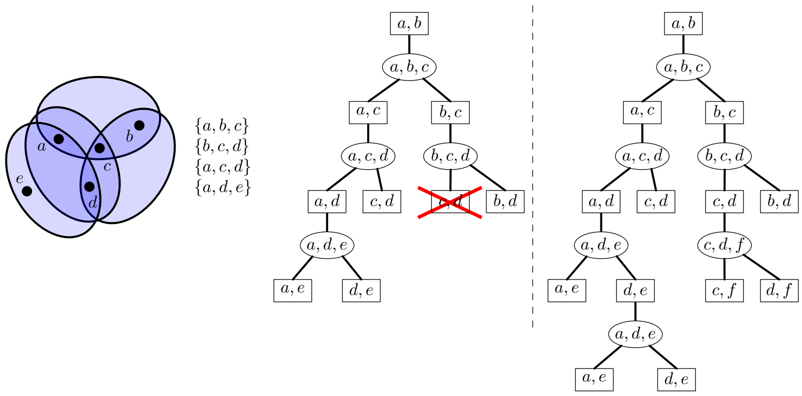


Figure 2.1: Examples with $k = 3$, $j = 2$. Left: a j-component with the two-type tree of its search process starting from $\{a, b\}$. Right: an instance of the branching process, giving a two-type tree that cannot come from the search process.

We also introduce auxiliary two-type graphs. A *two-type graph* is a *connected* bipartite graph on a set of vertices of type k and a set of vertices of type j, where each vertex of type k is connected to exactly $\binom{k}{j}$ vertices of type j. We only consider two-type graphs with at least one vertex of type k. Furthermore, all of the two-type graphs that we consider will be labelled in the sense that each vertex is distinguishable. However, we will also have a second level of labelling, which we refer to as *assignment* to distinguish it from the first level of labelling. Thus a two-type graph may be *assigned* or *unassigned*. More precisely, an *assigned two-type graph* with assignment set $[n]$ is a two-type graph with assignments on its vertices, where vertices of type j are assigned to j-sets of $[n]$ and vertices of type k to k-sets of $[n]$ in such a way that to each of the $\binom{k}{j}$ vertices of type j connected to a given vertex of type k with assignment $K \in \binom{[n]}{k}$ is assigned a distinct j-set $J \subset K$. Thus an unassigned two-type graph may be thought of as describing the local incidences of an abstract collection of j-sets and k-sets, while the assignments indicate exactly which vertices these j-sets and k-sets will be embedded onto. There is a natural bijection between the set of assigned two-type graphs with assignment set $[n]$ whose vertices all have distinct assignments and the set of *possible components*, i.e. pairs $(\mathcal{J}, \mathcal{K})$ which could potentially be the families of j-sets and k-sets in a j-component of some hypergraph on vertex set $[n]$. *Acyclic* two-type graphs are called *two-type trees*.

The *component search process* is defined as follows. We explore j-components in $\mathcal{H}$ via a breadth-first search algorithm: we keep track of disjoint sets of *neutral*, *active* and *explored* j-sets and k-sets. Furthermore, we refer to j-sets and k-sets that are either active or explored as *discovered*. Initially all j-sets and k-sets are neutral. We use a queue to stock active j-sets and k-sets, and at the start of the algorithm we choose an initial j-set J_0, label it as active and add it to the queue. When the queue is not empty, an element pops out from the queue (in a first-in first-out fashion). If the popped element is a j-set J_*, then we consider every k-set K containing J_* in arbitrary order, and if K is a hyperedge in $\mathcal{H}$ but not yet discovered, then we call it active and add it to the queue. Once we have done this for every K, we label J^* as explored; if the popped element is a k-set K_*, then we consider every j-set $J \subset K_*$ in arbitrary order, and if J is not yet discovered, then we call it *active* and add it to the queue, and label K^* as explored. We continue until the queue is empty, when we have found the j-component containing J_0. To obtain all the j-components, we only need to perform the same procedure on neutral j-sets in $\mathcal{H}$ until exhaustion. (An example of the search process is given in Figure 2.1.)

To prove Lemma 2.2.1, we provide in Lemma 2.3.1 an upper coupling on the search process with the following *two-type branching process*. Each vertex of the branching process is either of type j or of type k, and is assigned a j-set or k-set of $[n]$, respectively. Given a vertex u_0 of type j, the branching process begins with u_0 (as the root).

- For each vertex u of type j with assignment J, let $\mathcal{K}_J := \{K \in \binom{[n]}{k} \mid J \subset K\}$ be the set of possible assignments of children of u. For each assignment $K \in \mathcal{K}_J$, independently with probability p, we generate a new vertex v of type k and assign K to v.
- For each such new vertex v of type k, we then attach $\binom{k}{j} - 1$ many new vertices of type j as children of v, with distinct assignments from $\binom{K}{j} \setminus \{J\}$ (note that different vertices may have the same assignment).

Thus in particular the number of children of a vertex of type j is distributed as $\mathrm{Bi}\left(\binom{n-j}{k-j}, p\right)$, while each vertex of type k has $\binom{k}{j} - 1$ children deterministically.

We denote this branching process by $\mathcal{T}$, by which we mean the random process used to produce the assigned two-type tree. A particular tree generated by this process is called an *instance* of $\mathcal{T}$. Note that each instance of $\mathcal{T}$ is an assigned two-type tree rooted at a vertex of type j, which we call a *rooted assigned two-type tree*. An example of the two-type branching process is given in Figure 2.1. The *size* of an instance of $\mathcal{T}$ is the number of vertices of type k that it contains.

We will use the standard notion of *coupling* of random variables. Formally, a *coupling* of two random variables X and Y is a probability space in which there exist random variables X', Y' such that

- $X' \sim X$
- $Y' \sim Y$.

We say that Y is an *upper coupling* for X if furthermore there exists a natural partial order $\preceq$ in which $X' \preceq Y'$.

In our case, the random variables X, Y will be the two-type search tree generated by the component search process and the branching process $\mathcal{T}$ respectively, and the partial order $\preceq$ will be containment. Denoting the size (i.e. number of vertices of type k) of such a tree T by $|T|$, it is clear that $X' \subset Y' \Rightarrow |X'| \leq |Y'|$, and therefore we obtain a *dominance* inequality for integer-valued random variables: $\Pr(|X| \geq s) \leq \Pr(|Y| \geq s)$ for any integer s. More precisely, we have the following lemma.

Lemma 2.3.1. *There exists an upper coupling of the component search process on j-components of $\mathcal{H}$ with $\binom{n}{j}$ copies of $\mathcal{T}$. In particular, for any given $s \in \mathbb{N}$, the expected number of j-components of $\mathcal{H}$ of size at least s is not larger than the expected number of instances of $\mathcal{T}$ of size at least s.*

Proof. In the component search process, whenever a j-set J becomes active, the set of k-sets we may query is certainly contained in $\mathcal{K}_J$ (some may not be permissible since they have already been queried), and for each k-set K discovered in this way, the j-sets that become active are all in $\binom{K}{j} \setminus \{J\}$ (some may not become active if they were already discovered). Thus, in $\mathcal{T}$ we may have made some additional queries which are not made in the component search process, and we may have added some vertices of type j whose assignments correspond to j-sets not added in the search process. Hence, one component search process terminated when the component is fully discovered can certainly be coupled with one instance of $\mathcal{T}$.

Since we need at most $\binom{n}{j}$ component search process to discover all j-components, we can couple with $\binom{n}{j}$ branching processes starting from vertices of type j with all possible assignments $J \in \binom{[n]}{j}$. More precisely, whenever we start exploring a new j-component from a j-set J, we upper couple this portion of the component search process by the branching process starting from a vertex of type j with assignment J; using in total $\binom{n}{j}$ branching processes (although some of them may be left unused) we have an upper coupling for the search process on all the j-components. $\square$

2.4 Upper bound on L_1: Proof of Lemma 2.2.1

To prove Lemma 2.2.1 we first bound the number of possible rooted two-type trees that can be constructed by the two-type branching process $\mathcal{T}$. Let $\mathcal{B}$ be the set of all possible instances of $\mathcal{T}$ and thus also of all rooted assigned two-type trees. For each $s \in \mathbb{N}$, let $\mathcal{B}_s$ be the set of elements in $\mathcal{B}$ of size s, which is equal to the set of all rooted assigned two-type trees with s vertices of type k. Let B_s be the cardinality of $\mathcal{B}_s$. In the following lemma we determine the order of B_s.

Lemma 2.4.1. *For $s \in \mathbb{N}$, we have*

$$\binom{n}{j}\binom{n-j}{k-j}^s \frac{c_0^{s-1}s^{s-1}}{s!} \leq B_s \leq \binom{n}{j}\binom{n-j}{k-j}^s \frac{c_0^{s-1}s^{s-1}}{s!} e^{1/c_0}.$$

In particular, we have

$$B_s = \Theta\left(\binom{n}{j}\binom{n-j}{k-j}^s \frac{(c_0 e)^s}{s^{3/2}}\right).$$

Proof. We first consider the class of rooted (unassigned) two-type trees with a distinguished vertex J of type j as the root and at least one vertex of type k. (Recall that even in an unassigned two-type tree, the vertices are distinguishable.) We will consider the generating function $T_J = T_J(z)$ of this class, where z indicates the number of vertices of type k. In other words, $T_J(z)$ is a formal power series

$$T_J(z) = \sum_{i=0}^{\infty} F_i z^i,$$

where F_i is the number of rooted unassigned two-type trees with i vertices of type k. Then T_J satisfies the following equation:

$$T_J(z) = \exp\left(z(1+T_J(z))^{c_0}\right) - 1. \tag{2.4}$$

Let $W(z)$ denote the Lambert W-function defined by the equation

$$z = W(z)\exp(W(z)).$$

We have

$$T_J(z) = \exp\left(-\frac{W(-c_0 z)}{c_0}\right) - 1. \tag{2.5}$$

By the Lagrange Inversion Theorem (see [33, Appendix A.6]), we have

$$W^r(z) = \sum_{i\geq r} \frac{-r(-i)^{i-r-1}}{(i-r)!} z^i. \tag{2.6}$$

Since $W(z)$ is analytic in a neighbourhood of $z=0$ and $W(0)=0$, for any function $F(z)$ analytic in a neighbourhood of $z=0$, the composition $F(W(z))$ is still analytic near $z=0$. Thus, using Taylor expansion we have

$$\begin{aligned} T_J(z) = \exp\left(-\frac{W(-c_0 z)}{c_0}\right) - 1 &= \sum_{r\geq 1} \frac{1}{r!}\left(-\frac{W(-c_0 z)}{c_0}\right)^r \\ &= \sum_{r\geq 1} \frac{1}{r!} \sum_{i\geq r} \frac{r c_0^{i-r} i^{i-r-1}}{(i-r)!} z^i = \sum_{i\geq 1} z^i \left(\sum_{r=1}^{i} \frac{c_0^{i-r} i^{i-r-1}}{(r-1)!(i-r)!}\right). \end{aligned}$$

The number F_s of rooted unassigned two-type trees with s vertices of type k is $[z^s]T_J(z)$, i.e. the coefficient of z^s in $T_J(z)$. We obtain

$$F_s = [z^s]T_J(z) = \sum_{r=1}^{s} \frac{c_0^{s-r} s^{s-r-1}}{(r-1)!(s-r)!} = \frac{c_0^{s-1} s^{s-2}}{(s-1)!} \sum_{r=0}^{s-1} \frac{c_0^{-r} s^{-r} (s-1)_{(r)}}{r!}.$$

As an upper bound, we have

$$F_s \leq \frac{c_0^{s-1} s^{s-1}}{s!} \sum_{r=0}^{s-1} \frac{c_0^{-r}}{r!} \leq \frac{c_0^{s-1} s^{s-1}}{s!} e^{1/c_0}.$$

As a lower bound, we have

$$F_s \geq \frac{c_0^{s-1} s^{s-1}}{s!}.$$

In each instance of $\mathcal{T}$, we have $\binom{n}{j}$ choices for the assignment of the initial vertex of type j and for each vertex of type k discovered from a vertex of type j, we must choose $k-j$ new elements among $n-j$ (excluding those already in the parent vertex of type j). Then the assignments of the next c_0 vertices of type j are already determined. Therefore, we have the following relation between B_s and F_s:

$$B_s = \binom{n}{j}\binom{n-j}{k-j}^s F_s.$$

We thus deduce the claimed bounds of B_s using bounds on F_s. Moreover, by Stirling's approximation, we obtain the asymptotic behaviour of B_s:

$$B_s = \Theta\left(\binom{n}{j}\binom{n-j}{k-j}^s \frac{(c_0e)^s}{s^{3/2}}\right). \qquad \square$$

Proof of Lemma 2.2.1. We consider $\binom{n}{j}$ independent instances of the branching process $\mathcal{T}$. Let R_s be the random variable which counts the number of vertices of type j present in total in the instances that have size s. In each such instance, the s vertices of type k are present with probability p^s. For the number of absent vertices of type k (i.e. which are not selected during the process) in an instance of size s, we observe that the j-set assigned to the starting vertex of type j is contained in $\binom{n-j}{k-j}$ many k-sets, and subsequently for each of the s vertices of type k we discover c_0 further vertices of type j, whose assignments are each contained in $\left(\binom{n-j}{k-j}-1\right)$ further k-sets. However, we have to consider the s vertices of type k that are indeed discovered. The total number of absent vertices of type k is therefore

$$\binom{n-j}{k-j} + c_0s\left(\binom{n-j}{k-j}-1\right) - s = (1+c_0s)\binom{n-j}{k-j} - s(1+c_0).$$

Therefore, we have

$$\begin{aligned}\mathbb{E}(R_s) &\le B_sp^s(1-p)^{(1+c_0s)\binom{n-j}{k-j}-s(1+c_0)}\\ &= \Theta(1)\binom{n}{j}\left(\binom{n-j}{k-j}c_0ep(1-p)^{c_0\binom{n-j}{k-j}}\right)^s(1-p)^{\binom{n-j}{k-j}-s(1+c_0)}s^{-3/2}.\end{aligned}$$

The last equality is due to Lemma 2.4.1. Recall that $p = (1-\varepsilon)c_0^{-1}\binom{n-j}{k-j}^{-1}$. Using $(1-p) \le e^{-p}$ we have

$$\begin{aligned}\mathbb{E}(R_s) &\le \Theta(1)\binom{n}{j}\left((1-\varepsilon)\cdot e\cdot e^{-(1-\varepsilon)}\right)^s s^{-3/2}\\ &\qquad\cdot\exp\left((1-\varepsilon)(1+c_0^{-1})s\binom{n-j}{k-j}^{-1} - \frac{1-\varepsilon}{c_0}\right)\\ &\le \Theta(1)\binom{n}{j}\exp\left(s(\log(1-\varepsilon)+\varepsilon)\right)s^{-3/2}\exp\left(2s\binom{n-j}{k-j}^{-1}\right).\end{aligned}$$

Since $\delta = -\varepsilon - \log(1-\varepsilon)$, we have

$$\mathbb{E}(R_s) \le \Theta(1)\binom{n}{j}\exp\left(-s\delta\right)s^{-3/2}\exp\left(2s\binom{n-j}{k-j}^{-1}\right).$$

Now, let D_s be the random variable which counts the number of components of $\mathcal{H}$ of size s. By Lemma 2.3.1, $\mathbb{E}(D_s)$ is bounded above by the expected number of instances of the two-type branching process $\mathcal{T}$ of size s. In each such instance every vertex of type k is connected to exactly c_0+1 vertices of type j, therefore their expected number is equal to $\mathbb{E}(R_s)(c_0+1)^{-1}s^{-1}$, where R_s is the number of vertices of type j present in total in all the instances. Denoting by $D_{\ge s} = \sum_{t\ge s} D_t$ the random variable counting the number of components of size at least s, we have

$$\mathbb{E}(D_{\ge s}) \le \sum_{t\ge s}\mathbb{E}(R_t)(c_0+1)^{-1}t^{-1}.$$

Let $\hat{s} = \delta^{-1}(\log\lambda - \frac{5}{2}\log\log\lambda + K(n))$ and $\delta^- = \delta - 2\binom{n-j}{k-j}^{-1}$. Recalling that

$$\delta := -\varepsilon - \log(1-\varepsilon) = \varepsilon^2/2 + O(\varepsilon^3)$$

and that $\varepsilon^2 n^{k-j}(\log n)^{-1} \to \infty$ (by condition (2.3)), it holds that $\delta = \omega(n^{j-k}\log n)$, which means $\delta^- = \delta - 2\binom{n-j}{k-j}^{-1} = (1-o(1))\delta$. Thus, we have $\delta^- > 0$ for n large enough and we obtain

$$\begin{aligned}\mathbb{E}(D_{\ge \hat{s}}) &\le \sum_{t\ge \hat{s}}\mathbb{E}(R_t)(c_0+1)^{-1}t^{-1} \le \Theta(1)\binom{n}{j}(\hat{s})^{-5/2}\sum_{t\ge\hat{s}}(e^{-\delta^-})^t \\ &\le \Theta(1)\binom{n}{j}(\hat{s})^{-5/2}\frac{e^{-\delta^-\hat{s}}}{1-e^{-\delta^-}} \le \Theta(1)\binom{n}{j}(\hat{s})^{-5/2}\frac{e^{-\delta^-\hat{s}}}{\delta^-}.\end{aligned}$$

Without loss of generality, we may assume that $K(n) = o(\log\lambda)$. Therefore, since $\frac{\varepsilon^2 n^{k-j}}{\log n} \to \infty$, we have

$$\begin{aligned}\hat{s}\binom{n-j}{k-j}^{-1} &= (1+o(1))\delta^{-1}(\log\lambda)\binom{n-j}{k-j}^{-1} = O(1)\frac{\log(\varepsilon^3 n^j)}{\varepsilon^2 n^{k-j}} \\ &\le O(1)(\log n)\varepsilon^{-2}n^{j-k} = o(1).\end{aligned}$$

This leads to

$$e^{-\delta^-\hat{s}} = e^{-\delta\hat{s}}\exp\left(2\hat{s}\binom{n-j}{k-j}^{-1}\right) = (1+o(1))e^{-\delta\hat{s}}.$$

Since $\varepsilon^3 n^j \to \infty$ (by condition (2.2)), we have $\lambda = \varepsilon^3 \binom{n}{j} \to \infty$. We thus obtain

$$\begin{aligned}
\mathbb{E}(D_{\geq \hat{s}}) &\leq \Theta(1) \binom{n}{j} (\hat{s})^{-5/2} \frac{(1+o(1))e^{-\delta \hat{s}}}{(1+o(1))\delta} = \Theta(1) \binom{n}{j} (\hat{s})^{-5/2} \frac{e^{-\delta \hat{s}}}{\delta} \\
&\leq \Theta(1) \exp\left(\log \binom{n}{j} - \frac{5}{2} \log\left((1+o(1)) \frac{\log \lambda}{\delta} \right) - \log \lambda \right. \\
&\qquad \left. + \frac{5}{2} \log \log \lambda - K(n) - \log \delta \right) \\
&= \Theta(1) \exp\left(\log \binom{n}{j} + \frac{5}{2} \log \delta - \log\left(\varepsilon^3 \binom{n}{j} \right) - K(n) - \log \delta \right) \\
&= \Theta(1) \exp\left(\frac{3}{2} \left(\log(\varepsilon^2) - \log 2 + \log\left(1 + O(\varepsilon)\right) \right) - \log(\varepsilon^3) - K(n) \right) \\
&= O\left(\exp\left(-K(n) \right) \right).
\end{aligned}$$

Since $K(n) \to \infty$, by Markov's inequality, whp we have $D_{\geq \hat{s}} = 0$, meaning that there is no j-component of size larger than $\hat{s} = \delta^{-1}(\log \lambda - \frac{5}{2} \log \log \lambda + K(n))$. □

2.5 Lower bound on L_1: Proof of Lemma 2.2.2

In this section we will prove that $\mathcal{H}$ contains a hypertree component larger than a certain size, which provides a lower bound on L_1 (Lemma 2.2.2). To this end, we study the only obstacle for a j-component to be a hypertree, which are the *wheels*, a hypergraph analogue of cycles. We first give an upper bound on the number of possible wheels of length $\ell \in \mathbb{N}$ (Lemma 2.5.1). This then implies that almost all possible (not too large) instances of the two-type branching process correspond to hypertrees (Lemma 2.5.2). Lemma 2.5.1 also implies that whp there are no large non-hypertree components in $\mathcal{H}$ (Lemma 2.5.4). The proofs of these auxiliary lemmas will be delayed until Section 2.6.

Before giving these proofs, in this section we use the auxiliary lemmas to determine the asymptotic first and second moments of the number of large components in $\mathcal{H}$, which will be denoted by S_+. Since the second moment is approximately the square of the first (i.e. $\mathbb{E}(S_+^2) \approx \mathbb{E}(S_+)^2$), Chebyshev's inequality will imply that whp there are many such large components, which by Lemma 2.5.4 are hypertrees, as required. This will be made more precise later.

Recall that a hypertree component (i.e. j-component that contains as many j-sets as possible given its size) corresponds to an assigned two-type tree with no repeated assignments. An important structure that may appear in $\mathcal{H}$ is the so-called *wheel*. A *wheel* of length $\ell \geq 2$ is a pair of sequences, one of ℓ distinct hyperedges $K_1, K_2, \ldots, K_\ell, K_{\ell+1} = K_1$ and the other of ℓ distinct j-sets $J_0, J_1, \ldots, J_{\ell-1}, J_\ell = J_0$ such that $J_i \subset K_i \cap K_{i+1}$ for all $1 \leq i \leq \ell$ (see Figure 2.2). Two wheels are considered identical if they only differ by a cyclic rotation or order reversion of the elements of the sequences. Given a wheel, it lies in a single j-connected component, and the presence

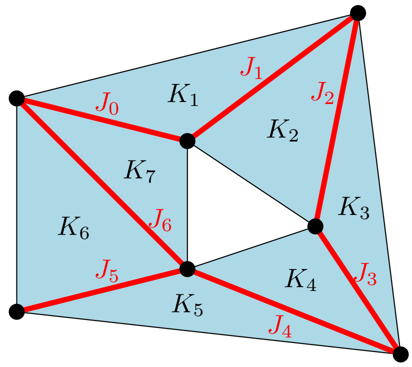


Figure 2.2: An example of a wheel in the case $k = 3$ and $j = 2$.

of a wheel is the only obstacle for a component to be a hypertree. The reason is that a component ceases to be a hypertree if and only if we encounter the same j-set or k-set at least twice in the component search process, which makes a wheel. We have the following enumeration result on wheels.

Lemma 2.5.1. *Let $w_\ell = w_\ell(n)$ be the number of possible wheels of length $\ell \geq 2$, with vertices chosen from $[n]$. We have*

$$w_\ell \leq \frac{c_w n^{k-1}}{p_0^{\ell-1}\ell}, \quad \text{where} \quad c_w = \frac{1}{j!(k-j-1)!} \prod_{m=1}^{j-1} \left(1 - c_0^{-1}\left(\binom{k-m}{j-m} - 1\right)\right)^{-1}.$$

We note that c_w does not depend on n or ℓ. The proof is postponed until Section 2.6.1.

Recall that we denote by $\mathcal{B}$ the set of possible instances of $\mathcal{T}$ and by $\mathcal{B}_s$ the elements in $\mathcal{B}$ with s vertices of type k (bounds on $B_s = |\mathcal{B}_s|$ were given in Lemma 2.4.1). We now consider the subset $\mathcal{B}^-$ of $\mathcal{B}$ formed by rooted assigned two-type trees in which all assignments are distinct, i.e. that correspond to hypertrees, and we denote by $\mathcal{B}_s^-$ the set of elements in $\mathcal{B}^-$ with s vertices of type k.

Lemma 2.5.2. *For $s \geq 2304$, we have*

$$B_s^- := |\mathcal{B}_s^-| = (1 - O(s^2 n^{-1}))B_s.$$

In particular, if $s \to \infty$ and also $s^2 n^{-1} \to 0$, we have $B_s^- = (1 - o(1))B_s$.

Lemma 2.5.2 will be proved in Section 2.6.2.

Let C_s be the number of j-sets in hypertree components of size s in $\mathcal{H}$. It is clear that C_s is a lower bound for the number of j-sets in components of size s.

Lemma 2.5.3. *For $s \geq 2304$ such that $s^2 n^{-1} \to 0$, we have*

$$\mathbb{E}(C_s) \geq \Theta(1) \exp\left(\log\binom{n}{j} - s\delta - \frac{3}{2}\log s\right).$$

Proof. To give a bound on $\mathbb{E}(C_s)$, we bound the probability of each element of $\mathcal{B}_s^-$ occurring in the hypergraph $\mathcal{H}$. Given a hypertree of size s, the probability that it occurs as a component in $\mathcal{H}$ is at least $p^s(1-p)^{(1+sc_0)\binom{n-j}{k-j}}$, since each j-set implies the absence of at most $\binom{n-j}{k-j}$ hyperedges, and since the number of j-sets in a hypertree component of size s is exactly sc_0+1. By Lemma 2.5.2, we have $B_s^- = (1-o(1))B_s$. Using Lemma 2.4.1 and the fact that

$$1-p = e^{-p+O(p^2)} = e^{-p+o\left(\frac{1}{sn^{k-j}}\right)},$$

we have

$$\begin{aligned}
\mathbb{E}(C_s) &\geq B_s^- p^s (1-p)^{(1+sc_0)\binom{n-j}{k-j}} \\
&= (1-o(1))B_s p^s (1-p)^{sc_0\binom{n-j}{k-j}+\binom{n-j}{k-j}} \\
&= \Theta(1)\binom{n}{j}\binom{n-j}{k-j}^s \frac{(c_0 e)^s}{s^{3/2}} \cdot (1-\varepsilon)^s c_0^{-s}\binom{n-j}{k-j}^{-s} e^{-(1-\varepsilon)s} e^{-(1-\varepsilon)c_0^{-1}} \\
&= \Theta(1)\binom{n}{j}\frac{e^s}{s^{3/2}}(1-\varepsilon)^s e^{-(1-\varepsilon)s} \\
&= \Theta(1)\exp\left(\log\binom{n}{j} + s - \frac{3}{2}\log s + s\log(1-\varepsilon) - s(1-\varepsilon)\right) \\
&= \Theta(1)\exp\left(\log\binom{n}{j} - s\left(-\varepsilon - \log(1-\varepsilon)\right) - \frac{3}{2}\log s\right).
\end{aligned}$$

We conclude the proof by recalling that $\delta = -\varepsilon - \log(1-\varepsilon)$ by definition. □

We also consider components in $\mathcal{H}$ that are not hypertree components, and therefore must contain a wheel. We have the following lemma, which will be proved in Section 2.6.3.

Lemma 2.5.4. *Let $s^\circ = s^\circ(n)$ satisfy $s^\circ\delta \geq (j-1)\log n$. Then whp there is no non-hypertree component in $\mathcal{H}$ of size larger than s°.*

In other words, whp any component containing a wheel has size at most s°. We can now combine the results of this section to prove Lemma 2.2.2, using a second moment argument.

Proof of Lemma 2.2.2. We set $s^* = \delta^{-1}\log\lambda$ and $s_* = \delta^{-1}(\log\lambda - \frac{5}{2}\log\log\lambda - K(n))$ where $K(n)\to\infty$. We assume here $K(n) = o(\log\lambda)$ without loss of generality. Furthermore, we set $s_0 = s_* + \delta^{-1}K(n)/2 = \delta^{-1}(\log\lambda - \frac{5}{2}\log\log\lambda - K(n)/2)$. Firstly, we know from Lemma 2.2.1 that whp there is no component larger than s^*. Let S_+ denote the number of j-sets in components of size between s_* and s^*. Let us observe that the range of ε implies that s_* and s_0 satisfy the conditions of Lemma 2.5.3. More precisely, if $s = s_*, s_0$, we have

$$s^2 n^{-1} = (1+o(1))\left(\frac{\log\lambda}{\delta}\right)^2 n^{-1} = (1+o(1))\left(\frac{2\log(\varepsilon^3\binom{n}{j})}{\varepsilon^2}\right)^2 n^{-1} \leq \frac{(2j\log n)^2}{\varepsilon^4 n} \to 0,$$

by our choice of ε satisfying condition (2.1). Therefore we can apply Lemma 2.5.3 and we have the following lower bound by counting only hypertree components:

$$
\begin{aligned}
\mathbb{E}(S_+) &\geq \sum_{s_* \leq s \leq s_0} \mathbb{E}(C_s) \geq (s_0 - s_*)\Theta(1) \exp\left(\log\binom{n}{j} - s_0\delta - \frac{3}{2}\log s_0\right) \\
&\geq \Theta(1)\frac{K(n)}{2\delta} \exp\left(\log\binom{n}{j} - \log\lambda + \frac{5}{2}\log\log\lambda + \frac{K(n)}{2} - \frac{3}{2}\log\left(\delta^{-1}\log\lambda\right)\right) \\
&\geq \Theta(1)\frac{K(n)}{2\delta} \exp\Big(- \log\varepsilon^3 + \frac{5}{2}\log\log\lambda + \frac{K(n)}{2} \\
&\qquad + \frac{3}{2}\log\frac{\varepsilon^2}{2} + \log(1 + O(\varepsilon)) - \frac{3}{2}\log\log\lambda\Big) \\
&\geq \Theta(1)\frac{K(n)}{2\delta} \exp(K(n)/2)\log\lambda = \Theta(1)K(n)\exp(K(n)/2)s^* = \omega(s^*). \qquad (2.7)
\end{aligned}
$$

We now show that $\mathbb{E}(S_+^2)$ is approximately $\mathbb{E}(S_+)^2$. Let q be the probability that a given j-set is in a component of size between s_* and s^*, so $\mathbb{E}(S_+) = q\binom{n}{j}$. For two j-sets J_1, J_2 (not necessarily distinct), let κ_1, κ_2 be the components in which they lie, respectively. Let s_1, s_2 be the sizes of κ_1 and κ_2 respectively. We have

$$
\begin{aligned}
\mathbb{E}(S_+^2) &= \sum_{J_1, J_2} \Pr(s_* \leq s_1 \leq s^*,\ s_* \leq s_2 \leq s^*) \\
&\leq \sum_{J_1} \Pr(s_* \leq s_1 \leq s^*) \sum_{J_2} \Pr(s_2 \geq s_* \mid s_* \leq s_1 \leq s^*).
\end{aligned}
$$

Given that κ_1 is of size between s_* and s^*, we want to bound the probability that κ_2 is of size at least s_*. Given the j-set J_2, if $J_2 \in \kappa_1$, then $\kappa_2 = \kappa_1$ is of size at least s_*; otherwise, we start a modified search process in the hypergraph, where we do not query any k-set which contains a j-set in κ_1. More precisely, we modify the component search process defined in Section 2.3 in the following way: whenever a j-set J_* pops out of the queue, rather than considering every k-set K containing J_*, we consider only the k-sets of $\binom{[n]}{k} \setminus \mathcal{K}$ which contain J_*, where $\mathcal{K}$ is the set of k-sets which contain some j-set of κ_1. This modified search process can be upper coupled by the unmodified search process $\mathcal{T}$, in which all k-sets are still available, but which is now independent of κ_1. Therefore, we have

$$
\begin{aligned}
\mathbb{E}(S_+^2) &\leq \sum_{J_1} \Pr(s_* \leq s_1 \leq s^*)\left(s_1 + \left(\binom{n}{j} - s_1\right)q\right) \\
&\leq \mathbb{E}(S_+)\left(s^* + (1 + o(1))q\binom{n}{j}\right) \\
&= \mathbb{E}(S_+)^2\left(1 + o(1) + \frac{s^*}{\mathbb{E}(S_+)}\right) \\
&\overset{(2.7)}{=} \mathbb{E}(S_+)^2(1 + o(1)),
\end{aligned}
$$

By Lemma 2.2.1 and Chebyshev's inequality, we have

$$\begin{aligned}\Pr(\mathcal{H} \text{ contains no component of size at least } s_*) &\leq \Pr(S_+ = 0) + o(1) \\ &\leq \frac{\mathbb{E}(S_+^2) - \mathbb{E}(S_+)^2}{\mathbb{E}(S_+)^2} + o(1) = o(1).\end{aligned}$$

Similarly, for any fixed constant m the probability that $\mathcal{H}$ contains at most m components of size at least s_* is bounded by $\Pr(S_+ \leq ms^*) + o(1)$. Again by Chebyshev's inequality, we have

$$\Pr(S_+ \leq ms^*) + o(1) \leq \frac{\mathbb{E}(S_+^2) - \mathbb{E}(S_+)^2}{(\mathbb{E}(S_+) - ms^*)^2} + o(1) = o(1).$$

The latter equality is due to the fact that $\mathbb{E}(S_+) \overset{(2.7)}{=} \omega(s^*)$.

We observe that $s_*\delta = (1 + o(1)) \log \lambda = (1 + o(1))(3 \log \varepsilon + j \log n)$, and the fact that $\varepsilon^4 n \geq \frac{\varepsilon^4 n}{(\log n)^2} \to \infty$ (by condition (2.1)) implies that $4 \log \varepsilon + \log n > 0$. We thus have $s_*\delta \geq (j - 1) \log n$ for n large enough. By Lemma 2.5.4, we know that whp all components of size at least s_* are hypertree components, including the m largest components. □

2.6 Wheels: Proofs of auxiliary results

2.6.1 Proof of Lemma 2.5.1

To construct a wheel with ℓ distinct hyperedges $K_1, K_2, \ldots, K_\ell$ and ℓ distinct j-sets $J_0, J_1, \ldots, J_{\ell-1}$ (and we will set $J_\ell = J_0$), we first pick a j-set J_0 and choose the other $k - j$ vertices for K_1. Subsequently, if we have constructed K_i, we pick a j-set J_i in K_i that is different from any $J_{i'}$ picked out before, and choose another $k - j$ vertices to construct K_{i+1}. To make sure that K_ℓ constructed in the end includes the initial j-set J_0, we keep track of vertices in J_0. For each vertex v in $J_0 = J_\ell$, we say that J_i *freezes* v if $v \in J_{i'}$ for every i' such that $i \leq i' \leq \ell$. We will denote by a_i the number of vertices frozen by J_i for $1 \leq i \leq \ell$, and for each choice of the a_i we aim to bound the number of wheels that achieve these values of a_i.

Observe that in order to be an achievable sequence, $(a_1, \ldots, a_\ell)$ must be non-decreasing. Furthermore, by definition, it is clear that only J_ℓ freezes all the j vertices in J_0, thus $a_\ell = j$ and $a_i \leq j - 1$ for all $1 \leq i \leq \ell - 1$. However, it is important to note that we do *not* necessarily have $a_1 = 0$, since it may be that some vertices are present in *every* j-set of the wheel. We will, however, use the bound $a_\ell - a_1 \geq 1$.

If J_i freezes a_i many vertices, then J_i must be chosen to contain all these a_i vertices, and if J_{i+1} freezes a_{i+1} many vertices, then a further $a_{i+1} - a_i$ many vertices must be chosen from J_0 when selecting the $k - j$ vertices to construct K_{i+1} from J_i. Note that every wheel can be obtained in our construction, which gives an over-counting, since for fixed a_i we could inadvertently choose vertices such that some are frozen earlier than necessary.

Let $\tau_0, \tau_1, \tau_2, \dots, \tau_{j-1}$ be integers such that τ_d is the number of hyperedges in the constructed wheel that freeze d vertices in J_0, i.e. the number of a_i's of value d. Since the a_i's are non-decreasing by definition, we can deduce the a_i's from the τ_i's and vice versa. We now consider the number of choices in each step. There are $\binom{n}{j}$ choices for J_0 and $\binom{n-j}{k-j}$ choices for the remaining vertices of K_1. Subsequently for each J_i that freezes a_i many vertices, there are $\binom{k-a_i}{j-a_i} - 1$ choices. Now, to obtain from J_i a k-set K_{i+1} that can contain a J_{i+1} which freezes a_{i+1} many vertices, there are $\binom{n-j-a_{i+1}+a_i}{k-j-a_{i+1}+a_i}$ choices.

The number of constructions $w^\star_\ell(a_1, a_2, \dots, a_\ell)$ with $a_1, \dots, a_\ell$ many vertices frozen by $J_1, \dots, J_\ell$ respectively, is bounded by (with τ_i's computed from the a_i's)

$$\begin{aligned}
w^\star_\ell(a_1, a_2, \dots, a_\ell) &\le \binom{n}{j}\binom{n-j}{k-j}\left[\prod_{i=1}^{\ell-1}\binom{n-j-a_{i+1}+a_i}{k-j-a_{i+1}+a_i}\right]\prod_{m=0}^{j-1}\left(\binom{k-m}{j-m}-1\right)^{\tau_m} \\
&\le \binom{n}{j}\binom{n-j}{k-j}\left[c_0\binom{n-j}{k-j}\right]^{\ell-1}\left[\prod_{i=1}^{\ell-1}\left(\frac{k-j}{n-j-a_{i+1}+a_i}\right)^{a_{i+1}-a_i}\right] \\
&\quad \prod_{m=1}^{j-1}\left(\frac{\binom{k-m}{j-m}-1}{c_0}\right)^{\tau_m} \\
&\le \frac{n^k}{j!(k-j)!p_0^{\ell-1}} \cdot \frac{k-j}{n-2j}\prod_{m=1}^{j-1}\left(\frac{\binom{k-m}{j-m}-1}{c_0}\right)^{\tau_m}.
\end{aligned}$$

Noting that wheels are considered identical up to rotation and reversed order, and that $\binom{k-m}{j-m} - 1 < c_0$ for $1 \le m \le j-1$, by summing over all possible τ_i's (and thus also a_i's) we have

$$\begin{aligned}
w_\ell &\le \frac{1}{2\ell}\sum_{0 \le a_1 \le a_2 \le \dots \le a_\ell = j} w^\star_\ell(a_1, a_2, \dots, a_\ell) \\
&\le \frac{n^k}{2\ell \cdot j!(k-j)!p_0^{\ell-1}} \cdot \frac{k-j}{n-2j}\prod_{m=1}^{j-1}\sum_{\tau_m \ge 0}\left(\frac{\binom{k-m}{j-m}-1}{c_0}\right)^{\tau_m} \\
&\le \frac{n^{k-1}\left(1+\frac{3j^2}{n}\right)}{2\ell \cdot j!(k-j-1)!p_0^{\ell-1}}\prod_{m=1}^{j-1}\left(1 - \frac{\binom{k-m}{j-m}-1}{c_0}\right)^{-1} \le \frac{c_w n^{k-j}}{p_0^{\ell-1}\ell},
\end{aligned}$$

as required. $\square$

2.6.2 Proof of Lemma 2.5.2

As in the proof of Lemma 2.4.1, we will use the generating functions of certain combinatorial classes: given a class $\mathcal{A}$ equipped with a *size function* $|\cdot| : \mathcal{A} \to \mathbb{N}_0$, its generating function

$$A(z) = \sum_{i=0}^{\infty} a_i z^i$$

is a formal power series in z, in which the coefficient a_i of z^i is the number of elements of the class $\mathcal{A}$ with size i, i.e.

$$a_i = |\{A \in \mathcal{A} : |A| = i\}|.$$

We introduce the notion of dominance for comparing generating functions. Recall that given a generating function $F(z)$, we denote by $[z^n]F(z)$ the coefficient of z^n in $F(z)$. For two generating functions $F(z)$ and $G(z)$, we say that $F(z)$ is *dominated* by $G(z)$ (denoted by $F(z) \preceq G(z)$) if for all $n \in \mathbb{N}$ we have $[z^n]F(z) \leq [z^n]G(z)$. For generating functions whose coefficients are non-negative integers, the dominance relation is clearly preserved under by addition, multiplication, differentiation by z and composition.

We denote by $\mathcal{B}^\circ$ the set $\mathcal{B} \setminus \mathcal{B}^-$, i.e. the set of all rooted assigned two-type trees that do not correspond to hypertrees, and $\mathcal{B}_s^\circ$ the corresponding set of elements with s vertices of type k. We first look at the structure of elements in $\mathcal{B}_s^\circ$. We define a *two-type unicycle* to be an assigned two-type graph obtained by attaching rooted assigned two-type trees or nothing to every vertex of type j in an assigned two-type graph generated by hyperedges in a wheel and the j-sets they contain. Note that vertices may share the same assignment.

Proposition 2.6.1. *There is an injection from $\mathcal{B}_s^\circ$ to the disjoint union of the four sets $Q_s^1, Q_s^2, Q_s^3, Q_s^4$, where*

(i) Q_s^1 is the set of tuples $(T^{\bullet,k}, (T_1, T_2, \dots, T_{c_0}), T_+)$ with $s-1$ vertices of type k in total, where $T^{\bullet,k} \in \mathcal{B}$ contains a marked vertex u of type k, then T_+ is either empty or in $\mathcal{B}$, with the assignment of its root the same as the parent of u, and the T_i's are either empty or in $\mathcal{B}$, but the assignments of their roots are fixed to the c_0 many j-subsets of the assignment of u, except that of its parent;

(ii) Q_s^2 is the set of tuples $(T^{\bullet,j}, (T_1, T_2, \dots, T_{c_0}))$ with $s-1$ vertices of type k in total, where $T^{\bullet,j} \in \mathcal{B}$ contains a marked non-root vertex w of type j, and the T_i's are either empty or in $\mathcal{B}$, but the assignments of their roots are fixed to the c_0 many j-subsets of the assignment of the parent of w, excluding that of w;

(iii) Q_s^3 is the set of tuples $(C^{\bullet,k}, (T_1, T_2, \dots, T_{c_0}))$ with $s-1$ vertices of type k in total, where $C^{\bullet,k}$ is a two-type unicycle with a marked vertex u of type k in the cycle and a marked vertex of type j, and T_i's are either empty or in $\mathcal{B}$, but the assignments of their roots are fixed to the c_0 many j-subsets of the assignment of the marked vertex u, excluding that of the parent of u in a breadth-first search starting from the marked vertex of type j using the order provided by assignments;

(iv) Q_s^4 is the set of tuples $(C^{\bullet,j}, T_0)$ with s vertices of type k in total, where $C^{\bullet,j}$ is a two-type unicycle with a marked vertex u of type j in the cycle and another marked vertex of type j, and T_0 is either empty or in $\mathcal{B}$, with the assignment of its root the same as that of u.

Proof. Let C be an element of $\mathcal{B}_s^\circ$. Since C does not correspond to a hypertree, we know that C has at least a pair of vertices with the same assignment. We perform a breadth-first search (with the order provided by assignments), until we meet a vertex v with the same assignment as some other vertex u that comes before. There are three possibilities:

(1) u, v are of type k, sharing the same parent;

(2) u, v are of type k, and u is the grandparent of v;

(3) u, v are of type k, and are at distance at least 4;

(4) u, v are of type j.

In case (1), let w be the common parent of u and v. We first take out v and its c_0 sub-trees $T_1, \ldots, T_{c_0}$ rooted at j-sets, then detach all the children (if any) of w that comes after v, and attach them to a duplicate w' of w to form a new tree T_+. We finally mark the vertex u and get a rooted assigned two-type tree $T^{\bullet,k}$ with one marked vertex of type k along with the T_i's and T_+. Since $C \in \mathcal{B}_s^\circ$, with v removed, we have $s-1$ vertices of type k. Therefore, the tuple $(T^{\bullet,k}, (T_1, T_2, \ldots, T_{c_0}), T_+)$ is in Q_s^1.

In case (2), we take out v and its c_0 sub-trees $T_1, \ldots, T_{c_0}$ rooted at j-sets, and we mark the vertex of type j between u and v. We thus get a rooted assigned two-type tree $T^{\bullet,j}$ with one marked vertex of type j along with the T_i's. Since $C \in \mathcal{B}_s^\circ$ but v has been removed, we have $s-1$ vertices of type k. Therefore, the tuple $(T^{\bullet,j}, (T_1, T_2, \ldots, T_{c_0}))$ is in Q_s^2.

In case (3), we take out the c_0 sub-trees $T_1, \ldots, T_{c_0}$ of v, we merge v and u and we mark u. The merging gives a cycle formed by the path from u and v to their lowest common ancestor in the tree, which corresponds to a wheel because no other vertices have the same assignment in the explored part. We also have two-type trees attached to its type j vertices. We also regard the original root as a marked vertex of type j. This is the two-type unicycle $C^{\bullet,k}$. Since $C \in \mathcal{B}_s^\circ$ and v has been merged with u, we have $s-1$ vertices of type k. The assignments of the roots of the T_i's (when not empty) are fixed to be subsets of size j of the assignment of v, excluding the assignment of its (original) parent. Therefore, the tuple $(C^{\bullet,k}, (T_1, T_2, \ldots, T_{c_0}))$ is in Q_s^3.

In case (4), by construction, the parents of u and v are different. We now separate off the sub-tree T_0 rooted at v, and merge u and v. Similarly to the second case, we have a wheel with a marked vertex of type j and two-type trees attached to its j-sets, and also the root as a marked vertex of type j that can be anywhere. This is the two-type unicycle $C^{\bullet,j}$. No type k vertex has been removed or merged, so there are s in total. The assignment of the root of T_0 is fixed to be the assignment of u. Therefore, the tuple $(C^{\bullet,j}, T_0)$ is in Q_s^4.

To show that this indeed defines an injection, we only need to observe that for each construction, the reverse direction has only one possibility. We can identify which of the three cases we are in from the first element of the resulting tuple. In case (1), given $(T^{\bullet,k}, (T_1, T_2, \ldots, T_{c_0}), T_+) \in Q_s^2$, we attach all the T_i's to a new vertex u' with the same assignment as u, then attach u' to the parent w of u as the last child, and identify the root of T_+ with w from the right. The root of the new tree is that of $T^{\bullet,j}$. In case (2), given $(T^{\bullet,j}, (T_1, T_2, \ldots, T_{c_0})) \in Q_s^2$, we add a new child to the marked vertex w of type j with the same assignment as the parent of w, and attach all T_i's as its sub-trees. The root of the new tree is the root of $T^{\bullet,j}$. In case (3), given $(C^{\bullet,k}, (T_1, T_2, \dots, T_{c_0})) \in Q_s^3$, the marked vertex u of type k is adjacent to two vertices v_1, v_2 of type j on the cycle (where v_1 has the smaller of the two assignments). We replace u by u_1 and u_2, and connect u_1 to v_1 and u_2 to v_2 (thus breaking up the cycle into a path). Additionally, all other vertices which were adjacent to u in $C^{\bullet,k}$ are connected to u_1. Finally the roots of the T_i's are connected to u_2. We thus obtain, a tree which we root at the marked vertex of type j. The construction for case (4) is similar to that of case (3). □

Before computing $B_s^- = |\mathcal{B}_s^-|$ (Lemma 2.5.2), we give a technical lemma in the spirit of Laplace's method.

Lemma 2.6.2. *Given any fixed integer $a \geq 1$, for $s \geq (16a)^2$, we have*

$$\sum_{i=1}^{s} \frac{i^a s_{(i)}}{s^i} \leq 5(2a)^{a/2} s^{(a+1)/2}.$$

Proof. We first observe that

$$\frac{i^a s_{(i)}}{s^i} = \exp\left(a \log i + \sum_{b=0}^{i-1} \log\left(1 - \frac{b}{s}\right)\right) \leq \exp\left(a \log i - \frac{i^2}{4s}\right).$$

Let $S(i) = a \log i - i^2/(4s)$. The maximum of $S(i)$ (viewed as a function on $\mathbb{R}$ rather than $\mathbb{N}$) occurs at $i_{opt} = (2as)^{1/2}$, with value $S(i_{opt}) = \frac{a}{2}(\log(2as) - 1)$. Since $S''(i) = -a/i^2 - 1/(2s) < 0$ for $1 \leq i \leq s$, we know that $S(i)$ is concave in this range. We also observe that, for $\alpha > 0$, we have

$$S(\alpha i_{opt}) = a \log i_{opt} + a \log \alpha - a\frac{(\alpha i_{opt})^2}{2i_{opt}^2} = S(i_{opt}) + a\left(\log \alpha + \frac{1}{2} - \frac{\alpha^2}{2}\right). \tag{2.8}$$

Let $i^* = \lceil (16as \log s)^{1/2} \rceil$. It is clear that $i^* \geq i_{opt} + 1$ for $s \geq (16a)^2$ with $a \geq 1$. Since $S(i)$ is concave and $i^* \geq i_{opt} + 1$, we have

$$\begin{aligned} S(i^*) \leq S((16as \log s)^{1/2}) &= a \log\left((16as \log s)^{1/2}\right) - \frac{16as \log s}{4s} \\ &= \frac{a}{2}\left(\log s + \log(16a) + \log\log(s)\right) - 4a \log s \\ &\leq \frac{a}{2}(\log\log s - 7\log s) + \frac{a}{2}\log(16a), \end{aligned}$$

which is decreasing in s. In the case $s = (16a)^2$, we have $S(i^*) = a \log\log(16a) - \frac{13a}{2}\log(16a)$, which is clearly negative for $a \geq 1$. Therefore, $S(i^*) < 0$ holds for all $s \geq (16a)^2$. Therefore, by the concavity of $S(i)$ and the fact that $i_{opt} < i^*$ for our range of s, we have

$$\begin{aligned}
\sum_{i=1}^{s} \frac{i^a s_{(i)}}{s^i} &\le \sum_{i=1}^{i^*} \exp\left(a\log i - \frac{i^2}{4s}\right) + s \le s + \exp(S(i_{opt})) + \int_1^{i^*} \exp(S(x))dx \\
&\overset{(2.8)}{\le} s + \exp(S(i_{opt})) + i_{opt}\exp(S(i_{opt}))e^{a/2}\int_0^{+\infty} \exp(a(\log\alpha - \alpha^2/2))d\alpha \\
&\le s + e^{-a/2}(2as)^{a/2} + (2as)^{(a+1)/2}\int_0^{+\infty}\exp(-a\alpha^2/4)d\alpha \\
&\le s + (2as)^{a/2} + 2\sqrt{\pi}(2a)^{a/2}s^{(a+1)/2} \\
&\le (\sqrt{1/2} + 1 + 2\sqrt{\pi})(2a)^{a/2}s^{(a+1)/2} \le 5(2a)^{a/2}s^{(a+1)/2}.
\end{aligned}$$

In the third line we used $\log\alpha < \alpha^2/4$, which holds for all $\alpha > 0$. This is because $\log\alpha - \alpha^2/4$ takes its maximum at $\alpha = \sqrt{2}$, where it has a negative value. □

We now consider the generating function of two-type graphs corresponding to wheels

$$w(z) = \sum_{\ell\ge 2} w_\ell z^\ell, \tag{2.9}$$

and that of the set $\mathcal{B}_s$ and $\mathcal{B}_s^\circ$, denoted by $B(z)$ and $B^\circ(z)$ respectively. By Proposition 2.6.1, we can partition $\mathcal{B}_s^\circ$ into four disjoint subsets $\mathcal{B}_s^{\circ,1}$, $\mathcal{B}_s^{\circ,2}$, $\mathcal{B}_s^{\circ,3}$ and $\mathcal{B}_s^{\circ,4}$ (i.e. the preimages of Q_s^1, Q_s^2, Q_s^3 and Q_s^4 under the injection) with generating functions $B^{\circ,1}(z)$, $B^{\circ,2}(z)$, $B^{\circ,3}(z)$, $B^{\circ,4}(z)$ respectively. We thus have $B^\circ(z) = B^{\circ,1}(z) + B^{\circ,2}(z) + B^{\circ,3}(z) + B^{\circ,4}(z)$, where $[z^s]B^{\circ,i}(z) = |\mathcal{B}_s^{\circ,i}|$ for $i = 1, 2, 3, 4$.

From Lemma 2.5.1 and the fact that a wheel consists of at least 2 hyperedges, we have

$$w(z) \preceq c_w n^{k-j} p_0 \left(\log\frac{1}{1 - p_0^{-1}z} - p_0^{-1}z\right). \tag{2.10}$$

The generating function $w^\bullet(z)$ of two-type graphs corresponding to wheels with a marked k-set is therefore

$$w^\bullet(z) = \frac{zd}{dz}w(z) \preceq \frac{c_w n^{k-1}p_0^{-1}z^2}{1 - p_0^{-1}z}. \tag{2.11}$$

The generating function of two-type graphs corresponding to wheels with a marked vertex of type j is dominated by $(c_0+1)w^\bullet(z)$, since vertex of type k in such two-type graphs is adjacent to (c_0+1) vertices of type j.

We recall that $T_J(z)$ is the generating function of unassigned two-type trees (defined in (2.4)). The generating function of assigned two-type trees with a given j-set as root assignment is given by $T_J(p_0^{-1}c_0^{-1}z)$, since for each vertex of type k, there are $\binom{n-j}{k-j} = p_0^{-1}c_0^{-1}$ choices to complete its assignment from that of its parent, which is of type j and has a j-set as assignment. We have an extra factor of $\binom{n}{j}$ when the root assignment is not given. We recall (2.5), where $T_J(z)$ is expressed with the Lambert W-function $W(z)$, which satisfies the equation $z = W(z)\exp(W(z))$:

$$T_J(z) = \exp\left(-c_0^{-1}W(-c_0 z)\right) - 1.$$

Hence,

$$1 + T_J(p_0^{-1}c_0^{-1}z) = \exp(-c_0^{-1}W(-p_0^{-1}z)). \tag{2.12}$$

By differentiating the equation $z = W(z)\exp(W(z))$, we have the following expression of $\frac{d}{dz}W(z)$:

$$\frac{d}{dz}W(z) = \frac{1}{\exp(W(z))(1+W(z))} = \frac{W(z)}{z(1+W(z))}. \tag{2.13}$$

We can now also compute the derivative of $T_J(z)$ as

$$\begin{aligned}\frac{d}{dz}T_J(z) &= \exp(-c_0^{-1}W(-c_0z)) \cdot (-c_0^{-1}) \cdot \frac{d}{dz}(W(-c_0z)) \\ &= -\exp(-c_0^{-1}W(-c_0z))\frac{W(-c_0z)}{c_0z(1+W(-c_0z))}.\end{aligned} \tag{2.14}$$

We can then give the following dominance relation for the derivative of $w^\bullet(z)$, which will be used later. To simplify the notation, we define

$$W_0 = W_0(z) = W(-p_0^{-1}z).$$

Since $W(z) = z\exp(-W(z))$, we have

$$\exp(-W_0) = \frac{W_0}{-p_0^{-1}z}. \tag{2.15}$$

Furthermore, by (2.13), we also have

$$\frac{d}{dz}W_0 = \frac{W_0}{z(1+W_0)}. \tag{2.16}$$

We thus have

$$\begin{aligned}&\frac{zd}{dz}\left[w^\bullet(z(1+T_J(c_0^{-1}p_0^{-1}z))^{c_0}\right] \overset{(2.5)}{=} \frac{zd}{dz}\left[w^\bullet\left(z\exp\left(-W(-p_0^{-1}z)\right)\right)\right] \\ &\overset{(2.15)}{=} \frac{zd}{dz}\left[w^\bullet(-p_0W_0)\right] \overset{(2.11)}{\preceq} zc_wn^{k-1}p_0^{-1}\frac{d}{dz}\left(\frac{p_0^2W_0^2}{1+W_0}\right) \\ &\overset{(2.16)}{=} \frac{zc_wn^{k-1}p_0W_0\,(2+W_0)}{(1+W_0)^2} \cdot \frac{W_0}{z(1+W_0)} = \frac{c_wn^{k-1}p_0W_0^2\,(2+W_0)}{(1+W_0)^3}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\frac{zd}{dz}\left[w^\bullet(z(1+T_J(c_0^{-1}p_0^{-1}z))^{c_0}\right] &= \frac{c_wn^{k-1}p_0W_0^2\,(2+W_0)}{(1+W_0)^3} \\ &= c_wn^{k-1}p_0\sum_{i\geq 2}\frac{(-1)^i(i-1)(i+2)}{2}W_0^i,\end{aligned} \tag{2.17}$$

where we have used the expansion

$$\frac{x^2(2+x)}{(1+x)^3} = \sum_{i\geq 2} \frac{(-1)^i(i-1)(i+2)}{2} x^i.$$

We now consider the generating functions $B^{\circ,1}(z), B^{\circ,2}(z), B^{\circ,3}(z), B^{\circ,4}(z)$ arising from Proposition 2.6.1. For $B^{\circ,1}(z)$, using (2.13) we have

$$\begin{aligned}
B^{\circ,1}(z) &\preceq z\left(\frac{zd}{dz}\binom{n}{j}T_J(p_0^{-1}c_0^{-1}z)\right)(1+T_J(p_0^{-1}c_0^{-1}z))^{c_0+1}\\
&\overset{(2.12),(2.14)}{=} z^2\binom{n}{j}(p_0^{-1}c_0^{-1})(-\exp(-c_0^{-1}W_0))\frac{W_0}{p_0^{-1}z(1+W_0)}\exp(-(1+c_0^{-1}W_0)\\
&= -c_0^{-1}z\binom{n}{j}\frac{W_0}{(1+W_0)}\exp(-(1+2c_0^{-1})W_0))\\
&= c_0^{-1}\binom{n}{j}z\sum_{i\geq 1}\sum_{r\geq 0}\frac{(-1)^{i+r}(1+2c_0^{-1})^r}{r!}W_0^{i+r}.
\end{aligned}$$

The extra factor z in the initial domination comes from the change of the number of vertices of type k in the injection. Therefore, by (2.6) and using the fact that $i+r \leq i(r+1)$ for integers $i \geq 1$, $r \geq 0$, we have

$$\begin{aligned}
|\mathcal{B}_s^{\circ,1}| = [z^s]B^{\circ,1}(z) &\leq c_0^{-1}\binom{n}{j}\sum_{i\geq 1}\sum_{r\geq 0}\frac{(-1)^{i+r+s}(i+r)(1+2c_0^{-1})^r(-s+1)^{s-i-r-2}}{p_0^{s-1}(s-1-r-i)!r!}\\
&= c_0^{s-2}\binom{n}{j}\binom{n-j}{k-j}^{s-1}\sum_{i\geq 1}\sum_{r\geq 0}\frac{i(r+1)(1+2c_0^{-1})^r(s-1)^{s-i-r-2}}{(s-1-r-i)!r!}\\
&= c_0^{s-2}\binom{n}{j}\binom{n-j}{k-j}^{s-1}\sum_{i\geq 1}\frac{i(s-1)^{s-i-2}}{(s-1-i)!}\sum_{r\geq 0}\frac{(r+1)(1+2c_0^{-1})^r(s-1-i)_{(r)}}{(s-1)^r r!}\\
&\leq \binom{n}{j}\binom{n-j}{k-j}^{s-1}\frac{c_0^{s-2}(s-1)^{s-2}}{(s-1)!}\sum_{i=1}^{s-1}\frac{i(s-1)_{(i)}}{(s-1)^i}\sum_{r\geq 0}\frac{(r+1)2^r}{r!}\\
&\leq (1+3e^2)\binom{n}{j}\binom{n-j}{k-j}^{s-1}\frac{c_0^{s-2}s^{s-1}}{s!}\sum_{i=1}^{s-1}\frac{is_{(i)}}{s^i}.
\end{aligned}$$

Now using Lemma 2.6.2 with $a=1$, for $s \geq 256$ we have

$$|\mathcal{B}_s^{\circ,1}| \leq (1+3e^2)\binom{n}{j}\binom{n-j}{k-j}^{s-1} 5\cdot 2^{1/2}\frac{c_0^{s-2}s^s}{s!} = O(sn^{j-k})B_s,$$

where the last equality follows from Lemma 2.4.1.

For $B^{\circ,2}$, observing that in any two-type tree the number of non-root type j vertices is $c_0 s$ times the number of type k vertices. We can now reduce the computation to that of $B^{\circ,1}$ by observing that $1+T_J(p_0^{-1}c_0^{-1}z)$ is a series with positive coefficients:

$$\begin{aligned}B^{\circ,2}(z) &\preceq c_0 z\left(\frac{zd}{dz}\binom{n}{j}T_J(p_0^{-1}c_0^{-1}z)\right)(1+T_J(p_0^{-1}c_0^{-1}z))^{c_0}\\ &= c_0 B^{\circ,1}(z)(1+T_J(p_0^{-1}c_0^{-1}z))^{-1} \preceq c_0 B^{\circ,1}(z).\end{aligned}$$

Now using the bound on $|\mathcal{B}_s^{\circ,1}|$, we have

$$|\mathcal{B}_s^{\circ,1}| \le c_0|\mathcal{B}_s^{\circ,1}| = O(sn^{j-k})B_s.$$

Now for $B^{\circ,3}$ we have

$$\begin{aligned}B^{\circ,3}(z) &\preceq (c_0+1)z\frac{zd}{dz}\left[w^{\bullet}(z(1+T_J(c_0^{-1}p_0^{-1}z))^{c_0})\right](1+T_J(c_0^{-1}p_0^{-1}z))^{c_0}\\ &\overset{(2.17)}{\preceq} 2c_0 z c_w n^{k-1}p_0\sum_{i\ge 2}\frac{(-1)^i(i-1)(i+2)}{2}W_0^i\exp(-W_0)\\ &= 2zc_0c_wn^{k-1}p_0\sum_{i\ge 2}\sum_{r\ge 0}\frac{(-1)^{r+i}(i-1)(i+2)}{2\cdot r!}W_0^{i+r}.\end{aligned}$$

Therefore, by (2.6), again using the bound $i+r\le i(r+1)$, we have

$$\begin{aligned}|\mathcal{B}_s^{\circ,3}| &\le [z^s]B^{\circ,3}(z) \le 2c_0c_wn^{k-1}p_0\sum_{i\ge 2}\sum_{r\ge 0}\frac{(i-1)(i+2)(i+r)(s-1)^{s-r-i-2}}{2p_0^{s-1}r!(s-r-i-1)!}\\ &\le 2(k-j)!c_0^{s-1}c_wn^{j-1}\binom{n-j}{k-j}^{s-1}\sum_{i\ge 2}\frac{i^2(i+1)(s-1)^{s-i-2}}{(s-i-1)!}\\ &\quad\cdot\sum_{r\ge 0}\frac{(r+1)(s-i-1)_{(r)}}{(s-1)^r r!}\\ &\le 2(k-j)!c_0^{s-1}c_wn^{j-1}\binom{n-j}{k-j}^{s-1}\frac{(s-1)^{s-2}}{(s-1)!}\sum_{i=2}^{s-1}\frac{2i^3(s-1)_{(i)}}{(s-1)^i}\sum_{r\ge 0}\frac{r+1}{r!}\\ &\le 4(k-j)!c_0^{s-1}c_wn^{j-1}\binom{n-j}{k-j}^{s-1}\frac{s^{s-1}}{s!}\sum_{i=2}^{s}\frac{i^3s_{(i)}}{s^i}(1+2e).\end{aligned}$$

Now using Lemma 2.6.2 with $a=3$ and Lemma 2.4.1, for $s\ge 2304$ we have

$$|\mathcal{B}_s^{\circ,3}| \le 4(k-j)!(1+2e)c_wn^{j-1}\binom{n-j}{k-j}^{s-1}5\cdot 6^{3/2}\frac{c_0^s s^{s+1}}{s!} = O(s^2n^{j-1-k})B_s.$$

For $B^{\circ,4}$, using (2.17), we have

$$\begin{aligned}
B^{\circ,4}(z) &\preceq (c_0+1)\frac{zd}{dz}\left[(c_0+1)w^{\bullet}(z(1+T_J(c_0^{-1}p_0^{-1}z))^{c_0})\right](1+T_J(c_0^{-1}p_0^{-1}z)) \\
&\overset{(2.17)}{\preceq} (c_0+1)^2 c_w n^{k-1} p_0 \sum_{i\ge 2} \frac{(-1)^i(i-1)(i+2)}{2} W_0^i \exp(-c_0^{-1}W_0) \\
&= (c_0+1)^2 c_w n^{k-1} p_0 \sum_{i\ge 2}\sum_{r\ge 0} \frac{(-1)^{r+i}(i-1)(i+2)}{2c_0^r r!} W_0^{i+r}.
\end{aligned}$$

Therefore, by (2.6) and the bound $i+r \le i(r+1)$, we have

$$\begin{aligned}
|\mathcal{B}_s^{\circ,4}| = [z^s]B^{\circ,4}(z) &\le (c_0+1)^2 c_w n^{k-1} p_0 \sum_{i\ge 2}\sum_{r\ge 0} \frac{(i-1)(i+2)(i+r)s^{s-r-i-1}}{2c_0^r r!(s-r-i)!p_0^s} \\
&\le \frac{4c_0^2\cdot 2(k-j)!\cdot c_0^{s-1} c_w n^{j-1} s^{s-1}}{s!}\binom{n-j}{k-j}^s \sum_{i\ge 2}\sum_{r\ge 0}\frac{i^2(i+1)(r+1)s_{(r+i)}}{2c_0^r r! s^{r+i}} \\
&\le \frac{8(k-j)!c_0^{s+1} c_w n^{j-1} s^{s-1}}{s!}\binom{n-j}{k-j}^s \sum_{i=2}^{s}\frac{i^3 s_{(i)}}{s^i}\sum_{r\ge 0}\frac{(r+1)(s-i)_{(r)}}{c_0^r r! s^r} \\
&\le \frac{8(k-j)!c_0^{s+1} c_w n^{j-1} s^{s-1}}{s!}\binom{n-j}{k-j}^s \sum_{i=2}^{s}\frac{i^3 s_{(i)}}{s^i}\sum_{r\ge 0}\frac{r+1}{r!} \\
&= \frac{8(k-j)!(1+2e)c_0^{s+1} c_w n^{j-1} s^{s-1}}{s!}\binom{n-j}{k-j}^s \sum_{i=2}^{s}\frac{i^3 s_{(i)}}{s^i}.
\end{aligned}$$

Again, using Lemma 2.6.2 with $a=3$ and Lemma 2.4.1, for $s \ge 2304$ we have

$$|\mathcal{B}_s^{\circ,4}| \le \frac{8(k-j)!(1+2e)c_0^{s+1} c_w n^{j-1}\cdot 5\cdot 6^{3/2} s^{s+1}}{s!}\binom{n-j}{k-j}^s = O(s^2 n^{-1})B_s.$$

Putting everything together,

$$|\mathcal{B}_s^{\circ}| = \left(O(sn^{j-k}) + O(s^2 n^{j-1-k}) + O(s^2 n^{-1})\right)B_s = O(s^2 n^{-1})B_s.$$

Since $B_s^- = B_s - |\mathcal{B}_s^{\circ}|$, this completes the proof. Note that we only need $s \ge 2304$ for all conditions concerning Lemma 2.6.2 to be fulfilled. $\square$

2.6.3 Proof of Lemma 2.5.4

Any non-hypertree j-component contains a wheel, since this is the only obstacle for a component to be a hypertree. We denote by $N_{\ge s}$ the number of non-hypertree components of size at least s. We consider the following *wheel-based branching process*: we start with a family of ℓ many k-sets that is the family of k-sets in a possible wheel and check if they exist as hyperedges in $\mathcal{H}$; if so, we perform $c_0\ell$ branching process, starting from each j-set contained in k-sets of the wheel (with possible duplications if some j-set belongs to more than two hyperedges or if hyperedges intersect in more than j vertices). Since a wheel of length ℓ has at most $c_0\ell$ many j-sets, by the same argument as in Lemma 2.3.1, we know that the expected number of non-hypertree j-component of size at least s is bounded from above by the expected number of wheel-based branching processes of size at least s. Let $N'_{\geq s}$ be the number of instances of wheel-based branching processes of size at least s, starting from every possible family of k-sets that could form a wheel. Then we have

$$\mathbb{E}(N_{\geq s}) \leq \mathbb{E}(N'_{\geq s}).$$

Let u_s be the number of possible two-type unicycles with s vertices of type k and with a marked vertex of type j. We first bound u_s.

Lemma 2.6.3. *For $s \geq 1024$, we have*

$$u_s \leq 122 c_0^2 c_w n^{k-1} p_0^{1-s} \frac{s^{s+1/2}}{s!}.$$

Proof. Let $U(z)$ be the generating function of two-type unicycles with a marked vertex of type j, with z indicating their sizes. Recall that $w(z)$ is the generating function of wheels with z indicating the number of hyperedges, defined in (2.9). We recall that $W_0 = W_0(z) = W(-p_0^{-1}z)$. Using arguments from the proof of Lemma 2.5.2 in Section 2.6.2, from the correspondence above, we have

$$\begin{aligned}
U(z) &\preceq (c_0+1)\frac{zd}{dz}\left[w(z(1+T_J(c_0^{-1}p_0^{-1}z))^{c_0})\right] \overset{(2.12)}{\preceq} 2c_0\frac{zd}{dz}\left[w(z\exp(-W_0))\right] \\
&\overset{(2.15)}{=} 2c_0\frac{zd}{dz}\left[w(-p_0W_0)\right] \overset{(2.16),(2.11)}{\preceq} \frac{2c_0^2 z c_w n^{k-1} p_0 W_0^2}{1+W_0} \cdot \frac{1}{z(1+W_0)} \\
&= \frac{2c_0^2 c_w n^{k-1} p_0 W_0^2}{(1+W_0)^2} = 2c_0^2 c_w n^{k-j} p_0 \sum_{r\geq 2}(-1)^r(r-1)W_0^r.
\end{aligned}$$

The initial domination comes from the definition of two-type unicycle and the fact that there are at most $(c_0+1)s$ vertices of type j in a two-type unicycle of size s. The last equality follows from the substitution of the Taylor expansion of $x^2(1+x)^{-2}$ with $x = W_0$.

We can now use (2.6) to estimate u_s.

$$\begin{aligned}
u_s = [z^s]U(z) &\leq 2c_0^2 c_w n^{k-1} p_0 \sum_{r\geq 2} \frac{r(r-1)s^{s-r-1}}{(s-r)!} p_0^{-s} \\
&= 2c_0^2 c_w n^{k-1} p_0^{1-s} \frac{s^{s-3}}{(s-3)!}\left(2 + \sum_{i=1}^{s-2} \frac{(i+1)(i+2)(s-3)_{(i)}}{s^i}\right) \\
&\leq 2c_0^2 c_w n^{k-1} p_0^{1-s} \frac{s^{s-1}}{s!}\left(2 + \sum_{i=1}^{s} \frac{6i^2 s_{(i)}}{s^i}\right) \\
&\leq 2c_0^2 c_w n^{k-1} p_0^{1-s} \frac{s^{s-1}}{s!}\left(2 + 6\cdot 20 s^{3/2}\right) \leq 244 c_0^2 c_w n^{k-1} p_0^{1-s} \frac{s^{s+1/2}}{s!}.
\end{aligned}$$

The penultimate inequality comes from the application of Lemma 2.6.2 with $a = 2$, which holds for $s \geq 1024$. $\square$

We now estimate $\mathbb{E}(N'_{\geq s})$. Given a two-type unicycle of size s with a wheel of length ℓ, the probability for it to appear in the corresponding wheel-based branching process is at most $p^s(1-p)^{sc_0\binom{n-j}{k-j}-sc_0}$, with p^s accounting for the existences of hyperedges, and $(1-p)^{sc_0\binom{n-j}{k-j}-sc_0}$ for the absences of hyperedges. Indeed, since every assigned two-type tree that corresponds to a hypertree with s' vertices of type k has $s'c_0+1$ vertices of type j, in all ℓc_0 branching processes, $s-\ell$ hyperedges are discovered, meaning that there are in total $(s-\ell)c_0 + \ell c_0 = sc_0$ many j-sets in the branching process. Then, each j-set makes at least $\left(\binom{n-j}{k-j}-1\right)$ queries to k-sets, receiving $s-\ell$ affirmatives in total, which makes the number of absences at least

$$sc_0\left(\binom{n-j}{k-j}-1\right) - s + \ell \geq sc_0\binom{n-j}{k-j} - s(c_0+1).$$

Let P_s be the number of j-sets in wheel-based branching processes of size exactly s. We recall that $p = p_0(1-\varepsilon)$. With Lemma 2.6.3, we have

$$\begin{aligned}
\mathbb{E}(P_s) &\leq u_s p^s (1-p)^{sc_0\binom{n-j}{k-j}-s(c_0+1)} \\
&\leq \Theta(1) n^{k-1} p_0 (1-\varepsilon)^s \frac{s^{s+1/2}}{s!}(1-p)^{sc_0\binom{n-j}{k-j}-s(c_0+1)} \\
&\leq \Theta(1) n^{k-1} p_0 (1-\varepsilon)^s e^s \exp\left(-p\left(sc_0\binom{n-j}{k-j} - s(c_0+1)\right)\right).
\end{aligned}$$

Using $p_0 = c_0^{-1}\binom{n-j}{k-j}^{-1}$ and $\delta = -\varepsilon - \log(1-\varepsilon)$, we have

$$\begin{aligned}
\mathbb{E}(P_s) \leq \Theta(1) n^{j-1} &\exp\left(s\log(1-\varepsilon) + s - (1-\varepsilon)s\right) \\
&\cdot \exp\left((1-\varepsilon)s(c_0^{-1}+1)\binom{n-j}{k-j}^{-1}\right) \\
\leq \Theta(1) n^{j-1} &\exp(-s\delta)\exp\left(2s\binom{n-j}{k-j}^{-1}\right).
\end{aligned}$$

We are now interested in $N'_{\geq s^\circ}$. We can assume that $s^\circ \leq \binom{n}{k}$, i.e. the total number of possible hyperedges. Let $\delta^- = \delta - 2\binom{n-j}{k-j}^{-1}$. Since $\delta = \varepsilon^2/2 + O(\varepsilon^3)$, with $\varepsilon^2 n^{k-j}(\log n)^{-1} \to \infty$ (by condition (2.3)), we have $\delta^- = (1-o(1))\delta$, therefore $\delta^- > 0$. As each wheel-based branching process with s hyperedges has at least sc_0 many j-sets, we have

$$\begin{aligned}
\mathbb{E}(N'_{\geq s^\circ}) &\leq \sum_{s \geq s^\circ} c_0^{-1} s^{-1}\mathbb{E}(P_s) \leq \sum_{s\geq s^\circ} \Theta(1) s^{-1} n^{j-1}\exp(-s\delta^-) \\
&\leq \Theta(1) n^{j-1}(s^\circ)^{-1}\sum_{s\geq s^\circ}\exp(-s\delta^-) \leq \Theta(1) n^{j-1}\frac{\exp(-s^\circ\delta^-)}{s^\circ\delta^-}.
\end{aligned}$$

Since $s^\circ(\delta - \delta^-) = s^\circ\binom{n-j}{k-j}^{-1} = o(\delta^{-1}n^{j-k}\log n) = o(1)$, and $s^\circ\delta \geq (j-1)\log n \to \infty$, we have

$$\mathbb{E}(N_{\geq s^\circ}) \leq \mathbb{E}(N'_{\geq s^\circ}) \leq \Theta(1)n^{j-1}\frac{\exp(-s^\circ\delta)\exp(o(1))}{s^\circ\delta + o(1)} = \Theta(1)n^{j-1}\frac{\exp(-s^\circ\delta)}{s^\circ\delta}.$$

We conclude that $\mathbb{E}(N_{\geq s^\circ}) \to 0$, using that $s^\circ\delta \geq (j-1)\log n$. By Markov's inequality, whp we have $N_{\geq s^\circ} = 0$, meaning that there is no non-hypertree component in $\mathcal{H}$ of size at least $(j-1)\log n\delta^{-1}$. □

2.7 Concluding remarks

2.7.1 The critical window

We note that the proofs of Lemmas 2.2.1 and 2.2.2, and thus also of Theorem 2.1.3, impose three restrictions on ε, namely conditions (2.1)–(2.3) – we discuss each of these in turn.

Condition (2.1): $\frac{\varepsilon^4 n}{(\log n)^2} \to \infty$.

This condition is required for the lower bound since our proof strategy is to show that whp the largest components are hypertrees. It seems likely that the lower bound on the component sizes also holds for smaller ε – however, heuristically the largest components will no longer be hypertrees, but will likely contain wheels. This means that new techniques are required to estimate the number of potential components containing wheels.

For the upper bound, however, this condition on ε is not required, and can be replaced by two weaker conditions.

Condition (2.2): $\varepsilon^3 n^j \to \infty$.

This condition was conjectured in [20] to give precisely the critical window. Our results strengthen the evidence for this conjecture, since the size of the largest components in the subcritical case (approximately $\frac{2\log(\varepsilon^3 n^j)}{\varepsilon^2}$) and in the supercritical case (approximately $2\varepsilon\binom{n}{j}$, see [20]) would be of the same order when $\varepsilon^3 n^j = \Theta(1)$.

Condition (2.3): $\frac{\varepsilon^2 n^{k-j}}{\log n} \to \infty$.

This condition is also required to prove the upper bound, in particular to guarantee that the sizes of the largest components (approximately $\frac{2\log(\varepsilon^3 n^j)}{\varepsilon^2}$) is negligible compared to $\binom{n-j}{k-j}$. It is not clear whether the condition is simply an artefact of our proof methods or a necessary restriction (implying that the critical window is not necessarily as conjectured). Further study is needed to clarify the situation.

2.7.2 Global structure

In this paper we investigated the structure of the largest components, but it would be interesting to characterise the structure of the *whole* subcritical hypergraph. To this end, one would need to investigate wheels more carefully.

Wheels can be classified by the size of their *centre*, i.e. the set of vertices shared by every hyperedge in the same wheel. If the centre of the wheel contains i vertices, we call it an i-wheel.

Heuristically, the behaviour of the hypergraph with respect to i-wheels is fundamentally dependent on i. In particular, at what probability we can expect i-wheels to appear and how frequently changes as i ranges from 0 to $j-1$ – those with a larger centre appear first. Furthermore, the length of a typical i-wheel seems to be constant if $i \geq 1$, but logarithmic in n for $i = 0$, which makes the analysis of this case rather different.

It would be interesting to study these results further and rigorously prove these heuristics, as well as tackling further questions, such as

- How many i-wheels are there in total, and of which length?
- How many j-sets are in components containing i-wheels?
- What is the size of the largest component that contains an i-wheel?
- Do any components contain more than one wheel?

2.7.3 Symmetry phenomenon

As mentioned in the introduction, random graphs display a symmetry phenomenon around the phase transition: the subcritical random graph with $p = (1-\varepsilon)p_0$ has approximately the same distribution as the supercritical random graph with $p = (1+\varepsilon)p_0$ with the giant component removed, in particular regarding the orders of components and their structure (they are all trees or unicyclic). More generally, the same is true in k-uniform hypergraphs for the case $j = 1$ (see for example [8], Lemma 15).

However, the analogous result for any $j \geq 2$ has not yet been proved. The order of the giant component was asymptotically determined in [20], but in contrast to the case $j = 1$, the distribution of the remaining hypergraph fundamentally depends not only on the number of j-sets in the giant component, but also on how they are distributed across the vertices, making this case rather more challenging.

3 Vanishing of cohomology groups of random simplicial complexes

3.1 Introduction

3.1.1 Motivation

In their seminal paper [28], Erdős and Rényi introduced the uniform random graph $G(n, m)$, chosen uniformly at random from the set of all labelled graphs on n vertices with exactly m edges and, among other results, addressed the problem of determining the probability of this graph being connected. This classical result is usually stated for the binomial random graph $G(n, p)$ on n vertices, in which each edge is present with a given probability p independently: the property of $G(n, p)$ being connected undergoes a *phase transition* around the sharp threshold $p = \frac{\log n}{n}$ [60]. Throughout the paper, we denote the natural logarithm by log and we say that an event holds *with high probability* (*whp* for short) if it holds with probability tending to 1 as n tends to infinity.

Theorem 3.1.1 ([28, 60]). *Let ω be any function of n which tends to infinity as $n \to \infty$. Then* with high probability *the following holds.*

(i) If $p = \frac{\log n - \omega}{n}$, then $G(n, p)$ is not connected.

(ii) If $p = \frac{\log n + \omega}{n}$, then $G(n, p)$ is connected.

As an even stronger result, Erdős and Rényi [28] determined the limiting probability of $G(n, p)$ being connected around the point of the phase transition. More precisely, this result can be stated for $G(n, p)$ as follows.

Theorem 3.1.2 (see e.g. [34, Theorem 4.1]). *Let $c \in \mathbb{R}$ be a constant and suppose that $(c_n)_{n \geq 1}$ is a sequence of real numbers that converges to c as $n \to \infty$. If*

$$p = \frac{\log n + c_n}{n},$$

then

$$\Pr\left(G(n, p) \textit{ is connected}\right) \xrightarrow{n \to \infty} e^{-e^{-c}}.$$

We note that while [34, Theorem 4.1] is stated for the uniform random graph, it is actually proved via the binomial model $G(n, p)$ and thus immediately translates into Theorem 3.1.2.

Subsequently, Bollobás and Thomason [15] proved a *hitting time* result for the random graph process, in which edges are added one at a time uniformly at random. This result relates the connectedness of the random graph process to the disappearance of the last *smallest obstruction*, an isolated vertex.

Theorem 3.1.3 ([15]). *With high probability, the random graph process becomes connected at exactly the moment when the last isolated vertex disappears.*

Since then, many higher-dimensional analogues of both random graphs and connectedness have been analysed and in particular two different approaches have received considerable attention. A first natural generalisation for dimension $k \geq 1$ is the binomial random $(k+1)$-uniform hypergraph $G_p = G(k; n, p)$ in which each $(k+1)$-tuple of vertices forms a hyperedge with probability p independently. There are several natural ways of defining connectedness of G_p, which have been extensively studied, including vertex-connectedness [7, 8, 13, 14, 45, 57, 59] and high-order connectedness (also known as j-tuple-connectedness) [19, 20, 22, 44]. Another topic which has received particular attention is generalisations of the ℓ-core of a random graph (i.e. the maximum subgraph with minimum degree at least ℓ) [16, 23, 27, 54], which itself may be viewed as a generalisation of the giant component of a random graph [9, 29, 38, 50, 52].

A more recent approach concerns random simplicial complexes, of which a first model for the 2-dimensional case was introduced by Linial and Meshulam [48]. They considered the concept of $\mathbb{F}_2$-*homological* 1-*connectivity* of the random 2-complex as the vanishing of its first homology group with coefficients in the two-element field $\mathbb{F}_2$, which is equivalent to the vanishing of the first cohomology group. More precisely, the model $\mathcal{Y}_p = \mathcal{Y}(k; n, p)$ considered by Linial and Meshulam [48] for $k = 2$ and subsequently by Meshulam and Wallach [53] for general $k \geq 2$ is defined as follows. Starting from the full $(k-1)$-*dimensional skeleton* on $[n] := \{1, \ldots, n\}$, that is, all simplices from dimension zero up to $k-1$, each $(k+1)$-set forms a k-simplex with probability p independently. They showed that the property of the vanishing of the $(k-1)$-th cohomology group $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)$ with coefficients in $\mathbb{F}_2$ has a sharp threshold at $p = \frac{k \log n}{n}$.

Theorem 3.1.4 ([48, 53]). *Let ω be any function of n which tends to infinity as $n \to \infty$. Then with high probability,*

(i) if $p = \frac{k \log n - \omega}{n}$, then $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) \neq 0$;

(ii) if $p = \frac{k \log n + \omega}{n}$, then $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) = 0$.

Meshulam and Wallach [53] further proved that the same statement remains true if the coefficients of the cohomology group are taken from any finite abelian group.

Later, Kahle and Pittel [44] derived a hitting time result for $\mathcal{Y}_p$ (analogous to Theorem 3.1.3) in the case $k = 2$. Moreover, they determined the limiting distribution of $\dim\left(H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2)\right)$ for general $k \geq 2$ and for p inside the critical window.

Theorem 3.1.5 ([44, Theorem 1.10]). *Let $k \geq 2$ and $c \in \mathbb{R}$ be a constant. If*

$$p = \frac{k \log n + c}{n},$$

then $\dim\left(H^{k-1}(\mathcal{Y}_p;\mathbb{F}_2)\right)$ *converges in distribution to a Poisson random variable with expectation* $e^{-c}/k!$. *In particular, we have*

$$\Pr\left(H^{k-1}(\mathcal{Y}_p;\mathbb{F}_2)=0\right)\xrightarrow{n\to\infty} e^{-e^{-c}/k!}.$$

Observe that Theorem 3.1.5 can be generalised to hold for $p=(k\log n+c_n)/n$, where $(c_n)_{n\geq 1}$ is a sequence of real numbers that converges to c as $n\to\infty$ (cf. Theorem 3.1.2), because $\dim\left(H^{k-1}(\mathcal{Y}_p;\mathbb{F}_2)\right)$ is a monotone function in p.

In this paper, we aim to bridge the gap between random hypergraphs and random simplicial complexes, considering random simplicial k-complexes that arise as the downward-closure of binomial random $(k+1)$-uniform hypergraphs (Definition 3.1.7). Unlike $\mathcal{Y}_p$, in this model the presence of the full $(k-1)$-dimensional skeleton is not guaranteed, thus the vanishing of the cohomology groups of dimensions lower than $k-1$ does not hold trivially. Therefore, for each $1\leq j\leq k-1$, we introduce $\mathbb{F}_2$-*cohomological* j*-connectedness* of a k-dimensional simplicial complex (Definition 3.1.8) as the vanishing of *all* cohomology groups with coefficients in $\mathbb{F}_2$ from dimension one up to j and the zero-th cohomology group being isomorphic to $\mathbb{F}_2$.

Although this notion of connectedness is not deterministically monotone for our model, we prove that $\mathbb{F}_2$-cohomological j-connectedness has a sharp threshold. Furthermore, we derive a hitting time result and determine the limiting probability for $\mathbb{F}_2$-cohomological j-connectedness inside the critical window. As a corollary, we deduce a hitting time result for $\mathcal{Y}_p$ in *general dimension*, thus extending the hitting time result of Kahle and Pittel [44].

3.1.2 Model

Throughout the paper let $k\geq 2$ be a fixed integer. For positive integers ℓ and $1\leq i\leq\ell$, write $[\ell]:=\{1,\dots,\ell\}$ and denote by $\binom{[\ell]}{i}$ the family of i-element subsets of $[\ell]$.

Definition 3.1.6. A family $\mathcal{G}$ of non-empty finite subsets of a vertex set V is called a *simplicial complex* 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 $\mathcal{G}$ of cardinality $k+1$ are called k*-simplices* of $\mathcal{G}$. If $\mathcal{G}$ has no $(k+1)$-simplices, then we call it k*-dimensional*, or k*-complex*.[1] If $\mathcal{G}$ is a k-complex, then for each $j=0,\dots,k-1$ the j*-skeleton* of $\mathcal{G}$ is the j-complex formed by all i-simplices in $\mathcal{G}$ with $0\leq i\leq j$.

We define a model of random k-complexes starting from the binomial random $(k+1)$-uniform hypergraph G_p on vertex set $[n]$: the 0-simplices are the vertices of G_p, the k-simplices are the hyperedges of G_p, but there is more than one way to guarantee the downward-closure property to obtain a simplicial complex. In the model $\mathcal{Y}_p$ considered

[1] Note that we do not require $\mathcal{G}$ to contain any k-simplices in order to be k-dimensional. This is in contrast to the usual terminology, but we adopt this convention for technical convenience – in fact, in all interesting cases of the model of random k-complex which we consider, with high probability this distinction makes no difference.

by Meshulam and Wallach in [53], the full $(k-1)$-skeleton on $[n]$ is always included. In contrast, we only include those simplices that are *necessary* to ensure the downward-closure property.

Definition 3.1.7. We denote by $\mathcal{G}_p = \mathcal{G}(k; n, p)$ the random k-dimensional simplicial complex on vertex set $[n]$ such that:

- the 0-simplices are the singletons of $[n]$;
- the k-simplices are the hyperedges of the binomial random $(k+1)$-uniform hypergraph G_p;
- for each $j \in [k-1]$, the j-simplices are exactly the $(j+1)$-subsets of hyperedges of G_p.

In other words, $\mathcal{G}_p$ is the random k-complex on $[n]$ obtained from G_p by taking the downward-closure of each hyperedge. For instance, denote by F_p the set of hyperedges of the binomial random 4-uniform hypergraph $G_p = G(3; n, p)$. Then the corresponding two models of random 3-dimensional simplicial complexes are given by

$$\mathcal{Y}_p = \mathcal{Y}(3; n, p) = \binom{[n]}{1} \cup \binom{[n]}{2} \cup \binom{[n]}{3} \cup F_p \qquad \text{and}$$
$$\mathcal{G}_p = \mathcal{G}(3; n, p) = \binom{[n]}{1} \cup \partial(\partial F_p) \cup \partial F_p \cup F_p,$$

where ∂E for a set E of j-simplices, $j \geq 1$, denotes the set of all $(j-1)$-simplices that are contained in elements of E.

Given a simplicial complex $\mathcal{G}$, let $H^i(\mathcal{G}; \mathbb{F}_2)$ be its i-th cohomology group with coefficients in $\mathbb{F}_2$ (see (3.4) in Section 3.2.3 for the definition). We define a notion of connectedness for a simplicial complex via the *vanishing* of its cohomology groups. Since the 0-th cohomology group $H^0(\mathcal{G}; \mathbb{F}_2)$ cannot vanish, we require this group to be "as small as possible".

Definition 3.1.8. Given a positive integer j, a simplicial complex $\mathcal{G}$ is called $\mathbb{F}_2$-*cohomologically j-connected* (*j-cohom-connected* for short) if

- $H^0(\mathcal{G}; \mathbb{F}_2) = \mathbb{F}_2$;
- $H^i(\mathcal{G}; \mathbb{F}_2) = 0$ for all $i \in [j]$.

Observe that $H^0(\mathcal{G}; \mathbb{F}_2)$ being isomorphic to $\mathbb{F}_2$ is equivalent to connectedness of $\mathcal{G}$ in the topological sense, which we call *topological connectedness* in order to distinguish it from other notions of connectedness. For $\mathcal{G} = \mathcal{G}_p$, this is also equivalent to *vertex-connectedness* of the associated $(k+1)$-uniform hypergraph. The threshold for this property is well-known, and will be discussed in Section 3.4.1.

Moreover, one might define an analogous version of connectedness via the vanishing of *homology* groups, which would be equivalent to our definition of $\mathbb{F}_2$-cohomological j-connectedness by the Universal Coefficient Theorem (see e.g. [55]).

A significant difference between $\mathcal{G}_p$ and $\mathcal{Y}_p$ is that for $\mathcal{Y}_p$ the only requirement for $\mathbb{F}_2$-cohomological $(k-1)$-connectedness is the vanishing of the $(k-1)$-th cohomology group, since the presence of the full $(k-1)$-skeleton guarantees topological connectedness and the vanishing of the j-th cohomology groups for all $j \in [k-2]$.

Moreover, it is important to observe that $\mathbb{F}_2$-cohomological j-connectedness is *not* necessarily a monotone increasing property of $\mathcal{G}_p$: adding a k-simplex to a j-cohom-connected complex might yield a complex without this property (see Example 3.3.2). Thus, the existence of a single threshold for j-cohom-connectedness is *not* guaranteed, but one of our main results shows that such a threshold indeed exists (Theorem 3.1.11).

3.1.3 Main results

The main contributions of this paper are fourfold. Firstly, we prove (Theorem 3.1.11) that for each $j \in [k-1]$, the probability

$$p_j := \frac{(j+1)\log n + \log\log n}{(k-j+1)n^{k-j}}(k-j)! \tag{3.1}$$

is a sharp threshold[2] for $\mathbb{F}_2$-cohomological j-connectedness. Secondly, we prove a hitting time result (also Theorem 3.1.11), relating the j-cohom-connectedness threshold to the disappearance of all copies of the *minimal obstruction* M_j (Definition 3.1.10). Thirdly, our results directly imply an analogous hitting time result for $\mathcal{Y}_p$ (Corollary 3.1.12), which Kahle and Pittel [44] proved for $k = 2$. Lastly, we analyse the critical window given by the threshold p_j, showing that inside the window the dimension of the j-th cohomology group converges in distribution to a Poisson random variable (Theorem 3.1.13).

Proving that p_j is indeed a (sharp) threshold turns out to be considerably more challenging than might be expected, largely because $\mathbb{F}_2$-cohomological j-connectedness of $\mathcal{G}_p$ is *not* a monotone increasing property. In particular, the subcritical case is much more involved than it would be for a monotone property, where often a simple second moment argument suffices. In order to circumvent the difficulties arising from the non-monotonicity, we introduce auxiliary structures called *local obstacles* (Definition 3.4.8), showing that whp $\mathcal{G}_p$ evolves in a monotone way regarding those (Lemma 3.4.9). In the supercritical case we must guarantee that whp there are no more obstructions to j-cohom-connectedness. In order to bound the number of potential "large" obstructions, basic calculations are not sufficient and therefore we define a suitable search process, which gives us more precise bounds on their number (Lemma 3.5.7).

Before defining the minimal obstruction M_j (Definition 3.1.10), we introduce the following necessary concepts.

Definition 3.1.9. Given a k-simplex K in a k-dimensional simplicial complex $\mathcal{G}$, a collection $\mathcal{F} = \{P_0, \ldots, P_{k-j}\}$ of j-simplices forms a *j-flower in K* (see Figure 3.1) if $K = \bigcup_{i=0}^{k-j} P_i$ and $C := \bigcap_{i=0}^{k-j} P_i$ satisfies $|C| = j$. We call the j-simplices P_i the *petals* and the set C the *centre* of the j-flower $\mathcal{F}$.

[2] In a strong sense, see Theorem 3.1.11 for details.

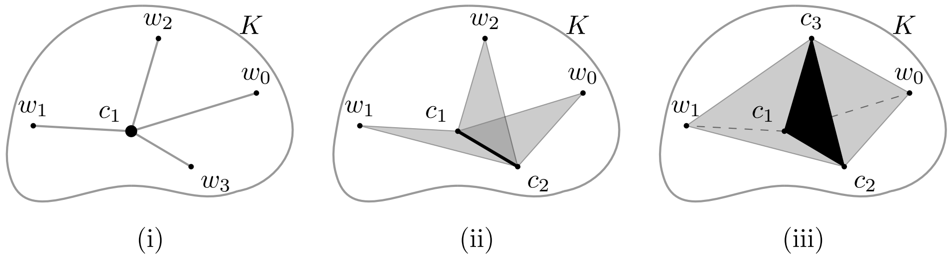


Figure 3.1: Examples of j-flowers in a k-simplex K, for $k = 4$ and $j = 1, 2, 3$.
(i) The 1-flower in K with centre $C = \{c_1\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1, 2, 3$ (grey).
(ii) The 2-flower in K with centre $C = \{c_1, c_2\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1, 2$ (grey).
(iii) The 3-flower in K with centre $C = \{c_1, c_2, c_3\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1$ (grey).

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

$$\mathcal{F}(K, C) := \{C \cup \{w\} \mid w \in K \setminus C\}. \tag{3.2}$$

When j is clear from the context, we simply refer to a j-flower as a *flower*.

A *j-cycle* in a simplicial complex $\mathcal{G}$ is a set J of j-simplices such that every $(j-1)$-simplex of $\mathcal{G}$ is contained in an *even* number of j-simplices in J.

Definition 3.1.10. A *copy of M_j* (see Figure 3.2) in a k-complex $\mathcal{G}$ is a triple (K, C, J) where

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

(M2) C is a $(j-1)$-simplex in K such that each petal of the flower $\mathcal{F} = \mathcal{F}(K, C)$ is contained in *no other* k-simplex of $\mathcal{G}$;

(M3) J is a j-cycle in $\mathcal{G}$ that contains *exactly one petal* of the flower $\mathcal{F}$, i.e. there exists a vertex $w_0 \in K \setminus C$ such that

$$J \cap \mathcal{F} = \Big\{C \cup \{w_0\}\Big\}.$$

We will see in Section 3.3.1 that a copy of M_j can be interpreted as a minimal obstruction for $\mathbb{F}_2$-cohomological j-connectedness.

The random k-complex $\mathcal{G}_p$ can be viewed as a *process*, by assigning a *birth time* to each k-simplex. More precisely, for each $(k+1)$-set of vertices in $[n]$ independently, sample a birth time uniformly at random from $[0, 1]$. (With probability 1 no two $(k+1)$-sets have the same birth time.) Then $\mathcal{G}_p$ is exactly the complex generated by the $(k+1)$-sets with birth times at most p, by taking the downward-closure. If p is

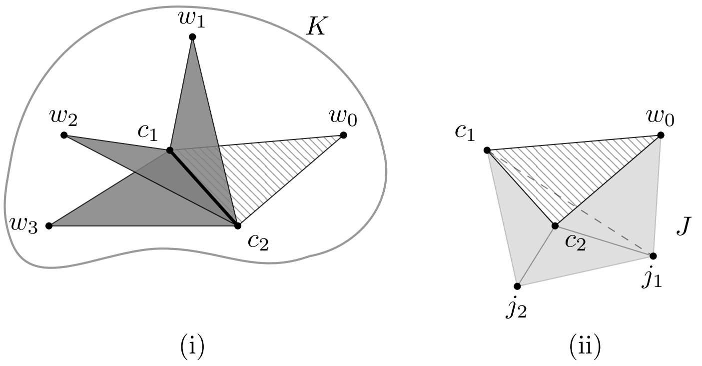


Figure 3.2: A copy of M_j, for $k = 5$ and $j = 2$. The striped j-simplices are identified. (i) The k-simplex K that contains the flower $\mathcal{F}(K, C)$ with centre $C = \{c_1, c_2\}$ and petals $P_i = C \cup \{w_i\}$, for $i = 0, 1, 2, 3$. Each petal P_i is contained in no other k-simplex except K.
(ii) The j-cycle J that consists of the j-simplices $P_0 = \{c_1, c_2, w_0\}$, $\{c_2, w_0, j_1\}$, $\{c_2, j_1, j_2\}$, $\{c_1, c_2, j_2\}$, $\{c_1, j_1, j_2\}$ and $\{c_1, w_0, j_1\}$. It intersects the flower $\mathcal{F}(K, C)$ only in the petal P_0.

gradually increased from 0 to 1, we may interpret $\mathcal{G}_p$ as a process. Thus, we can define p_{M_j} as the birth time of the k-simplex whose appearance causes the last copy of M_j to disappear. More formally, let

$$p_{M_j} := \sup\{p \in [0, 1] \mid \mathcal{G}_p \text{ contains a copy of } M_j\}. \tag{3.3}$$

Our first main result states that the value p_{M_j} is the hitting time for j-cohom-connectedness of $\mathcal{G}_p$ and is "close" to p_j defined in (3.1), implying that p_j is in fact a sharp threshold for $\mathbb{F}_2$-cohomological j-connectedness.

Theorem 3.1.11. *Let $k \geq 2$ be an integer and let ω be any function of n which tends to infinity as $n \to \infty$. For each $j \in [k-1]$, with high probability the following statements hold.*

(i) p_{M_j} satisfies

$$\frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)! < p_{M_j} < \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!.$$

(ii) For all $p < p_{M_j}$, $\mathcal{G}_p$ is not $\mathbb{F}_2$-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_p; \mathbb{F}_2) \neq \mathbb{F}_2 \quad \text{or} \quad H^i(\mathcal{G}_p; \mathbb{F}_2) \neq 0 \text{ for some } i \in [j].$$

(iii) For all $p \geq p_{M_j}$, $\mathcal{G}_p$ is $\mathbb{F}_2$-cohomologically j-connected, i.e.

$$H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2 \quad \text{and} \quad H^i(\mathcal{G}_p; \mathbb{F}_2) = 0 \text{ for all } i \in [j].$$

For the case $j = k-1$, Theorem 3.1.11 gives a threshold $p_{k-1} = \frac{k \log n + \log\log n}{2n}$ for $\mathbb{F}_2$-cohomological $(k-1)$-connectedness, which is about half as large as the threshold $\frac{k \log n}{n}$ in Theorem 3.1.4 for $\mathcal{Y}_p$. The reason for this is that the minimal obstructions are different: in $\mathcal{Y}_p$ the minimal obstruction is a $(k-1)$-simplex which is not contained in any k-simplex of the complex (such a $(k-1)$-simplex is called *isolated*). By definition, isolated $(k-1)$-simplices do not exist in $\mathcal{G}_p$, because $\mathcal{G}_p$ contains only those $(k-1)$-simplices that lie in some k-simplex.

Observe that Theorem 3.1.11 (ii) and (iii) provide a hitting time result for the process described above. A similar result was proved by Kahle and Pittel [44] for $\mathcal{Y}_p$, but only for the two-dimensional case. They considered the random complex process associated with $\mathcal{Y}_p$ and related the vanishing of the first cohomology group to the disappearance of the last isolated edge (i.e. 1-simplex). As a corollary of Theorem 3.1.11, we obtain a hitting time result for $\mathcal{Y}_p$ for general $k \geq 2$. To this end, let

$$p_{\mathrm{isol}} := \sup\{p \in [0,1] \mid \mathcal{Y}_p \text{ contains isolated } (k-1)\text{-simplices}\}$$

be the birth time of the k-simplex whose appearance causes the last isolated $(k-1)$-simplex in $\mathcal{Y}_p$ to disappear and let

$$p_{\mathrm{conn}} := \sup\{p \in [0,1] \mid H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) \neq 0\}$$

be the time when $\mathcal{Y}_p$ becomes $\mathbb{F}_2$-cohomologically $(k-1)$-connected.

Corollary 3.1.12. *Let $k \geq 2$ be an integer. Then, with high probability*

$$p_{conn} = p_{isol}.$$

In other words, with high probability the random process associated with $\mathcal{Y}_p$ becomes $\mathbb{F}_2$-cohomologically $(k-1)$-connected at exactly the moment when the last isolated $(k-1)$-simplex disappears.

Our last main result gives an explicit expression for the limiting probability of the random complex $\mathcal{G}_p$ being $\mathbb{F}_2$-cohomologically j-connected inside the critical window given by the threshold p_j (cf. Theorems 3.1.2 and 3.1.5). More generally, we prove that the *dimension* of the j-th cohomology group with coefficients in $\mathbb{F}_2$ converges in distribution to a Poisson random variable.

Theorem 3.1.13. *Let $k \geq 2$ be an integer, $j \in [k-1]$ and $c \in \mathbb{R}$ be a constant. Suppose that $(c_n)_{n\geq 1}$ is a sequence of real numbers that converges to c as $n \to \infty$. If*

$$p = \frac{(j+1)\log n + \log\log n + c_n}{(k-j+1)n^{k-j}}(k-j)!,$$

then $\dim\left(H^j(\mathcal{G}_p; \mathbb{F}_2)\right)$ converges in distribution to a Poisson random variable with expectation

$$\lambda_j := \frac{(j+1)e^{-c}}{(k-j+1)^2 j!},$$

while whp $H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2$ and $H^i(\mathcal{G}_p; \mathbb{F}_2) = 0$ for all $i \in [j-1]$. In particular,

$$\Pr\left(\mathcal{G}_p \text{ is } j\text{-cohom-connected}\right) \xrightarrow{n\to\infty} e^{-\lambda_j}.$$

Indeed, in the proof we will see that whp $\dim\left(H^j(\mathcal{G}_p;\mathbb{F}_2)\right)$ equals the number of pairs (K,C) for which there exists a j-cycle J such that (K,C,J) is a copy of M_j in $\mathcal{G}_p$.

Let us note that Hoffman, Kahle and Paquette [35] proved a similar result for the $\mathcal{Y}_p$ model relating the dimension of the cohomology group to the number of isolated $(k-1)$-simplices, and as a corollary obtained a hitting time result for general dimension analogous to Corollary 3.1.12, but for cohomology groups over $\mathbb{Q}$ rather than $\mathbb{F}_2$.

3.1.4 Related work

This paper draws inspiration from [48] and [53], but the proof techniques are considerably different. We first note that in $\mathcal{Y}_p$ the presence of the full $(k-1)$-dimensional skeleton trivially yields the topological connectedness of $\mathcal{Y}_p$ and the vanishing of all the i-th cohomology groups with $i \in [k-2]$. This is not true in $\mathcal{G}_p$ and therefore we need to consider *all* cohomology groups up to dimension j, for each $j \in [k-1]$.

Moreover, in [48] and [53] one standard application of the second moment method is sufficient for the analysis of the subcritical case (i.e. statement (i)) of Theorem 3.1.4. By contrast, $\mathbb{F}_2$-cohomological j-connectedness of $\mathcal{G}_p$ is *not* a monotone increasing property (see Example 3.3.2). This makes the subcritical case far from trivial. More precisely, it does not suffice to prove that $\mathcal{G}_p$ is not j-cohom-connected at $p_- = \frac{(j+1)\log n+\log\log n-\omega}{(k-j+1)n^{k-j}}(k-j)!$; rather we need to show that whp the property is not satisfied for *any* p up to and including p_-. Also observe that in terms of our hitting time result, it is *not* enough to show that for each "small" p whp $\mathcal{G}_p$ is not j-cohom-connected. Rather, we need to know that $\mathcal{G}_p$ is not j-cohom-connected whp *for all such p simultaneously.*

The proof of the supercritical case $p \geq p_{M_j}$ is also more challenging than for $\mathcal{Y}_p$; we are forced to derive stronger bounds for the number of *bad functions* (see Definition 3.2.4), due to the fact that for $j = k-1$, the threshold in Theorem 3.1.11 is about half as large as the corresponding threshold in [53]. To this end, we define a breadth-first search process that makes use of the new notion of *traversability* (Definition 3.5.3). Moreover, non-monotonicity of j-cohom-connectedness forces us to prove that for all $p \geq p_{M_j}$, the probability of $\mathcal{G}_p$ *not* being j-cohom-connected is small enough that we can apply a union bound over all relevant values of p.

3.1.5 Paper overview

This paper is structured as follows.

In Section 3.2 we present some preliminary results that we will use throughout the paper and we provide an overview of cohomology theory, which will allow us to define the concept of a *bad function* (see Definition 3.2.4), a configuration in a complex $\mathcal{G}$ that is a witness for $H^j(\mathcal{G};\mathbb{F}_2)$ not vanishing. Section 3.3 is devoted to the main concepts and the proof ideas used in this paper. After explaining why a copy of M_j is a minimal obstruction to j-cohom-connectedness, we heuristically show why the value p_j defined in (3.1) should be the threshold for j-cohom-connectedness and give an outline of the proofs of our main theorems.

In Section 3.4, we provide the auxiliary results needed for the proofs of Theorem 3.1.11 (i) and (ii). We analyse the subcritical case when $p < p_{M_j}$ and determine the approximate value of p_{M_j}, i.e. when the last minimal obstruction disappears. In Section 3.5 we define a breadth-first search process which will allow us to examine the supercritical case when $p \geq p_{M_j}$ and to obtain results necessary for the proofs of Theorem 3.1.11 (iii) and Theorem 3.1.13.

We prove the main results Theorems 3.1.11 and 3.1.13 and Corollary 3.1.12 in Section 3.6, using the auxiliary results from Sections 3.4 and 3.5. Finally, in Section 3.7, we discuss some open problems.

3.2 Preliminaries

3.2.1 Birth times

We mentioned in Section 3.1.3 how to use the standard birth times interpretation to describe the binomial model $\mathcal{G}_p$ as a process. In this setting, it is useful to introduce the operation of "adding a simplex".

Definition 3.2.1. 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 $\mathcal{G} + B = \mathcal{G}$. With this operation, $\mathcal{G}_p$ (interpreted as a process) may also be described in the following way. If p_K is the smallest birth time larger than p of any k-simplex K, then $\mathcal{G}_{p_K} = \mathcal{G}_p + K$.

A property $\mathcal{P}$ of k-complexes is called *monotone increasing* if $\mathcal{P}$ is closed under adding k-simplices. The complement of a monotone increasing property is called *monotone decreasing*. Finally, $\mathcal{P}$ is *monotone* if it is monotone increasing or decreasing.

Considering the birth times interpretation, we shall take union bounds over finite sets of birth times. With a slight abuse of terminology, sometimes we will talk about taking "union bounds over p" in some interval, which makes little sense if we think of p as being able to take any value within the interval, but indeed we are conditioning on the set of birth times and taking the union bound over all birth times in the relevant interval.

We also note that conditioned on a k-simplex not being present at time $p = q_1$, the probability that it is present at time q_2 is $\frac{q_2-q_1}{1-q_1}$. Thus we may obtain $\mathcal{G}_{q_2}$ from $\mathcal{G}_{q_1}$ by *exposing* an additional probability of $\frac{q_2-q_1}{1-q_1}$. Since we will only ever want to consider such a situation with $q_1 = o(1)$, we often simply take $q_2 - q_1$ as an approximation (and lower bound) for $\frac{q_2-q_1}{1-q_1}$, or use q_2 as an upper bound.

3.2.2 Probabilistic tools

We frequently use the following Chernoff bound.

Lemma 3.2.2 (see e.g. [39, Theorem 2.1]). *Given a binomial random variable X with expectation μ and a real number $a > 0$,*

$$\Pr(X \geq \mu + a) \leq \exp\left(-\frac{a^2}{2(\mu + a/3)}\right);$$
$$\Pr(X \leq \mu - a) \leq \exp\left(-\frac{a^2}{2\mu}\right).$$

For the analysis of the critical window (cf. Theorem 3.1.13), we will need the method of moments, as presented in the following lemma.

Lemma 3.2.3 (see e.g. [34, Theorem 20.11]). *Let $(S_n)_{n\geq 1}$ be a sequence of sums of indicator random variables. Suppose that there exists $\lambda > 0$ such that for every fixed integer $t \geq 1$*

$$\lim_{n\to\infty} \mathbb{E}\binom{S_n}{t} = \frac{\lambda^t}{t!}.$$

Then, for every integer $s \geq 0$,

$$\lim_{n\to\infty} \Pr(S_n = s) = e^{-\lambda}\frac{\lambda^s}{s!},$$

i.e. S_n converges in distribution to a Poisson random variable with expectation λ. We write $S_n \xrightarrow{d} Po(\lambda)$.

3.2.3 Cohomology terminology

We formally introduce cohomology groups with coefficients in $\mathbb{F}_2$ for a simplicial complex. The following notions are all standard, except the definition of a bad function (Definition 3.2.4).

Given a k-complex $\mathcal{G}$, for each $j \in \{0, \ldots, k\}$ denote by $C^j(\mathcal{G})$ the set of *j-cochains*, that is, the set of 0-1 functions on the j-simplices. The *support* of a function in $C^j(\mathcal{G})$ is the set of j-simplices mapped to 1. Each $C^j(\mathcal{G})$ forms a group with respect to pointwise addition modulo 2. We define the *coboundary operators* $\delta^j \colon C^j(\mathcal{G}) \to C^{j+1}(\mathcal{G})$ for $j = 0, \ldots, k-1$ as follows: for $f \in C^j(\mathcal{G})$, the $(j+1)$-cochain $\delta^j f$ assigns to each $(j+1)$-simplex σ the value

$$\delta^j f(\sigma) := \sum_{\tau\subset\sigma,\ |\tau|=j+1} f(\tau) \pmod 2.$$

In addition, we denote by δ^{-1} the unique group homomorphism $\delta^{-1} \colon \{0\} \to C^0(\mathcal{G})$. The j-cochains in $\operatorname{im}\delta^{j-1}$ are called *j-coboundaries*, and the j-cochains in $\ker\delta^j$ are called *j-cocycles*. A straightforward calculation shows that each coboundary operator is a group homomorphism and that every j-coboundary is also a j-cocycle, i.e. $\operatorname{im}\delta^{j-1}$ is a subgroup of $\ker\delta^j$. Therefore, we can define the j-th cohomology group of $\mathcal{G}$ with coefficients in $\mathbb{F}_2$ as the quotient group

$$H^j(\mathcal{G}; \mathbb{F}_2) := \ker\delta^j / \operatorname{im}\delta^{j-1}. \tag{3.4}$$

By definition, $H^j(\mathcal{G};\mathbb{F}_2)$ vanishes if and only if every j-cocycle is a j-coboundary. This motivates the following definition of a *bad function*.

Definition 3.2.4. For a k-complex $\mathcal{G}$ and $j \in [k-1]$, we say that a function $f \in C^j(\mathcal{G})$ is *bad* if

(i) f is a j-cocycle, i.e. it assigns an even number of 1's to the j-simplices on the boundary of each $(j+1)$-simplex;

(ii) f is not a j-coboundary, i.e. it is not induced by a $(j-1)$-cochain.

Thus, $H^j(\mathcal{G};\mathbb{F}_2)$ vanishes if and only if no bad function in $C^j(\mathcal{G})$ exists.

Recall that a set of j-simplices is a j-cycle if every $(j-1)$-simplex is contained in an *even* number of j-simplices of the set. It is easy to see that if f is a j-cocycle and J is a j-cycle such that the restriction $f|_J$ has support of odd size, then f is not a j-coboundary and thus is a bad function.

3.3 Intuition and outline of proofs

For the rest of the paper, let $j \in [k-1]$ be fixed.

3.3.1 Minimal obstructions

Let us explain why M_j (Definition 3.1.10) can be interpreted as the (unique) minimal obstruction to j-cohom-connectedness. Given a triple (K, C, J) which forms a copy of M_j in a k-complex $\mathcal{G}$, it is easy to define a bad function $f \in C^j(\mathcal{G})$ (see Definition 3.2.4): let f take value 1 on the petals of the flower $\mathcal{F}(K, C)$ (see (3.2)) and 0 everywhere else. Since the petals are all in the k-simplex K but in no further k-simplices, every $(j+1)$-simplex L in $\mathcal{G}$ is even, because L contains either two petals (if $C \subseteq L \subseteq K$) or none (otherwise). However, J would be a j-cycle containing precisely one j-simplex, namely the petal $C \cup \{w_0\}$, on which f takes value 1, ensuring that f is not a j-coboundary. Thus f is bad and has support of size $k - j + 1$, which is the number of petals of $\mathcal{F}(K, C)$.

In the following lemma we show that in fact such a bad function is the only possibility for an obstruction which is minimal with respect to the size of the support. Given a k-simplex K and a collection S of j-simplices, define S_K to be the set of j-simplices of S contained in K.

Lemma 3.3.1. *Let $\mathcal{G}$ be a k-complex and let S be the support of a j-cocycle. Then for each k-simplex K,*

(i) either $S_K = \emptyset$ or both $|S_K| \geq k - j + 1$ and $\bigcup_{\sigma \in S_K} \sigma = K$;

(ii) if $|S_K| = k - j + 1$, then S_K forms a j-flower in K.

Proof. (i) Suppose $S_K \neq \emptyset$ and let $\sigma_0 \in S_K$. Let the vertices of $K \setminus \sigma_0$ be denoted by $v_1, \ldots, v_{k-j}$. Each $(j+1)$-simplex $\sigma_0 \cup \{v_i\}$ has to be even with respect to f and thus contains some j-simplex $\sigma_i \in S_K \setminus \{\sigma_0\}$, 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 $|S_K| \geq k - j + 1$ and

$$K \supseteq \bigcup_{\sigma \in S_K} \sigma \supseteq \sigma_0 \cup \{v_1, \dots, v_{k-j}\} = K.$$

(ii) Suppose now that $S_K = \{\sigma_0, \dots, \sigma_{k-j}\}$, with $\sigma_0, \dots, \sigma_{k-j}$ defined as above. For $2 \leq i \leq k-j$, the $(j+1)$-simplex $\tau := \sigma_1 \cup \{v_i\}$ contains σ_1, but no σ_ℓ with $\ell \notin \{1, i\}$. By the choice of S as the support of a j-cocycle, τ is even and thus $\sigma_i \subset \tau$. This means that

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

As this holds for all i, S_K forms a flower in K with centre $\sigma_0 \cap \sigma_1$. □

Both the presence of a copy of M_j and j-cohom-connectedness in $\mathcal{G}_p$ are *not* monotone properties, as the following example shows.

Example 3.3.2. Let $\mathcal{G}$ be the 2-complex on vertex set $\{1, 2, 3, 4, 5\}$ generated by the 3-uniform hypergraph with hyperedges $\{1, 2, 3\}$ and $\{1, 4, 5\}$, see Figure 3.3. Then $\mathcal{G}$ is 1-cohom-connected and thus contains no copies of M_1. Adding to $\mathcal{G}$ the 2-simplex $\{2, 3, 4\}$ (and its downward-closure) creates several copies of M_1 and thus yields a complex $\mathcal{G}'$ which is not 1-cohom-connected. If we further add the 2-simplex $\{1, 3, 4\}$ to $\mathcal{G}'$, we obtain a 2-complex $\mathcal{G}''$ which is 1-cohom-connected and thus contains no copies of M_1.

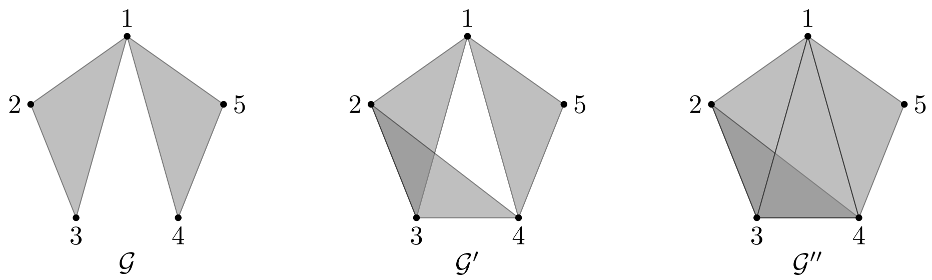


Figure 3.3: Adding simplices might create new copies of M_j or destroy existing ones.

3.3.2 Finding the threshold

In this section we provide a heuristic argument for why the threshold for the disappearance of the last copy of M_j should be around p_j. To do this, we will make use of a simplified version of the obstruction M_j.

Definition 3.3.3. A *copy of* M_j^- (see Figure 3.4) in a k-complex $\mathcal{G}$ is a pair (K, C) where

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

(M2) C is a $(j-1)$-simplex in K such that each petal of the flower $\mathcal{F}(K, C)$ is contained in *no other* k-simplex of $\mathcal{G}$.

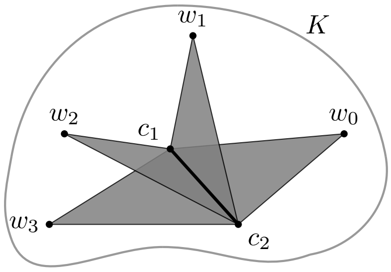


Figure 3.4: A copy of M_j^-, for $k = 5$ and $j = 2$. The k-simplex K contains the flower $\mathcal{F}(K, C)$ with centre $C = \{c_1, c_2\}$ and petals $P_i = C \cup \{w_i\}$, for $i = 0, 1, 2, 3$. Each petal P_i is contained in no other k-simplex except K.

In other words, a copy of M_j^- can be viewed as a copy of M_j without the condition (M3), i.e. without the j-cycle J containing one of the petals (see Figures 3.2 and 3.4). Therefore,

$$M_j^- \not\subset \mathcal{G}_p \quad \Rightarrow \quad M_j \not\subset \mathcal{G}_p.$$

Moreover, we will show (Lemma 3.4.6) that, for p approaching the value p_j, the j-cycle J needed to extend a copy of M_j^- to a copy of M_j is very likely to exist. Hence in this range the existence of M_j^- and M_j are essentially equivalent events.

Let us estimate the expected number of copies of M_j^- in $\mathcal{G}_p$. The probability of $k+1$ arbitrary vertices with a fixed centre C forming a copy of M_j^- is about $p(1-p)^{(k-j+1)\binom{n}{k-j}}$, which we can approximate by

$$pe^{-\frac{(k-j+1)n^{k-j}}{(k-j)!}p},$$

so the expected number of copies of M_j^- is of order $n^{k+1}pe^{-\frac{(k-j+1)n^{k-j}}{(k-j)!}p}$. We seek p such that

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

This holds when

$$(k+1)\log n + \log p - \frac{(k-j+1)n^{k-j}}{(k-j)!}p = 0,$$

which implies

$$\begin{aligned}
p &= \frac{(k+1)\log n + \log p}{(k-j+1)n^{k-j}}(k-j)! \\
&= \frac{(k+1)\log n + \log\left(\frac{(k+1)\log n+\log p}{k-j+1}(k-j)!\right) - (k-j)\log n}{(k-j+1)n^{k-j}}(k-j)! \\
&= \frac{(j+1)\log n + \log\log n + O(1)}{(k-j+1)n^{k-j}}(k-j)!,
\end{aligned}$$

which corresponds to the stated threshold p_j defined in (3.1).

3.3.3 Outline of the proofs

We now give an outline of the proofs of our main theorems. Let us begin with Theorem 3.1.11. To analyse the zero-th cohomology group, we define the probabilities

- $p_0 := \frac{\log n}{n^k}k!$;
- $p_T := \sup\{p \in [0,1] \mid \mathcal{G}_p \text{ is not topologically connected}\}$.

In other words, p_T is the birth time of the k-simplex whose appearance causes the complex $\mathcal{G}_p$ to become topologically connected. Recall that topological connectedness is equivalent to the random hypergraph G_p becoming vertex-connected. It is known (see e.g. [19, 56, 57]) that p_0 is the threshold for vertex-connectedness of the random $(k+1)$-uniform hypergraph, that is whp $p_T = (1+o(1))p_0$ (Lemma 3.4.1).

Recall from (3.1) and (3.3) that for each $j \in [k-1]$ we have

- $p_j = \dfrac{(j+1)\log n + \log\log n}{(k-j+1)n^{k-j}}(k-j)!$;
- $p_{M_j} = \sup\{p \in [0,1] \mid \mathcal{G}_p \text{ contains a copy of } M_j\}$.

In other words, p_{M_j} is the birth time of the k-simplex whose appearance causes the last copy of M_j to disappear.

In Section 3.4, we study the subcritical case when $p < p_{M_j}$, providing results needed for the proof of Theorem 3.1.11 (ii). Moreover, we show that whp the value of p_{M_j} is "close" to p_j (Corollary 3.4.11), thus proving Theorem 3.1.11 (i).

In order to prove Theorem 3.1.11 (ii), we aim to show that whp $H^j(\mathcal{G}_p; \mathbb{F}_2) \neq 0$ throughout the interval $[p_{M_{j-1}}, p_{M_j})$. A direct argument based on determining the dimensions of $C^{j-1}(\mathcal{G}_p)$, $C^j(\mathcal{G}_p)$ and $C^{j+1}(\mathcal{G}_p)$ may be considered, but it would work only for some values of j and some ranges of p (see Section 3.7.1). We actually prove a stronger result (Lemma 3.3.4), for which we define the following probabilities: for each $j \in [k-1]$, set

- $p_j^- := \left(1 - \dfrac{1}{\sqrt{\log n}}\right)\dfrac{(j+1)\log n}{(k-j+1)n^{k-j}}(k-j)!$;

- $p_j^{(1)} := \dfrac{1}{10(j+1)\binom{k+1}{j+1}n^{k-j}}$.

We will also need the value
$$p_0^- := \frac{\log n}{n^k} = \frac{p_0}{k!}.$$
The motivation behind these seemingly arbitrary definitions will become clear as the argument develops. We will prove that three copies of M_j suffice to cover the interval $[p_{j-1}^-, p_{M_j})$, which whp contains the interval $[p_{M_{j-1}}, p_{M_j})$ by Theorem 3.1.11 (i).

Lemma 3.3.4. *Let $j \in [k-1]$. With high probability, there exist three triples (K_ℓ, C_ℓ, J_ℓ), $\ell = 1, 2, 3$, such that for all $p \in [p_{j-1}^-, p_{M_j})$, (K_ℓ, C_ℓ, J_ℓ) forms a copy of M_j in $\mathcal{G}_p$ for some ℓ. In particular, whp $H^j(\mathcal{G}_p; \mathbb{F}_2) \neq 0$ for all $p \in [p_{j-1}^-, p_{M_j})$.*

This will in particular imply that whp $\mathcal{G}_p$ is not j-cohom-connected in the interval $[p_{j-1}^-, p_{M_j})$. By Lemma 3.3.4 applied with j replaced by i for each $i \in [j]$ and by the fact that $\mathcal{G}_p$ is not topologically connected in $[0, p_T)$ by definition, whp $\mathcal{G}_p$ is not j-cohom-connected in the range
$$[0, p_T) \cup \bigcup_{i=1}^{j} [p_{i-1}^-, p_{M_i}) \overset{\text{(whp)}}{=} [0, p_{M_j}).$$
This completely covers the subcritical case (Theorem 3.1.11 (ii)).

In order to prove Lemma 3.3.4, we divide the interval $[p_{j-1}^-, p_{M_j})$ into smaller subintervals
$$[p_{j-1}^-, p_{M_j}) = [p_{j-1}^-, p_j^{(1)}] \cup [p_j^{(1)}, p_j^-] \cup [p_j^-, p_{M_j})$$
and show that for each of these subintervals, whp there is one copy of M_j which exists in $\mathcal{G}_p$ throughout this interval, using the following strategy.

(I) At around p_{j-1}^-, whp there exist "many" copies of M_j (Lemma 3.4.4) and whp at least one of these survives until probability $p_j^{(1)}$ (Lemma 3.4.12).

(II) For any $p \geq p_j^{(1)}$, whp all copies of M_j^- give rise to copies of M_j, thus the existence of M_j^- and M_j are essentially equivalent events (Lemma 3.4.6). In particular, the last M_j to disappear corresponds to the last M_j^- (Corollary 3.4.10).

(III) At around p_j^-, whp there are "many" copies of M_j^- (Lemma 3.4.7) and whp one of these already existed at $p_j^{(1)}$ (Lemma 3.4.13).

(IV) The last M_j^- to disappear whp already existed at p_j^- (Lemma 3.4.14).

In Section 3.5 we study the supercritical case, i.e. the case $p \geq p_{M_j}$, and derive auxiliary results, necessary to prove Theorem 3.1.11 (iii). By the definition of p_{M_j}, we know that $\mathcal{G}_p$ contains no M_j in this range, so by Lemma 3.3.1 it remains to show that whp there are no bad functions with support of size $s > k - j + 1$. In other words, we need to prove that each j-cocycle with support of size s is also a j-coboundary.

To this end, we prove (Corollaries 3.5.8 and 3.5.10) that from slightly before the threshold p_j onwards, every j-cocycle can be written as the sum of functions arising from copies of M_j^- (see Definition 3.5.1). We first show (Lemma 3.5.4) that the support of any smallest j-cocycle *not* generated by copies of M_j^- satisfies a property which we call *traversability* (Definition 3.5.3). We then bound the probability that such a support of size s exists. For constant s, simple bounds will suffice (Lemma 3.5.5); for larger values of s, traversability will allow us to define a breadth-first search process that we use to track the construction of a traversable support and thus count the number of such supports much more accurately (Lemma 3.5.7).

Combining the results from Sections 3.4 and 3.5, we prove Theorem 3.1.11 in Section 3.6. We then apply Theorem 3.1.11 to derive Corollary 3.1.12, which provides a hitting time result for $\mathcal{Y}_p$, relating the vanishing of $H^{k-1}(\mathcal{Y}_p;\mathbb{F}_2)$ to the disappearance of the last isolated $(k-1)$-simplex.

Finally, we prove Theorem 3.1.13 in Section 3.6.3. We analyse $\mathbb{F}_2$-cohomological j-connectedness of $\mathcal{G}_p$ within the critical window given by the threshold for this property, i.e. we consider $p = \frac{(j+1)\log n+\log\log n+O(1)}{(k-j+1)n^{k-j}}(k-j)!$. In this range, whp all j-cocycles arise from copies of M_j^- (Corollary 3.5.8). Using the method of moments (Lemma 3.2.3), we will show that the number of copies of M_j^- converges in distribution to a Poisson random variable and that whp this number equals the dimension of the j-th cohomology group of $\mathcal{G}_p$. Thus, in particular we derive an explicit expression for the limiting probability of $\mathcal{G}_p$ being j-cohom-connected.

3.4 Subcritical regime

In this section we study the subcritical case $p < p_{M_j}$ and derive the necessary results for the proofs of statements (i) and (ii) of Theorem 3.1.11.

3.4.1 Topological connectedness

We begin with a result stating that

$$p_0 = \frac{\log n}{n^k}k!$$

is a sharp threshold for topological connectedness of $\mathcal{G}_p$. Recall that p_T is the birth time of the k-simplex whose appearance causes the complex $\mathcal{G}_p$ to become topologically connected.

Lemma 3.4.1. *Let ω be any function of n which tends to infinity as $n \to \infty$. Then with high probability*

$$\frac{\log n - \omega}{n^k}k! < p_T < \frac{\log n + \omega}{n^k}k!$$

and thus in particular $p_T > p_0^-$.

Observe that Lemma 3.4.1 is equivalent to p_0 being a sharp threshold for vertex-connectedness of the random $(k+1)$-uniform hypergraph, which follows for instance from [19] or [57] as a special case of each (see also [56] for a stronger result). The proof relies on standard applications of the first and second moment methods and is an easy generalisation of the graph case (see e.g. [45]).

3.4.2 Counting obstructions

In this section we provide several results concerning the number of minimal obstructions that exist in $\mathcal{G}_p$ whp. First we define a special case of M_j (Definition 3.4.3), which will be useful in the subsequent arguments.

Definition 3.4.2. For any $(j+2)$-set A in a complex $\mathcal{G}$, the collection of all $(j+1)$-subsets of A is called a *j-shell* if each of them forms a j-simplex in $\mathcal{G}$. The j-shell is called *hollow* if A does not form a $(j+1)$-simplex in $\mathcal{G}$.

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.

Definition 3.4.3. Given a k-complex $\mathcal{G}$ on vertex set $[n]$, a $(k+1)$-set K in $\mathcal{G}$, a j-set $C \subseteq K$, and two vertices $w \in K \setminus C$ and $a \in [n] \setminus K$, we say that the 4-tuple (K, C, w, a) forms a *copy of* M_j^* (see Figure 3.5) if

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

(M2) C is a $(j-1)$-simplex in K such that each petal of the flower $\mathcal{F}(K,C)$ is contained in *no other* k-simplex of $\mathcal{G}$;

(M3*) $C \cup \{w\} \cup \{a\}$ is a j-shell in $\mathcal{G}$.

Recall that (M1) and (M2) mean that (K, C) forms a copy of M_j^- (see Definition 3.3.3). 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 $C \cup \{w\} \cup \{a\}$ is called a *side* of the j-shell.

Observe that given a 4-tuple (K, C, w, a) which forms a copy of M_j^* in $\mathcal{G}_p$, the j-shell $C \cup \{w\} \cup \{a\}$ is hollow by (M2) and the fact that every $(j+1)$-simplex in $\mathcal{G}_p$ is contained in a k-simplex. Moreover, since the j-simplices of a j-shell form a j-cycle, a copy of M_j^* is in particular a copy of M_j (Definition 3.1.10). Therefore, the following implications hold.

$$M_j^* \subset \mathcal{G}_p \quad \Rightarrow \quad M_j \subset \mathcal{G}_p \quad \Rightarrow \quad M_j^- \subset \mathcal{G}_p. \tag{3.5}$$

We will see later (Lemma 3.4.6) that for "large" p, whp every copy of M_j^- is extendable to several copies of M_j^*. Therefore, the existence of copies of M_j^-, M_j^* and M_j in $\mathcal{G}_p$ are essentially equivalent events in that range.

Define X_* to be the number of copies of M_j^* in $\mathcal{G}_p$. We need a general expression for its expectation for certain possible values of the probability p. To this end, consider the family $\mathcal{T}^*$ of 4-tuples $T^* = (K, C, w, a)$, where $K \subseteq [n]$ with $|K| = k+1$, where

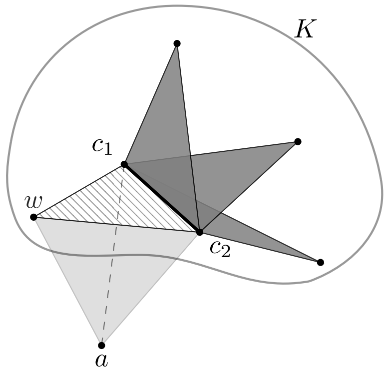


Figure 3.5: A copy of M_j^*, for $k = 5$ and $j = 2$. The pair (K, C), with K a k-simplex and $C = \{c_1, c_2\}$, forms a copy of M_j^-. The $(j+2)$-set $C \cup \{w\} \cup \{a\}$ is a j-shell with base $C \cup \{w\}$ and apex vertex a.

C is a j-subset of K, where $w \in K \setminus C$, and where $a \in [n] \setminus K$. Each of these tuples may form a copy of M_j^* with K as k-simplex, C as the centre, and $C \cup \{w\} \cup \{a\}$ as the j-shell with base $C \cup \{w\}$ and apex vertex a. For each such tuple T^*, let X_{T^*} be the indicator random variable of the event that T^* forms a copy of M_j.

We next show that at probability p_{j-1}^- the number of copies of M_j^* is concentrated around its expectation, whose order we also determine.

Lemma 3.4.4. *If $p = p_{j-1}^-$, then $\mathbb{E}(X_*) = \Theta\left((\log n)^{j+2}\right)$. Furthermore, with high probability $X_* = (1 + o(1))\mathbb{E}(X_*)$.*

Proof. Let $T^* = (K, C, w, a) \in \mathcal{T}^*$ be a fixed 4-tuple. Recall that T^* forms a copy of M_j^* in $\mathcal{G}_p$ if conditions (M1), (M2), and (M3*) of Definition 3.4.3 hold.

Clearly, (M1) holds with probability p. In order to determine the probability that (M2) holds, consider a fixed petal. The probability that this petal lies in no other k-simplex is

$$r = r(p, n, k, j) := (1 - p)^{\binom{n-j-1}{k-j} - 1}. \tag{3.6}$$

For $p = p_{j-1}^- = \Theta\left(\frac{\log n}{n^{k-j+1}}\right)$, we have

$$r \geq 1 - \binom{n-j-1}{k-j} p = 1 - o(1),$$

and thus each petal lies in no other k-simplices whp. Therefore, taking a union bound, (M2) holds with probability at least $1 - (k - j + 1)(1 - r) = 1 - o(1)$.

Now consider (M3*), conditioned on the event that both (M1) and (M2) hold. The base $C \cup \{w\}$ of $C \cup \{w\} \cup \{a\}$ already lies in K, so it remains to prove that all other $(j+1)$-sets in $C \cup \{w\} \cup \{a\}$, i.e. the sides of this (potential) j-shell, are j-simplices in $\mathcal{G}_p$. Denote the sides of $C \cup \{w\} \cup \{a\}$ by $L_1, \ldots, L_{j+1}$. The number of $(k+1)$-sets containing L_i is $\binom{n-j-1}{k-j}$, but some of these $(k+1)$-sets might not be allowed to be

k-simplices because they contain a petal of the flower $\mathcal{F}(K,C)$ (see (3.2)). However, the number of $(k+1)$-sets for which this is the case is $O(n^{k-j-1})$. All other $(k+1)$-sets meet $C\cup\{w\}\cup\{a\}$ only in L_i. In particular, for $i=1,\dots,j+1$, let $\mathcal{K}_i$ denote the set of $(k+1)$-sets which meet $C\cup\{w\}\cup\{a\}$ only in L_i. Then $|\mathcal{K}_i| = \binom{n-j-1}{k-j} - O(n^{k-j-1})$, and the $\mathcal{K}_i$ are pairwise disjoint. Furthermore, conditional on (M1) and (M2), L_i is a j-simplex if and only if at least one $(k+1)$-set of $\mathcal{K}_i$ is present as a k-simplex. In particular, conditional on (M1) and (M2) the events that $L_1,\dots,L_{j+1}$ are j-simplices are independent. Thus, conditional on (M1) and (M2) each L_i forms a j-simplex independently with probability

$$1-(1-p)^{\binom{n-j-1}{k-j}-O(n^{k-j-1})} = (1+o(1))q, \tag{3.7}$$

where

$$q := \frac{pn^{k-j}}{(k-j)!} = \Theta\left(\frac{\log n}{n}\right). \tag{3.8}$$

Therefore, conditional on (M1) and (M2) holding, (M3*) holds with probability $(1+o(1))q^{j+1}$. The probability that T^* forms a copy of M_j^* is thus

$$(1+o(1))pq^{j+1}.$$

The number of 4-tuples $(K,C,w,a)\in\mathcal{T}^*$ is

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

and thus we have

$$\mathbb{E}(X_*) = (1+o(1))\frac{pq^{j+1}n^{k+2}}{j!(k-j)!} = \Theta\left((\log n)^{j+2}\right), \tag{3.9}$$

as required.

In order to prove the second statement of the lemma, we will show that $\mathbb{E}(X_*^2) = (1+o(1))\mathbb{E}(X_*)^2$ and then apply Chebyshev's inequality. We have

$$\mathbb{E}(X_*^2) = \sum_{T_1^*,T_2^*\in\mathcal{T}^*} \Pr\left(\{X_{T_1^*}=1\}\cap\{X_{T_2^*}=1\}\right).$$

Given two 4-tuples $T_1^* = (K_1,C_1,w_1,a_1)$ and $T_2^* = (K_2,C_2,w_2,a_2)$, we define

- $I = I(T_1^*,T_2^*) := (K_1\cup\{a_1\})\cap(K_2\cup\{a_2\})$ and $i := |I|$;
- $s = s(T_1^*,T_2^*) := \begin{cases} 1 & \text{if } K_1 = K_2, \\ 2 & \text{otherwise;} \end{cases}$
- $\mathcal{L}_\ell$ to be the set of all $(j+1)$-subsets of $\{C_\ell\cup\{a_\ell\}\cup\{w_\ell\}\}$ for $\ell = 1,2$ and

  $$t = t(T_1^*,T_2^*) := |(\mathcal{L}_1\cup\mathcal{L}_2)\setminus\{C_1\cup\{w_1\},C_2\cup\{w_2\}\}|,$$

  i.e. the number of $(j+1)$-sets that are sides of the (potential) j-shells of T_1^* and T_2^*, but not a base of either j-shell.

If $s = 2$ and the intersection of the two simplices contains a petal, then T_1^* and T_2^* cannot both form an M_j^*, because (M2) would be violated. In the following, we therefore assume that this is not the case.

The probability that both T_1^* and T_2^* satisfy (M1) is p^s. As before, (M2) holds whp. Conditioned on (M1) and (M2) holding, we claim that (M3*) holds for both tuples simultaneously with probability $(1+o(1))q^t$. In order to prove this, denote the relevant sides of the two j-shells by $L_1, \ldots, L_t$. As before, no k-simplex can contain more than one side of the same j-shell, because otherwise it would also contain the base of the j-shell, which would contradict (M2). In particular, no k-simplex contains at least three of the L_i. Each L_i lies in a k-simplex with probability $(1 + o(1))q$ by (3.7). Moreover, the number of $(k + 1)$-sets that contain $L_i \cup L_{i'}$ for some $i' \neq i$ is $O(n^{k-j-1})$. Thus, the probability that L_i lies in such a k-simplex is

$$1 - (1 - p)^{O(n^{k-j-1})} = O\left(\frac{\log n}{n^2}\right) \overset{(3.8)}{=} o(q^2).$$

This means that the probability that $L_1, \ldots, L_t$ all lie in k-simplices is $(1 + o(1))q^t$. This in turn yields

$$\Pr(\{X_{T_1^*} = 1\} \cap \{X_{T_2^*} = 1\}) = (1 + o(1))p^s q^t. \tag{3.10}$$

Define $\mathcal{T}^2(i, s, t)$ to be the set of pairs $(T_1^*, T_2^*) \in \mathcal{T}^* \times \mathcal{T}^*$ with parameters i, s and t. Denote by $\mathcal{S}$ the set of triples (i, s, t) for which $\mathcal{T}^2(i, s, t)$ is non-empty. With this notation, (3.10) implies that

$$\mathbb{E}(X_*^2) = (1 + o(1)) \sum_{(i,s,t)\in\mathcal{S}} \; \sum_{(T_1^*,T_2^*)\in\mathcal{T}^2(i,s,t)} p^s q^t.$$

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

Case 1: s=1. This means that $K_1 = K_2$ and thus $i \geq k + 1$.

- $i = k+1$. 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

$$O\left(pq^{2j+2}n^{2k+4-(k+1)}\right) \overset{(3.9)}{=} O\left(\frac{\mathbb{E}(X_*)^2}{pn^{k+1}}\right) = o(\mathbb{E}(X_*)^2).$$

- $i = k + 2$. The two j-shells have the same apex vertex but may have different bases. Thus $t \geq j + 1$ (with equality if the two bases are identical), which gives a contribution of order

$$O(pq^{j+1}n^{2k+4-(k+2)}) \overset{(3.9)}{=} O\left(\mathbb{E}(X_*)\right) = o\left(\mathbb{E}(X_*)^2\right).$$

Case 2: s=2.

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

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

choices for T_1^*. For any fixed T_1^*, the number of choices for T_2^* that yield $i = 0$ is

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

Thus, the contribution of all such pairs is

$$(1+o(1))\frac{p^2q^{2j+2}n^{2k+4}}{(j!(k-j)!)^2} \overset{(3.9)}{=} (1+o(1))\mathbb{E}(X_*)^2.$$

- $1 \leq i \leq j$. In this case T_1^* and T_2^* cannot share a j-simplex of their shells, i.e. $t = 2j + 2$. Therefore the contribution is

$$O\left(p^2q^{2j+2}n^{2k+4-i}\right) \overset{(3.9)}{=} O\left(\frac{\mathbb{E}(X_*)^2}{n^i}\right) = o(\mathbb{E}(X_*)^2).$$

- $i = j+1$. Here, T_1^* and T_2^* can share at most one j-simplex of their shells, which means $t \geq 2j + 1$ and we have a contribution of order

$$O\left(p^2q^{2j+1}n^{2k+4-(j+1)}\right) \overset{(3.9)}{=} O\left(\frac{\mathbb{E}(X_*)^2}{qn^{j+1}}\right) \overset{(3.8)}{=} o(\mathbb{E}(X_*)^2).$$

- $j + 2 \leq i \leq k + 2$. In this case $t \geq j$, because T_1^* and T_2^* may share their j-shells but have different bases, i.e. two j-simplices of the (potential) j-shells may be automatically present because of K_1 and K_2. Therefore the contribution is

$$O\left(p^2q^jn^{2k+4-i}\right) \overset{(3.9)}{=} O\left(\frac{\mathbb{E}(X_*)^2}{q^{j+2}n^i}\right) \overset{(3.8)}{=} o(\mathbb{E}(X_*)^2).$$

Summing over all cases shows that $\mathbb{E}(X_*^2) = (1 + o(1))\mathbb{E}(X_*)^2$, as desired. Thus, Chebyshev's inequality implies that $X_* = (1 + o(1))\mathbb{E}(X_*)$ whp. $\square$

Remark 3.4.5. The case $s = 1$, $i = k + 1$ in the proof of Lemma 3.4.4 also gives the expected number of pairs of copies of M_j^* coming from a common M_j^-. Since this is of order

$$\Theta\left(pq^{2j+2}n^{k+3}\right) \overset{(3.8)}{=} \Theta\left(\frac{(\log n)^{2j+3}}{n^j}\right) = o(1),$$

by Markov's inequality we deduce that whp in $\mathcal{G}_{p_{j-1}^-}$ each copy of M_j^- can be extended to *at most* one copy of M_j^*. We will make use of this observation in Lemma 3.4.12.

In contrast to Remark 3.4.5, the following lemma ensures that at around $p = p_j^{(1)}$, whp every j-simplex in $\mathcal{G}_p$ is the base of "many" j-shells. Thus it is very likely that each copy of M_j^- gives rise to several copies of M_j^*, allowing us to consider just copies of M_j^- as obstructions to j-cohom-connectedness. In other words,

$$\text{whp} \quad \text{for each } p \geq p_j^{(1)}, \qquad M_j^- \subset \mathcal{G}_p \;\; \Rightarrow \;\; M_j^* \subset \mathcal{G}_p.$$

Combining this with (3.5), the existence of copies of M_j^-, M_j^* and M_j are essentially equivalent for $p \geq p_j^{(1)}$. Recall from Definition 3.2.1 that for a complex $\mathcal{G}$ and a set B, $\mathcal{G} + B$ is the complex obtained by adding the set B and its downward-closure to $\mathcal{G}$.

Lemma 3.4.6. *Let $p = p_j^{(1)}$. Then there exists a positive constant γ such that with high probability for every $(j+1)$-set B the complex $\mathcal{G}_p + B$ contains at least γn many j-shells that contain B.*

Proof. Recall that

$$p = p_j^{(1)} = \frac{1}{10(j+1)\binom{k+1}{j+1} n^{k-j}}.$$

Let $L_1, \ldots, L_{j+1}$ denote the $(j-1)$-simplices contained in B. We are interested in the number of vertices a such that $B \cup \{a\}$ forms a j-shell, i.e. the number of $a \notin B$ such that $L_i \cup \{a\}$ is a j-simplex in $\mathcal{G}_p + B$ for all $i \in [j+1]$. To ensure independence in the following calculations, we will only consider a certain type of such j-shells, giving us a lower bound on their total number. Pick two disjoint sets A and D both of size $\lceil n/3 \rceil$ such that $A \cap B = D \cap B = \emptyset$. We will consider only (potential) j-shells formed in the following way.

- The vertex a is in A;
- for each $i = 1, \ldots, j+1$, the j-simplex $L_i \cup \{a\}$ is present in $\mathcal{G}_p$ (and thus also in $\mathcal{G}_p + B$) as a subset of the k-simplex $R_i \cup L_i \cup \{a\}$, for some (not necessarily distinct) $(k-j)$-sets $R_1, \ldots, R_{j+1}$ in D.

In this way all the required j-simplices would come from different k-simplices, ensuring independence.

Fix $a \in A$ and let E_a be the event that $B \cup \{a\}$ is a j-shell. Observe that for each L_i, the probability that there is no suitable set $R_i \subseteq D$ is

$$(1-p)^{\binom{|D|}{k-j}} \leq (1-p)^{\frac{n^{k-j}}{4^{k-j}(k-j)!}}.$$

Therefore, setting $\beta := 10(j+1)\binom{k+1}{j+1}4^{k-j}(k-j)!$, by independence we have

$$\begin{aligned}
\Pr(E_a) &\geq \left(1 - (1-p)^{\frac{n^{k-j}}{4^{k-j}(k-j)!}}\right)^{j+1} \\
&\geq \left(\frac{n^{k-j}}{4^{k-j}(k-j)!}p - \frac{1}{2}\left(\frac{n^{k-j}}{4^{k-j}(k-j)!}p\right)^2\right)^{j+1} \\
&= \left(\frac{1}{\beta} - \frac{1}{2\beta^2}\right)^{j+1} =: \lambda > 0.
\end{aligned}$$

The events E_a are independent for distinct a, so the number of j-shells we count in this way dominates $\mathrm{Bi}(\lceil n/3\rceil, \lambda)$. Fixing a constant $0 < \gamma < \lambda/3$, we can apply the Chernoff bound (Lemma 3.2.2) to deduce that

$$\Pr\left(\mathrm{Bi}(\lceil n/3\rceil, \lambda) < \gamma n\right) \leq \exp\left(-\frac{(n\lambda/3 - \gamma n)^2}{2n\lambda/3}\right) = \exp\left(-\frac{n(\lambda/3-\gamma)^2}{2\lambda/3}\right).$$

Finally, taking a union bound over all $\binom{n}{j+1}$ possible choices for the set B, we can bound the probability that the desired property does not hold by

$$\binom{n}{j+1}\exp\left(-\frac{n(\lambda/3-\gamma)^2}{2\lambda/3}\right) = o(1),$$

as required. □

We now also prove that shortly before the (claimed) critical threshold for $\mathbb{F}_2$-cohomological j-connectedness, the number of copies of M_j^- is concentrated around its expectation, using similar techniques as in Lemma 3.4.4.

Lemma 3.4.7. *Let $\omega = o(\log n)$ be a function of n which tends to infinity as $n \to \infty$. Let*

$$p \in \left[p_j^-, \frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)!\right],$$

and let X_- be the number of copies of M_j^- in $\mathcal{G}_p$. Then $\mathbb{E}(X_-) = \Omega(e^\omega)$ and with high probability $X_- = (1+o(1))\mathbb{E}(X_-)$.

Proof. Let K be a $(k+1)$-set and let C be a j-set in K. In order for (K, C) to form a copy of M_j^-, we need K to be a k-simplex and each petal of the flower $\mathcal{F}(K, C) = \{C \cup \{w\} \mid w \in K \setminus C\}$ to lie in no other k-simplex. For a fixed petal, the probability of this event is equal to $r = (1-p)^{\binom{n-j-1}{k-j}-1}$ defined in (3.6). Moreover, there are $O(n^{k-j-1})$ many $(k+1)$-sets that contain more than one petal. Now since

$$(1-p)^{O(n^{k-j-1})} = 1 - o(1),$$

whp there are no k-simplices containing more than one petal. Thus,

$$\begin{aligned} \mathbb{E}(X_-) &= (1+o(1))\binom{n}{k+1}\binom{k+1}{j}pr^{k-j+1} \\ &= (1+o(1))\binom{n}{k+1}\binom{k+1}{j}p(1-p)^{(k-j+1)\binom{n-j-1}{k-j}}. \end{aligned} \tag{3.11}$$

The derivative of the right hand side of (3.11) with respect to p is negative throughout the considered interval. Therefore the upper extreme of p gives the smallest expectation, which is of order

$$\Theta(n^{k+1})\Theta\left(\frac{\log n}{n^{k-j}}\right)\Theta\left(\exp\left(-(j+1)\log n - \log\log n + \omega\right)\right) = \Theta\left(e^\omega\right) \to \infty.$$

In order to apply a second moment argument, we will now show that

$$\mathbb{E}(X_-^2) = (1 + o(1))\mathbb{E}(X_-)^2,$$

implying that whp X_- is concentrated around its expectation. Let $\mathcal{T}^-$ denote the family of pairs $T^- = (K, C)$, where $K \subseteq [n]$ with $|K| = k + 1$ and C is a j-subset of K. Each of these pairs may form a copy of M_j^- 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

- $s = s(T_1^-, T_2^-) := \begin{cases} 1 & \text{if } K_1 = K_2, \\ 2 & \text{otherwise;} \end{cases}$
- $\mathcal{F}_\ell := \mathcal{F}(K_\ell, C_\ell)$ for $\ell = 1, 2$;
- $t = t(T_1^-, T_2^-) := |\mathcal{F}_1 \cup \mathcal{F}_2|$, i.e. the total number of (potential) petals.

The probability of two pairs in $\mathcal{T}^-$ both forming a copy of M_j^- is $(1+o(1))p^s r^t$. With this observation, we can determine the contribution to $\mathbb{E}(X_-^2)$ of the pairs with a fixed value of s.

- $s = 1$. Petals can be shared, but certainly $t \geq k - j + 1$ and the contribution is at most of order

  $$O\left(n^{k+1} p r^{k-j+1}\right) \overset{(3.11)}{=} O(\mathbb{E}(X_-)) = o(\mathbb{E}(X_-)^2).$$

- $s = 2$. 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^-. The number of such pairs is

  $$\binom{n}{k+1}\binom{n-k-1}{k+1}\binom{k+1}{j}^2 + O(n^{2k+1}) = (1 + o(1))\binom{n}{k+1}^2\binom{k+1}{j}^2.$$

  Thus these pairs provide a contribution of

  $$(1 + o(1))\binom{n}{k+1}^2\binom{k+1}{j}^2 p^2 r^{2(k-j+1)} \overset{(3.11)}{=} (1 + o(1))\mathbb{E}(X_-)^2.$$

In total, we have $\mathbb{E}(X_-^2) = (1 + o(1))\mathbb{E}(X_-)^2$, and Chebyshev's inequality implies that $X_- = (1 + o(1))\mathbb{E}(X_-)$ whp. $\square$

3.4.3 Excluding obstructions and determining the hitting time

The goal of this section is to determine when there are no more copies of M_j in $\mathcal{G}_p$ whp. This result, together with Lemmas 3.4.6 and 3.4.7, will enable us to prove that whp the birth time p_{M_j} is close to p_j, the (claimed) threshold for j-cohom-connectedness (Corollary 3.4.11).

Consider the probability

$$\bar{p}_j := \frac{(j+1)\log n + \frac{1}{2}\log\log n}{(k-j+1)n^{k-j}}(k-j)!. \tag{3.12}$$

Define $\bar{p}_{M_j}$ as the first birth time p larger than $\bar{p}_j$ such that there are no copies of M_j in $\mathcal{G}_p$. By Lemmas 3.4.6 and 3.4.7, whp $\mathcal{G}_{\bar{p}_j}$ contains a growing number of copies of M_j. By definition of p_{M_j}, conditioned on this high probability event we have $\bar{p}_{M_j} \leq p_{M_j}$. In the next lemma we show that in fact they are equal whp. To do so, we need the following definition.

Definition 3.4.8. Given a k-complex $\mathcal{G}$, a k-simplex K is a *local obstacle* if K contains at least $k-j+1$ many j-simplices which are not contained in any other k-simplex of $\mathcal{G}$.

Note that this definition is similar to that of M_j^- (Definition 3.3.3), but without the restriction that the $k-j+1$ many j-simplices must form a flower.

Lemma 3.4.9. *With high probability, for all $p \geq \bar{p}_j$ every local obstacle that exists in $\mathcal{G}_p$ also exists in $\mathcal{G}_{\bar{p}_j}$. In particular, we have $p_{M_j} = \bar{p}_{M_j}$ whp.*

Proof. Suppose that $\mathcal{G}_p$ contains a local obstacle which is not present in $\mathcal{G}_{\bar{p}_j}$ and let K be the $(k+1)$-set realising this obstacle. Then its birth time p_K satisfies $p_K \in (\bar{p}_j, p]$. The set K can become a local obstacle only if

(i) K contains a collection $\mathcal{L}$ of (at least) $k-j+1$ many $(j+1)$-sets which are not yet j-simplices in $\mathcal{G}_{\bar{p}_j}$;

(ii) p_K is smaller than the birth time of any other $(k+1)$-set containing at least one of the $(j+1)$-sets in $\mathcal{L}$.

If K satisfies (i), then for any $(j+1)$-set $L \in \mathcal{L}$, no $(k+1)$-set intersecting K precisely in L is allowed to be a k-simplex in $\mathcal{G}_{\bar{p}_j}$ and thus there are at least $\binom{n-k-1}{k-j}(k-j+1)$ many $(k+1)$-sets which are not k-simplices. Hence, given K and $k-j+1$ fixed $(j+1)$-sets within K, the probability of satisfying property (i) in $\mathcal{G}_{\bar{p}_j}$ is bounded from above by

$$\begin{aligned}(1-\bar{p}_j)^{\binom{n-k-1}{k-j}(k-j+1)} &= (1+o(1))\exp\left(-\log\left(n^{j+1}\right) - \log\left((\log n)^{1/2}\right)\right)\\ &= O\left(\frac{1}{n^{j+1}\sqrt{\log n}}\right).\end{aligned}$$

On the other hand, each $(j+1)$-set in $\mathcal{L}$ is contained in $\binom{n-j-1}{k-j}$ potential k-simplices. Let $\mathcal{K}$ denote the union of these sets of $(k+1)$-sets over all $(j+1)$-sets in $\mathcal{L}$. Observe that in particular, $K \in \mathcal{K}$. In order for K to satisfy (ii), it would have to have the smallest birth time of any $(k+1)$-set in $\mathcal{K}$. Conditional on all $(k+1)$-sets of $\mathcal{K}$ not being present as k-simplices in $\mathcal{G}_{\bar{p}_j}$, their birth times are independent and uniformly distributed in $(\bar{p}_j, 1]$. Thus the order in which the k-simplices of $\mathcal{K}$ are born is uniformly random, and in particular the probability that K is born first is $|\mathcal{K}|^{-1} = O\left(\frac{1}{n^{k-j}}\right)$. Thus, the expected number of sets K satisfying (i) and (ii) is at most

$$\binom{n}{k+1} 2^{\binom{k+1}{j+1}} O\left(\frac{1}{n^{j+1}\sqrt{\log n}}\right) O\left(\frac{1}{n^{k-j}}\right) = O\left(\frac{1}{\sqrt{\log n}}\right) = o(1)$$

and the conclusion follows by Markov's inequality. □

Observe that in particular each copy of M_j^- is a local obstacle. Thus, we derive the following corollary.

Corollary 3.4.10. *Whp for all $p \geq p_{M_j}$, there are no copies of M_j^- in $\mathcal{G}_p$.*

We can now easily deduce that the birth time p_{M_j} at which the last copy of M_j disappears is close to p_j. Observe that the following corollary is exactly Theorem 3.1.11 (i).

Corollary 3.4.11. *Let ω be any function of n which tends to infinity as n tends to infinity. Then whp*

$$\frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)! < p_{M_j} < \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!.$$

Proof. We may assume without loss of generality that $\omega = o(\log n)$. By Lemmas 3.4.6 and 3.4.7, $p_{M_j} > \frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)!$ whp. On the other hand, setting $p = \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!$ and arguing as in Lemma 3.4.7 (see (3.11)), the expected number of copies of M_j^- is bounded from above by

$$\begin{aligned}
&(1+o(1))n^{k+1} p \exp\left(-\frac{(n-j-1)^{k-j}}{(k-j)!}(k-j+1)p\right) \\
&\quad = \Theta\left(n^{j+1}\log n \exp\left(-(j+1)\log n - \log\log n - \omega\right)\right) \\
&\quad = \Theta\left(e^{-\omega}\right) = o(1).
\end{aligned}$$

So by Markov's inequality, whp there are no copies of M_j^- and thus also no copies of M_j in $\mathcal{G}_p$, i.e.

$$\bar{p}_{M_j} < \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!$$

and by Lemma 3.4.9 we have $\bar{p}_{M_j} = p_{M_j}$ whp. □

3.4.4 Covering the intervals: proof of Lemma 3.3.4

In order to prove Lemma 3.3.4, we show that for each $j \in [k-1]$, whp there exist three minimal obstructions which survive throughout each of the intervals $[p_{j-1}^-, p_j^{(1)}]$, $[p_j^{(1)}, p_j^-]$ and $[p_j^-, p_{M_j})$, respectively.

Recall that

$$p_j^{(1)} = \frac{1}{10(j+1)\binom{k+1}{j+1}n^{k-j}}. \tag{3.13}$$

The first step is to show that at least one of the $X_* = \Theta((\log n)^{j+2})$ copies of M_j^* which are present whp at probability p_{j-1}^- (Lemma 3.4.4) survives until time $p_j^{(1)}$. To do so, we will count the number of *dangerous sets*, that is $(k+1)$-sets which, if they are selected as k-simplices, make one or more of copies of M_j^* disappear. Then we show that whp up to probability $p_j^{(1)}$ the number of copies of M_j^* destroyed by dangerous sets which became k-simplices is less than X_*.

Lemma 3.4.12. *With high probability one copy of M_j^* exists in $\mathcal{G}_p$ throughout the range $[p_{j-1}^-, p_j^{(1)}]$.*

Proof. Define $x = 2\mathbb{E}(X_*)$ at time $p = p_{j-1}^-$. By Lemma 3.4.4, we know that whp

$$\frac{x}{3} \leq X_* \leq x,$$

so let us condition on this high probability event occurring.

We know that we can generate $\mathcal{G}_{p_j^{(1)}}$ from $\mathcal{G}_{p_{j-1}^-}$ by exposing an additional probability of $\frac{p_j^{(1)} - p_{j-1}^-}{1 - p_{j-1}^-} \leq p_j^{(1)}$, therefore we will use the upper bound $p_j^{(1)}$ in the following calculations. Set $p = p_j^{(1)}$ and let Y be the number of dangerous sets selected as k-simplices in $\mathcal{G}_p$. A $(k+1)$-set can contain at most $\binom{k+1}{j+1}$ petals, each of which can be part of at most $j+1$ different copies of M_j^-, since by definition a petal belongs to exactly one k-simplex and within this petal we have $\binom{j+1}{j} = j+1$ choices for the centre which then uniquely defines the copy of M_j^-. So each of the k-simplices counted by Y can destroy at most $c := \binom{k+1}{j+1}(j+1)$ copies of M_j^-. Moreover, by Remark 3.4.5, whp c is also the maximum number of copies of M_j^* that can disappear by adding a dangerous set to the complex. Therefore, we now show that

$$\Pr\left(cY \geq \frac{x}{3}\right) = o(1).$$

This will imply that whp $cY < X_*$, so at least one of the copies of M_j^* counted by X_* will survive throughout the considered probability interval.

A dangerous $(k+1)$-set makes one or more copies of M_j^* disappear if it becomes a k-simplex and contains at least one petal of each of their flowers. For a copy of M_j^*, the number of $(k+1)$-sets that intersect it in at least one petal is at most $(k-j+1)\binom{n-j-1}{k-j}$. Therefore, whp the number of dangerous $(k+1)$-sets is at most

$$(k-j+1)\binom{n-j-1}{k-j}x \leq \frac{k-j+1}{(k-j)!}n^{k-j}x \leq 2n^{k-j}x =: N.$$

Due to the independence of the chosen k-simplices, Y is dominated by $\mathrm{Bi}(N, p)$. Since

$$\mathbb{E}(\mathrm{Bi}(N,p)) = Np \overset{(3.13)}{=} \frac{x}{5c},$$

by the Chernoff bound (Lemma 3.2.2) we have

$$\begin{aligned}\Pr\left(Y \geq \frac{x}{3c}\right) &\leq \Pr\left(\mathrm{Bi}(N,p) \geq \frac{x}{3c}\right) \\ &\leq \exp\left(-\frac{\left(\frac{x}{3c} - Np\right)^2}{2\left(Np + \left(\frac{x}{3c} - Np\right)/3\right)}\right) \\ &= \exp\left(-\frac{2x}{55c}\right) = o(1),\end{aligned}$$

because $x \xrightarrow{n\to\infty} \infty$ by Lemma 3.4.4. $\square$

We now consider the second subinterval $[p_j^{(1)}, p_j^-]$. In this range, we will show that whp one of the "many" copies of M_j^- which exist whp at time p_j^- (Lemma 3.4.7) was already present at the beginning of the interval. Together with the fact that whp each M_j^- gives rise to a copy of M_j (Lemma 3.4.6), this will imply that whp one copy of M_j exists throughout this interval.

Lemma 3.4.13. *With high probability one copy of M_j^- exists in $\mathcal{G}_p$ throughout the range $[p_j^{(1)}, p_j^-]$.*

Proof. Set

$$p = p_j^- = \left(1 - \frac{1}{\sqrt{\log n}}\right)\frac{(j+1)\log n}{(k-j+1)n^{k-j}}(k-j)!.$$

By Lemma 3.4.7, at probability p the number X_- of copies of M_j^- is concentrated around its expectation

$$\mathbb{E}(X_-) \overset{(3.11)}{=} \Theta\left(n^{k+1}p(1-p)^{(k-j+1)\binom{n-j-1}{k-j}}\right) = \Theta\left(n^{\frac{j+1}{\sqrt{\log n}}}\log n\right),$$

which is growing with n. Note that a fixed k-simplex can give rise to only $\binom{k+1}{j} = \Theta(1)$ different copies of M_j^-. Therefore whp there are $\Theta\left(n^{\frac{j+1}{\sqrt{\log n}}}\log n\right)$ many copies of M_j^- that arise from different k-simplices, and whose birth times are thus independent. Given that these copies exist at time p_j^-, the birth times of the corresponding k-simplices are uniformly distributed in the interval $[0, p_j^-]$. The probability that any fixed such copy already existed at time $p_j^{(1)}$ is therefore

$$\frac{p_j^{(1)}}{p_j^-} = \Theta\left(\frac{1}{\log n}\right).$$

Thus, because of the independence, the probability that none of them was present at $p_j^{(1)}$ is at most

$$\left(1 - \Theta\left(\frac{1}{\log n}\right)\right)^{\Theta\left(n^{\frac{j+1}{\sqrt{\log n}}}\log n\right)} \leq \exp\left(-\Theta\left(n^{\frac{j+1}{\sqrt{\log n}}}\right)\right) = o(1),$$

as required. $\square$

We now conclude the argument by covering the third interval $[p_j^-, p_{M_j})$ of the subcritical range.

Lemma 3.4.14. *With high probability one copy of M_j^- exists in $\mathcal{G}_p$ throughout the range $[p_j^-, p_{M_j})$.*

Proof. By the definition of p_j^- and Corollary 3.4.11, we know that whp $p_j^- = (1 - o(1))p_{M_j}$. So, conditioning on this high probability event and arguing as in the proof of Lemma 3.4.13, the final minimal obstruction to disappear at time p_{M_j} already existed at time p_j^- with probability at least

$$\frac{p_j^-}{p_{M_j}} = 1 - o(1),$$

as required. □

Proof of Lemma 3.3.4. By Lemma 3.4.6, the copies of M_j^- from Lemmas 3.4.13 and 3.4.14 whp give rise to copies of M_j^*, and thus in particular to copies of M_j. Therefore, Lemmas 3.4.12, 3.4.13 and 3.4.14 together imply Lemma 3.3.4. □

3.5 Critical window and supercritical regime

3.5.1 Overview

In this section, we study obstructions around the point of the claimed phase transition and in the supercritical regime, that is, for $p = (1 + o(1))p_j$ and $p \geq p_{M_j}$, respectively. The results of this section will form the foundation of the proof of Theorem 3.1.11 (iii). Furthermore, they will be an essential ingredient in the proof of Theorem 3.1.13.

By the definition of p_{M_j}, there are no copies of M_j in $\mathcal{G}_p$ (and whp also no copies of M_j^- by Corollary 3.4.10) for any $p \geq p_{M_j}$. It remains to show that there are no other obstructions either. In fact, we shall even prove (Corollary 3.5.10) that from slightly before p_{M_j} onwards, all j-cocycles are generated by copies of M_j^- (recall that a j-cocycle is a j-cochain in $\ker \delta^j$, see Section 3.2.3). To make this more precise, we need the following terminology.

Definition 3.5.1. Let (K, C) be a copy of M_j^- in a k-complex $\mathcal{G}$. We say that a j-cochain $f_{K,C}$ *arises from* (K, C) if its support is the j-flower $\mathcal{F}(K, C)$. (Observe that $f_{K,C}$ is then a j-cocycle.)

We say that a j-cocycle f in $\mathcal{G}$ is *generated by copies of* M_j^- if it lies in the same cohomology class as a sum of j-cocycles that arise from copies of M_j^-. We denote by $\mathcal{N}_{\mathcal{G}}$ the set of j-cocycles that are *not* generated by copies of M_j^-.

Our goal is to show that whp $\mathcal{N}_{\mathcal{G}_p} = \emptyset$ for $p \geq p_j^-$ (Corollaries 3.5.8 and 3.5.10), which in particular will imply that whp each j-cocycle in $\mathcal{G}_p$ is also a j-coboundary (i.e. there are no bad functions, see Definition 3.2.4) for all $p \geq p_{M_j}$. Furthermore, it will enable us to directly relate the number of copies of M_j^- to the dimension of $H^j(\mathcal{G}_p; \mathbb{F}_2)$ (cf. Theorem 3.1.13).

Definition 3.5.2. For each $p \in [0,1]$, let f_p be a function in $\mathcal{N}_{\mathcal{G}_p}$ with smallest support S_p, if such a function exists.

In order to prove that whp $\mathcal{N}_{\mathcal{G}_p}$ is empty, we show (Lemma 3.5.4) that for any k-complex $\mathcal{G}$, a smallest support of elements of $\mathcal{N}_{\mathcal{G}}$ (and so in particular S_p in $\mathcal{G}_p$) would have to be *traversable* (see Definition 3.5.3). We then show that whp no $\mathcal{G}_p$ with $p \geq p_{j-1}^-$ can contain a traversable support S_p. For "small" sizes of S_p and $p = (1+o(1))p_j$, basic estimates and a union bound argument will suffice (Lemma 3.5.5); for larger size, we will make use of traversability to define a breadth-first search process that finds all possible supports. In this way, we can bound the number of possibilities for S_p more carefully, thus allowing us to prove that whp for all relevant p simultaneously, S_p cannot be "large" (Lemma 3.5.7). Finally, we complete the argument proving that whp no new elements of $\mathcal{N}_{\mathcal{G}_p}$ with "small" support size can appear if we increase p (Lemma 3.5.9).

3.5.2 Traversability

Definition 3.5.3. Let $\mathcal{G}$ be a k-complex in which each simplex is contained in a k-simplex, and let S be a collection of j-simplices of $\mathcal{G}$. For $\sigma_1, \sigma_2 \in S$, we set

$$\sigma_1 \sim \sigma_2 \quad \text{if} \quad \sigma_1 \text{ and } \sigma_2 \text{ lie in a common } k\text{-simplex.}$$

We say that the set S is *traversable* if the transitive closure of $\sim$ is $S \times S$.

In other words, a set of j-simplices in such a k-complex is traversable if it *cannot* be partitioned into two non-empty subsets such that each k-simplex (and thus also each $(j+1)$-simplex) contains j-simplices in at most one of the two subsets.

Lemma 3.5.4. *Let $\mathcal{G}$ be a k-complex in which each simplex is contained in a k-simplex, and let f be an element of $\mathcal{N}_{\mathcal{G}}$ with smallest support S. Then S is traversable. In particular, S_p is traversable in $\mathcal{G}_p$, if it exists, for each $p \in [0,1]$.*

Proof. Suppose S is not traversable. Then we can find a partition $S = T_1 \dot\cup T_2$, with T_1 and T_2 non-empty such that each $(j+1)$-simplex of $\mathcal{G}$ contains j-simplices in at most one of the two parts. Define g_1 and g_2 to be j-cochains with supports T_1 and T_2, respectively. By the choice of T_1 and T_2, both g_1 and g_2 are j-cocycles. Moreover, neither of them lies in $\mathcal{N}_{\mathcal{G}}$ by the minimality of S. As the property of being generated by copies of M_j^- is closed under summation, $f = g_1 + g_2$ is generated by copies of M_j^-, a contradiction to $f \in \mathcal{N}_{\mathcal{G}}$. $\square$

3.5.3 Small supports

The following counting argument shows that whp, at around time p_j traversable supports of j-cocycles of constant size do not exist. This implies in particular that S_p (if it exists) has to be "large".

Lemma 3.5.5. *For $p = (1 + o(1))p_j$ and for any constant $d \geq k - j + 2$, with high probability there is no j-cocycle in $\mathcal{G}_p$ with traversable support of size s with*

$k - j + 2 \leq s \leq d$. *In particular, with high probability either S_p does not exist or $|S_p| > d$.*

Proof. Consider a traversable support S of a j-cocycle of size s with $k-j+2 \leq s \leq d$. Suppose that S covers v vertices and denote by ℓ the number of k-simplices that make S traversable. These quantities are easily bounded by

$$\frac{s}{\binom{k+1}{j+1}} \leq \ell \leq s \leq d, \tag{3.14}$$

and

$$v \leq j + 1 + (k-j)\ell. \tag{3.15}$$

We know by Lemma 3.3.1 that if a k-simplex contains a j-simplex in S, then all its $k+1$ vertices are covered by S. Therefore, all $s\binom{n-v}{k-j}$ many $(k+1)$-sets consisting of the vertices of one j-simplex in S and $k-j$ vertices not covered by S cannot be k-simplices in $\mathcal{G}_p$. Thus, the probability that a fixed such S exists is at most

$$\begin{aligned}
p^\ell (1-p)^{s\binom{n-v}{k-j}} &= p^\ell (1-p)^{s\left(\frac{n^{k-j}}{(k-j)!} + O(n^{k-j-1})\right)} \\
&= O\left(\left(\frac{\log n}{n^{k-j}}\right)^\ell \exp\left(-\frac{s(j+1)}{k-j+1}\log n + o(\log n)\right)\right) \\
&= O\left(n^{-\ell(k-j) - \frac{s(j+1)}{k-j+1} + o(1)} (\log n)^\ell\right).
\end{aligned}$$

Denote by $E_{s,v,\ell}$ the event that a traversable support S with parameters s, v, and ℓ exists. There are $O(n^v)$ different ways of choosing S, thus

$$\Pr(E_{s,v,\ell}) = O\left(n^{v - \ell(k-j) - \frac{s(j+1)}{k-j+1} + o(1)} (\log n)^\ell\right).$$

Using (3.15) and the fact that $s \geq k - j + 2$, we obtain

$$v - \ell(k-j) - \frac{s(j+1)}{k-j+1} + o(1) \leq -\frac{j+1}{k-j+1} + o(1) \leq -\frac{j}{k-j+1}$$

and thus

$$\Pr(E_{s,v,\ell}) = o(1).$$

Finally, observe that by (3.14) and (3.15), there is only a constant number of possible values for s, v, and ℓ. Therefore, the probability that any such support S exists is $o(1)$, as required. $\square$

Note that a similar argument also works for s up to $O\left(\frac{\log n}{\log\log n}\right)$, but we only need it for constant size, since we will cover the range between constant size and size $O\left(\frac{\log n}{\log\log n}\right)$ with a different argument that we use for all large s.

3.5.4 Large supports

For larger support sizes, the previous calculations do not work anymore and we will need a more careful technique for bounding the number of possible supports, namely a breadth-first search process. We will also make use of the following proposition due to Meshulam and Wallach [53].

Proposition 3.5.6 ([53, Proposition 3.1])**.** *Let Δ be the downward-closure of the $(n-1)$-simplex on vertex set $[n]$, where $n \geq j+2$. For $f \in C^j(\Delta)$, define $w(f)$ to be the smallest size of a support of a j-cochain of the type $f + \delta^{j-1} g$, where $g \in C^{j-1}(\Delta)$. Furthermore, denote by $b(f)$ the size of the support of $\delta^j f$, i.e. the number of $(j+1)$-simplices in Δ containing an odd number of j-simplices of the support in f. Then*

$$b(f) \geq \frac{w(f)n}{j+2}.$$

In the next lemma we show that whp in the supercritical range, a smallest support of elements of $\mathcal{N}_{\mathcal{G}_p}$ cannot be "large".

Lemma 3.5.7. *There exists a positive constant $\bar{d}$ such that with high probability for all $p \geq p_j^-$, either S_p does not exist or $|S_p| < \bar{d}$.*

Proof. Write $s := |S_p|$. By Lemma 3.5.4, S_p (if it exists) is traversable and thus we can discover it via the following breadth-first search process: start from any j-simplex in S_p and query all $(k+1)$-sets containing it. Since S_p is the support of the j-cocycle f_p, any of these sets which forms a k-simplex must contain at least one other j-simplex in S_p. From all j-simplices in S_p found in this way, we can continue the process according to some pre-determined order of j-simplices, but we explore only $(k+1)$-sets which would give us some previously undiscovered j-simplex in S_p. By the traversability of S_p, we discover all of S_p in this process.

Let us bound the number of traversable supports of size s which are contained in $\ell \leq s$ many k-simplices (recall (3.14)), which we can find via the described search process. Define the sequence $\underline{b} = (b_1, \ldots, b_s)$, where $b_i \geq 0$ is the number of k-simplices we discover from the i-th j-simplex in this process. From the i-th j-simplex we may query up to $\binom{n}{k-j}$ many $(k+1)$-sets and for each of the b_i discovered k-simplices we can find at most $\binom{k+1}{j+1} - 1$ new j-simplices of the support, so this can happen in at most $\binom{\binom{n}{k-j}}{b_i} 2^{\binom{k+1}{j+1} b_i}$ different ways. Thus, if we condition on the sequence $\underline{b}$, the number of supports of size s we can find is bounded from above by

$$\binom{n}{j+1} \prod_{i=1}^{s} \binom{\binom{n}{k-j}}{b_i} 2^{\binom{k+1}{j+1} b_i} \leq n^{j+1} \frac{\left(\binom{n}{k-j} 2^{\binom{k+1}{j+1}}\right)^{\ell}}{\prod_{i=1}^{s} b_i!},$$

where we are using that $\sum_{i=1}^{s} b_i = \ell$.

In order to apply Proposition 3.5.6 to f_p (which is possible, because $\mathcal{G}_p$ is a subcomplex of Δ), let us determine the value $w(f_p)$. First observe that for $p \geq p_j^-$, whp $\mathcal{G}_p$ has a complete $(j-1)$-dimensional skeleton, which can be proved by a simple first moment calculation. Thus, if we consider $f_p + \delta^{j-1}g$ with $g \in C^{j-1}(\Delta)$, then whp also $g \in C^{j-1}(\mathcal{G}_p)$ and thus $f_p + \delta^{j-1}g$ lies in the same cohomology class of $H^j(\mathcal{G}_p; \mathbb{F}_2)$ as f_p. By the minimality of S_p, this implies that $w(f_p) = |S_p| = s$ whp. For the rest of the proof, let us condition on this high probability event.

Now Proposition 3.5.6 tells us that at least $\frac{sn}{j+2}$ many $(j+2)$-sets would form odd $(j+1)$-simplices if they were present in $\mathcal{G}_p$. The fact that f_p is a j-cocycle implies that no such $(j+2)$-set is allowed to be in a k-simplex. Each $(j+2)$-set is contained in $\binom{n-j-2}{k-j-1}$ many $(k+1)$-sets, each of which contains $\binom{k+1}{j+2}$ many $(j+2)$-sets. Therefore the number of $(k+1)$-sets that cannot be chosen as k-simplices in $\mathcal{G}_p$ is at least

$$\frac{sn\binom{n-j-2}{k-j-1}}{(j+2)\binom{k+1}{j+2}} \geq \alpha_0 s n^{k-j} \geq \alpha_0 \ell n^{k-j},$$

for some constant $\alpha_0 = \alpha_0(k,j) > 0$. Thus, the probability that a fixed support exists together with the ℓ many k-simplices that make it traversable, but that no odd $(j+1)$-simplices are present is at most

$$\left(p(1-p)^{\alpha_0 n^{k-j}}\right)^{\ell}.$$

The derivative of this expression with respect to p is negative throughout the range $p \geq p_j^-$, therefore in the following calculations involving p we can use the lower bound p_j^-. Given the sequence $\underline{b}$, the probability $q_{\underline{b}}$ that some such support exists and that the connecting k-simplices have no odd $(j+1)$-simplices satisfies

$$\begin{aligned}
q_{\underline{b}} \prod_{i=1}^{s} b_i! &\leq n^{j+1}\left(2^{\binom{k+1}{j+1}}\binom{n}{k-j}p(1-p)^{\alpha_0 n^{k-j}}\right)^{\ell} \\
&\leq n^{j+1}\left(2^{\binom{k+1}{j+1}}\frac{(j+1)\log n}{k-j+1}e^{-(1-o(1))\alpha_0\frac{j+1}{k-j+1}(k-j)!\log n}\right)^{\ell} \\
&\leq n^{j+1}\left(n^{-\alpha_0\frac{j}{k-j+1}(k-j)!}\right)^{\ell} \\
&\leq n^{j+1}n^{-\alpha_1\ell} \leq n^{-\frac{\alpha_1}{2}\ell},
\end{aligned}$$

where $\alpha_1 = \alpha_1(k,j) > 0$ and the last inequality holds for $\ell \geq \frac{2(j+1)}{\alpha_1}$. Moreover, since $\ell \geq \frac{s}{\binom{k+1}{j+1}}$, we can find another positive constant α_2 such that

$$q_{\underline{b}} \prod_{i=1}^{s} b_i! \leq n^{-\alpha_2 s}. \tag{3.16}$$

For each sequence $\underline{b} = (b_1, \ldots, b_s)$ define

$$t(\underline{b}) := |\{i : b_i \geq n^{\alpha_2/2}\}|$$

and let B_t be the set of all sequences $\underline{b}$ such that $t(\underline{b}) = t$. We can crudely bound $|B_t|$, the number of sequences in B_t, by

$$s^t \binom{n}{k-j}^t (n^{\alpha_2/2})^{s-t}.$$

On the other hand, if $\underline{b} \in B_t$, then

$$\prod_{i=1}^{s} b_i! \geq \left(\left(n^{\alpha_2/2}\right)!\right)^t \geq \left(\left(n^{\alpha_2/2}\right)^{n^{\alpha_2/3}}\right)^t \geq n^{tn^{\alpha_2/4}}.$$

Summing over all possible sequences $\underline{b}$, we obtain

$$\begin{aligned}
\sum_{\underline{b}} \frac{1}{\prod_{i=1}^{s} b_i!} &= \sum_{t=0}^{s} \sum_{\underline{b} \in B_t} \frac{1}{\prod_{i=1}^{s} b_i!} \\
&\leq \sum_{t=0}^{s} \frac{s^t \binom{n}{k-j}^t (n^{\alpha_2/2})^{s-t}}{n^{tn^{\alpha_2/4}}} \\
&= n^{\alpha_2 s/2} \sum_{t=0}^{s} \left(\frac{s\binom{n}{k-j}}{n^{\alpha_2/2} n^{n^{\alpha_2/4}}} \right)^t \\
&\leq (s+1) n^{\alpha_2 s/2}.
\end{aligned} \tag{3.17}$$

Combining (3.16) and (3.17), the probability that some support of fixed size s exists is at most

$$(s+1)n^{\alpha_2 s/2} n^{-\alpha_2 s} \leq n^{-\alpha_2 s/3}.$$

Let $\bar{d} > \frac{4(k+1)}{\alpha_2}$ be a constant. If we sum over all $s \geq \bar{d}$, we see that the probability that S_p exists and $|S_p| \geq \bar{d}$ is at most $n^{-\alpha_2 \bar{d}/4}$. This holds for every $p \geq p_j^-$ and thus, taking a union bound over all $O(n^{k+1})$ birth times in this range, the probability for S_p of size at least $\bar{d}$ to exist for any $p \geq p_j^-$ is $O\left(n^{k+1-(\alpha_2 \bar{d}/4)}\right)$, which tends to zero for our choice of $\bar{d}$. □

We can now show that whp for p "close" to p_j each j-cocycle in $\mathcal{G}_p$ arises from copies of M_j^-.

Corollary 3.5.8. *For every $p = (1 + o(1))p_j$ with $p \geq p_j^-$, we have $\mathcal{N}_{\mathcal{G}_p} = \emptyset$ with high probability.*

Proof. By Lemma 3.3.1 and the definition of $\mathcal{N}_{\mathcal{G}_p}$ (Definition 3.5.1), whp either S_p does not exist or $|S_p| \geq k - j + 2$. Furthermore, Lemma 3.5.7 tells us that whp for all $p \geq p_j^-$, either S_p does not exist or it must be of constant size. For $p = (1 + o(1))p_j$, Lemma 3.5.5 implies that whp S_p does *not* have constant size, and thus whp S_p does not exist, meaning that whp $\mathcal{N}_{\mathcal{G}_p}$ is empty. □

3.5.5 Monotonicity with high probability

Although the existence of bad functions in $\mathcal{G}_p$ is not intrinsically a monotone property, in this section we show that in fact, from time p_{M_j} on, whp this property behaves in a monotone way.

By Corollary 3.4.11, whp we can apply Corollary 3.5.8 with $p = p_{M_j}$, therefore whp $\mathcal{N}_{\mathcal{G}_{p_{M_j}}}$ is empty. In other words, whp there are no bad functions in $\mathcal{G}_{p_{M_j}}$, i.e. $H^j(\mathcal{G}_{p_{M_j}}; \mathbb{F}_2) = 0$. However, we still need to prove that $\mathcal{G}_p$ does not lose this property for any larger p. More precisely, we already know by Lemma 3.5.7 that whp no $\mathcal{G}_p$ for $p \geq p_{M_j}$ contains a j-cocycle with "large" support, but "small" supports have been excluded by Lemma 3.5.5 only in the range $p = (1 + o(1))p_j$. In the next lemma we show that if a new obstruction appears, then the k-simplex whose birth causes this appearance must be a local obstacle (Definition 3.4.8). But we already know by Lemma 3.4.9 that whp no new local obstacles appear, which will complete the argument.

Lemma 3.5.9. *Whp either $\mathcal{N}_{\mathcal{G}_p} = \emptyset$ for all $p \geq p_{M_j}$ or the k-simplex K with smallest birth time $p_K \geq p_{M_j}$, for which $\mathcal{N}_{\mathcal{G}_{p_K}} \neq \emptyset$, forms a local obstacle in $\mathcal{G}_{p_K}$.*

Proof. The lemma is trivially true if whp $\mathcal{N}_{\mathcal{G}_p} = \emptyset$ for all $p \geq p_{M_j}$, we may thus assume that K exists with positive probability. Let $p < p_K$ be such that $\mathcal{G}_{p_K} = \mathcal{G}_p + K$.

Suppose first that $S_{p_K} \cap \mathcal{G}_p \neq \emptyset$. Let S be a maximal subset of S_{p_K} which is traversable in $\mathcal{G}_p$ and let f be the j-cochain in $\mathcal{G}_p$ with support S. Every k-simplex of $\mathcal{G}_p$ containing some j-simplex in S cannot contain j-simplices in $S_{p_K} \setminus S$ by the maximality of S. Therefore, every $(j+1)$-simplex of $\mathcal{G}_p$ is even with respect to f, because it is even with respect to f_{p_K}. This means that f is a j-cocycle in $\mathcal{G}_p$.

Lemma 3.5.7 implies that there exists a constant $\bar{d}$ such that whp $|S_{p_K}| < \bar{d}$ and thus also $|S| < \bar{d}$. But Lemma 3.4.6, together with the fact that $p > p_{M_j} > p_j^{(1)}$ whp, implies that whp each j-simplex in S lies in linearly many j-shells in $\mathcal{G}_p$, at most $|S| - 1$ of which can contain other elements of S. Thus, whp there are j-shells in $\mathcal{G}_p$ that contain precisely one element of S, which means that f is not a j-coboundary, i.e. f is a bad function in $\mathcal{G}_p$. Now recall that whp there are no copies of M_j^- in $\mathcal{G}_p$ by Corollary 3.4.10 and thus all bad functions lie in $\mathcal{N}_{\mathcal{G}_p}$. This means that $\mathcal{N}_{\mathcal{G}_p} \neq \emptyset$, a contradiction to the choice of K.

Thus, whp S_{p_K} is entirely contained in K and its simplices are not in other k-simplices of $\mathcal{G}_{p_K}$. Moreover, it follows from Lemma 3.3.1 that $|S_{p_K}| \geq k - j + 1$, implying that whp K forms a local obstacle in $\mathcal{G}_{p_K}$. □

The following corollary shows that in the supercritical regime $p \geq p_{M_j}$, whp no j-cocycle arises from copies of M_j^-.

Corollary 3.5.10. *With high probability $\mathcal{N}_{\mathcal{G}_p} = \emptyset$ for all $p \geq p_{M_j}$ simultaneously.*

Proof. Recall that by Corollaries 3.4.11 and 3.5.8, $\mathcal{N}_{\mathcal{G}_{p_{M_j}}} = \emptyset$ whp. If $\mathcal{N}_{\mathcal{G}_p} \neq \emptyset$ for some $p > p_{M_j}$, then whp the k-simplex whose birth creates a j-cocycle that is not generated by copies of M_j^- would form a local obstacle by Lemma 3.5.9. But Lemma 3.4.9 tells us that whp no new local obstacles appear after time $\bar{p}_j$, which whp is smaller than p_{M_j} by (3.12) and Corollary 3.4.11. $\square$

3.6 Proofs of main results

3.6.1 Proof of Theorem 3.1.11

Corollary 3.4.11 states that for any function ω of n which tends to infinity as $n \to \infty$, whp we have

$$\frac{(j+1)\log n + \log\log n - \omega}{(k-j+1)n^{k-j}}(k-j)! < p_{M_j} < \frac{(j+1)\log n + \log\log n + \omega}{(k-j+1)n^{k-j}}(k-j)!,$$

which is precisely Theorem 3.1.11 (i).

To prove (ii), recall that Lemma 3.3.4 states that for all $i \in [j]$, whp $H^i(\mathcal{G}_p;\mathbb{F}_2) \neq 0$ for all $p \in [p_{i-1}^-, p_{M_i})$. By (i), whp for all $i \in [j-1]$

$$p_{M_i} > \left(1 - \frac{1}{\sqrt{\log n}}\right)\frac{(i+1)\log n}{(k-i+1)n^{k-i}}(k-i)! = p_i^-,$$

and thus whp $\mathcal{G}_p$ is not j-cohom-connected throughout $\bigcup_{i=1}^{j}[p_{i-1}^-, p_{M_i}) = [p_0^-, p_{M_j})$.

Now observe that by Lemma 3.4.1 whp $p_T > p_0^-$ and that $\mathcal{G}_p$ is not topologically connected in $[0, p_T)$ by the definition of p_T. Therefore, whp $\mathcal{G}_p$ is not j-cohom-connected in

$$[0, p_{M_j}) = [0, p_T) \cup [p_0^-, p_{M_j}),$$

as required.

It remains to prove (iii). We have to show that whp there are no bad functions in $\mathcal{G}_p$ for every $p \geq p_{M_j}$. By Corollary 3.4.10, whp for all $p \geq p_{M_j}$, there are no copies of M_j^- in $\mathcal{G}_p$. Thus, if $H^j(\mathcal{G}_p;\mathbb{F}_2) \neq 0$, then any representative of a non-zero cohomology class cannot arise from copies of M_j^- and therefore lies in $\mathcal{N}_{\mathcal{G}_p}$ (Definition 3.5.1). But by Corollary 3.5.10, whp each such $\mathcal{N}_{\mathcal{G}_p}$ is empty and thus whp $H^j(\mathcal{G}_p;\mathbb{F}_2) = 0$ for all $p \geq p_{M_j}$. Analogously, whp all cohomology groups $H^i(\mathcal{G}_p;\mathbb{F}_2)$ for $i \in [j-1]$ vanish, because whp $p_{M_i} < p_{M_j}$ by (i). Finally, by (i) and Lemma 3.4.1 whp $p_T < p_{M_j}$, meaning that whp $\mathcal{G}_p$ is topologically connected for all $p \geq p_{M_j}$. This implies that whp each such $\mathcal{G}_p$ is $\mathbb{F}_2$-cohomologically j-connected. $\square$

3.6.2 Proof of Corollary 3.1.12

Let ω be any function of n which tends to infinity as $n \to \infty$. It is known (see e.g. [53]) that whp

$$\frac{k\log n - \omega}{n} < p_{\mathrm{isol}} < \frac{k\log n + \omega}{n}. \tag{3.18}$$

The proof is an easy application of the first and second moment methods.

In order to prove that $p_{\mathrm{conn}} = p_{\mathrm{isol}}$ whp, suppose that a $(k-1)$-simplex σ is isolated in $\mathcal{Y}_p$ for some p. The indicator function f_σ of σ is a $(k-1)$-cocycle, because σ is isolated. But f_σ is *not* a $(k-1)$-coboundary, because σ lies in $(n-k$ many$)$ $(k-1)$-shells. In particular, $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) \neq 0$. By the definitions of p_{conn} and p_{isol}, this implies that $p_{\mathrm{conn}} \geq p_{\mathrm{isol}}$.

For the opposite direction, fix the birth times of all k-simplices. Then for all $p \geq p_{\mathrm{isol}}$, we have $\mathcal{Y}_p = \mathcal{G}_p$ and therefore $\mathcal{Y}_p$ is $\mathbb{F}_2$-cohomologically $(k-1)$-connected whp for every $p \geq \max(p_{\mathrm{isol}}, p_{M_{k-1}})$ by Theorem 3.1.11 (iii). By (3.18) and Theorem 3.1.11 (i), whp for any (slowly) growing function ω

$$p_{\mathrm{isol}} > \frac{k \log n - \omega}{n} > \frac{k \log n + \log \log n + \omega}{2n} > p_{M_{k-1}},$$

hence whp for all $p \geq p_{\mathrm{isol}}$ we have $H^{k-1}(\mathcal{Y}_p; \mathbb{F}_2) = H^{k-1}(\mathcal{G}_p; \mathbb{F}_2) = 0$. This means that whp $p_{\mathrm{conn}} \leq p_{\mathrm{isol}}$ and thus $p_{\mathrm{conn}} = p_{\mathrm{isol}}$, as required. $\square$

3.6.3 Proof of Theorem 3.1.13

We are interested in the asymptotic distribution of $D_j := \dim\left(H^j(\mathcal{G}_p; \mathbb{F}_2)\right)$ for

$$p = \frac{(j+1)\log n + \log \log n + c_n}{(k-j+1)n^{k-j}}(k-j)!,$$

where $c_n \xrightarrow{n \to \infty} c \in \mathbb{R}$.

Recall that X_- is the random variable defined in Lemma 3.4.7 which counts the number of copies of M_j^-. We apply the method of moments (Lemma 3.2.3) to X_-, showing that it converges in distribution to a Poisson random variable with expectation

$$\lambda_j = \frac{(j+1)e^{-c}}{(k-j+1)^2 j!}.$$

Subsequently, we will prove that whp $X_- = D_j$. In particular this will imply that

$$D_j \xrightarrow{d} \mathrm{Po}(\lambda_j),$$

as required.

In order to determine the expectation of X_-, let $K \subset [n]$ be a $(k+1)$-set and let C be a j-subset of K. Recall that the probability that a (potential) petal $C \cup \{w\}$ with $w \in K \setminus C$ lies in no other k-simplex is given by

$$r = (1-p)^{\binom{n-j-1}{k-j}-1}$$

(see (3.6)). Arguing as in Lemma 3.4.7, we see that dependencies between the petals are negligible and thus

$$\mathbb{E}(X_-) = (1+o(1))\binom{n}{k+1}\binom{k+1}{j} p r^{k-j+1}. \tag{3.19}$$

We observe that

$$\begin{aligned}
r^{k-j+1} &= (1-p)^{\left(\binom{n-j-1}{k-j}-1\right)(k-j+1)} \\
&= \exp\left(-\frac{n^{k-j}}{(k-j)!}(k-j+1)p + O\left(n^{k-j-1}p\right) + O\left(n^{k-j}p^2\right)\right) \\
&= \exp\left(-(j+1)\log n - \log\log n - c_n + o(1)\right) \\
&= (1+o(1))\frac{e^{-c_n}}{n^{j+1}\log n}. \qquad (3.20)
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}(X_-) &= (1+o(1))\frac{n^{k+1}}{(k-j+1)!j!}\cdot\frac{(j+1)\log n + \log\log n + c_n}{(k-j+1)n^{k+1}\log n}(k-j)!e^{-c_n} \\
&= (1+o(1))\frac{(j+1)e^{-c_n}}{(k-j+1)^2 j!} \overset{c_n\to c}{=} (1+o(1))\lambda_j. \qquad (3.21)
\end{aligned}$$

Denote by $\mathcal{T}^-$ the set of all pairs (K, C) that can form a copy of M_j^- in $\mathcal{G}_p$. For each $T^- \in \mathcal{T}^-$, denote by X_{T^-} the indicator random variable of the event that T^- forms a copy of M_j^- in $\mathcal{G}_p$. For each fixed integer $t \geq 1$, we now determine the binomial moments

$$\mathbb{E}\binom{X_-}{t} = \sum_{\mathcal{S}\in\binom{\mathcal{T}^-}{t}} \Pr\left(\bigcap_{T^-\in\mathcal{S}}\{X_{T^-}=1\}\right).$$

Suppose first that all $T^- \in \mathcal{S}$ have different $(k+1)$-sets. In this case, if all $T^- \in \mathcal{S}$ form copies of M_j^-, none of the petals are shared (by property (M2) of M_j^-, see Definition 3.3.3). If we choose t distinct $(k+1)$-sets uniformly at random, whp they will be disjoint and in particular no two $T_1^-, T_2^- \in \mathcal{S}$ will share a petal. To choose t distinct $(k+1)$-sets, there are

$$\binom{\binom{n}{k+1}}{t} = (1+o(1))\frac{\binom{n}{k+1}^t}{t!}$$

choices. Therefore, the contribution to $\mathbb{E}\binom{X_-}{t}$ made by the sets $\mathcal{S}$ for which all $T^- \in \mathcal{S}$ have distinct $(k+1)$-set is

$$\begin{aligned}
(1+o(1))\binom{\binom{n}{k+1}}{t}\binom{k+1}{j}^t p^t r^{t(k-j+1)} &\overset{(3.19)}{=} (1+o(1))\frac{\mathbb{E}(X_-)^t}{t!} \\
&\overset{(3.21)}{=} (1+o(1))\frac{\lambda_j{}^t}{t!}, \qquad (3.22)
\end{aligned}$$

which is the desired asymptotic value.

We now show that the contribution coming from sets $\mathcal{S}$ whose elements use $u < t$ different $(k+1)$-sets is negligible. We have $\binom{\binom{n}{k+1}}{u}$ ways to select the $(k+1)$-sets and at most $u^{t-u}\binom{k+1}{j}^t$ different ways to locate the t potential M_j^- in them. Moreover, observe that two different copies of M_j^- in the same k-simplex share at most one petal (otherwise they would have the same centre and thus be identical) and in that case these two copies have $(k-j+1)+(k-j)$ petals in total. This means that each of the u many $(k+1)$-sets contains at least $k-j+1$ petals, and at least one $(k+1)$-set contains at least $(k-j+1)+(k-j)$ petals. Therefore the total number of petals required for such a set $\mathcal{S}$ is bounded from below by $u(k-j+1)+(k-j)$. In total, the contribution of such sets $\mathcal{S}$ to the binomial moment is at most

$$\binom{\binom{n}{k+1}}{u} u^{t-u} \binom{k+1}{j}^t p^u r^{u(k-j+1)} r^{k-j}.$$

Replacing t by u in (3.22), we deduce that

$$\binom{\binom{n}{k+1}}{u} u^{t-u} \binom{k+1}{j}^t p^u r^{u(k-j+1)} = (1+o(1))\frac{\lambda_j{}^u}{u!}\left(u\binom{k+1}{j}\right)^{t-u} = \Theta(1).$$

Furthermore, (3.20) yields $r^{k-j} = o(1)$. Together with (3.22), we deduce that

$$\mathbb{E}\binom{X_-}{t} = (1+o(1))\frac{\lambda_j^t}{t!}$$

for each fixed integer $t \geq 1$. Now Lemma 3.2.3 yields $X_- \xrightarrow{d} \mathrm{Po}(\lambda_j)$.

It remains to show that $X_- = D_j$ whp. To this end, denote by $f_1, \dots, f_{X_-}$ the j-cocycles arising from the copies of M_j^- in $\mathcal{G}_p$. Corollary 3.5.8 in particular implies that whp the cohomology classes of $f_1, \dots, f_{X_-}$ generate $H^j(\mathcal{G}_p; \mathbb{F}_2)$, which means that $X_- \geq D_j$.

In order to prove the opposite direction, we show that the cohomology classes of $f_1, \dots, f_{X_-}$ are linearly independent. Observe first that whp $X_- = o(n)$ by Markov's inequality, because X_- has bounded expectation. Let $I \subseteq [X_-]$ be non-empty and let S be the support of $\sum_{i \in I} f_i$. By the arguments above for $t = 2$ and $u = 1$, whp there are no two M_j^- that share the same k-simplex. Thus, whp the f_i's have disjoint support by property (M2) of an M_j^- (Definition 3.3.3), and in particular $S \neq \emptyset$. Pick $L \in S$. Lemma 3.4.6 and the fact that $p > p_j^{(1)}$ tell us that whp there are $\Theta(n)$ many j-shells in $\mathcal{G}_p$ that contain L. All these j-shells meet only in L, thus at most $|S| \leq (k-j+1)|I| = o(n)$ of them can contain another j-simplex in S. Thus, there are j-shells that meet S only in L, showing that $\sum_{i \in I} f_i$ is not a j-coboundary. Therefore the cohomology classes of $f_1, \dots, f_{X_-}$ are linearly independent whp. This shows that whp $X_- \leq D_j$ and thus $X_- = D_j$, as desired.

Together with $X_- \xrightarrow{d} \mathrm{Po}(\lambda_j)$, this proves that $D_j \xrightarrow{d} \mathrm{Po}(\lambda_j)$. By Theorem 3.1.11 (for $j-1$ instead of j) whp $H^0(\mathcal{G}_p; \mathbb{F}_2) = \mathbb{F}_2$ and $H^i(\mathcal{G}_p; \mathbb{F}_2) = 0$ for all $i \in [j-1]$. In particular,

$$\begin{aligned}
\Pr(\mathcal{G}_p \text{ is } j\text{-cohom-connected}) &= \Pr\left(H^j(\mathcal{G}_p; \mathbb{F}_2) = 0\right) + o(1) \\
&= (1+o(1)) \Pr\left(\mathrm{Po}(\lambda_j) = 0\right) \\
&= (1+o(1)) e^{-\lambda_j}.
\end{aligned}$$

This concludes the proof of Theorem 3.1.13. $\square$

3.7 Concluding remarks

3.7.1 Comparison of proof methods

Let us note that for the subcritical regime (Theorem 3.1.11 (ii)), one might try to use a different approach in order to prove that $H^j(\mathcal{G}_p; \mathbb{F}_2)$ does not vanish in the interval $[p_{j-1}^-, p_{M_j})$. If the dimension of $C^j(\mathcal{G}_p)$ (viewed as an $\mathbb{F}_2$-vector space) is larger than the sum of the dimensions of $C^{j-1}(\mathcal{G}_p)$ and $C^{j+1}(\mathcal{G}_p)$, then $H^j(\mathcal{G}_p; \mathbb{F}_2) \neq 0$ would follow. However, this behaviour only happens for "small" $p \in [p_{j-1}^-, p_{M_j})$ and, more importantly, only for $j \geq \frac{k-1}{2}$. In contrast, our proof method works for all values of j. Moreover, our result that $[p_{j-1}^-, p_{M_j})$ whp is covered by three copies of M_j (Lemma 3.3.4), together with the fact that whp $\mathcal{G}_{p_0^-}$ has isolated vertices (this can be proved using an easy second moment argument), implies the following slightly stronger statement.

Proposition 3.7.1. *With high probability for every $p < p_{M_j}$, the complex $\mathcal{G}_p$ contains an isolated vertex or a copy of M_i for some $i \in [j]$.*

In the supercritical regime (Theorem 3.1.11 (iii)), the counting methods used in [48, 53] for $\mathcal{Y}_p$ are *not* sufficient to prove the non-existence of j-cocycles in $\mathcal{G}_p$. This is due to the fact that these methods have been designed for the special case $j = k - 1$ and for a threshold which is about twice as large as p_{k-1}. For this reason, the more careful arguments used in Lemmas 3.5.4 to 3.5.9 become necessary.

3.7.2 Alternative models

There are several ways to define random k-complexes. If the k-simplices are chosen independently with probability p, then the models $\mathcal{Y}_p$ and $\mathcal{G}_p$ are somewhat extremal constructions, in the sense that $\mathcal{Y}_p$ contains *all* simplices of lower dimension, while $\mathcal{G}_p$ only comprises those simplices that are necessary in order to be a complex. What happens in between, i.e. when the complex contains all simplices in $\mathcal{G}_p$, but in addition, some simplices of dimensions $1, \dots, k - 1$ might be added in a random fashion? Depending on the choice of probabilities, such a complex might show behaviour that is different from both $\mathcal{Y}_p$ and $\mathcal{G}_p$.

Random complexes also arise naturally from random graphs. For instance, the random *clique complex* $\mathcal{X}_p(n)$ (also known as *flag complex*) on vertex set $[n]$ can be defined as the maximal complex whose 1-skeleton is the binomial random graph. Equivalently, a non-empty set $U \subseteq [n]$ forms a simplex in $\mathcal{X}_p(n)$ if and only if U is a clique in the binomial random graph. Topological properties of $\mathcal{X}_p(n)$ have been studied in [24, 41, 42]. Another example is the random *neighbourhood complex* arising from the binomial random graph by letting each non-empty set of vertices that have a common neighbour form a simplex [40]. See [43] for an overview of these and other models of random complexes.

3.7.3 Other notions of connectedness

The vanishing of cohomology groups with coefficients in $\mathbb{F}_2$ is just one possible way of defining the concept of "connectedness" of $\mathcal{G}_p$. An obvious alternative would be to consider coefficients from other groups or fields. For $\mathcal{Y}_p$, such notions of connectedness have been studied for coefficients in any finite abelian group, in $\mathbb{Z}$, or in any field [3,4,36,49,51,53]. In particular, the threshold for the vanishing of $H^{k-1}(\mathcal{Y}_p; R)$ for a finite abelian group R is independent of the choice of R [53].

For $\mathcal{G}_p$, it is not obvious whether the threshold for j-cohom-connectedness depends on the choice of the group of coefficients. An indication that the threshold might in fact be independent of the group is the observation that M_j remains the minimal obstruction when the coefficients are taken from any abelian group. It is possible that adapting the arguments in this paper would prove that the results remain true for cohomology groups with any abelian group of coefficients replacing $\mathbb{F}_2$.

A rather strong notion of connectedness would be to require the homotopy groups $\pi_1(\mathcal{G}_p), \ldots, \pi_j(\mathcal{G}_p)$ to vanish. For the 2-dimensional case, the vanishing of $\pi_1(\mathcal{Y}_p)$ was studied by Babson, Hoffman and Kahle [5]. In particular, they showed that whp $\pi_1(\mathcal{Y}_p) \neq 0$ at the time that $H^1(\mathcal{Y}_p; \mathbb{F}_2)$ becomes zero. From that time on, the models $\mathcal{Y}_p$ and $\mathcal{G}_p$ coincide. As $\pi_1(\mathcal{G}_p) \neq 0$ follows immediately from $H^1(\mathcal{G}_p; \mathbb{F}_2) \neq 0$, the range that should be of particular interest with respect to $\pi_1(\mathcal{G}_p)$ in the 2-dimensional case is

$$\frac{\log n + \frac{1}{2}\log\log n}{n} \leq p \leq \frac{2\log n + \omega}{n}.$$

A natural conjecture would be that whp $\pi_1(\mathcal{G}_p) \neq 0$ in this range.

3.7.4 Appearance of (co-)homology groups

Theorem 3.1.13 provides a limit result for the dimension $D_j = \dim(H^j(\mathcal{G}_p; \mathbb{F}_2))$ of the j-th cohomology group of $\mathcal{G}_p$ around the point of the phase transition. It would be interesting to know the behaviour of D_j also for earlier regimes. More precisely, how large is D_j in the interval $[p_{j-1}^-, p_{M_j})$? How far below p_{j-1}^- do we have $D_j > 0$ whp?

One might also ask when the j-th *homology group* first becomes non-trivial. If we consider the homology group over $\mathbb{F}_2$, this is equivalent to the j-th cohomology group becoming non-trivial, but with more general coefficients, this may no longer be true. In the case $j = k$, the k-th homology group of $\mathcal{G}_p$ is the same as the k-th homology group of $\mathcal{Y}_p$, and the threshold for the appearance of this homology group over $\mathbb{R}$ was determined by Linial and Peled [49].

Acknowledgements

The authors thank Penny Haxell for very helpful discussions regarding the analysis of the two-dimensional case. An extended abstract of the case $k = 2$ has appeared in the Proceedings of Eurocomb 2017 [18].

We would also like to thank the anonymous referees for their helpful suggestions.

4 Cohomology groups of non-uniform random simplicial complexes

4.1 Introduction

4.1.1 Motivation

One of the first and most famous results in the theory of random graphs, due to Erdős and Rényi [28], 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. Almost equivalenty, in modern terminology, with high probability the binomial random graph $G(n,p)$ becomes connected around $p = \frac{\log n}{n}$ (see [60]).

The result was subsequently strengthened by Bollobás and Thomason [15] 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 the notion of vertex-connectedness (see e.g. [7,8,13,14,25,45,57,59]) or of high-order connectedness (also known as j-tuple-connectedness, e.g. [19,20,22,44]).

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 [48], 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 [44].

Meshulam and Wallach [53] extended the result of [48] to random simplicial k-complexes with full $(k-1)$-skeleton (for any $k \geq 2$), 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 [17], 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).

Since then, many different models of random simplicial complexes have been introduced (see e.g. [24, 40–43]), as well as several notions of connectedness have been analysed (see e.g. [3, 4, 36, 51]). In this paper, we consider a model of random simplicial complexes generated from *non-uniform* random hypergraphs, and for 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 [17] as special cases, and therefore our main result extends and unifies the results of [17], [48], and [53]. We also note that our model is equivalent to the 'upper model' which was recently introduced independently in [32], although [32] considers different ranges of probabilities and different properties to the ones we focus on in this paper.

4.1.2 Model

Throughout the paper let $d \geq 2$ 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 its neutral element by 0_R. For an integer $k \geq 1$, we write $[k] := \{1, \ldots, k\}$ and $[k]_0 := \{0, \ldots, k\}$. 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 4.1.1. 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 random *non-uniform* hypergraph, in which sets of vertices have different probabilities of forming an edge depending on their size.

Definition 4.1.2. For each $k \in [d]$, let $p_k = p_k(n) \in [0,1] \subset \mathbb{R}$ be given and write $\mathbf{p} := (p_1, \ldots, p_d)$. Denote by $G(n, \mathbf{p})$ the binomial (non-uniform) 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 $G(n, \mathbf{p})$.

[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.

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 those are not already in the downward-closure).[2]

Denote by $H^i(\mathcal{G};R)$ the i-th cohomology group of a simplicial complex $\mathcal{G}$ with coefficients in R (see (4.5) in Section 4.2.2 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. [55, 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 $i \geq 1$, the *vanishing* of $H^i(\mathcal{G};R)$ can be viewed as a 'higher-order connectedness' of $\mathcal{G}$.

Definition 4.1.3. Given a non-negative integer j, a simplicial complex $\mathcal{G}$ is called *R-cohomologically j-connected* (*j-cohom-connected* for short) if

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

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

We note that the analogous definition of connectedness considered by Meshulam and Wallach in [53] was only for the case $j = d-1$, and only demanded the vanishing of the $(d-1)$-th cohomology group—this was reasonable since with the complete $(d-1)$-dimensional skeleton, the i-th cohomology group must always vanish for all $i \leq d-2$ (or equal R if $i = 0$).

4.1.3 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 $j \leq k \leq d$ (these will be defined later, see Definitions 4.4.2 and 4.4.3), and we will see in Section 4.4 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 4.4.9), which is 'minimal' in a natural sense (Lemma 4.4.10).

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 [15] 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.

[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.

If we fix a 'direction' $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$ 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 $\bar{\mathbf{p}}$ 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 $\mathcal{G}_{\tau_0}$. Therefore we will often give definitions or state and prove results for the random complex $\mathcal{G}_\tau$ for some appropriate value of τ, and subsequently apply them to the process at that time, meaning in particular that $\mathcal{G}_{\tau_0} \subset \mathcal{G}_{\tau_1}$ if $\tau_0 \leq \tau_1$, 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 (4.1) in Section 4.2 for the formal definition of scaled birth time).

Note that the evolution of the process $(\mathcal{G}_\tau)$ is unchanged if the direction $\bar{\mathbf{p}}$ 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 (Theorem 4.1.4) in particular states that this happens for a specific type of direction that we call *j-critical* and that will be formally defined in Section 4.2.1 (Definition 4.2.3).

Theorem 4.1.4 (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 $\mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$, let $\tau_{\max} = 1/\bar{p}_d$, and let*

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

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

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

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

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

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

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

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

Observe that in Theorem 4.1.4 we do not consider j-cohom-connectedness for the case $j = 0$. Indeed, the condition $H^0(\mathcal{G}_\tau; R) = R$ corresponds to the topological connectedness of $\mathcal{G}_\tau$, i.e. vertex-connectedness of the underlying (non-uniform) random hypergraph, that has been extensively studied, and for which much stronger results are known (see e.g. [19, 57]). However, topological connectedness is a necessary condition for the j-cohom-connectedness of $\mathcal{G}_\tau$ (see Definition 4.1.3), therefore in order to make this paper self-contained, this case is treated separately in Lemma 4.3.2.

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 4.4.11), 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 4.1.4 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 4.1.5 (Rank in the critical window)**.** *Let $c \in \mathbb{R}$ be a constant and suppose that $(c_n)_{n\geq 1}$ is a sequence of real numbers with $c_n \xrightarrow{n\to\infty} c$. Let $j \in [d-1]$, $\tau = 1+\frac{c_n}{\log n}$, 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*

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

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

The constant $\mathcal{E}$ will be explicitly defined in (4.4).

4.1.4 Proof techniques

The three statements of the Hitting Time Theorem (Theorem 4.1.4) follow from the auxiliary results presented in Section 4.3, which in turn are proved throughout the paper.

We show in Lemma 4.3.1 that the choice of a j-critical direction $\bar{\mathbf{p}}$ (Definition 4.2.3) implies that the last minimal obstruction disappears at around time $\tau = 1$, thus proving statement (i) of Theorem 4.1.4.

The main ingredient in the proof of Theorem 4.1.4 (ii) will be Lemma 4.3.4, 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 4.7 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 $j \leq k \leq d$ (Lemmas 4.7.1, 4.7.3 and 4.7.4), and thus $H^j(\mathcal{G}_\tau; R) \neq 0$ (Corollary 4.4.9). In addition, we show that there exists an appropriate scaling factor τ such that the vector $\tau\bar{\mathbf{p}}$ is an i-critical direction, for every $i \in [j-1]$ (Lemma 4.3.5). Thus we can apply Lemma 4.3.4

with j replaced by i and find intervals $I_i(\varepsilon)$ where whp $H^i(\mathcal{G}_\tau; R) \neq 0$ (Corollary 4.3.6). 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 4.3.2). In this way we can complete the argument, 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 $\tau \geq \tau_j^*$ there are no copies of the minimal obstruction $\hat{M}_{j,k}$, thus in order to prove statement (iii) of Theorem 4.1.4 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 4.3.7, which we prove in Section 4.8. We show that the smallest support of any non-zero element of the cohomology group must be *traversable* (Lemma 4.8.4), 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 4.8.5 and 4.8.7).

To prove the Rank Theorem (Theorem 4.1.5), in Section 4.9 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 4.8.8) and that indeed they are a *minimal* set of generators for $H^j(\mathcal{G}_\tau; R)$. We conclude by showing that the number of such obstructions converges in distribution to a Poisson random variable (Lemma 4.9.1).

4.1.5 Outline of the paper

The paper is structured as follows.

In Section 4.2, we introduce some preliminary concepts regarding the parametrisation of a j-critical direction, as well as some standard concepts of cohomology theory.

Section 4.3 contains the main auxiliary results that we combine to prove the Hitting Time Theorem (Theorem 4.1.4).

The proofs of the auxiliary results of Section 4.3 will follow in Sections 4.4–4.8. In particular, the results of Section 4.8 will also lay the foundation of the proof of the Rank Theorem (Theorem 4.1.5), which is presented in Section 4.9.

In Section 4.10 we explain in more detail why with the choice of a j-critical direction, Theorems 4.1.4 and 4.1.5 cover all interesting cases. The standard but technical proofs of many auxiliary results are included in Section 4.11, for completeness.

Finally, in Section 4.12 we discuss our main results and present some open problems.

4.2 Preliminaries

4.2.1 Parametrisation

In this section we will define the concept of j-critical direction, which appears in Theorems 4.1.4 and 4.1.5.

Given a direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, let k 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

$$\tau_K := \frac{t_K}{\bar{p}_k}. \tag{4.1}$$

(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[4].

The motivation of the following definitions will become apparent later (Lemma 4.4.13 and Section 4.10).

Definition 4.2.1. Given $j \in [d-1]$, a vector $\bar{\mathbf{p}} = (\bar{p}_1, \dots, \bar{p}_d)$ is called *j-admissible* if for each $1 \le k \le d$ there are real-valued constants $\bar{\alpha}_k, \bar{\gamma}_k$, and a function $\bar{\beta}_k = \bar{\beta}_k(n)$ such that

$$\bar{p}_k = \frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)!, \tag{4.2}$$

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 $j+1 \le k_0 \le d$ 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 4.2.2. If $\bar{\mathbf{p}}$ is j-admissible and $k \ge j+1$, 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 $p_k \le 1$.

This observation means that, for $k \ge j+1$ 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 4.2.1 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 4.10) 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 $\bar{\mathbf{p}} = (\bar{p}_1, \dots, \bar{p}_d)$, for every index

[4]With probability 1 no two simplices have the same birth time.

k with $j \le k \le d$ and $\bar{p}_k \neq 0$ we define the parameters

$$\begin{aligned}
\bar{\lambda}_k &:= j+1-\bar{\gamma}_k-(k-j+1)\sum\nolimits_{i=j+1}^{d}\bar{\alpha}_i,\\
\bar{\mu}_k &:= -(k-j+1)\sum\nolimits_{i=j+1}^{d}\frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}}+\begin{cases}0 & \text{if } \bar{p}_k>1,\\ \log\log n & \text{if } \bar{p}_k\le 1 \text{ and } \bar{\alpha}_k\neq 0,\\ \log(\bar{\beta}_k) & \text{if } \bar{p}_k\le 1 \text{ and } \bar{\alpha}_k=0,\end{cases}\\
\bar{\nu}_k &:= \begin{cases}-\log((j+1)!) & \text{if } k=j,\\ -\log(j!)-\log(k-j+1)+\log(\bar{\alpha}_k) & \text{if } \bar{\alpha}_k\neq 0,\\ -\log(j!)-\log(k-j+1) & \text{otherwise.}\end{cases}
\end{aligned} \tag{4.3}$$

Note that all $\bar{\lambda}_k$, $\bar{\nu}_k$ are constants (since the $\bar{\alpha}_i$ are constants), while the $\bar{\mu}_k$ are functions of n, with $\bar{\mu}_k = o(\log n)$ by Definition 4.2.1.

Definition 4.2.3. We say that a j-admissible vector $\bar{\mathbf{p}}$ is a j-*critical direction* if:

(C1) $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k \le 0$, for all indices k with $j \le k \le d$ 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 $j \le \bar{k} \le d$.

More generally, if we have a vector $\mathbf{p} = (p_1, \dots, p_d)$ (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 4.2.4. Given a vector $\mathbf{p} = (p_1, \dots, p_d)$, for each $j \le k \le d$, define

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

Furthermore, we define the parameters λ_k, μ_k, and ν_k analogously to (4.3), 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 4.2.5. 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 $1 \le k \le d$, at least one of α_k, γ_k is zero and neither of them is negative.

In order to prove Theorem 4.1.5, 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 4.1.5). We also need the following concepts.

Definition 4.2.6. Given a j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1, \ldots, \bar{p}_d)$, an index k with $j \leq k \leq d$ 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 $\bar{\mathbf{p}}$.

For any $\tau = 1 + O(1/\log n)$, the critical dimensions are precisely those indices k for which there is a positive probability of having copies of $\hat{M}_{j,k}$ 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 $\mathcal{E}$ appeared in Theorem 4.1.5 is precisely

$$\mathcal{E} := \exp(-c(j+1)) \sum_{k \in \mathcal{C}} \exp(\bar{\mu}_k + \bar{\nu}_k + c\gamma_k), \tag{4.4}$$

as we will define in the proof of Theorem 4.1.5 (Section 4.9). 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 $\hat{M}_{j,k}$ (Corollary 4.5.8).

4.2.2 Cohomology

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

Let $j \in [d]_0$. To define cohomology groups, one considers *ordered j-simplices*, that is, j-simplices with an ordering of their vertices.[5] We adopt the notation $[v_0, \ldots, v_j]$ for a j-simplex whose vertices are ordered $v_0, \ldots, v_j$. If $\sigma = [v_0, \ldots, v_j]$ is an ordered j-simplex and $i \in [j]_0$, then $[v_0, \ldots, \hat{v}_i, \ldots, v_j]$ denotes the ordered $(j-1)$-simplex obtained from σ by removing v_i (and preserving the order on the remaining vertices).

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* $\operatorname{supp}(f)$ 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

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

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. Thus, we can define the *j-th cohomology group of $\mathcal{G}$ with coefficients in R* as the quotient group

$$H^j(\mathcal{G}; R) := \ker \delta^j / \operatorname{im} \delta^{j-1}. \tag{4.5}$$

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

4.2.3 Non-vanishing of cohomology groups

In view of Theorems 4.1.4 and 4.1.5, 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 4.2.8. To this end, we give the following necessary definition.

Definition 4.2.7. 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 $\mathcal{G}$.

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 4.2.8. *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*

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

(ii) 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 (i). 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 (ii), this means that

$$f([v_0, \ldots, \hat{v}_i, \ldots, v_{j+1}]) = 0_R \quad \Leftrightarrow \quad i \neq 0.$$

Thus we have

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

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 $\mathcal{G}$. $\square$

4.3 Hitting Time Theorem: proof of Theorem 4.1.4

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 4.1.4). These auxiliary results are proved throughout the rest of the paper.

4.3.1 Hitting time and subcritical case

To prove Theorem 4.1.4 (i), recall that τ_j^* is the birth time of the simplex whose appearance causes the last copy of $\hat{M}_{j,k}$ for any $j \leq k \leq d$ to disappear. We want to show that this happens at around time $\tau = 1$. More precisely, we will prove the following.

Lemma 4.3.1. *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 (i) of Theorem 4.1.4 will be a corollary of Lemma 4.3.1, which is proved in Section 4.6. Indeed, we will prove a slightly stronger result (Lemma 4.6.1).

For the subcritical case (i.e. statement (ii)) of Theorem 4.1.4, we first determine the threshold for the topological connectedness of $\mathcal{G}(n, \mathbf{p})$, i.e. when $H^0(\mathcal{G}(n, \mathbf{p}); R) = R$.

Lemma 4.3.2. *There exist positive constants $c^- = c^-(d)$ and $c^+ = c^+(d)$ such that*

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

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

The proof of Lemma 4.3.2 (i) consists of an easy application of the second moment method, while Lemma 4.3.2 (ii) follows from [17, Lemma 4.1]. For completeness, we include the proof of both parts of Lemma 4.3.2 in Section 4.11.1.

Remark 4.3.3. In fact, with a slightly more careful extension of the argument, one could strengthen Lemma 4.3.2 to give the exact threshold for the non-uniform case. 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 proof of this stronger statement, which is a standard generalisation of the graph case.

In particular, Lemma 4.3.2 will imply that we can choose a positive constant ε such that the process $(\mathcal{G}_\tau)$ is whp 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 4.7, will be key.

Lemma 4.3.4. *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 $\tau \in I_j(\varepsilon)$.

In particular, we will show that for any $\tau \in I_j(\varepsilon)$ whp there is an index k with $j \leq k \leq d$ such that a copy of the minimal obstruction $\hat{M}_{j,k}$ exists in $\mathcal{G}_\tau$ (Lemmas 4.7.1, 4.7.3, and 4.7.4).

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 4.3.4 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 4.3.5. *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 4.3.5 is intuitively obvious, its proof is rather technical. We therefore delay the proof of Lemma 4.3.5 until Section 4.11.2.

Using Lemma 4.3.5, we can consider for general $i \in [j]$ 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 4.1.4 and $\mathcal{G}_\tau = \mathcal{G}(n, \tau\bar{\mathbf{p}})$ and for each $i \in [j]$, let

$$\tau_i^* := \sup\{\tau \in \mathbb{R}_{\geq 0} \mid \mathcal{G}_\tau \text{ contains a copy of } \hat{M}_{i,k} \text{ for some } k \text{ with } i \leq k \leq d\}. \tag{4.6}$$

Observe that for $i = j$, this hitting time matches with the definition of τ_j^* in Theorem 4.1.4. We derive the following result from Lemmas 4.3.1, 4.3.4, and 4.3.5.

Corollary 4.3.6. *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

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

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

Indeed, statement (ii) of Corollary 4.3.6 also holds for $i = j$, with $\eta = \eta_j = 1$ (Lemma 4.3.1).

Proof. If $i = j$, statement (i) is given by Lemma 4.3.4.

For any $i \in [j-1]$, by Lemma 4.3.5 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 4.3.4 and 4.3.1 applied with j replaced by i, we obtain (i) and (ii), respectively. $\square$

4.3.2 Supercritical case

In Theorem 4.1.4 (iii) 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 $j \leq k \leq d$, but we also have to exclude other type of obstructions. In Section 4.8 we indeed prove the following.

Lemma 4.3.7. *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 4.3.7 also holds when j is replaced by any $i \in [j-1]$.

4.3.3 Proof of Theorem 4.1.4

We now apply the auxiliary results of Sections 4.3.1 and 4.3.2 to prove the Hitting Time Theorem (Theorem 4.1.4).

Proof. (i) 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 4.3.1 with any function ω' which tends to infinity but satisfies $\omega' = o(\omega)$, e.g. picking $\omega' = \sqrt{\omega}$ will suffice.
(ii) For $\varepsilon > 0$, define

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

By definition of j-admissibility (Definition 4.2.1), for every $k \in [d]$ we have $\bar{p}_k = O(\log n/n^{k-j})$. Thus, by Lemma 4.3.2 we can choose ε small enough such that whp

$$\text{for every } \tau \in I_0(\varepsilon), \qquad H^0(\mathcal{G}_\tau; R) \neq R. \tag{4.7}$$

Now consider the intervals $I_i(\varepsilon)$. By Corollary 4.3.6 (ii), we can choose ε small enough (namely $\varepsilon < \eta_i$ for every $i \in [j]$) such that for each i

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

and thus

$$\bigcup\nolimits_{i=0}^{j} I_i(\varepsilon) = [0, \tau_j^*). \tag{4.8}$$

By Corollary 4.3.6 (i), for $\varepsilon > 0$ whp $H^i(\mathcal{G}_\tau; R) \neq 0$ for every $\tau \in I_i(\varepsilon)$. Thus, by choosing ε such that conditions (4.7) and (4.8) hold simultaneously, whp the process $(\mathcal{G}_\tau)$ is not j-cohom-connected for all $\tau \in [0, \tau_j^*)$, as required.
(iii) Recalling that $\bar{p}_k = O(\log n/n^{k-j})$ for any $k \in [d]$ by Definition 4.2.1, Lemma 4.3.2 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 4.3.1.

Furthermore, by Lemma 4.3.7 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 4.3.6 (ii). Thus, whp the process $(\mathcal{G}_\tau)$ is j-cohom-connected for every $\tau \in [\tau_j^*, \tau_{\max}]$, as required. □

4.4 Minimal obstructions

In this section we define copies of $\hat{M}_{j,k}$ (Definitions 4.4.2 and 4.4.3) 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 4.4.1. Let k be an integer with $j \le k \le d$. 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 4.1) 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 *flower*.

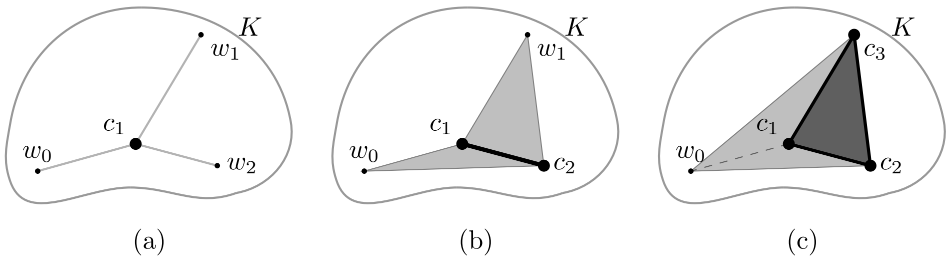


Figure 4.1: Examples of j-flowers in a k-simplex K, for $k = 3$ and $j = 1, 2, 3$.
(a) The 1-flower in K with centre $C = \{c_1\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1, 2$ (grey).
(b) The 2-flower in K with centre $C = \{c_1, c_2\}$ (bold black) and petals $P_i = C \cup \{w_i\}$, $i = 0, 1$ (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 $C \subseteq K$, there is a unique j-flower in K with centre C, namely

$$\mathcal{F}(K, C) := \{C \cup \{w\} \mid w \in K \setminus C\}. \tag{4.9}$$

Note that if $k = j$, then any choice of a centre $C \subseteq K$ produces the same flower $\mathcal{F}(K, C) = \big\{C \cup \{K \setminus C\}\big\} = \{K\}$, with the set K itself as unique petal.

Definition 4.4.2. Let k be an integer with $j + 1 \le k \le d$. We say that a 4-tuple (K, C, w, a) forms a *copy of* $\hat{M}_{j,k}$ (see Figure 4.2) 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 simplex of $\mathcal{G}$ that contains a petal of the flower $\mathcal{F} = \mathcal{F}(K, C)$ is itself contained in K;

(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.

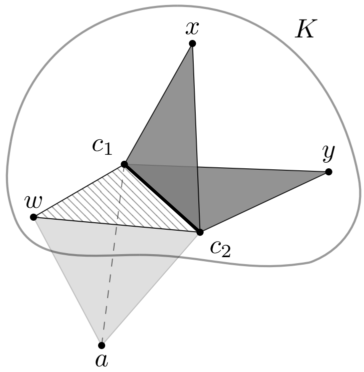


Figure 4.2: A copy of $\hat{M}_{j,k}$, for $k = 4$ and $j = 2$.
(a) The k-simplex K contains the flower $\mathcal{F}(K, C)$ with centre $C = \{c_1, c_2\}$, whose petals $C \cup \{w\}$, $C \cup \{x\}$, and $C \cup \{y\}$ are not present in any simplex which is not contained in K.
(b) The $(j+2)$-set $C \cup \{w\} \cup \{a\}$ is a j-shell whose base is the petal $C \cup \{w\}$ and with apex vertex $a \notin K$.

We want 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 4.4.3. 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 4.4.2.

It is easy to see that a copy of $\hat{M}_{j,j}$ in Definition 4.4.3 satisfies conditions (M1)–(M3) of Definition 4.4.2.

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 4.4.2 and 4.4.3.

Definition 4.4.4. Let k be an integer with $j+1 \le k \le d$. 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 simplex of $\mathcal{G}$ that contains a petal of the flower $\mathcal{F} = \mathcal{F}(K, C)$ is contained in K.

We need an analogous concept for the case $k = j$.

Definition 4.4.5. 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 of $[n]$.

We shall see later (Corollary 4.5.4) that the shell required for (M3) in Definition 4.4.2 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 4.4.13).

Definition 4.4.6. For $j \le k \le d$, 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 4.4.9). 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 of (K, C) of $M_{j,k}$.

Definition 4.4.7. Let $M = (K, C)$ be a copy of $M_{j,k}$ in a simplicial complex.

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

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

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

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

Proposition 4.4.8. *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.*

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

(ii) 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. (i) First observe that if $k = j$ then $\mathcal{F}(K, C) = \{K\}$ and K is an isolated j-simplex by Definition 4.4.5. 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 $k \geq j+1$. Let $K = \{v_0, \dots, v_k\}$ and $C = \{v_0, \dots, v_{j-1}\}$ 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 $\rho = \{v_0, \dots, v_{j-1}, v_{i_1}, v_{i_2}\}$, with $j \leq i_1 \leq i_2 \leq k$. Then we have

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

This implies that f is a j-cocycle if and only if it takes the same value $r \in R$ on each petal $[v_0, \dots, v_{j-1}, v_i]$ with $j \leq i \leq k$, i.e. if and only if $f = f_{M,r}$.

(ii) If f is a j-cocycle and not a j-coboundary, by (i) 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 (i). Furthermore, property (M3) in Definition 4.4.2 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 4.2.8 yields that f is not a j-coboundary. $\square$

Corollary 4.4.9. *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 4.4.7 is a j-cocyle but not a j-coboundary by Proposition 4.4.8 (ii), 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 4.4.10. *Let $\mathcal{S}$ be the support of a j-cocycle f in a d-complex $\mathcal{G}$. Then for each k with $j + 1 \leq k \leq d$ and each k-simplex K,*

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

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

Note in particular that the lemma 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;
- $|\mathcal{S}| \geq k - j + 2$;
- $|\mathcal{S}| = k - j + 1$ and $\mathcal{S}$ forms a j-flower in some k-simplex K.

Since in the latter case an 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 4.4.10. (i) 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 $|\mathcal{S}_K| \geq k - j + 1$ and

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

(ii) Suppose now that $\mathcal{S}_K = \{\sigma_0, \dots, \sigma_{k-j}\}$ with $\sigma_0, \dots, \sigma_{k-j}$ defined as above. For $2 \leq i \leq k-j$ (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 4.1.4 and 4.1.5) 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 4.4.2 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 4.4.11. 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 $\{1, 2\}$, $\{1, 3\}$, and $\{3, 4\}$, as in Figure 4.3. It is easy to see that $\mathcal{G}$ is 1-cohom-connected and thus contains no copies of $\hat{M}_{1,k}$ for any $k \geq 1$. If we add the 2-simplex $\{2, 3, 4\}$ (and its downward-closure) to $\mathcal{G}$, the 4-tuple $(\{2, 3, 4\}, \{3\}, 2, 1)$ 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 $k \geq 1$.

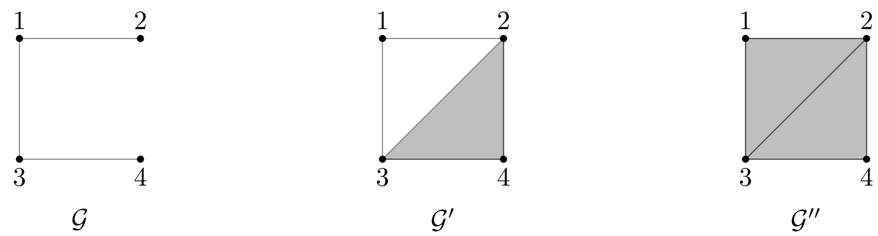


Figure 4.3: Adding simplices might create new copies of $\hat{M}_{j,k}$ or destroy existing ones.

In order to determine the critical range for the disappearance of copies of $M_{j,k}$, in Lemma 4.4.13 we will calculate the expectation of $X_{j,k}$, i.e. the number of copies of $M_{j,k}$ (Definition 4.4.6). We first estimate the probability of (M2). Define

$$\bar{q} = \bar{q}(\bar{\mathbf{p}}, n, j) := \prod\nolimits_{k=j+1}^{d} (1 - \bar{p}_k)^{\binom{n-j-1}{k-j}}. \tag{4.10}$$

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. $\tau = 1$) for any $k \geq j+1$. Moreover if $\bar{\mathbf{p}}$ is a j-admissible direction, by Definition 4.2.1 we have

$$\bar{q} = (1 + o(1)) \exp\left(-\sum\nolimits_{k=j+1}^{d} \left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right), \tag{4.11}$$

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 4.4.12. *Let $\tau = O(1)$ and let $\mathbf{p} = (p_1, \ldots, p_d) = \tau\bar{\mathbf{p}}$ 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 $j+1 \leq k \leq d$ and $K \notin \mathcal{S}$. Then*

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

Proof. We first observe that for $j+1 \leq k \leq d$, 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

$$\begin{aligned}
\Pr(A) &= \prod_{k=j+1}^{d} (1-p_k)^{|\mathcal{J}|\binom{n-j-1}{k-j} - O(n^{k-j-1})} \\
&= (1+o(1)) \prod_{k=j+1}^{d} (1-\tau\bar{p}_k)^{|\mathcal{J}|\binom{n-j-1}{k-j}} \\
&= (1+o(1)) \exp\left(-|\mathcal{J}|\tau \sum_{k=j+1}^{d} \binom{n-j-1}{k-j} \bar{p}_k\right) \\
&\overset{(4.11)}{=} (1+o(1)) \bar{q}^{\,\tau|\mathcal{J}|},
\end{aligned}$$

as claimed. □

We now apply Proposition 4.4.12 to calculate the expectation of $X_{j,k}$, for $\tau = O(1)$. Suppose first that $k \geq j+1$. 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 $p_k \leq 1$ by Remark 4.2.2). By Proposition 4.4.12 applied with $\mathcal{J} = \mathcal{F}(K, C)$ being the set of petals and with $\mathcal{S} = \{K\}$, the probability that (M2) holds is $(1+o(1))\bar{q}^{\,\tau(k-j+1)}$. Therefore,

$$\mathbb{E}(X_{j,k}) = (1+o(1)) \frac{n^{k+1} p_k}{j!(k-j+1)!} \bar{q}^{\,\tau(k-j+1)}. \tag{4.12}$$

The case $k = j$ is very similar, but for a pair (K, C) that forms 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 4.4.5). 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

$$\mathbb{E}(X_{j,j}) = (1+o(1)) \frac{n^{j+1} \min\{p_j, 1\}}{(j+1)!} \bar{q}^{\,\tau}. \tag{4.13}$$

In next lemma we use (4.12) and (4.13) to obtain an explicit asymptotic expression for $\log(\mathbb{E}(X_{j,k}))$, that we will need in Section 4.5. Recall that given a vector $\mathbf{p}$, the parameters λ_k, μ_k, and ν_k are as defined in Definition 4.2.4.

Lemma 4.4.13. *Let $\tau = O(1)$ and let $\mathbf{p} = (p_1, \ldots, p_d) = \tau\bar{\mathbf{p}}$ for a j-admissible direction $\bar{\mathbf{p}}$. Then the number $X_{j,k}$ of copies of $M_{j,k}$ in $\mathcal{G}_\tau$ satisfies*

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

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

The proof of Lemma 4.4.13 consists of (standard, but involved) technical calculations and therefore is deferred to Section 4.11.3.

Recall that in our main Theorems 4.1.4 and 4.1.5 we consider a j-critical direction $\bar{\mathbf{p}}$ which in particular is a j-admissible direction (cf. Definitions 4.2.1 and 4.2.3). Thus Proposition 4.4.12 and Lemma 4.4.13 are applicable. By Lemma 4.4.13, heuristically the critical range for the disappearance of copies of $M_{j,k}$ is when $\lambda_k \log n + \mu_k = \Theta(1)$, that is, $\lambda_k = 0$ and $\mu_k = \Theta(1)$. This justifies the conditions (C1) and (C2) in Definition 4.2.3, which together with Lemma 4.4.13 yield that for $\mathbf{p} = \bar{\mathbf{p}}$,

$$\mathbb{E}(X_{j,\bar{k}}) = 1 + o(1) \quad \text{and} \quad \mathbb{E}(X_{j,k}) \leq 1 + o(1) \text{ for all other indices } k. \tag{4.14}$$

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 4.5.4) 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 4.1.4, we consider

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

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

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

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

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

4.5 Finding minimal obstructions

To prove Lemma 4.3.4, the strategy is to show that whp a copy of $\hat{M}_{j,k}$ (for some $j \leq k \leq d$) exists in $\mathcal{G}_\tau$, for every $\tau \in I_j(\varepsilon) = [\varepsilon/n, \tau_j^*)$. Thus, 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 4.4.14 (ii).

Lemma 4.5.1. *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 4.5.1 is a standard but slightly technical application of the second moment method, and is therefore postponed to Section 4.11.4.

In Lemma 4.5.3 we shall 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 4.5.2. 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 4.5.3. *For every $\varepsilon > 0$ there exists a constant $\zeta > 0$ such that if $\tau \geq \frac{\varepsilon}{\log n}$, 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 $A, D \subseteq [n]$ be disjoint sets, both of size $\lceil n/3 \rceil$ and such that $A \cap B = D \cap B = \emptyset$. Recall (Definition 4.4.14 (ii)) that k_0 is an index with $j+1 \leq k_0 \leq d$ 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 $R_i \subseteq D$, 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 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

$$\begin{aligned}(1-p_{k_0})^{\binom{|D|}{k_0-j}} &\leq \exp\left(-\left(p_{k_0}\frac{n^{k_0-j}}{4^{k_0-j}(k_0-j)!}\right)\right) \\ &\leq (1+o(1))\exp\left(-\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0-j}}\right) \leq \exp\left(-\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0}}\right), \end{aligned} \tag{4.15}$$

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 (4.15), we obtain

$$\Pr(E_a) \geq \left(1 - \exp\left(\frac{\bar{\alpha}_{k_0}\varepsilon}{4^{k_0}}\right)\right)^{j+1} =: \lambda > 0.$$

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

$$\Pr\Big(\mathrm{Bi}\left(\lceil n/3 \rceil, \lambda\right) < \zeta n\Big) \leq \exp\left(-\frac{\left(\frac{n\lambda}{3} - \zeta n\right)^2}{\frac{2n\lambda}{3}}\right) = \exp\left(-\frac{n}{6\lambda}\left(\lambda - 3\zeta\right)^2\right).$$

Choosing $0 < \zeta < \lambda/3$ 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

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

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 4.5.4. *Let $\varepsilon > 0$ be constant. If $\tau \geq \frac{\varepsilon}{\log n}$, then whp for every copy (K, C) of $M_{j,k}$ in $\mathcal{G}_\tau$ (for any $j \leq k \leq d$), 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 $\mathcal{G}_\tau$.*

We now show that for τ slightly less than 1, whp we have many copies of $M_{j,k}$.

Lemma 4.5.5. *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 $k \geq j$, 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 4.5.1, although the second moment calculation is significantly simpler without the j-shell of $\hat{M}_{j,k}$, and can be found in Section 4.11.5.

The following proposition describes more precisely the parameters in Definition 4.2.4 for $\mathbf{p} = \tau\bar{\mathbf{p}}$ and τ 'close' to 1, in terms of the analogous parameters defined in (4.2) and (4.3) for $\bar{\mathbf{p}}$.

Proposition 4.5.6. *Let $\tau = 1 + \xi$ with $\xi = \xi(n) = o(1)$ and let $\mathbf{p} = \tau\bar{\mathbf{p}}$. Then for all $j \leq k \leq d$,*

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

The easy but technical proof of Proposition 4.5.6 appears in Section 4.11.6.

We derive the following corollaries, that we will use to prove Lemma 4.5.5 and Lemma 4.9.1, respectively.

Corollary 4.5.7. *Let $\tau = \left(1 - \frac{1}{\omega_0}\right)$ with $\omega_0 = \omega_0(n) = o(\log n)$ and let $\mathbf{p} = \tau\bar{\mathbf{p}}$. Then for all $j \leq k \leq d$,*

$$\mathbb{E}(X_{j,k}) \geq \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{3\omega_0}\right).$$

Corollary 4.5.8. *Let $c \in \mathbb{R}$ be a constant and suppose $(c_n)_{n\geq 1}$ 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 $j \leq k \leq d$,*

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

We delay the proofs of Corollaries 4.5.7 and 4.5.8 until Sections 4.11.7 and 4.11.8, respectively.

4.6 Determining the hitting time: proof of Lemma 4.3.1

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

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

as defined in Theorem 4.1.4. We will show that whp this happens at around the claimed threshold $\tau = 1$ (Lemma 4.3.1).

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 4.5.3 and 4.5.5 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 $\tau'' \leq \tau_j^*$. The following main result says that they are in fact equal whp, and indeed whp both are close to 1.

Lemma 4.6.1. *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 4.3.1 is an immediate corollary of Lemma 4.6.1. To prove Lemma 4.6.1, we will need some further concepts and some auxiliary results.

Definition 4.6.2. Given $k \geq j$ 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 4.6.3. Given an integer k with $j \leq k \leq d$, 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 $j \leq k \leq d$ is certainly a local j-obstacle, although a local j-obstacle is not necessarily an obstruction to j-cohom-connectedness.

Lemma 4.6.4. *Whp, for all $\tau \geq \tau'$, 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 $j \leq k \leq d$. The lemma then follows by applying a union bound over all k.

We first note that by Remark 4.2.2, $\bar{p}_k < 1$ if $k \geq j+1$. On the other hand, if $k = j$ and $\bar{p}_j \geq \frac{1}{\tau'} = 1 + o(1)$, 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

$$\text{either} \quad k \geq j+1 \qquad \text{or} \quad k = j \text{ and } \bar{p}_j < \frac{1}{\tau'} < 1 + \frac{\log\log n}{9d\log n}. \tag{4.16}$$

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

$$\tau^- := 1 - \frac{\log\log n}{5d\log n}. \tag{4.17}$$

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

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

by Remark 4.2.2 and (4.16).

Fix $k \geq j$, let K be a $(k+1)$-set, and recall that $\tau_K \in [0,1]$ 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 (4.1)). 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 $\mathcal{J}$ 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

$$\Pr(L_1) = (1 - \tau^-\bar{p}_k). \tag{4.18}$$

In order to compute $\Pr(L_2 \mid L_1)$, first observe that L_2 is independent of L_1. By

Proposition 4.4.12 applied with $\mathbf{p} = \tau^{-}\bar{\mathbf{p}}$, we have

$$\begin{aligned}
\Pr(L_2 \mid L_1) &= \Pr(L_2) \\
&= (1+o(1))\bar{q}^{\,\tau^{-}(k-j+1)} \\
&\overset{(4.12),(4.13)}{=} \left(\frac{\Theta(1)\cdot\mathbb{E}(\bar{X}_{j,k})}{n^{k+1}\min\{\bar{p}_k,1\}}\right)^{\tau^{-}} \overset{(\bar{p}_k<1+o(1))}{=} \left(\frac{\Theta(1)\cdot\mathbb{E}(\bar{X}_{j,k})}{n^{k+1}\bar{p}_k}\right)^{\tau^{-}},
\end{aligned} \tag{4.19}$$

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 $\mathbb{E}(\bar{X}_{j,k}) \le 1 + o(1)$ (see (4.14)).

We now want to bound $\Pr\left(L_3 \mid (L_1 \wedge L_2)\right)$. For any i such that $j+1 \le i \le d$ 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 $\mathcal{J}$. It will be convenient to pick $i = k_0$, recalling from Definition 4.4.14 (ii) that $k_0 \ge j+1$ is such that $\bar{\alpha}_{k_0} \ne 0$ and $\bar{\gamma}_{k_0} = 0$. Thus we have a family $\mathcal{Z}$ of

$$z := |\mathcal{Z}| = (k-j+1)\binom{n-k-1}{k_0-j} = \Theta\left(\frac{\log n}{\bar{p}_{k_0}}\right) \tag{4.20}$$

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 4.6.5. *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*

$$\Pr\left(L_3' \mid (L_1 \wedge L_2)\right) = (1+o(1))\frac{\bar{p}_k}{z\bar{p}_{k_0}(1-\tau^{-}\bar{p}_k)} \overset{(4.20)}{=} \Theta\left(\frac{\bar{p}_k}{(1-\tau^{-}\bar{p}_k)\log n}\right).$$

We will delay the proof of Claim 4.6.5 until after the proof of Lemma 4.6.4, which we now complete. Note that $L_3 \subseteq L_3'$, thus Claim 4.6.5 in particular implies that

$$\Pr\left(L_3 \mid (L_1 \wedge L_2)\right) \le \Pr\left(L_3' \mid (L_1 \wedge L_2)\right) = \Theta\left(\frac{\bar{p}_k}{(1-\tau^{-}\bar{p}_k)\log n}\right). \tag{4.21}$$

Putting (4.18), (4.19), and (4.21) together we have

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

Recalling that $\mathbb{E}(\bar{X}_{j,k}) \le 1 + o(1)$ by (4.14), we thus deduce that

$$\Pr(L_1 \wedge L_2 \wedge L_3) = O\left(\frac{\bar{p}_k^{1-\tau^-}}{n^{\tau^-(k+1)}\log n}\right). \tag{4.22}$$

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 4.2.1) 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

$$\begin{aligned}
\Theta(n^{k+1})\Pr(L_1 \wedge L_2 \wedge L_3) &\overset{(4.22)}{=} O\left(\frac{\left(n^{k+1}\bar{p}_k\right)^{1-\tau^-}}{\log n}\right)\\
&= O\left(\frac{n^{(j+1)(1-\tau^-)}}{(\log n)^{\tau^-}}\right)\\
&\overset{(4.17)}{=} O\left(\exp\left(\left(\frac{j+1}{5d} - (1-o(1))\right)\log\log n\right)\right)\\
&= O\left(\exp\left(\frac{-\log\log n}{2}\right)\right) \xrightarrow{n\to\infty} 0.
\end{aligned}$$

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

We now also prove the auxiliary claim.

Proof of Claim 4.6.5. 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 $k_0 \ge j+1$, by Remark 4.2.2 we have $\frac{1}{\bar{p}_{k_0}} = \omega(1)$, and thus

$$\frac{1}{\bar{p}_{k_0}} - \tau^- = (1+o(1))\frac{1}{\bar{p}_{k_0}}. \tag{4.23}$$

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

$$\Pr\left(L_3' \mid (L_1 \wedge L_2)\right) = \Pr\left(S \mid (L_1 \wedge L_2)\right)\Pr\left(L_3' \mid (L_1 \wedge L_2 \wedge S)\right). \tag{4.24}$$

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

$$\Pr\left(S \mid (L_1 \wedge L_2)\right) = \Pr(S \mid L_1) = \frac{1/\bar{p}_{k_0} - \tau^-}{1/\bar{p}_k - \tau^-} \overset{(4.23)}{=} (1+o(1))\frac{\bar{p}_k}{\bar{p}_{k_0}(1-\tau^-\bar{p}_k)}. \tag{4.25}$$

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

$$\Pr\left(L_3' \mid (L_1 \wedge L_2 \wedge S)\right) = \frac{1}{1+z} \overset{(4.20)}{=} (1+o(1))\frac{1}{z}. \tag{4.26}$$

Putting (4.24), (4.25), and (4.26) together, we have

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

as claimed.

Case 2: $\frac{1}{\bar{p}_k} < \frac{1}{\bar{p}_{k_0}}$. First observe that if $k = j$, by (4.16) we have $\bar{p}_j < 1+\frac{\log\log n}{9d\log n}$, while if $k \geq j+1$, then $\bar{p}_k < 1$ by Remark 4.2.2. Thus by the definition of τ^- (see (4.17)), for any $k \geq j$ we have

$$1-\tau^-\bar{p}_k > \frac{\log\log n}{12d\log n}. \tag{4.27}$$

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

$$\mathbb{E}(\zeta_k) = z\cdot\frac{1/\bar{p}_k-\tau^-}{1/\bar{p}_{k_0}-\tau^-} \overset{(4.23)}{=} (1+o(1))\frac{z\bar{p}_{k_0}}{\bar{p}_k}(1-\tau^-\bar{p}_k). \tag{4.28}$$

Since $z\bar{p}_{k_0} = \Theta(\log n)$ by (4.20) and $\bar{p}_k < 1+\frac{\log\log n}{9d\log n}$, we obtain

$$\mathbb{E}(\zeta_k) \overset{(4.27)}{=} \Omega\left(\log n\cdot\frac{\log\log n}{12d\log n}\right) = \Omega\left((\log\log n)^{1/2}\right) \to \infty.$$

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 at most $\exp(-\Theta(\mathbb{E}(\zeta_k)^{1/2}))$.

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

$$\begin{aligned}
\Pr\left(L_3' \mid (L_1 \wedge L_2)\right) &= \frac{1+o(1)}{1+\left(1\pm\frac{1}{\mathbb{E}(\zeta_k)^{1/4}}\right)\mathbb{E}(\zeta_k)} + \exp(-\Theta(\mathbb{E}(\zeta_k)^{1/2}))\\
&= \frac{1+o(1)}{\mathbb{E}(\zeta_k)}\\
&\overset{(4.28)}{=} (1+o(1))\frac{\bar{p}_k}{z\bar{p}_{k_0}(1-\tau^-\bar{p}_k)},
\end{aligned}$$

as claimed. □

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

Proof of Lemma 4.6.1. Observe that in particular a copy of $M_{j,k}$ (or, more precisely, the associated k-simplex) is a local j-obstacle. Lemma 4.6.4 shows that if a local j-obstacle is present in $\mathcal{G}_\tau$ for some $\tau \geq \tau'$, 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 the associated copy of $M_{j,k}$ would whp already exist at time τ'' and thus whp form a copy of $\hat{M}_{j,k}$ (Lemma 4.5.3) at that time too. This cannot happen by definition of τ'', which gives $\tau'' = \tau_j^*$ whp, as required.

To prove the second statement, observe that by Lemmas 4.5.3 and 4.5.5, 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 $\tau = 1 + \frac{\omega}{\log n}$ we have that for any $j+1 \le k \le d$ the expected number of copies of $M_{j,k}$ satisfies

$$\begin{aligned}
\mathbb{E}(X_{j,k}) &\overset{(4.12)}{=} (1+o(1))\frac{n^{k+1}\tau\bar{p}_k}{j!(k-j+1)!}\prod\nolimits_{i=j+1}^{d}(1-\tau\bar{p}_i)^{(k-j+1+O(\frac{1}{n}))\binom{n}{i-j}} \\
&= (1+o(1))\frac{n^{k+1}\bar{p}_k}{j!(k-j+1)!} \\
&\quad \cdot \exp\left(-\sum_{i=j+1}^{d}\left(k-j+1+O\left(\frac{1}{n}\right)\right)\binom{n}{i-j}\left(\bar{p}_i + \frac{\bar{p}_i\omega}{\log n} + O(\bar{p}_i^2)\right)\right) \\
&\le (1+o(1))\mathbb{E}(\bar{X}_{j,k})\exp(-\bar{\alpha}_{k_0}\omega) = o(1),
\end{aligned}$$

where we are using that $\mathbb{E}(\bar{X}_{j,k}) \le 1 + o(1)$ by (4.14) 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 hat whp $\tau_j^* = \tau''$. $\square$

4.7 Subcritical case: proof of Lemma 4.3.4

In this section we first derive some auxiliary results and combine them to prove Lemma 4.3.4, which plays a crucial role in the proof of the subcritical case (i.e. statement (ii)) of Theorem 4.1.4.

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 4.7.1, 4.7.3, and 4.7.4).

For the next lemma, let us recall that the index k_0 is such that $j+1 \le k_0 \le d$ and $\bar{\alpha}_{k_0} \neq 0$ (see Definition 4.4.14 (ii)).

Lemma 4.7.1. *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 4.7.2. Indeed, Lemma 4.7.1 would also hold with k_0 replaced by any index i with $j+1 \le i \le d$ such that $\bar{\alpha}_i \neq 0$.

Proof of Lemma 4.7.1. By Lemma 4.5.1, there exist constants $0 < c_1 < c_2$ (depending on ε) such that the number $\hat{X}_{j,k_0}$ of copies of $\hat{M}_{j,k_0}$ in $\mathcal{G}_{\frac{\varepsilon}{n}}$ whp satisfies

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

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 $j+1 \le k \le d$, 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 $\binom{n-j-1}{k-j} \le \frac{n^{k-j}}{(k-j)!}$ 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

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

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 $\binom{k+1}{j+1} \le \binom{d+1}{j+1} =: c_5$ 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

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

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. $\square$

Lemma 4.7.3. *For every constant $\delta > 0$ and every critical dimension k with $j \leq k \leq d$, 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 4.5.5 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 $\Theta(\exp(\sqrt{\log n}))$ many copies with distinct k-simplices. For each such copy (K, C), the scaled birth time of K (see (4.1)) 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

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

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

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

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 4.5.4 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. $\square$

Lemma 4.7.4. *Whp the minimal obstruction which vanishes at time τ_j^* (defined in Theorem 4.1.4) 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 4.4.14, the last minimal obstruction to vanish is a copy of $\hat{M}_{j,\ell}$. Similar to the proof of Lemma 4.7.3, this copy of $\hat{M}_{j,\ell}$ has a birth time that is uniformly distributed within $\left[0, \min\{\tau_j^*, 1/\bar{p}_\ell\}\right)$ and by Lemma 4.6.1, whp

$\tau_j^* \leq 1 + \frac{1}{(\log n)^{1/3}}$. Conditioned on this high probability event, the probability that this copy of $M_{j,\ell}$ already existed 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 the corresponding copy of $M_{j,\ell}$ was already present at time $\tau = 1 - \frac{1}{(\log n)^{1/3}}$. By Corollary 4.5.4, this means that whp it was part of a copy of $\hat{M}_{j,\ell}$ throughout the interval. □

Proof of Lemma 4.3.4. Lemmas 4.7.1, 4.7.3, and 4.7.4 imply that whp for any $\tau \in [\varepsilon/n, \tau_j^*)$ a copy of $\hat{M}_{j,k}$ (for some $j \leq k \leq d$) exists in $\mathcal{G}_\tau$. Therefore, for any τ in this range $H^j(\mathcal{G}_\tau; R) \neq 0$ by Corollary 4.4.9. □

4.8 Critical and supercritical cases: proof of Lemma 4.3.7

In this section we present some auxiliary results and prove Lemma 4.3.7, which we have used to show that whp the process $(\mathcal{G}_\tau) = (\mathcal{G}(n, \tau\bar{\mathbf{p}}))_\tau$ is j-cohom-connected for all $\tau \geq \tau_j^*$ (Theorem 4.1.4 (iii)). Furthermore, the results of this section will be fundamental for the proof of Theorem 4.1.5 (Section 4.9).

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 4.8.3, Lemma 4.8.4), and then use this property to show that whp the support cannot be small (Lemma 4.8.5). Subsequently, we use traversability and a result of Meshulam and Wallach [53] to show that whp the support cannot be large (Lemma 4.8.7), which is a contradiction.

However, so far this only proves that for any $\tau \geq \tau_j^*$ whp $H^j(\mathcal{G}_\tau; R) = 0$. We need to know that whp for any $\tau \geq \tau_j^*$ 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 4.8.9).

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 4.8.1. 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

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

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 $\tau \geq \tau_j^*$ we have $\mathcal{N}_\tau = \emptyset$. Since in this range there are no copies of $\hat{M}_{j,k}$ and by Corollary 4.5.4 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 4.8.2. 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 will first show (Lemma 4.8.4) that $\mathcal{S}_\tau$ must satisfy the following concept of *traversability*.

Definition 4.8.3. Let $(\mathcal{S}, \mathcal{H})$ be a pair where $\mathcal{S}$ is a collection of j-simplices in $\mathcal{G}_\tau$ and $\mathcal{H}$ is a collection of simplices in $\mathcal{G}_\tau$ of dimensions $j+1, \dots, d$.

We say that $\mathcal{S}$ is *$\mathcal{H}$-traversable* if it *cannot* be partitioned into two non-empty subsets such that every simplex of $\mathcal{H}$ contains elements of $\mathcal{S}$ in at most one of the two subsets. Equivalently, $\mathcal{S}$ is $\mathcal{H}$-traversable if for every $J, J' \in \mathcal{S}$ there exists a sequence $J = J_0, J_1, \dots, J_m = J'$ of j-simplices in $\mathcal{S}$ and a sequence $K_1, \dots, K_m$ of simplices in $\mathcal{H}$ (not necessarily all of the same dimension) such that $(J_{i-1} \cup J_i) \subset K_i$ for all $i \in [m]$.

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

Lemma 4.8.4. *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 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 $\mathcal{S}_{(1)}$ 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 $f_\tau \in \mathcal{N}_\tau$. $\square$

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

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 4.8.5. *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 $s \le h$, 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 $s \le h$.

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

$$\sum\nolimits_{k=j+1}^{d} t_k \le s \le h, \tag{4.29}$$

and

$$v \le (j+1) + \sum\nolimits_{k=j+1}^{d} (k-j)t_k. \tag{4.30}$$

Since $\mathcal{S}$ is the support of a j-cocycle, by Lemma 4.4.10 (i) 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

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

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. There are $O(n^v)$ ways of choosing such a pair, therefore we obtain

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

By (4.30), we have

$$v - \sum\nolimits_{k=j+1}^{d} (k-j)t_k \le j+1.$$

Moreover, for an index i such that $t_i \ge 1$ (such an index exists, because otherwise the support would be empty), by Lemma 4.4.10 and since the considered j-cocycle does not arise from a copy of $M_{j,k}$, it holds that $s \geq i - j + 2$. Recalling that

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

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

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

Thus,

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

Since by (4.29) and (4.30) 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 [53, 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 still works without the additional condition that R is finite and we include this proof in Section 4.11.9 for completeness.

Proposition 4.8.6 ([53, Proposition 3.1])**.** *Let $n \geq j+2$ and let Δ be the downward-closure of the $(n-1)$-simplex on $[n]$. Let f be a j-cochain with support $\mathcal{S}$ such that any other j-cochain of the form $f+g$, where g is a j-coboundary, has support of size larger than or equal to $|\mathcal{S}|$. Denote by $b(f)$ the size of the support of $\delta^j f$, i.e. the number of $(j+1)$-simplices 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*

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

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

Lemma 4.8.7. *There exists a positive constant $\tilde{h} \in \mathbb{R}^+$ such that whp for all $\tau \geq \tau_0 := 1 - \frac{1}{(\log n)^{1/3}}$ we have $|\mathcal{S}_\tau| < \tilde{h}$ (if $\mathcal{S}_\tau$ exists).*

Proof. We first note that if τ is large enough that there exists a $k \in \{j+1, \ldots, d\}$ with $p_k = \tau \bar{p}_k \geq 1$, 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 $\tau \leq \min_{k \in \{j+1,\ldots,d\}} 1/\bar{p}_k$, and in particular each $p_k \leq 1$ for $k \geq j+1$.

By Lemma 4.8.4 if $\mathcal{S}_\tau$ exists it is traversable, therefore it can be discovered by the following breadth-first search process. Start from any j-simplex in $\mathcal{S}_\tau$ and query all $(k+1)$-sets containing it, for every $k \in \{j+1, \ldots, d\}$. Any of these sets which forms a k-simplex in $\mathcal{T}(\mathcal{S}_\tau)$ must contain at least one other j-simplex of $\mathcal{S}_\tau$, because of the minimality of $\mathcal{T}(\mathcal{S}_\tau)$. From all j-simplices in $\mathcal{S}_\tau$ found in this way, we continue the process according to some pre-determined order of j-simplices, by only querying those $(k+1)$-sets which contain some previously undiscovered j-simplex in $\mathcal{S}_\tau$. The traversability of $\mathcal{S}_\tau$ yields that all j-simplices of $\mathcal{S}_\tau$ (and also all simplices of $\mathcal{T}(\mathcal{S}_\tau)$) are discovered in this process.

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 described search process. For such a pair, we define the *exploration matrix* $B = (b_{i,k})$ for $i \in [s]$ and $k \in \{j+1, \dots, d\}$, where $b_{i,k} \geq 0$ 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. Consider the j-simplex in step i of the exploration: from this we can query at most $\binom{n}{k-j}$ many $(k+1)$-sets and from each of the $b_{i,k}$ discovered 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 $k \in \{j+1, \dots, d\}$, 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

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

using that $\sum_{i=1}^{s} b_{i,k} = t_k$, for each $k \in \{j+1, \dots, d\}$.

In order to obtain a lower bound on the number of $(j+2)$-sets that are not allowed to be $(j+1)$-simplices, we aim to apply Proposition 4.8.6, for which we need to know that $\mathcal{G}_\tau$ has full $(j-1)$-skeleton.

The expected number of j-sets that do not form a $(j-1)$-simplex in $\mathcal{G}_\tau$ is bounded from above by

$$\begin{aligned}\binom{n}{j} (1 - p_{k_0})^{\binom{n-j}{k_0-j+1}} &\leq \binom{n}{j} (1 - \tau_0 \bar{p}_{k_0})^{\binom{n-j}{k_0-j+1}} \\ &\leq (1+o(1)) n^j \exp\left(-\tau_0 n \cdot \frac{\bar{\alpha}_{k_0} \log n + \bar{\beta}_{k_0}}{k_0 - j + 1} \right) = o(1),\end{aligned}$$

thus by Markov's inequality whp $\mathcal{G}_\tau$ has a complete $(j-1)$-dimensional skeleton. This means that, with Δ as in Proposition 4.8.6, whp $C^{j-1}(\Delta) = C^{j-1}(\mathcal{G}_\tau)$, implying that for every coboundary $g \in C^{j-1}(\Delta)$ the j-cochain $f_\tau + g$ lies in the same cohomology class of $\mathcal{G}_\tau$ as f_τ. Since $\mathcal{S}_\tau$ is minimal, we have that whp the hypotheses of Proposition 4.8.6 are satisfied.

Conditioning on this high probability event and since f_τ is a j-cocycle, Proposition 4.8.6 tells us that at least $\frac{sn}{j+2}$ many $(j+2)$-sets are not allowed to be part of k-simplices of $\mathcal{G}_\tau$, for every $k \in \{j+1, \dots, d\}$. Each of these $(j+2)$-sets is contained in $\binom{n-j-2}{k-j-1}$ many $(k+1)$-sets, each of which contains $\binom{k+1}{j+2}$ many $(j+2)$-sets. Thus we have at least

$$\frac{sn\binom{n-j-2}{k-j-1}}{(j+2)\binom{k+1}{j+2}} \geq h_0 s n^{k-j}$$

many $(k+1)$-sets that cannot be chosen as k-simplices in $\mathcal{G}_\tau$, for some positive constant h_0 and for every $k = j+1, \dots, d$. Given a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with $\mathbf{t}(\mathcal{S}) = (t_{j+1}, \dots, t_d)$, recalling that $s \geq \sum_{k=j+1}^{d} t_k =: Y_{\mathbf{t}}$ by (4.29), the probability that such a pair exists is at most

$$\prod_{k=j+1}^{d} p_k^{t_k}(1-p_k)^{h_0 n^{k-j} s} \leq \tau^{Y_{\mathbf{t}}} \prod_{k=j+1}^{d} \bar{p}_k^{t_k}(1-\tau\bar{p}_k)^{h_0 n^{k-j} Y_{\mathbf{t}}} =: x(\tau).$$

Differentiating the positive function $x(\tau)$ with respect to τ we have

$$\frac{dx}{d\tau} = \frac{x(\tau)Y_{\mathbf{t}}}{\tau}\left(1 - \tau h_0 \sum_k \frac{n^{k-j}\bar{p}_k}{1-\tau\bar{p}_k}\right).$$

Recalling that the index k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ and $\bar{\gamma}_{k_0} = 0$, we have

$$\frac{dx}{d\tau} \leq \frac{x(\tau)Y_{\mathbf{t}}}{\tau}\left(1 - \tau h_0 \frac{\bar{\alpha}_{k_0}\log n + \bar{\beta}_{k_0}}{1-\tau\bar{p}_{k_0}}\right) < 0$$

for $\tau = \omega(1/\log n)$. Thus, since the derivative of $x(\tau)$ is negative throughout the whole range $\tau \geq \tau_0 = 1 - o(1)$, we have $x(\tau) \leq x(\tau_0)$ for all $\tau \geq \tau_0$, and therefore in the following calculations we may substitute τ_0 for τ.

For a pair $(\mathcal{S}, \mathcal{T}(\mathcal{S}))$ with given exploration matrix B, let r_B denote the probability that such a pair exists and that the simplices in $\mathcal{T}(\mathcal{S})$ do not contain a $(j+1)$-simplex σ such that $(\delta^j f_\tau)(\sigma) \neq 0_R$. Then r_B satisfies

$$\begin{aligned}
r_B \prod_{i,k} b_{i,k}! &\leq n^{j+1} \prod_{k=j+1}^{d} \left(\binom{n}{k-j} 2^{\binom{k+1}{j+1}} \tau_0 \bar{p}_k\right)^{t_k} (1-\tau_0\bar{p}_k)^{h_0 n^{k-j} Y_{\mathbf{t}}} \\
&\leq n^{j+1} \prod_k \left(\Theta(1)\frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)^{t_k} \\
&\quad \cdot \exp\left(-(1+o(1))h_0 Y_{\mathbf{t}}\left(\frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{\bar{\gamma}_k}}(k-j)!\right)\right) \\
&\leq n^{j+1}\left(O(\log n)\right)^{Y_{\mathbf{t}}} n^{-(1+o(1))h_0\bar{\alpha}_{k_0}(k-j)! Y_{\mathbf{t}}} \\
&\leq n^{j+1} n^{-h_1 Y_{\mathbf{t}}},
\end{aligned}$$

where $h_1 := \frac{h_0\bar{\alpha}_{k_0}(k-j)!}{2}$. Now suppose that $s \geq \tilde{h} \geq \frac{2(j+1)\sum_k \binom{k+1}{j+1}}{h_1}$. Since $Y_{\mathbf{t}} \geq \frac{s}{\sum_k \binom{k+1}{j+1}}$, we can find another positive constant h_2 such that

$$r_B \leq \frac{n^{-h_2 s}}{\prod_{i,k} b_{i,k}!}.$$

Given an exploration matrix $B = (b_{i,k})$, define $\mathbf{u}(B) = (u_{j+1}, u_{j+2}, \ldots, u_d)$, where

$$u_k := |\{i : b_{i,k} \geq n^{h_2/(d-j+1)}\}|.$$

Conversely, given $\mathbf{u} = (u_{j+1}, \ldots, u_d)$ with $u_k \geq 0$, 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 $n^{h_2/(d-j+1)}$ possibilities for each of the small entries and, since the sum of all the entries is $\sum_k t_k \leq s$, at most s possibilities for each of the large entries. Thus we obtain the (rather crude) upper bound

$$
\begin{aligned}
|\mathcal{B}_{\mathbf{u}}| &\leq \left(\prod_k \binom{s}{u_k}\right) \left(n^{h_2/(d-j+1)}\right)^{s(d-j)-\sum_k u_k} s^{\sum_k u_k} \\
&\leq s^{2\sum_k u_k} \left(n^{h_2/(d-j+1)}\right)^{s(d-j)-\sum_k u_k}.
\end{aligned}
$$

Moreover, for $B \in \mathcal{B}_{\mathbf{u}}$

$$
\prod_{i,k} b_{i,k}! \geq \left(\left(n^{\frac{h_2}{d-j+1}}\right)!\right)^{\sum_k u_k} \geq n^{n^{\frac{h_2}{d-j+2}} \sum_k u_k}.
$$

Putting everything together, the probability r_s that $\mathcal{S}_\tau$ of fixed size $s \geq \tilde{h}$ exists (together with the simplices collection $\mathcal{T}(\mathcal{S}_\tau)$) satisfies

$$
\begin{aligned}
r_s &\leq \sum_{\mathbf{u}} \sum_{B \in \mathcal{B}_{\mathbf{u}}} r_B \leq \sum_{\mathbf{u}} |\mathcal{B}_{\mathbf{u}}| \frac{n^{-h_2 s}}{\prod_{i,k} b_{i,k}!} \\
&\leq \sum_{\mathbf{u}} \left[\prod_k \left(\frac{s^2}{n^{\frac{h_2}{d-j+1}} n^{n^{\frac{h_2}{d-j+2}}}}\right)^{u_k}\right] \frac{n^{\frac{h_2}{d-j+1} s(d-j)}}{n^{h_2 s}} \\
&\leq (s+1)^{d-j} \cdot 1 \cdot n^{-\frac{h_2}{d-j+1} s} \leq n^{-h_3 s},
\end{aligned}
$$

for some positive constant h_3.

Thus, the probability that $\mathcal{S}_\tau$ exists with $|\mathcal{S}_\tau| \geq \tilde{h}$ is at most

$$
\sum_{s \geq \tilde{h}} r_s \leq \sum_{s \geq \tilde{h}} n^{-h_3 s} \leq \frac{n^{-h_3 \tilde{h}}}{1 - n^{-h_3}} \leq n^{-h_3 \tilde{h}/2}.
$$

Provided $\tilde{h} > \frac{2(d+1)}{h_3}$, and taking a union bound over the $O(n^{d+1})$ birth times of simplices in the considered probability range, the probability that $\mathcal{S}_\tau$ exists and has size at least $\tilde{h}$ for any τ is $O(n^{d+1-h_3\tilde{h}/2})$, which tends to zero for our choice of $\tilde{h}$. $\square$

Lemma 4.8.7 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 4.8.5 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 4.8.8. *For every $\tau = 1 + O\left(\frac{1}{\log n}\right)$, we have $\mathcal{N}_\tau = \emptyset$ whp.*

Proof. For every such τ Lemma 4.8.7 holds and thus there exists a constant $\tilde{h}$ such that either $\mathcal{S}_\tau$ does not exist or $|\mathcal{S}_\tau| \leq \tilde{h}$. But by Lemma 4.8.5 (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 in this range does not exist by Lemma 4.6.4.

Lemma 4.8.9. *Let K be the simplex with smallest scaled birth time $\tau_K \geq \tau_j^*$ 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 4.6.1 and Corollary 4.8.8, 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 $|K| = k+1$ and let $\tau \geq \tau_j^*$ 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 4.8.7 there exists a positive constant $\tilde{h}$ such that whp $|\mathcal{S}| \leq |\mathcal{S}_{\tau_K}| < \tilde{h}$. Lemma 4.5.3 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 $\tau \geq \tau_j^*$ 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 $|\mathcal{S}_{\tau_K}| \geq k - j + 1$ by Lemma 4.4.10. $\square$

Corollary 4.8.10. *Whp for all $\tau \geq \tau_j^*$ we have $\mathcal{N}_\tau = \emptyset$.*

Proof. By Lemma 4.6.1 and Corollary 4.8.8 whp $\mathcal{N}_{\tau_j^*}$ is empty. If there is $\tau > \tau_j^*$ such that $\mathcal{N}_\tau \neq \emptyset$, then Lemma 4.8.9 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 4.6.1 whp $\tau > \tau_j^* \geq \tau' = 1 - \frac{\log\log n}{10d \log n}$, thus by Lemma 4.6.4 whp no new local j-obstacle can appear in $\mathcal{G}_\tau$. $\square$

We can now use Corollary 4.8.10 to prove Lemma 4.3.7.

Proof of Lemma 4.3.7. By the definition of τ_j^*, there are no copies of $\hat{M}_{j,k}$ in $\mathcal{G}_\tau$ for any $\tau \geq \tau_j^*$, 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 4.8.10. $\square$

4.9 Rank in the critical window: proof of Theorem 4.1.5

In order to prove the Rank Theorem (Theorem 4.1.5), 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 4.9.1, for which we need the following notation. Given a sequence of random variables $(X_n)_{n\geq 1}$ we denote by $\mathcal{L}(X_i)$ the probability distribution of X_i and by $\mathcal{L}\left((X_n)_{n\geq 1}\right)$ the joint probability distribution of the sequence $(X_n)_{n\geq 1}$.

Lemma 4.9.1. *Let $c \in \mathbb{R}$ be a constant and $(c_n)_{n\geq 1}$ be a sequence of real numbers such that $c_n \xrightarrow{n\to\infty} c$. For any $j \leq k \leq d$ define*

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

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

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

where $\mathrm{Po}(0) \equiv 0$.

To prove Lemma 4.9.1 we use a multivariate Poisson approximation technique from [6], which will be presented in Section 4.11.10. The proof of Lemma 4.9.1 then appears in Section 4.11.11.

Proof of Theorem 4.1.5. Consider $\mathcal{X} = \sum_{k=j}^{d} X_{j,k}$ and define

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

where $\mathcal{C} = \mathcal{C}(\bar{\mathbf{p}}, j)$ is the set of the critical dimensions for the j-critical direction $\bar{\mathbf{p}}$ (see Definition 4.2.6). By Lemma 4.9.1, we have

$$\mathcal{L}(\mathcal{X}) \xrightarrow{\mathrm{d}} \sum_{k=j}^{d} \mathrm{Po}(\mathcal{E}_k) = \mathrm{Po}\left(\sum_{k\in\mathcal{C}} \exp\left(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)\right)\right) = \mathrm{Po}(\mathcal{E}). \tag{4.31}$$

We first show that $H^j(\mathcal{G}_\tau; R) \cong R^{\mathcal{X}}$ whp. Let $M_1, M_2, \dots, M_{\mathcal{X}}$ denote the copies of $M_{j,k}$, for every $j \leq k \leq d$ that are present in $\mathcal{G}_\tau$. By Corollary 4.8.8 we have $\mathcal{N}_\tau = \emptyset$ whp and by Proposition 4.4.8 (i) 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 the cohomology classes of those elements generates $H^j(\mathcal{G}_\tau; R)$.

We now need to show that if we take two tuples $(r_1, \dots, r_{\mathcal{X}}) \neq (r'_1, \dots, r'_{\mathcal{X}})$ 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 $(r_1, \dots, r_{\mathcal{X}})$ 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 $\mathcal{X}$ has bounded expectation by Corollary 4.5.8. We further claim that for each $j \leq k \leq d$ whp no two copies of $M_{j,k}$ share the same k-simplex. Indeed, for $k = j$ by Definition 4.4.5 all copies of $M_{j,j}$ come from different j-simplices. If $k \geq j+1$, 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 4.4.12.

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

$$O\left(\binom{n}{k+1}\binom{k+1}{j}^2 p_k \bar{q}^{\,\tau(2k-2j+1)}\right) \overset{(4.12)}{=} O\left(\mathbb{E}(X_{j,k})\bar{q}^{\,\tau(k-j)}\right) = O(\bar{q}^{\,\tau(k-j)}),$$

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

$$\bar{q} \overset{(4.11)}{=} O\left(\exp\left(-\sum_{i=j+1}^{d}\left(\bar{\alpha}_i \log n + \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}}\right)\right)\right) = O(n^{-\bar{\alpha}_{k_0}/2}) = o(1), \tag{4.32}$$

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 4.2.1). Since $k \geq j+1$ 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 4.4.4, 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 4.5.3 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 $|S| \leq \sum_{k=j}^{d}(k-j+1)X_{j,k} \leq (d-j+1)\mathcal{X} = o(n)$ 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 4.2.8.

We therefore have $H^j(\mathcal{G}_\tau; R) \cong R^{\mathcal{X}}$ whp. Since $\mathcal{L}(\mathcal{X}) \xrightarrow{\mathrm{d}} \mathrm{Po}(\mathcal{E})$ by (4.31), there exists a coupling $Y \sim \mathrm{Po}(\mathcal{E})$ such that $\mathcal{X} = Y$ whp. Thus, whp

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

as required. $\square$

4.10 More about parametrisation

In this section we clarify the parametrisation of $\bar{\mathbf{p}}$ and the assumptions made for j-admissibility and j-criticality in Definitions 4.2.1 and 4.2.3. Note that the arguments here are independent of the proof of Theorem 4.1.4—rather, they justify why the assumptions made in the theorem 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 R-cohomological j-connectedness occurs around time $\tau = 1$, i.e. when $\mathbf{p} = \bar{\mathbf{p}}$.

First 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 first 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 $p_j, \dots, p_d$ into account.

4.10.1 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 $\bar{\gamma}_k \geq 0$, once $\bar{\gamma}_k$ is defined (see (4.35)).

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

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

For $R = \mathbb{F}_2$ and $j+1 \leq k \leq d$, [17, 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 $\mathbf{p} = (p_1, \dots, p_d)$ to lie around

$$p_k = \frac{\alpha_k \log n + r_k}{n^{k-j}}(k-j)!, \quad \forall j \leq k \leq d, \tag{4.33}$$

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 $p_k \geq \frac{c_k \log n}{n^{k-j}}(k-j)!$ 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 $j \leq k \leq d$ will not change this. Furthermore, it follows from the results of [17], and indeed also from Theorem 4.1.4, 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 $j+1 \leq k \leq d$.

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

$$\bar{\alpha}_k := \lim_{n\to\infty} \left(\frac{\bar{p}_k n^{k-j}}{(k-j)! \log n} \right) \tag{4.34}$$

exists. Observe that if $\bar{p}_k = o(\log n / n^{k-j})$, then $\bar{\alpha}_k = 0$. Indeed, we next argue that by scaling the direction $\bar{\mathbf{p}}$ appropriately, we may also assume that at least one $\bar{\alpha}_k$ is non-zero.

4.10.2 Existence of k_0: Justifying (A4)

So far we have only guaranteed certain properties of $\bar{\mathbf{p}}$ by scaling appropriately. We now argue that the case when $\bar{p}_j \geq 1$ and $\bar{\alpha}_k = 0$ for all $k = j+1, \ldots, d$ can be easily reduced to the case when $\bar{p}_j \geq 1$ and for some $j+1 \leq k \leq d$ we have $\bar{\alpha}_k \neq 0$, and therefore we may assume that there exists $k_0 \geq j+1$ with $\bar{\alpha}_{k_0} \neq 0$ as stated in (A4).

Indeed, suppose we have the slightly more general case that $\bar{p}_j \geq 1$ and $\bar{p}_k \leq \frac{\log n}{Cn^{k-j}}$ 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 $0 \leq k \leq j-1$, which guarantees the existence of an isolated simplex of dimension at most k for any $\tau \leq cn^{-j}$, 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 $k \leq j-1$ 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 $k \geq j+1$ 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 $k_0 \geq j+1$.

4.10.3 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'. 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 $\bar{p}_k \leq n^{-(k+k_0-j+1)}$, 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

$$\bar{\gamma}_k := \sup\{\gamma \in \mathbb{R} \mid \bar{p}_k n^{k-j+\gamma} = o(1)\} \tag{4.35}$$

exists for every k with $j \leq k \leq d$ and $\bar{p}_k \neq 0$. By the existence of the limit in (4.34), we have $\bar{\gamma}_k \geq 0$.

4.10.4 Fine-tuning

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

$$\bar{p}_k = \frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)!. \tag{4.36}$$

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

4.10.5 j-admissibility

Let us note that by our definitions of $\bar{\alpha}_k, \bar{\beta}_k, \bar{\gamma}_k$ for a j-admissible direction (Definition 4.2.1), properties (A1)–(A3) certainly hold for $j \leq k \leq d$, and we have also argued that we may assume that (A4) holds. Indeed, we will also assume that the parametrisation in (4.36) and the properties (A1)–(A3) are valid for $1 \leq k \leq j-1$—we will need this to show that we can rescale every $\bar{p}_k$ to obtain an i-critical direction for every $i \in [j-1]$ (Lemma 4.3.5).

4.10.6 j-criticality

It remains only to justify the assumptions of Definition 4.2.3. 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 $k \geq j$. 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 $\bar{\lambda}_{k_0} \geq j+1-\frac{1}{2} > 0$, 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.

4.11 Proofs of auxiliary results

4.11.1 Proof of Lemma 4.3.2

(i) 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

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

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

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

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

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

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

(ii) 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 $p_k \geq \frac{c^+\log n}{n^k}$. Then (ii) follows from the fact that we can choose c^+ such that $\mathcal{G}(n,\tilde{\mathbf{p}})$ is whp topologically connected by [17, Lemma 4.1]. □

We note that the proof idea of Lemma 4.3.2 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 4.1.4 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 [17] for the proof of (ii), also follows as a special case of earlier and much stronger results from [19] and from [57]. The proof in [17] has the advantage that it is a simple and elementary extension of the standard graph argument.

4.11.2 Proof of Lemma 4.3.5

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

Recall from Definition 4.2.3 that the j-critical direction $\bar{\mathbf{p}} = (\bar{p}_1,\ldots,\bar{p}_d)$ is in particular a j-admissible direction (Definition 4.2.1), i.e. for every $k\in[d]$

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

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

Given a positive constant η and a function $\epsilon = \epsilon(n) = o(1)$, consider the vector

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

We will show that we can choose $\eta = \eta_{j-1}$ and $\epsilon = \epsilon_{j-1}(n)$ such that $\bar{\mathbf{p}}'$ is a $(j-1)$-critical direction (Definition 4.2.3).

For every $k \in [d]$, we have

$$\begin{aligned}\bar{p}'_k = \frac{\eta + \epsilon}{n}\bar{p}_k &= \frac{\eta + \epsilon}{n} \cdot \frac{\bar{\alpha}_k \log n + \bar{\beta}_k}{n^{k-j+\bar{\gamma}_k}}(k-j)! \\ &= \frac{\bar{\alpha}'_k \log n + \bar{\beta}'_k}{n^{k-(j-1)+\bar{\gamma}'_k}}(k-(j-1))!,\end{aligned}$$

where

$$\bar{\alpha}'_k := \frac{\eta\bar{\alpha}_k}{k-j+1}, \qquad \bar{\beta}'_k := \frac{\eta\bar{\beta}_k + \epsilon\bar{\alpha}_k \log n + \epsilon\bar{\beta}_k}{k-j+1}, \qquad \bar{\gamma}'_k := \bar{\gamma}_k. \tag{4.37}$$

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 4.2.1 and thus $\bar{\mathbf{p}}'$ is a $(j-1)$-admissible direction.

We now want to prove that $\bar{\mathbf{p}}'$ is also a $(j-1)$-critical direction, i.e. that conditions (C1) and (C2) in Definition 4.2.3 both hold. Thus, we need to compute the parameters $\bar{\lambda}'_k$, $\bar{\mu}'_k$, and $\bar{\nu}'_k$ as in (4.3), but with j replaced by $j-1$.

We have

$$\begin{aligned}\bar{\lambda}'_k \; &= \; j - \bar{\gamma}'_k - (k-j+2)\sum\nolimits_{i=j}^{d} \bar{\alpha}'_i \\ &\overset{(4.37)}{=} \; j - \bar{\gamma}_k - \eta(k-j+2)\sum\nolimits_{i=j}^{d} \frac{\bar{\alpha}_i}{i-j+1}.\end{aligned}$$

In order for (C1) and (C2) to hold, we want in particular that

$$\begin{aligned} &\bar{\lambda}'_k \le 0 \qquad \text{for every } j-1 \le k \le d \\ \text{and} \quad &\bar{\lambda}'_{\bar{k}} = 0 \qquad \text{for some } j-1 \le \bar{k} \le d. \end{aligned} \tag{4.38}$$

Recall that the index $j+1 \le k_0 \le d$ was such that $\bar{\alpha}_{k_0} \ne 0$ and $\bar{\gamma}_{k_0} = 0$ (cf. (A1) and (A4) for $\bar{\mathbf{p}}$). We observe that for $\eta \to 0$ it holds that $\bar{\lambda}'_{k_0} \to j > 0$, while for $\eta \to \infty$ we have $\bar{\lambda}'_k \to -\infty$ for every k. Hence, by continuity we can choose the positive constant η such that (4.38) holds.

Furthermore we observe that

$$\begin{aligned}\bar{\nu}'_k \;&=\; \begin{cases} -\log((j+1)!) & \text{if } k = j, \\ -\log(j!) - \log(k-j+1) + \log(\bar{\alpha}'_k) & \text{if } \bar{\alpha}'_k \ne 0, \\ -\log(j!) - \log(k-j+1) & \text{otherwise} \end{cases} \\ &\overset{(4.37)}{=} O(1)\end{aligned}$$

and

$$\begin{aligned}
\bar{\mu}'_k &= -(k-j+2)\sum_{i=j}^{d}\frac{\bar{\beta}'_i}{n^{\bar{\gamma}'_i}} + \begin{cases} 0 & \text{if } \bar{p}'_k > 1, \\ \log\log n & \text{if } \bar{p}'_k \leq 1 \text{ and } \bar{\alpha}'_k \neq 0, \\ \log(\bar{\beta}'_k) & \text{if } \bar{p}'_k \leq 1 \text{ and } \bar{\alpha}'_k = 0, \end{cases} \\
&\overset{(4.37)}{=} -(k-j+2)\epsilon\sum_{i=j}^{d}\frac{\bar{\alpha}_i\log n + O(\bar{\beta}_i)}{i-j+1} + \begin{cases} O(\log\log n) & \text{if } \bar{\alpha}_k \neq 0, \\ \log(\bar{\beta}_k) + O(1) & \text{if } \bar{\alpha}'_k = 0. \end{cases}
\end{aligned}$$

Since we fixed η such that (4.38) holds, to guarantee that (C1) and (C2) are both satisfied we need that

$$\begin{aligned}
& \bar{\mu}'_k \leq -\bar{\nu}'_k && \text{for every } j-1 \leq k \leq d \\
\text{and} \quad & \bar{\mu}'_{\bar{k}} = -\bar{\nu}'_{\bar{k}} && \text{for } \bar{k} \text{ such that } \bar{\lambda}'_{\bar{k}} = 0.
\end{aligned} \tag{4.39}$$

Observe that in the expression for $\bar{\mu}'_k$ we have $\sum_{i=j}^{d}\frac{\bar{\alpha}_i\log n + O(\bar{\beta}_i)}{i-j+1} = o(\log n)$, thus we can find a function ω of n with $\omega \to \infty$ such that $-\frac{1}{\omega} \leq o(\log n) \leq \frac{1}{\omega}$. If we choose $\epsilon = -\frac{1}{\sqrt{\omega}}$ then $\bar{\mu}'_k \to \infty$, whereas for $\epsilon = \frac{1}{\sqrt{\omega}}$ then $\bar{\mu}'_k \to \infty$. Therefore, by continuity we can choose $\epsilon = o(1)$ such that (4.39) holds.

Finally, observe that for any η and ϵ such that (4.38) and (4.39) simultaneously hold, we have that (C1) and (C2) are both satisfied, i.e. $\bar{\mathbf{p}}'$ is a $(j-1)$-critical direction, as required. □

4.11.3 Proof of Lemma 4.4.13

Suppose first that $k \geq j+1$. We observe that

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

by Definition 4.2.4. Moreover, since $\bar{\alpha}_i/n^{\bar{\gamma}_i} = \bar{\alpha}_i$ for each $j+1 \leq i \leq d$ by (A1) in Definition 4.2.1, we have

$$\begin{aligned}
\bar{q}^{\,\tau} &\overset{(4.11)}{=} (1+o(1))\exp\left(-\tau\sum_{i=j+1}^{d}\left(\bar{\alpha}_i\log n + \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}}\right)\right) \\
&\overset{(4.2)}{=} (1+o(1))\exp\left(-\sum_{i=j+1}^{d}\tau\bar{p}_i\frac{n^{i-j}}{(i-j)!}\right) \\
&= (1+o(1))\exp\left(-\sum_{i=j+1}^{d}p_i\frac{n^{i-j}}{(i-j)!}\right) \\
&= (1+o(1))\exp\left(-\sum_{i=j+1}^{d}\left(\alpha_i\log n + \frac{\beta_i}{n^{\gamma_i}}\right)\right).
\end{aligned} \tag{4.40}$$

Substituting these values into (4.12) and taking the logarithm, we obtain

$$\begin{aligned}
\log\left(\mathbb{E}(X_{j,k})\right) &= (j+1-\gamma_k)\log n + \log(\alpha_k \log n + \beta_k) - \log(j!) - \log(k-j+1) \\
&\qquad - \sum\nolimits_{i=j+1}^{d} \left((k-j+1) \left(\alpha_i \log n + \frac{\beta_i}{n^{\gamma_i}} \right) + o(1) \right) \\
&= \left(j+1-\gamma_k - (k-j+1) \sum\nolimits_{i=j+1}^{d} \alpha_i \right) \log n \\
&\qquad - (k-j+1) \sum\nolimits_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} + \log(\alpha_k \log n + \beta_k) \\
&\qquad - \log(j!) - \log(k-j+1) + o(1) \\
&= \lambda_k \log n + \mu_k + \nu_k + o(1),
\end{aligned}$$

as required.

If $k = j$, substituting (4.40) into (4.13) 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

$$\begin{aligned}
\log(\mathbb{E}(X_{j,j})) = \left(j+1 - \sum\nolimits_{i=j+1}^{d} \alpha_i \right) \log n + \sum\nolimits_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} - \log((j+1)!) \\
+ \log\left(\min\{p_j, 1\}\right) + o(1). \qquad (4.41)
\end{aligned}$$

We first consider the case when $p_j \leq 1$, in which case by Definition 4.2.4 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 (4.41), we obtain

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

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 (4.41) gives

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

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

4.11.4 Proof of Lemma 4.5.1

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

$$p_{k_0} = \frac{\varepsilon}{n}\bar{p}_{k_0} \overset{(4.2)}{=} \Theta\left(\frac{\log n}{n^{k_0-j+1}}\right), \tag{4.42}$$

since the choice of k_0 is such that $\bar{\alpha}_{k_0} \neq 0$ (see Definition 4.4.14 (ii)). By Proposition 4.4.12 the probability that (M2) holds is $(1+o(1))\bar{q}^{\,\tau(k-j+1)}$, where since $\tau = \varepsilon/n$ we have

$$\bar{q}^{\,\tau} \overset{(4.11)}{=} (1+o(1))\exp\left(-\frac{\varepsilon}{n}\cdot\sum\nolimits_{k=j+1}^{d}\left(\bar{\alpha}_k \log n + \frac{\bar{\beta}_k}{n^{\bar{\gamma}_k}}\right)\right) = 1+o(1),$$

and therefore (M2) holds 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

$$r := \sum\nolimits_{k=j}^{d}\frac{p_k n^{k-j}}{(k-j)!} = \Theta\left(\frac{\log n}{n}\right). \tag{4.43}$$

Combining all the probabilities, we obtain

$$\Pr(\hat{T} \text{ forms a copy of } \hat{M}_{j,k_0}) = (1+o(1))p_{k_0}r^{j+1}. \tag{4.44}$$

Recall that $\hat{X}_{j,k_0}$ denotes the number of copies of $\hat{M}_{j,k_0}$ in $\mathcal{G}_\tau$. Now (4.44) implies that

$$\begin{aligned}\mathbb{E}(\hat{X}_{j,k_0}) &= \binom{n}{k_0+1}\binom{k_0+1}{j}(k_0-j+1)(n-k_0-1)p_{k_0}(1+o(1))r^{j+1} \\ &= (1+o(1))\frac{p_{k_0}r^{j+1}n^{k_0+2}}{j!(k_0-j)!} \overset{(4.42)}{=} \Theta((\log n)^{j+2}).\end{aligned} \tag{4.45}$$

We now aim to calculate the 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 4.4.12, 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 i-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

$$1 - \prod_{k=j+1}^{d} (1-p_k)^{O(n^{k-j-1})} = O\left(\frac{\log n}{n^2}\right) \overset{(4.43)}{=} o(r^2).$$

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

$$\Pr(\hat{T}_1, \hat{T}_2 \text{ both form copies of } \hat{M}_{j,k_0}) = (1+o(1))(p_{k_0})^s r^t. \tag{4.46}$$

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. Denote by $\mathcal{S}$ the set of triples (i,s,t) for which $\mathcal{T}^2(i,s,t)$ is non-empty. With this notation, (4.46) implies that

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

Observe that $|\hat{\mathcal{T}}^2(i,s,t)| = O(n^{2k+4-i})$. 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 $k_0 + 1 \leq i \leq k_0 + 2$.

- $i = k_0 + 1$. 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 $t \geq j+1$, which gives a contribution of order

$$O(p_{k_0} r^{j+1} n^{2k_0+4-(k_0+2)}) \overset{(4.45)}{=} O\left(\mathbb{E}(\hat{X}_{j,k_0})\right) = o\left(\mathbb{E}(\hat{X}_{j,k_0})^2\right).$$

- $i = k_0 + 2$. 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

$$O\left(p_{k_0} r^{2j+2} n^{2k_0+4-(k_0+1)}\right) \overset{(4.45)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{p_{k_0} n^{k_0+1}}\right) = o\left((\mathbb{E}(\hat{X}_{j,k_0})^2\right).$$

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

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

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.

Case 2: s=2.

- $i = 0$. We show that this case represents the dominant contribution to the expected value $\mathbb{E}((\hat{X}_{j,k_0})^2)$. 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

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

Thus, the contribution of all such pairs is

$$(1 + o(1))\frac{p_{k_0}^2 r^{2j+2} n^{2k_0+4}}{(j!(k_0 - j)!)^2} \overset{(4.45)}{=} (1 + o(1))\mathbb{E}(\hat{X}_{j,k_0})^2.$$

- $1 \le i \le j$. 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

$$O\left(p_{k_0}^2 r^{2j+2} n^{2k_0+4-i}\right) \overset{(4.45)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{n^i}\right) = o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

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

$$O\left(p_{k_0}^2 r^{2j+1} n^{2k_0+4-(j+1)}\right) \overset{(4.45)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{rn^{j+1}}\right) \overset{(4.43)}{=} o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

- $j+2 \le i \le k_0+2$. In this case $t \ge j$, 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

$$O\left(p_{k_0}^2 r^j n^{2k_0+4-i}\right) \overset{(4.45)}{=} O\left(\frac{\mathbb{E}(\hat{X}_{j,k_0})^2}{r^{j+2}n^i}\right) \overset{(4.43)}{=} o(\mathbb{E}(\hat{X}_{j,k_0})^2).$$

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. □

4.11.5 Proof of Lemma 4.5.5

We will prove the lemma with $c = \bar{\alpha}_{k_0}/4$, where k_0 is as defined in Definition 4.4.14 (ii). We first bound the expected number of copies of $M_{j,k}$.

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 $k \geq j+1$.

Let $\mathcal{T}_k$ denote the family of pairs $T = (K, C)$, where $K \subseteq [n]$ 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 $\mathcal{F}(K, C)$. For $T \in \mathcal{T}_k$, denote by X_T the indicator random variable of the event that T forms a copy of $M_{j,k}$. Thus $X_{j,k} = \sum_{T \in \mathcal{T}_k} X_T$.

Given two pairs $T_1 = (K_1, C_1)$ and $T_2 = (K_2, C_2)$, we define

- $s = s(T_1, T_2) := \begin{cases} 1 & \text{if } K_1 = K_2, \\ 2 & \text{otherwise;} \end{cases}$
- $\mathcal{F}_h := \{C_h \cup \{w\} \mid w \in K_h \setminus C_h\}$ 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 4.4.12, 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 $t \geq k - j + 1$ and the contribution is at most of order
$$O\left(n^{k+1} p_k \bar{q}^{\,\tau(k-j+1)}\right) \overset{(4.12)}{=} O(\mathbb{E}(X_{j,k})) = o(\mathbb{E}(X_{j,k})^2).$$
- $s = 2$. 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
$$\binom{k+1}{j}^2 = (1+o(1))\binom{n}{k+1}^2\binom{k+1}{j}^2.$$
Thus these pairs provide a contribution of
$$(1+o(1))\binom{n}{k+1}^2\binom{k+1}{j}^2 p_k^2 \bar{q}^{\,\tau 2(k-j+1)} \overset{(4.12)}{=} (1+o(1))\mathbb{E}(X_{j,k})^2.$$

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 4.4.5). 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 4.2.6). 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 4.4.12, we obtain

$$\begin{aligned}
\mathbb{E}((X_{j,j})^2) &= \mathbb{E}(X_{j,j}) + \sum_{i=0}^{j} \binom{n}{j+1}\binom{j+1}{i}\binom{n-j-1}{j+1-i} \min\{p_j, 1\}^2 \\
&\quad \cdot (1+o(1))\bar{q}^{\,2\tau} \\
&\overset{(4.13)}{\leq} \mathbb{E}(X_{j,j})^2 \left(1 + o(1) + \sum_{i=1}^{j} O(n^{-i})\right) \\
&= \mathbb{E}(X_{j,j})^2 \,(1+o(1))\,.
\end{aligned}$$

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 and Corollary 4.5.7 whp

$$X_{j,k} = (1+o(1))\mathbb{E}(X_{j,k}) > \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{4\omega_0}\right),$$

as required. $\square$

4.11.6 Proof of Proposition 4.5.6

Observe that for each $k \geq j$ 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 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 $k \geq j+1$, recall that $p_k \leq 1$ by Remark 4.2.2 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

$$\sum_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} = \sum_{i=j+1}^{d} \frac{(1+\xi)\,\bar{\beta}_i + \bar{\alpha}_i \xi \log n}{n^{\bar{\gamma}_i}}$$

$$\overset{\text{(A1)}}{=} \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \xi \cdot \left(\sum_{i=j+1}^{d} \bar{\alpha}_i \log n + \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} \right)$$

$$\overset{\text{(A2)-(A3)}}{=} \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + \xi \cdot \left(\sum_{i=j+1}^{d} \bar{\alpha}_i \log n + \sum_{i:\bar{\gamma}_i \neq 0} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + o(\log n) \right)$$

$$= \sum_{i=j+1}^{d} \frac{\bar{\beta}_i}{n^{\bar{\gamma}_i}} + (1+o(1))\xi \sum_{i=j+1}^{d} \bar{\alpha}_i \log n, \tag{4.47}$$

and for all k such that $\alpha_k = 0$ we have the additional term

$$\log(\beta_k) = \log((1+\xi)\bar{\beta}_k) = \log(\bar{\beta}_k) + o(1).$$

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 $p_j \leq 1$, then $\alpha_j = \bar{\alpha}_j = 0$. Furthermore, if $p_j \leq 1 \leq \bar{p}_j$, then $\bar{\alpha}_j = \bar{\gamma}_j = 0$ and $\bar{p}_j = \bar{\beta}_j = 1 + o(1)$. Therefore, we have

$$\begin{aligned}
\mu_j &= -\sum\nolimits_{i=j+1}^{d} \frac{\beta_i}{n^{\gamma_i}} + \begin{cases} 0 & \text{if } p_j > 1, \\ \log(\beta_j) & \text{if } p_j \leq 1 \end{cases} \\
&= -\sum\nolimits_{i=j+1}^{d} \frac{(1+\xi)\,\bar{\beta}_i + \bar{\alpha}_i \xi \log n}{n^{\bar{\gamma}_i}} + \begin{cases} 0 & \text{if } p_j > 1, \\ \log(\bar{\beta}_j) + \log(1+\xi) & \text{if } p_j \leq 1 \end{cases} \\
&= \bar{\mu}_j - \xi \sum\nolimits_{i=j+1}^{d} \frac{\bar{\beta}_i + \bar{\alpha}_i \log n}{n^{\bar{\gamma}_i}} + o(1) + \begin{cases} \log(\bar{\beta}_j) & \text{if } p_j \leq 1 \leq \bar{p}_j, \\ 0 & \text{otherwise} \end{cases} \\
&\overset{(4.47)}{=} \bar{\mu}_j - (1+o(1))\,\xi \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n,
\end{aligned}$$

as required. □

4.11.7 Proof of Corollary 4.5.7

By Proposition 4.5.6 applied with $\xi = -1/\omega_0$, for any $j \leq k \leq d$ we have $\lambda_k = \bar{\lambda}_k$, $\nu_k = \bar{\nu}_k$, and

$$\mu_k = \bar{\mu}_k + (1+o(1))(k-j+1)\frac{1}{\omega_0} \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i \log n \geq \bar{\mu}_k + \frac{\bar{\alpha}_{k_0} \log n}{2\omega_0},$$

where we are using that $k - j + 1 \geq 1$. Thus we have

$$\lambda_k \log n + \mu_k + \nu_k \geq \bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k + \frac{\bar{\alpha}_{k_0} \log n}{2\omega_0},$$

therefore Lemma 4.4.13 and the fact that $\bar{\lambda}_k \log n + \bar{\mu}_k + \bar{\nu}_k = O(1)$ imply that

$$\mathbb{E}(X_{j,k}) \geq \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{2\omega_0} + O(1)\right) \geq \exp\left(\frac{\bar{\alpha}_{k_0} \log n}{3\omega_0}\right),$$

as claimed. □

4.11.8 Proof of Corollary 4.5.8

For any $j \leq k \leq d$, Proposition 4.5.6 applied with $\xi = \frac{c_n}{\log n}$ tells us that

$$\lambda_k = \bar{\lambda}_k, \qquad \mu_k = \bar{\mu}_k - (1 + o(1))(k - j + 1)c_n \sum\nolimits_{i=j+1}^{d} \bar{\alpha}_i, \qquad \nu_k = \bar{\nu}_k. \tag{4.48}$$

If k is not a critical dimension, by Definition 4.2.3 and Lemma 4.4.13 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 4.4.13 at time τ we have

$$\mathbb{E}(X_{j,k}) \overset{(4.48)}{=} (1 + o(1))\mathbb{E}(\bar{X}_{j,k}) = o(1),$$

as required.

If k is a critical dimension, we have $\bar{\lambda}_k = 0$ and $\bar{\mu}_k = O(1)$ (see Definition 4.2.6). Thus by Lemma 4.4.13 we have

$$\begin{aligned}
\mathbb{E}(X_{j,k}) &= \exp(\lambda_k \log n + \mu_k + \nu_k + o(1)) \\
&\overset{(4.48)}{=} (1 + o(1)) \frac{\exp(\bar{\mu}_k + \bar{\nu}_k)}{\exp\left((k - j + 1)c_n \sum_{i=j+1}^{d} \bar{\alpha}_i\right)} \\
&= (1 + o(1)) \exp(\bar{\mu}_k + \bar{\nu}_k + c(\bar{\gamma}_k - j - 1)),
\end{aligned}$$

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

4.11.9 Proof of Proposition 4.8.6

Given an ordered m-simplex $\Phi = [v_0, \ldots, v_m]$ and a vertex $u \notin \Phi$, define the ordered $(m + 1)$-simplex

$$[u, \Phi] := [u, v_0, \ldots, v_m],$$

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 $\sigma = [v_0, \ldots, v_j]$ we have

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) = f(\sigma) - \sum\nolimits_{i=0}^{j} (-1)^i f_v\left([v_0, \ldots, \hat{v}_i, \ldots, v_j]\right). \tag{4.49}$$

If $u \notin \sigma$, then

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) \overset{(4.49)}{=} f(\sigma) - \sum\nolimits_{i=0}^{j} (-1)^i f\left([v, v_0, \ldots, \hat{v}_i, \ldots, v_j]\right) = (\delta^j f)([v,\sigma]),$$

by definition of the operator δ^j.

If $v \in \sigma$ then $v = v_\ell$ for some $\ell \in [j]_0$, implying that $f_v[v_0, \ldots, \hat{v}_i, \ldots, v_j] = 0$ for every $i \neq \ell$ and $f_v[v_0, \ldots, \hat{v}_\ell, \ldots, v_j] = (-1)^\ell f(\sigma)$. Thus

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) \overset{(4.49)}{=} f(\sigma) - (-1)^{2\ell} f(\sigma) = 0_R.$$

Putting everything together

$$f(\sigma) - (\delta^{j-1} f_v)(\sigma) = \begin{cases} (\delta^j f)([v,\sigma]) & \text{if } v \notin \sigma; \\ 0_R & \text{otherwise.} \end{cases} \tag{4.50}$$

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

$$n|\mathcal{S}| \le \sum_{v \in [n]} |\mathrm{supp}(f - \delta^{j-1} f_v)| = |\{(v,\sigma) : v \in [n],\ \sigma \in \mathrm{supp}(f - \delta^{j-1} f_v)\}|. \tag{4.51}$$

For a pair (v,σ), by (4.50) 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

$$\begin{aligned} n|\mathcal{S}| &\overset{(4.51)}{\le} |\{(v,\rho) : v \in \rho,\ \rho \in \mathrm{supp}(\delta^j f)\}| \\ &= (j+2)|\mathrm{supp}(\delta^j f)| \\ &= (j+2) b(f), \end{aligned}$$

as required. □

4.11.10 Multivariate Poisson approximation

In order to prove Lemma 4.9.1, we will use a method from [6] that we briefly explain in this section.

Given a discrete set H, the *total variation distance* between the distributions of two H-valued random variables Y and Z is defined by

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

Lemma 4.11.1 ([6, 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*

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

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

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

where $\mathrm{Po}(\mathbf{m})$ *denotes the joint Poisson distribution* $(\mathrm{Po}(m_1)\dots, \mathrm{Po}(m_r))$ *and* $\mathrm{Po}(0) \equiv 0$.

It is easy to see that if there exists $\tilde{\mathbf{m}} := (\tilde{m}_1, \dots, \tilde{m}_r)$ 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

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

4.11.11 Proof of Lemma 4.9.1

We will first show that we can apply Lemma 4.11.1 with $W_k = X_{j,k}$ and $m_k = \mathbb{E}(X_{j,k})$ for $k = j, \dots, d$. 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 $j \le k \le d$. As in the proof of Lemma 4.5.5, we consider the sets

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

for each $j+1 \le k \le d$. 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 4.4.5). Following the notation of Lemma 4.11.1, we set $\Gamma := \dot\cup_{k=j}^d \mathcal{T}_k$ 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 $M_{j,k}$;
- $\pi_a := \Pr(I_a = 1) = \mathbb{E}(I_a)$;
- $\mathcal{B}_a$ is the collection of *forbidden sets* for a, i.e.

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

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 4.4.1 and (4.9)). 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 4.4.4).

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 the subsets of $[n]$ (of cardinality at most $d+1$) that contain K_a.

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 4.4.4 and 4.4.5 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

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

To apply Lemma 4.11.1, we now aim to define the random variables $(J_{ba})_{b \in \Gamma}$. 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

$$J_{ba} = \mathbb{1}\{E^1_{ba} \wedge E^2_{ba} \wedge E^3_{ba}\}. \tag{4.52}$$

We claim that

$$\mathcal{L}\left((J_{ba})_{b \in \Gamma}\right) = \mathcal{L}\left((J_{ba})_{b \in \Gamma} | I_a = 1\right). \tag{4.53}$$

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 (4.52) and the definition of E^3_{ba} we have $J_{ba} = 0$ deterministically and we set $\mathcal{D}_{ba} := \emptyset$. Otherwise, by (4.52) 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 (4.53).

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

$$\big((I_a = 1) \wedge (J_{ba} = 1)\big) \quad \Longleftrightarrow \quad \big((I_a = 1) \wedge (I_b = 1)\big). \tag{4.54}$$

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_a and I_b, clearly the events E^1_{ba} and E^2_{ba} hold. Moreover, I_a and I_b can be both equal to 1 only 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 (4.54).

Hence, conditioned on $I_a = 1$, for every $b \in \Gamma$ (4.54) yields that J_{ba} and I_b are the same random variable, and thus in particular

$$\mathcal{L}\big((J_{ba})_{b\in\Gamma} | I_a = 1\big) = \mathcal{L}\big((I_b)_{b\in\Gamma} | I_a = 1\big). \tag{4.55}$$

In total, we have

$$\mathcal{L}\big((J_{ba})_{b\in\Gamma}\big) \overset{(4.53)}{=} \mathcal{L}\big((J_{ba})_{b\in\Gamma} | I_a = 1\big) \overset{(4.55)}{=} \mathcal{L}\big((I_b)_{b\in\Gamma} | I_a = 1\big). \tag{4.56}$$

Since $X_{j,k} = \sum_{a\in\mathcal{T}_k} I_a$ for any $j \leq k \leq d$, we can thus apply Lemma 4.11.1. Setting $Z_k := \mathrm{Po}(\mathbb{E}(X_{j,k}))$ for each k, we obtain

$$d_{TV}\left(\mathcal{L}(\mathbf{X}), (Z_j, \ldots, Z_d)\right) \leq \sum_{a\in\Gamma} \pi_a \left(\pi_a + \sum_{b\neq a} \mathbb{E}|J_{ba} - I_b|\right). \tag{4.57}$$

We want to show that the right-hand side of (4.57) is $o(1)$. Recall that for every $b \in \Gamma$ by (4.10) and Proposition 4.4.12 we have

$$\mathbb{E}(I_b) = \pi_b = (1 + o(1)) p_{k_b} \bar{q}^{\,\tau(k_b - j + 1)}, \tag{4.58}$$

and therefore (cf. (4.12))

$$\mathbb{E}(X_{j,k}) = \sum\nolimits_{b\in\mathcal{T}_k} \mathbb{E}(I_b) = \Theta(n^{k+1} p_k \bar{q}^{\,\tau(k-j+1)}) = O(1), \tag{4.59}$$

where the last equality holds because we are considering $\mathcal{G}_\tau$ within the critical window. Furthermore, by (4.56) we have

$$\mathbb{E}(J_{ba}) = \Pr(I_b = 1 | I_a = 1). \tag{4.60}$$

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 $k_b = k_a \geq j+1$ 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

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

we have

$$\begin{aligned}
\sum\nolimits_{b \in B_1} \mathbb{E}|J_{ba} - I_b| &\leq \sum\nolimits_{b \in B_1} \big(\mathbb{E}(J_{ba}) + \mathbb{E}(I_b)\big) \\
&\overset{(4.60)}{=} \sum\nolimits_{b \in B_1} \big(\Pr(I_b = 1 | I_a = 1) + \pi_b\big) \\
&\overset{(4.58)}{=} O(1) \cdot \big(O(\bar{q}^{\,\tau(k_a-j)}) + O(p_{k_a} \bar{q}^{\,\tau(k_a-j+1)})\big) \\
&= O(\bar{q}^{\,\tau(k_a-j)}) = o(1),
\end{aligned}$$

where the last equality follows from the facts that $\tau(k_a - j) > 1 + o(1)$ and $\bar{q} = o(1)$ (cf. (4.32)).

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_{ba}^3 does not happen, thus $J_{ba} = 0$ deterministically by (4.52).

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

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

we have

$$\begin{aligned}
\sum\nolimits_{b \in B_{2.1}} \mathbb{E}|J_{ba} - I_b| &\leq \sum\nolimits_{b \in B_{2.1}} \mathbb{E}(I_b) \\
&\overset{(4.58)}{=} \sum\nolimits_{k=j}^{d} O\left(n^{k-j} p_k \bar{q}^{\,\tau(k-j+1)}\right) \\
&\overset{(4.59)}{=} \sum\nolimits_{k=j}^{d} O\left(\frac{\mathbb{E}(X_{j,k})}{n^{j+1}}\right) = o(1).
\end{aligned}$$

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

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

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} hold, yielding that if $I_b = 1$ then also $J_{ba} = 1$, that is $J_{ba} \geq I_b$ deterministically and therefore

$$\mathbb{E}|J_{ba} - I_b| = \mathbb{E}(J_{ba}) - \mathbb{E}(I_b). \tag{4.61}$$

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 $j + 1 \leq k \leq d$. 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

$$\begin{aligned}
\mathbb{E}(J_{ba}) &\overset{(4.60)}{=} \Pr(I_b = 1 | I_a = 1) \\
&= p_{k_b} \prod_{k=j+1}^{d} (1 - \tau \bar{p}_k)^{(k_b-j+1)\binom{n-j-1}{k-j} - O(n^{k-j-1})} \\
&\overset{(4.10)}{=} (1 + o(1)) p_{k_b} \bar{q}^{\,\tau(k_b-j+1)} \prod_{k=j+1}^{d} \exp\left(O(p_k) \cdot O(n^{k-j-1})\right) \\
&= (1 + o(1)) p_{k_b} \bar{q}^{\,\tau(k_b-j+1)} \exp\left(\sum_{k=j+1}^{d} O\left(\frac{\log n}{n^{k-j}} n^{k-j-1}\right)\right) \\
&= (1 + o(1)) p_{k_b} \bar{q}^{\,\tau(k_b-j+1)},
\end{aligned}$$

and thus

$$\mathbb{E}|J_{ba} - I_b| \overset{(4.61)}{=} \mathbb{E}(J_{ba}) - \mathbb{E}(I_b) = \mathbb{E}(J_{ba}) - \pi_b \overset{(4.58)}{=} o(p_{k_b} \bar{q}^{\,\tau(k_b-j+1)}). \tag{4.62}$$

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

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

we have

$$\begin{aligned}
\sum_{b \in B_3} \mathbb{E}|J_{ba} - I_b| &\overset{(4.62)}{=} \sum_{k=j}^{d} O(n^{k+1}) \cdot o(p_k \bar{q}^{\,\tau(k-j+1)}) \\
&\overset{(4.59)}{=} o(1) \cdot \sum_{k=j}^{d} O\left(\mathbb{E}(X_{j,k})\right) = o(1).
\end{aligned}$$

Since $\{b \in \Gamma : b \neq a\} = B_1 \cup B_{2.1} \cup B_{2.3} \cup B_3$, 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

$$\begin{aligned}
\sum\nolimits_{a\in\Gamma} \pi_a \sum\nolimits_{b\neq a} \mathbb{E}|J_{ba} - I_b| &= \sum\nolimits_{k=j}^{d} o(1) \cdot \sum\nolimits_{a\in\Gamma_k} \pi_a \\
&= \sum\nolimits_{k=j}^{d} o(1) \cdot \mathbb{E}(X_{j,k}) \\
&\overset{(4.59)}{=} \sum\nolimits_{k=j}^{d} o(1) \cdot O(1) = o(1). \qquad (4.63)
\end{aligned}$$

The right-hand side of (4.57) is therefore

$$\begin{aligned}
\sum\nolimits_{a\in\Gamma} \pi_a \left(\pi_a + \sum\nolimits_{b\neq a} \mathbb{E}|J_{ba} - I_b|\right) &\overset{(4.63)}{=} \left(\sum\nolimits_{a\in\Gamma} \pi_a^2\right) + o(1) \\
&\leq \left(\max_{a\in\Gamma} \pi_a\right) \left(\sum\nolimits_{a\in\Gamma} \pi_a\right) + o(1) \\
&= \left(\max_{a\in\Gamma} \pi_a\right) \left(\sum\nolimits_{k=j}^{d} \mathbb{E}(X_{j,k})\right) + o(1) \\
&\overset{(4.59)}{=} \left(\max_{a\in\Gamma} \pi_a\right) \cdot O(1) + o(1),
\end{aligned}$$

and

$$\begin{aligned}
\max_{a\in\Gamma} \pi_a &\overset{(4.58)}{=} \max_{j\leq k\leq d} \left((1+o(1))p_k \bar{q}^{\,\tau(k-j+1)}\right) \\
&\overset{(4.59)}{=} \max_{j\leq k\leq d} \frac{\mathbb{E}(X_{j,k})}{\Theta(n^{k+1})} = O\left(\frac{1}{n^{j+1}}\right) = o(1).
\end{aligned}$$

In conclusion, we have

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

Since by Corollary 4.5.8, $\lim_{n\to\infty} \mathbb{E}(X_{j,k}) = \mathcal{E}_k$ for every $j \leq k \leq d$, we have

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

as required. □

4.12 Concluding remarks

4.12.1 Non-triviality of cohomology groups

To prove Theorem 4.1.4 (ii), 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 $i \leq k \leq d$ throughout the interval $I_i(\varepsilon)$ (Corollary 4.3.6). However, it is likely that the i-th cohomology group would be non-trivial already in earlier regimes. 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 4.1.5.

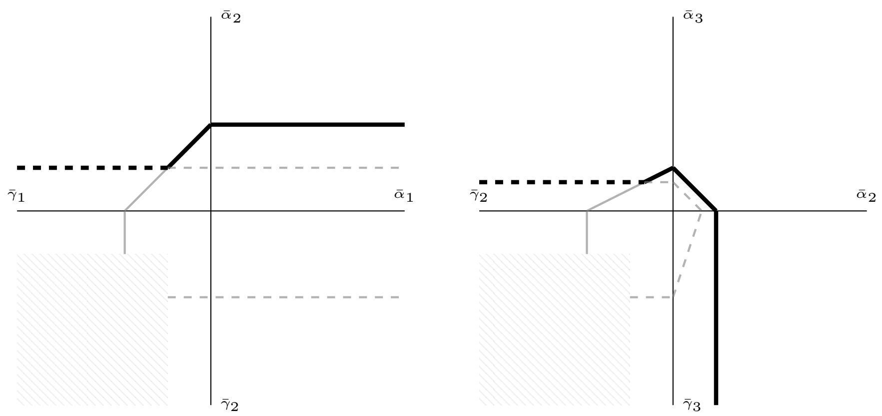
4.12.2 Dimension of the last minimal obstruction

In Theorem 4.1.5 we obtain an asymptotic description of the j-th cohomology group in the critical window. More strongly, in this regime Lemma 4.9.1 yields the asymptotic (joint) distribution of the number of copies of $M_{j,k}$, for every index k with $j \leq k \leq d$. This leads to the natural question: what is the dimension of the last copy of $M_{j,k}$ that vanishes?

4.12.3 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 $j \leq k \leq d$ represent critical dimensions for $\bar{\mathbf{p}}$. Recall from Definition 4.2.6 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 $o(\log n)$. The constant $\bar{\lambda}_k$ depends on the parameters $\bar{\gamma}_k$, $\bar{\alpha}_{j+1}$, …, $\bar{\alpha}_d$. Ignoring the lower order terms, we can therefore plot the ranges where $\bar{\lambda}_k = 0$ as a function of those parameters, for $j \leq k \leq d$.

We present some examples of these plots in Figure 4.4. Further examples can be found at: `https://www.wolframcloud.com/obj/delgiudice/CriticalDimensions`



(a) (b)

Figure 4.4: Plots of the equations $\bar{\lambda}_k = 0$ ($k = 1, 2$ in (a) and $k = 2, 3$ in (b), respectively). For each $i \in [k]$, we use the same axis for $\bar{\alpha}_i$ and $\bar{\gamma}_i$, because property (A1) in Definition 4.2.1 states that only one of $\bar{\alpha}_i, \bar{\gamma}_i$ is non-zero. The black bold parts mark the ranges for which $\bar{\mathbf{p}}$ is a j-critical direction. The striped portions in the lower left corners indicate the regions in which whp $\mathcal{G}_\tau$ has no simplices of positive dimensions.
(a) Plots of $\bar{\lambda}_1 = 0$ (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 edges in $\mathcal{G}_\tau$.
(b) Plots of $\bar{\lambda}_2 = 0$ (plain) and $\bar{\lambda}_3 = 0$ (dashed), for $d = 3$, $j = 1$. In this case, the plots are under the condition that $\bar{p}_1 = 0$.

Acknowledgements

First, I would like to thank my supervisor Mihyun Kang, for her constant presence throughout my Ph.D. studies and for always raising the bar. Thanks to her, I feel a new Nicola inside me, more precise and resolute, perhaps on the right track towards self-confidence. I also thank all the members of my working group, for having been kind and friendly companions of my doctoral adventure. In particular, I thank Oliver Cooley and Philipp Sprüssel, to whom I am deeply grateful for all the mathematical discussions and inputs which led to this thesis.

Moreover, I thank the TU Graz and the doctoral program "Discrete Mathematics", which represented a fruitful learning environment and provided me with unique opportunities, such as a 6-month research stay at ETH Zürich and the participation in several international conferences. They also enabled me to meet many Ph.D. students and more experienced colleagues, who now I consider true friends. I thank them for sharing with me both joyful and frustrating moments of these first steps in the research career. I want to thank in particular Stefan, who often represented a light in the darkness of my uncertainty, either as DK coordinator or as a friendly psychologist.

Most importantly, I am grateful to my family for their incessant moral support, which I always concretely felt despite the geographical distance. Talking about geography, I thank my friends spread all around the world, whose precious love has been the engine which made me achieve this important goal. I mention in particular all my friends in Vienna, together with Claudia and Raffaele in Bologna, who hosted me during (too?) many weekends when I needed to recharge my batteries, and always made me feel at home. After charging these batteries, each time it has been an honour to use the renewed energy with the people I consider my Graz family: Alessandro, Angelica, Francesca, Giulia, Leonardo, Marc, Matthias, Nicole, and Roberta. I thank them for the dinners, the bike rides, the parties, the games, the music, and because they "have been like water when I was feeling so thirsty".

Finally, although I might regret this choice forever, I must end this section in the following way. I need to thank the Italian TV personality and singer Elettra Lamborghini. Perhaps you are not the most gifted artist of all time and you are not well-received by the critics, but I personally believe that with your Latin beat you taught us all a valuable lesson: no matter how insurmountable a problem can appear, just never give up "tanto poi resta la musica e il resto scompare".

Bibliography

[1] F. Ajazi, G. M. Napolitano, and T. Turova, *Phase transition in random distance graphs on the torus*, J. Appl. Probab. **54** (2017), no. 4, 1278–1294.

[2] T. Andriamampianina and V. Ravelomanana, *Enumeration of connected uniform hypergraphs*, Proceedings of 17th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2005), 2005, pp. 387 – 398.

[3] L. Aronshtam and N. Linial, *When does the top homology of a random simplicial complex vanish?*, Random Struct. & Alg. **46** (2015), no. 1, 26–35.

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

[5] E. Babson, C. Hoffman, and M. Kahle, *The fundamental group of random 2-complexes*, J. Amer. Math. Soc. **24** (2011), no. 1, 1–28.

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

[7] M. Behrisch, A. Coja-Oghlan, and M. Kang, *The order of the giant component of random hypergraphs*, Random Struct. & Alg. **36** (2010), no. 2, 149–184.

[8] ———, *Local limit theorems for the giant component of random hypergraphs*, Combin. Probab. Comput. **23** (2014), no. 3, 331–366.

[9] B. Bollobás, *The evolution of random graphs*, Trans. Amer. Math. Soc. **286** (1984), no. 1, 257–274.

[10] ———, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[11] B. Bollobás, S. Janson, and O. Riordan, *The phase transition in inhomogeneous random graphs*, Random Struct. & Alg. **31** (2007), no. 1, 3–122.

[12] B. Bollobás and O. Riordan, *Mathematical results on scale-free random graphs*, Handbook of graphs and networks, Wiley-VCH, Weinheim, 2003, pp. 1–34.

[13] B. Bollobás and O. Riordan, *Asymptotic normality of the size of the giant component in a random hypergraph*, Random Struct. & Alg. **41** (2012), no. 4, 441–450.

[14] ———, *Exploring hypergraphs with martingales*, Random Struct. & Alg. **50** (2017), no. 3, 325–352.

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

[16] F. C. Botelho, N. Wormald, and N. Ziviani, *Cores of random r-partite hypergraphs*, Inform. Process. Lett. **112** (2012), no. 8-9, 314–319.

[17] O. Cooley, N. Del Giudice, M. Kang, and P. Sprüssel, *Vanishing of cohomology groups of random simplicial complexes*, Random Struct. & Alg. **56** (2020), 461–500.

[18] O. Cooley, P. Haxell, M. Kang, and P. Sprüssel, *Homological connectedness of random hypergraphs*, Electronic Notes in Discrete Mathematics **61** (2017), 279–285, The European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB'17).

[19] O. Cooley, M. Kang, and C. Koch, *Threshold and hitting time for high-order connectedness in random hypergraphs*, Electron. J. Combin. **23** (2016), no. 2, Paper 2.48, 14.

[20] ———, *The size of the giant high-order component in random hypergraphs*, Random Struct. & Alg. **53** (2018), no. 2, 238–288.

[21] ———, *The size of the giant component in random hypergraphs: a short proof*, Electron. J. Combin. **26** (2019), no. 3, Paper 3.6, 17.

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

[23] C. Cooper, *The cores of random hypergraphs with a given degree sequence*, Random Struct. & Alg. **25** (2004), no. 4, 353–375.

[24] A. Costa, M. Farber, and D. Horak, *Fundamental groups of clique complexes of random graphs*, Trans. London Math. Soc. **2** (2015), no. 1, 1–32.

[25] R. W. R. Darling and J. R. Norris, *Structure of large random hypergraphs*, Ann. Appl. Probab. **15** (2005), no. 1A, 125–152.

[26] É. de Panafieu, *Phase transition of random non-uniform hypergraphs*, J. Discrete Algorithms **31** (2015), 26–39.

[27] A. Dembo and A. Montanari, *Finite size scaling for the core of large random hypergraphs*, Ann. Appl. Probab. **18** (2008), no. 5, 1993–2040.

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

[29] ———, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61.

[30] ———, *On the strength of connectedness of a random graph*, Acta Math. Acad. Sci. Hungar. **12** (1961), 261–267.

[31] ———, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368.

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

[33] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[34] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016.

[35] C. Hoffman, M. Kahle, and E. Paquette, *Spectral gaps of random graphs and applications*, arXiv:1201.0425.

[36] ———, *The threshold for integer homology in random d-complexes*, Discrete Comput. Geom. **57** (2017), no. 4, 810–823.

[37] S. Janson, *The largest component in a subcritical random graph with a power law degree distribution*, Ann. Appl. Probab. **18** (2008), no. 4, 1651–1668.

[38] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel, *The birth of the giant component*, Random Struct. & Alg. **4** (1993), no. 3, 231–358, With an introduction by the editors.

[39] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[40] M. Kahle, *The neighborhood complex of a random graph*, J. Combin. Theory Ser. A **114** (2007), no. 2, 380–387.

[41] ———, *Topology of random clique complexes*, Discrete Math. **309** (2009), no. 6, 1658–1671.

[42] ———, *Sharp vanishing thresholds for cohomology of random flag complexes*, Ann. of Math. (2) **179** (2014), no. 3, 1085–1107.

[43] ———, *Topology of random simplicial complexes: a survey*, Algebraic topology: applications and new directions, Contemp. Math., vol. 620, Amer. Math. Soc., Providence, RI, 2014, pp. 201–221.

[44] M. Kahle and B. Pittel, *Inside the critical window for cohomology of random k-complexes*, Random Struct. & Alg. **48** (2016), no. 1, 102–124.

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

[46] ———, *The number of connected sparse edged uniform hypergraphs*, Discrete Math. **171** (1997), 153–168.

[47] ———, *The phase transition in a random hypergraph*, J. Comput. Appl. Math. **142** (2002), no. 1, 125–135, Probabilistic methods in combinatorics and combinatorial optimization.

[48] N. Linial and R. Meshulam, *Homological connectivity of random 2-complexes*, Combinatorica **26** (2006), no. 4, 475–487.

[49] N. Linial and Y. Peled, *On the phase transition in random simplicial complexes*, Ann. of Math. (2) **184** (2016), no. 3, 745–773.

[50] T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Struct. & Alg. **1** (1990), no. 3, 287–310.

[51] T. Łuczak and Y. Peled, *Integral homology of random simplicial complexes*, Discrete Comput. Geom. **59** (2018), no. 1, 131–142.

[52] T. Łuczak, B. Pittel, and J. C. Wierman, *The structure of a random graph at the point of the phase transition*, Trans. Amer. Math. Soc. **341** (1994), no. 2, 721–748.

[53] R. Meshulam and N. Wallach, *Homological connectivity of random k-dimensional complexes*, Random Struct. & Alg. **34** (2009), no. 3, 408–417.

[54] M. Molloy, *Cores in random hypergraphs and Boolean formulas*, Random Struct. & Alg. **27** (2005), no. 1, 124–135.

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

[56] D. Poole, *On weak hamiltonicity of a random hypergraph*, arXiv:1410.7446.

[57] ———, *On the strength of connectedness of a random hypergraph*, Electron. J. Combin. **22** (2015), no. 1, Paper 1.69, 16.

[58] V. Ravelomanana and A. L. Rijamamy, *Creation and growth of components in a random hypergraph process*, 12th International Computing and Combinatorics Conference (COCOON) (2006), 350–359.

[59] J. Schmidt-Pruzan and E. Shamir, *Component structure in the evolution of random hypergraphs*, Combinatorica **5** (1985), no. 1, 81–94.

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

[61] T. S. Turova, *The largest component in subcritical inhomogeneous random graphs*, Combin. Probab. Comput. **20** (2011), no. 1, 131–154.

[62] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge University Press, Cambridge, 1999, pp. 239–298.