

Threshold and hitting time for high-order connectedness in random hypergraphs

Oliver Cooley Mihyun Kang Christoph Koch*

Institute of Discrete Mathematics
Graz University of Technology
Steyrergasse 8010, Graz 8010, Austria
{cooley,kang,ckoch}@math.tugraz.at

Submitted: Feb 26, 2015; Accepted: May 31, 2016; Published: Jun 10, 2016
Mathematics Subject Classifications: 05C80, 05C65

Abstract

We consider the following definition of connectedness in k -uniform hypergraphs: two j -sets (sets of j vertices) are j -connected if there is a walk of edges between them such that two consecutive edges intersect in at least j vertices. The hypergraph is j -connected if all j -sets are pairwise j -connected. We determine the threshold at which the random k -uniform hypergraph with edge probability p becomes j -connected with high probability. We also deduce a hitting time result for the random hypergraph process – the hypergraph becomes j -connected at exactly the moment when the last isolated j -set disappears. This generalises the classical hitting time result of Bollobás and Thomason for graphs.

Keywords: random hypergraphs; connectedness; hitting time

1 Introduction

1.1 Preliminaries and main results

In the study of random graphs, one of the most well-known results concerns the *hitting time* for connectedness. More precisely, if we add randomly chosen edges one by one to an initially empty graph on n vertices, then with high probability at the moment the last isolated vertex gains its first edge, the whole graph will also become connected (this classical result was first proved by Bollobás and Thomason in [3]). This interplay between local and global properties is an example of the common phenomenon relating graph properties with their smallest obstruction; the graph can certainly not be connected

*The authors are supported by Austrian Science Fund (FWF): P26826, W1230.

while an isolated vertex still exists, but this smallest obstruction is also the critical one which is last to disappear.

In this paper we generalise the result of Bollobás and Thomason to random k -uniform hypergraphs. For an integer $k \geq 2$, a k -uniform hypergraph consists of a set V of vertices together with a set E of (hyper-)edges, each consisting of k vertices. (A 2-uniform hypergraph is simply a graph.) We need to define the notion of connectedness, for which there is a whole family of possible definitions. For any $1 \leq j \leq k - 1$, we say that two j -sets (sets of j vertices) J_1, J_2 are j -connected if there is a sequence of edges E_1, \dots, E_m such that

- $J_1 \subset E_1$ and $J_2 \subset E_m$;
- $|E_i \cap E_{i+1}| \geq j$ for all $1 \leq i \leq m - 1$.

In other words, we may walk from J_1 to J_2 using edges which consecutively intersect in at least j vertices. A j -component is a maximal set of pairwise j -connected j -sets. A k -uniform hypergraph is j -connected if there is one j -component which contains all j -sets.¹

Note that in the case $k = 2, j = 1$ this is simply the usual definition of connectedness for graphs. More generally, for arbitrary $k \geq 2$ the case $j = 1$ is by far the most well-studied. The definition for general j is also entirely natural, albeit harder to visualise and often requires more complex analysis. In this paper we will be interested in arbitrary $1 \leq j \leq k - 1$ and $k \geq 3$.

There is also more than one model for random hypergraphs. We first define the *uniform model*, the counterpart of the uniform model of Erdős and Rényi for graphs: given any natural numbers k, M, n such that $M \leq \binom{n}{k}$, the random hypergraph $\mathcal{H}^k(n, M)$ is a hypergraph chosen uniformly at random from all k -uniform hypergraphs on vertex set $\{1, \dots, n\}$ which have M edges. This is closely related to the *random hypergraph process* $\{\mathcal{H}^k(n, M)\}_M$ which is defined as follows:

- $\mathcal{H}^k(n, 0)$ is the hypergraph on vertex set $\{1, \dots, n\}$ with no edges;
- For $1 \leq M \leq \binom{n}{k}$, $\mathcal{H}^k(n, M)$ is obtained from $\mathcal{H}^k(n, M - 1)$ by adding an edge chosen uniformly at random from among those k -sets which do not already form an edge.

Note that the distribution of the random hypergraph obtained in the M -th step of the process is the same as in the uniform model $\mathcal{H}^k(n, M)$, so the notation is consistent.

We consider asymptotic properties of random hypergraphs and throughout this paper any asymptotics are as $n \rightarrow \infty$. In particular we say *with high probability* (or *whp*) to mean with probability tending to 1 as $n \rightarrow \infty$.

¹We remark that j -connectedness of a k -uniform hypergraph H is equivalent to vertex-connectedness of the auxiliary $\binom{k}{j}$ -uniform hypergraph H' whose vertices are the j -sets of H and with an edge joining $\binom{k}{j}$ such j -sets if they lie in a common edge of H .

We say that a j -set is *isolated* if it is not contained in any edges. It is trivial to see that if a hypergraph contains isolated j -sets, then it is not j -connected (assuming it has more than j vertices). Our main result is that this trivial smallest obstruction is also the critical one in a random hypergraph.

Let $\tau_c = \tau_c(n, j, k)$ denote the time step in the hypergraph process $\{\mathcal{H}^k(n, M)\}_M$ at which the hypergraph becomes j -connected. Similarly, let τ_i denote the time at which the last isolated j -set disappears. Note that the properties of being j -connected or of having no isolated j -set are certainly monotone increasing properties, so these two variables are well-defined. Furthermore, as noted above, $\tau_i \leq \tau_c$ deterministically.

Theorem 1. *For any $1 \leq j \leq k - 1$ and $k \geq 3$, with high probability in the random hypergraph process $\{\mathcal{H}^k(n, M)\}_M$ we have $\tau_c = \tau_i$.*

The case $j = 1$ of this theorem was already proved by Poole as a special case of the results in [7]. The case $j = k - 1$ was previously proved by Kahle and Pittel in [6]. For all other j , this result was previously unknown.

The uniform model and the associated hypergraph process allow us to formulate exact hitting time results such as Theorem 1. However, the drawback is that the analysis of the model can become tricky due to the fact that the presence of different edges is not independent (the total number is fixed). For this reason, it is often easier to analyse the *binomial model*: $\mathcal{H}^k(n, p)$ is a random k -uniform hypergraph on vertex set $\{1, \dots, n\}$ in which each k -set is an edge with probability p independently of all other k -sets. In Section 2 we will show that if $p = M/\binom{n}{k}$, then the two models are very similar and we can transfer results from one model to the other.

For the proof of Theorem 1 we will also make use of the following result (Theorem 2), which is interesting in itself and is therefore stated in a significantly more general form than we need for Theorem 1. For integer valued random variables Z and Z' we denote their *total variation distance* by $d_{TV}(Z, Z')$, i.e.

$$d_{TV}(Z, Z') = \frac{1}{2} \sum_{i \in \mathbb{Z}} |\mathbb{P}(Z = i) - \mathbb{P}(Z' = i)|.$$

For integer-valued random variables X_n and Y , we say X_n *converges in distribution to* Y , denoted by $X_n \xrightarrow{d} Y$, if for every integer i we have $\mathbb{P}(X_n = i) \rightarrow \mathbb{P}(Y = i)$.

Theorem 2. *For any $k \geq 3$ and $1 \leq j \leq k - 1$ and for any integer $s \geq 0$, let*

$$p_s = p_s(n, k, j) = \frac{j \log n + s \log \log n + c_n}{\binom{n}{k-j}},$$

where $c_n = o(\log n)$, and let D_s be the number of j -sets of degree precisely s in $\mathcal{H}^k(n, p_s)$ (i.e. which lie in s edges). Then we have

$$d_{TV}(D_s, \text{Po}(\mathbb{E}(D_s))) = O(n^{-j}(\log n)^{s+1}). \tag{1}$$

In particular

- (i) $D_s = 0$ whp if $c_n \rightarrow \infty$;
- (ii) $D_s \xrightarrow{d} \text{Po}\left(\frac{j^s e^{-c}}{j!s!}\right)$ if $c_n \rightarrow c$ for any $c \in \mathbb{R}$;
- (iii) $D_s \rightarrow \infty$ whp if $c_n \rightarrow -\infty$.

These two theorems together give the following corollary.

Theorem 3. Let $k \geq 3$ and $1 \leq j \leq k - 1$, and let $p_0 = \frac{j \log n + c_n}{\binom{n}{k-j}}$.

1. If $c_n \rightarrow -\infty$, then with high probability $\mathcal{H}^k(n, p)$ contains isolated j -sets (and is therefore not j -connected).
2. If $c_n \rightarrow \infty$, then with high probability $\mathcal{H}^k(n, p)$ is j -connected (and therefore contains no isolated j -sets).

In other words, the properties of being j -connected and having no isolated j -sets both undergo a (sharp) phase transition at threshold p_{conn} , defined as

$$p_{\text{conn}} = \frac{j \log n}{\binom{n}{k-j}}.$$

1.2 Methods

The main contribution of this paper is to deduce Theorem 1 from Theorem 2. Attempting to prove this directly using standard techniques generalised from the graph case does not work because j -components in a hypergraph may be strangely and non-intuitively distributed. To overcome this problem we quote a powerful result from [4], which guarantees one component with a large subset which is in some sense *smoothly distributed*. We then show that whp all non-trivial components are connected to this smooth subset.

1.3 Notation and definitions

We introduce a few more definitions before we proceed with the proofs. We fix $k \geq 3$ and $1 \leq j \leq k - 1$ for the remainder of the paper. The *order* $|H|$ of a hypergraph H is the number of vertices it contains, while its *size* $e(H)$ is the number of edges. Since a j -component consists of j -sets of vertices, we may view it as a j -uniform hypergraph in which the edges are the j -sets in the component. In particular, the size of a j -component is the number of j -sets it contains. In the remainder of the paper we will use *component* to mean j -component.

We will sometimes need to relate the j -sets of a component to the edges of the hypergraph which connect them. To allow us to do this, for a k -uniform hypergraph H we define the j -size of H to be the number of j -sets contained in edges of H .

We ignore floors and ceilings whenever this does not significantly affect the argument.

2 Contiguity of $\mathcal{H}^k(n, M)$ and $\mathcal{H}^k(n, p)$

We need to know that $\mathcal{H}^k(n, p)$ and $\mathcal{H}^k(n, M)$ are roughly equivalent, which is a generalisation of a standard fact about the corresponding graph models (see [2, 5]). In fact, [5] considers a more general setting than we require here, but what we state is an immediate corollary of the results there (see [5], Corollary 1.16). Let $N = \binom{n}{k}$ and to ease notation, for some property Q we will denote by $\mathbb{P}_M(Q)$ the probability that $\mathcal{H}^k(n, M)$ has property Q . $\mathbb{P}_p(Q)$ is defined similarly.

Lemma 4. *Let Q be some monotone increasing property of k -uniform hypergraphs and let $M = Np \rightarrow \infty$. Then*

1. $\mathbb{P}_p(Q) \rightarrow 1$ implies $\mathbb{P}_M(Q) \rightarrow 1$;
2. $\mathbb{P}_p(Q) \rightarrow 0$ implies $\mathbb{P}_M(Q) \rightarrow 0$.

This lemma allows us to transfer properties from $\mathcal{H}^k(n, p)$ to $\mathcal{H}^k(n, M)$ (transferring in the other direction is also possible, with some small modifications, but we will not need to do this here). However, this only works for monotonically increasing properties. This is fine for the properties of being j -connected or of having no isolated j -sets, but in the proof of Theorem 1 we will need to consider the probability of having a component of size r , for various fixed r . This property is not even convex (and nor is its complement) and so for this case we will need some more careful arguments.

The following standard argument allows us to transfer high probability events from the binomial to the uniform model provided that the failure probability is small enough.

Lemma 5. *Let Q be an arbitrary property. Suppose $M \rightarrow \infty$ and $p = M/N \rightarrow 0$. Then*

$$\mathbb{P}_M(Q) \leq \frac{\mathbb{P}_p(Q)}{\mathbb{P}(e(\mathcal{H}^k(n, p)) = M)} = \Theta(M^{1/2})\mathbb{P}_p(Q).$$

Proof. The inequality follows from the fact that

$$\begin{aligned} \mathbb{P}_p(Q) &= \sum_{m=0}^N \mathbb{P}_m(Q) \mathbb{P}(e(\mathcal{H}^k(n, p)) = m) \\ &\geq \mathbb{P}_M(Q) \mathbb{P}(e(\mathcal{H}^k(n, p)) = M). \end{aligned}$$

For the equality we use Stirling's approximation to deduce that

$$\begin{aligned} \mathbb{P}(e(\mathcal{H}^k(n, p)) = M) &= \binom{N}{M} p^M (1-p)^{N-M} \\ &= \Theta(1) \sqrt{\frac{N}{M(N-M)}} \frac{N^N}{M^M (N-M)^{N-M}} p^M (1-p)^{N-M} \\ &= \Theta(M^{-1/2}). \end{aligned} \quad \square$$

3 Proof of Theorem 2

Let $C = \binom{k}{j} - 1$. Fix an integer $s \geq 0$ and suppose $p = p_s = \frac{j \log n + s \log \log n + c_n}{\binom{n}{k-j}}$, where $c_n = o(\log n)$. Then the expected number of j -sets of degree s in $\mathcal{H}^k(n, p)$ satisfies

$$\begin{aligned} \mathbb{E}(D_s) &= \binom{n}{j} \binom{\binom{n-j}{k-j}}{s} p^s (1-p)^{\binom{n-j}{k-j}-s} \\ &= (1 + o(1)) \frac{n^j}{j!} \frac{\left(\frac{n^{k-j}}{(k-j)!}\right)^s}{s!} p^s \exp\left(-p \binom{n}{k-j}\right) \\ &= (1 + o(1)) \frac{1}{j!s!} e^{s(k-j) \log n - s \log((k-j)!) + s \log p - s \log \log n - c_n} \\ &= (1 + o(1)) \frac{j^s}{j!s!} e^{-c_n}, \end{aligned} \tag{2}$$

since

$$\log p = -(k-j) \log n + \log \log n + \log(j(k-j)!) + O\left(\frac{\log \log n + |c_n|}{\log n} + \frac{1}{n}\right).$$

For the Poisson-approximation we use the Chen-Stein method (see [1]). For any j -set J we denote its degree in $\mathcal{H}^k(n, p)$ by $\deg(J)$ and analyse how D_s changes by conditioning on the event $\{\deg(J_0) = s\}$ for an arbitrary j -set J_0 .

First we construct $\mathcal{H}^k(n, p)$ and denote by E_0 the set of edges containing J_0 , then we distinguish three cases:

- (a) If $\deg(J_0) < s$, add $s - \deg(J_0)$ distinct k -sets chosen uniformly at random from $\{K \in \binom{V}{k} \mid J_0 \subset K\} \setminus E_0$ to the hypergraph;
- (b) If $\deg(J_0) = s$, do nothing;
- (c) If $\deg(J_0) > s$, delete a set of $\deg(J_0) - s$ edges chosen uniformly at random from E_0 .

We denote the resulting hypergraph by $\mathcal{H}^* = \mathcal{H}^*(J_0)$. For any j -set J we write $\deg^*(J)$ for its degree in \mathcal{H}^* and $D_s^*(J_0)$ for the number of j -sets $J \neq J_0$ such that $\deg^*(J) = s$. Furthermore observe that this construction provides a coupling of $\mathcal{H}^k(n, p)$ and \mathcal{H}^* such that removing all edges containing J_0 in either one of them yields the same random hypergraph $\mathcal{H}^- = \mathcal{H}^-(J_0)$. For any j -set J we write $\deg^-(J)$ for its degree in \mathcal{H}^- .

We use the following form of the Chen-Stein approximation given by Theorem 1.B in [1].

Theorem 6 (Chen-Stein approximation [1]). *Given a finite index set \mathcal{I} and a random variable $W = \sum_{i \in \mathcal{I}} Z_i$, where Z_i is a Bernoulli random variable with parameter $p_i \in [0, 1]$, denote by $\lambda = \sum_{i \in \mathcal{I}} p_i$ the expectation of W . Assume that for each $i \in \mathcal{I}$ there is a pair of coupled random variables (U_i, V_i) such that U_i has the distribution of W and $V_i + 1$ has the distribution of W conditioned on $\{Z_i = 1\}$. Then we have*

$$d_{TV}(W, \text{Po}(\lambda)) \leq \min\{1, \lambda^{-1}\} \sum_{i \in \mathcal{I}} p_i \mathbb{E}(|U_i - V_i|).$$

For the proof of Theorem 2, we let \mathcal{I} be the set of all j -sets and for all J let $Z_J = \mathbb{1}_{[\deg(J)=s]}$, $p_J = \mathbb{P}(\deg(J) = s)$, $U_J = W = D_s$ and $V_J = D_s^*(J_0)$. We observe that \mathcal{H}^* has the same distribution as \mathcal{H} conditioned on the event $\{\deg(J_0) = s\}$, so V_J has the same distribution as W conditioned on $\{Z_{J_0} = 1\}$.

Now applying Theorem 6 and using the upper bound $\min\left\{1, \frac{1}{\mathbb{E}(D_s)}\right\} \leq \frac{1}{\mathbb{E}(D_s)}$, we obtain

$$d_{TV}(D_s, \text{Po}(\mathbb{E}(D_s))) \leq \frac{\sum_J \mathbb{P}(\deg(J) = s) \mathbb{E}(|D_s - D_s^*(J_0)|)}{\mathbb{E}(D_s)} = \mathbb{E}(|D_s - D_s^*(J_0)|). \quad (3)$$

Hence it suffices to estimate the random variable $|D_s - D_s^*(J_0)|$. We first observe that

$$\begin{aligned} |D_s - D_s^*(J_0)| &= \mathbb{1}_{[\deg(J_0)=s]} + \sum_{J \neq J_0} |\mathbb{1}_{[\deg(J)=s]} - \mathbb{1}_{[\deg^*(J)=s]}| \\ &\leq \mathbb{1}_{[\deg(J_0)=s]} + \sum_{t=1}^s \sum_{\substack{J \neq J_0 \\ \deg^*(J) > \deg(J)}} \mathbb{1}_{[\deg(J_0)=s-t]} \\ &\quad + \sum_{t=1}^{\binom{n-j}{k-j}-s} \sum_{\substack{J \neq J_0 \\ \deg^*(J) < \deg(J) \\ \deg^-(J) \leq s}} \mathbb{1}_{[\deg(J_0)=s+t]}. \end{aligned}$$

To justify the inequality, first note that if $\deg(J_0) = s$, then $\mathcal{H} = \mathcal{H}^*$ and only the first term contributes. Furthermore, if $\deg(J_0) < s$, say $\deg(J_0) = s - t$ for some $t \in [1, s]$, then the only contribution to $|D_s - D_s^*(J_0)|$ comes from j -sets $J \neq J_0$ whose degree increased, i.e. $\deg^*(J) > \deg(J)$. Similarly, if $\deg(J_0) = s + t$ for some $t \in [1, \binom{n-j}{k-j} - s]$, observe that for a j -set J to contribute it is necessary to have either $\deg(J) = s$ or $\deg^*(J) = s$. Note that these cannot hold unless $\deg^-(J) \leq s$, and we will simply bound the probability of this (more likely) event.

Moreover, each inner sum has at most Ct terms, since we certainly only sum over j -sets J whose degree has changed, and adding or deleting an edge influences the degree of at most C j -sets (other than J_0).

Note also that $\deg^-(J)$ has distribution

$$\text{Bin}\left(\binom{n-j}{k-j} - \binom{n-|J_0 \cup J|}{k-|J_0 \cup J|}, p\right),$$

independently of $\deg(J_0)$, and the probability that $\deg^-(J) \leq s$ is maximised when $|J_0 \cup J|$ is minimised. Hence for an upper bound we will assume that $|J_0 \cup J| = j + 1$. This motivates the definition

$$q = \mathbb{P}\left(\text{Bin}(\tilde{N}, p) \leq s\right),$$

where $\tilde{N} = \binom{n-j}{k-j} - \binom{n-j-1}{k-j-1} = (1 + o(1))\binom{n}{k-j}$. Combining these two arguments we obtain the upper bound

$$|D_s - D_s^*(J_0)| \leq \mathbb{1}_{[\deg(J_0)=s]} + \sum_{t=1}^s \mathbb{1}_{[\deg(J_0)=s-t]} Ct + \sum_{t=1}^{\binom{n-j}{k-j}-s} \mathbb{1}_{[\deg(J_0)=s+t]} \text{Bin}(Ct, q).$$

Therefore, using the notation $x^+ = \max\{x, 0\}$ for any $x \in \mathbb{R}$, we have

$$\mathbb{E}(|D_s - D_s^*(J_0)|) \leq \mathbb{P}(\deg(J_0) = s) + C\mathbb{E}((s - \deg(J_0))^+) + Cq\mathbb{E}((\deg(J_0) - s)^+). \quad (4)$$

We can estimate both probabilities in (4) using

$$\begin{aligned} \mathbb{P}(\deg(J_0) = s) &\leq q = \sum_{i=0}^s \binom{\tilde{N}}{i} p^i (1-p)^{\tilde{N}-i} \\ &\leq O(1)(\tilde{N}p)^s \exp(-\tilde{N}p) \\ &= O((\log n)^s n^{-j}), \end{aligned}$$

where the second and third lines follow because s is bounded. Moreover, we have

$$\mathbb{E}(s - \deg(J_0))^+ \leq s\mathbb{P}(\deg(J_0) \leq s) \leq sq = O((\log n)^s n^{-j})$$

and furthermore

$$\mathbb{E}(\deg(J_0) - s)^+ \leq \mathbb{E}(\deg(J_0)) + s = O(\log n).$$

Therefore (3) and (4) provide (1), i.e.

$$d_{TV}(D_s, \text{Po}(\mathbb{E}(D_s))) = O(n^{-j}(\log n)^{s+1}). \quad (5)$$

Now assume $\lim_{n \rightarrow \infty} c_n = c$. By (2) we know that $\mathbb{E}(D_s) \rightarrow \frac{j^s e^{-c}}{j!s!}$ and by the continuity in λ of the function $\mathbb{P}(\text{Po}(\lambda) = i)$ for each i

$$\text{Po}(\mathbb{E}(D_s)) \xrightarrow{d} \text{Po}\left(\frac{j^s e^{-c}}{j!s!}\right),$$

hence by the triangle inequality and (1), case (ii) in the second claim follows. Cases (i) and (iii) can be easily deduced from case (ii).

4 Proof of Theorem 1

The proof which we present is largely elementary except for the use of Theorem 2, which relies on Theorem 6, and one powerful result from [4] (Lemma 7 below). This result is stated for a much smaller probability than we have in this setting, which is therefore not the optimal range for its application, but nevertheless it will turn out to be strong enough.

We first collect two preliminary results which we will need later (Corollary 8 and Proposition 9 below).

4.1 Smooth subset

The following result from [4] will be key.

Lemma 7. *Suppose $n^{-1/3} \ll \varepsilon \ll 1$ and let $p^* = \frac{1+\varepsilon}{\binom{k}{j-1}\binom{n}{k-j}}$. Then whp there is a j -component of $\mathcal{H}^k(n, p^*)$ with a subset S of at least $\varepsilon^3 n^j$ j -sets satisfying the following property:*

$$\text{Every } (j-1)\text{-set of vertices in } \mathcal{H}^k(n, p^*) \text{ is contained in } (1 \pm o(1)) \frac{|S|}{\binom{n}{j}} n \text{ } j\text{-sets of } S.$$

In other words, we can find a reasonably large subset S of a component which is *smooth* in the sense that all $(j-1)$ -sets are in about the “right” number of j -sets of S . More precisely, we say that a set S of j -sets is *smooth* if every $(j-1)$ -set is contained in $(1 \pm o(1)) \frac{|S|}{\binom{n}{j}} n$ j -sets of S .

We note that Lemma 7 is not stated explicitly in this form in [4], but is implicit in the proof. We therefore give a brief outline of how it can be deduced from the results in that paper. The casual reader who is unfamiliar with [4] may skip the following proof.

Proof. Starting from some j -set J , we explore the j -component containing J using a breadth-first search process $\text{BFS}(J)$. This partitions the j -sets of the j -component into *generations*, which can be numbered according to the order they were discovered in.

We fix a starting j -set J which lies in the largest component of $\mathcal{H}^k(n, p)$, let ∂C_g denote the g -th generation of this search process $\text{BFS}(J)$, and $C_g = \cup_{g' \leq g} \partial C_{g'}$. Then there are generations g_0 and g_1 such that the following statements hold whp.

1. Either $|\partial C_{g_1}| \geq \varepsilon^3 n^j$ or $|C_{g_1}| \geq \varepsilon^{3/2} n^j$;
2. $|C_{g_0}| = o(|C_{g_1}|)$ (and in particular $g_0 < g_1$);
3. Every generation ∂C_g with $g_0 \leq g \leq g_1$ is smooth.

We set $g_0 = i_1(j-1)$, where $i_1(j-1)$ is defined immediately before Lemma 16 in [4], and set $g_1 = i_1$, where i_1 is defined in [4] to be the generation at which one of three stopping conditions (S1), (S2) or (S3) is invoked (see Section 5.4 of [4]). These stopping conditions contain a parameter λ , which we choose to be $\lambda = \varepsilon^{3/2}$.

Property (1) follows from these stopping conditions. We use here the fact that J is in the largest component of $\mathcal{H}^k(n, p)$, which is a giant component whp by Theorem 2 of [4], therefore whp either (S2) or (S3) is invoked at time i_1 (as (S2) would be invoked before (S1)).

Property (2) follows from Lemmas 10, 24 and 25 in [4]. More precisely, whp $g_0 = i_1(j-1) = i_0(j-1) + O(\log n)$, while $g_1 = i_1 \geq i_0(j-1) + \Omega(\varepsilon^{-1} \log n)$ by Lemma 25 of [4]. Furthermore, whp $|C_{i_0(j-1)}| = o(\varepsilon^{3/2} n^j)$ by Lemma 24 of [4] applied with $\ell = 0$. Finally, by Lemma 10 of [4], whp the generations between g_0 and g_1 are at least as large

as those between $i_0(j-1)$ and g_0 , and there are significantly more of them ($\Omega(\varepsilon^{-1} \log n)$ compared to $O(\log n)$).

Finally, Property (3) is given by Lemma 16 in [4] (with $\ell = j-1$ and using the fact that $i_1 \geq i_1(j-1)$).

We now use these three properties to prove the existence of the set S . We make a case distinction based on Property (1). If $|\partial C_{g_1}| \geq \varepsilon^3 n^j$, then we simply set $S = \partial C_{g_1}$, and S is smooth by (3).

On the other hand, if $|C_{g_1}| \geq \varepsilon^{3/2} n^j$, then we let $S = C_{g_1} \setminus C_{g_0-1}$. Then since every generation from g_0 to g_1 is smooth, and since a union of smooth sets is also smooth, we have that S is smooth. Furthermore, $|S| = (1 - o(1))|C_{g_1}| \geq \varepsilon^3 n^j$. \square

Lemma 7 has the following corollary which we will apply later.

Corollary 8. *Suppose $n^{-1/3} \ll \varepsilon \ll 1$ and $p \geq \frac{1+\varepsilon}{\binom{k}{j}-1 \binom{n}{k-j}}$. Then $\mathcal{H}^k(n, p)$ has a j -component containing a set S of size at least $\varepsilon^3 n^j$ such that every $(j-1)$ -set is contained in $(1 \pm o(1)) \frac{|S|}{\binom{n}{j}} n$ j -sets of S .*

Proof. We set $p^* = \frac{1+\varepsilon}{\binom{k}{j}-1 \binom{n}{k-j}}$ and $p' = \frac{p-p^*}{1-p^*}$ and let $\mathcal{H}_1 = \mathcal{H}(n, p^*)$ and $\mathcal{H}_2 = \mathcal{H}(n, p')$ independently. Observe that we may couple in such a way that $\mathcal{H}^k(n, p) = \mathcal{H}_1 \cup \mathcal{H}_2$. Furthermore, by Lemma 7, whp \mathcal{H}_1 has a component containing a smooth set S of the appropriate size. In $\mathcal{H}^k(n, p)$ this component may be bigger than in \mathcal{H}_1 , but certainly still contains S . \square

4.2 Well-constructed hypergraphs

We will also use the following proposition. We say that a hypergraph is *well-constructed* if it can be generated from an initial j -set via a search process, i.e. by successively adding edges such that each edge contains at least one previously discovered j -set and also contains at least one previously undiscovered j -set.

Proposition 9. *Up to isomorphism, the number of well-constructed k -uniform hypergraphs of j -size s is at most 2^{ks^2} .*

Proof. We explore the hypergraph by adding the edges one by one in the order in which it is well-constructed. The resulting hypergraph is uniquely determined, up to isomorphism, by the intersection of each edge with the previous vertices (though we will multiple count the isomorphism classes, this is permissible for an upper bound). When adding the i -th edge, we certainly have at most $(i-1)k$ vertices so far, and so the number of possible intersections is at most $2^{(i-1)k}$. Multiplying over all edges, of which there are certainly at most s (each edge gives at least one new j -set), we have that the number of such hypergraphs is at most $2^{\sum_{i=1}^s (i-1)k} \leq 2^{ks^2}$. \square

4.3 Hitting time

We now proceed with the proof of Theorem 1. Let us consider any p, M satisfying

$$\frac{j \log n - \omega}{\binom{n}{k-j}} \leq p = M/N \leq \frac{j \log n + \omega}{\binom{n}{k-j}}$$

where $\omega = \log \log n$. We apply Theorem 2 (with $s = 0$ and $c_n = \pm\omega$) and Lemma 4 to observe that in both $\mathcal{H}^k(n, p)$ and $\mathcal{H}^k(n, M)$, whp there are isolated j -sets at the lower end of this range but not at the upper end. We will now prove that other than these isolated j -sets, there is just one very large component whp. We certainly know by Corollary 8 that there is a component containing a large smooth set S . For the rest of the proof we fix this component and the set S .

We consider the possibility that there is a second non-trivial component containing r j -sets, and make a case distinction on the size of r . Note that for any j -component of size r in a k -uniform hypergraph, there is a well-constructed subhypergraph of every (up to a constant $\binom{k}{j}$ error) j -size up to r . (More precisely, for any integer $r' \leq r$ there exists an integer r'' with $|r'' - r'| \leq \binom{k}{j}$ and a well-constructed subhypergraph of j -size r'' .)

Let us set $r_0 = \log \log n$.

Case 1: $2 \leq r \leq r_0$. We first observe that in a component of size $r \geq 2$ we must have at least one edge, and therefore at least $\binom{k}{j} \geq k \geq 3$ j -sets, i.e. we automatically have $r \geq 3$.

We show that the expected number of components of size r is very small and apply Markov's inequality. Any component of size r can be associated with a well-constructed hypergraph H of j -size r which is isolated from the remaining j -sets of $\mathcal{H}^k(n, p)$. Then $e(H) \leq r$ and furthermore $|H| \leq j + (k - j)e(H)$, since each new edge of H gives at most $k - j$ new vertices. For each j -set of H , we have at least $\binom{n-j}{k-j} - r \binom{n-j-1}{k-j-1}$ non-edges (any k -set containing this j -set but no other j -sets of H). Thus the expected number of isolated copies of H in $\mathcal{H}^k(n, p)$ satisfies

$$\mathbb{E}(X_H) \leq n^{j+(k-j)e(H)} p^{e(H)} (1-p)^{r \left(\binom{n-j}{k-j} - r \binom{n-j-1}{k-j-1} \right)} \quad (6)$$

and so

$$\begin{aligned} \log(\mathbb{E}(X_H)) &\leq (j + (k - j)e(H)) \log n + O(r \log \log n) \\ &\quad - (k - j)e(H) \log n - (1 - O(r/n) - O(\omega/\log n)) r j \log n \\ &= (1 - r + o(1)) j \log n \leq (-3rj/5) \log n. \end{aligned}$$

Note that this bound does not depend on the specific structure of H , only on the number of j -sets r . Let X_r be the number of components of size r . Then by Proposition 9 we have

$$\mathbb{E}(X_r) \leq 2^{kr^2} n^{-3rj/5} \leq n^{-4rj/7}$$

where for the last inequality we use the fact that $r \leq r_0 = o(\log n)$.

By taking a union bound over all $3 \leq r \leq r_0$, we conclude that with probability at least $1 - 2n^{-12j/7}$ there are no j -components of this size.

Case 2: $r \geq r_0$.

In this case, rather than looking at the full component we look at a well-constructed subgraph H of j -size r_0 . Such a subgraph certainly exists up to an additive $\binom{k}{j}$ error term in the j -size, which will not affect calculations significantly. Most of the calculations which lead to (6) are still valid, replacing r by r_0 . However, since we are no longer considering a full component, we must be more careful about the number of non-edges.

At this point we make use of the set S of j -sets guaranteed by Corollary 8, which lie in a different component to H . For each of the r_0 j -sets of H , pick an arbitrary $(j-1)$ -set within it and by Corollary 8, this $(j-1)$ -set is contained in $(1 \pm o(1))\varepsilon^3 n$ j -sets of S . For each such pair of j -sets intersecting in $j-1$ vertices, there are $\binom{n-j-1}{k-j-1}$ k -sets containing both of them, all of which must be non-edges, since the j -sets lie in different components.

It may be that we multiple count the non-edges in this way. However, each k -set may only be counted from a pair of j -sets it contains, and therefore the number of times it is counted is certainly at most $\binom{k}{j}(k-j)$. Thus in total the number of non-edges is at least

$$\frac{r_0 \varepsilon^3 n}{2 \binom{k}{j} (k-j)} \binom{n}{k-j-1} = \Theta(r_0 \varepsilon^3 n^{k-j}).$$

We may thus calculate the expected number of such structures H (cf. (6)):

$$\mathbb{E}(X_H) \leq n^{j+(k-j)e(H)} p^{e(H)} (1-p)^{\Theta(r_0 \varepsilon^3 n^{k-j})}$$

and so, letting Y be the number of such well-constructed hypergraphs of j -size r_0 which are not in the same component as S , we have

$$\log(\mathbb{E}(Y)) \leq k r_0^2 \log 2 + j \log n + O(r_0 \log \log n) - \Theta(r_0 \varepsilon^3 \log n).$$

Now observe that in Corollary 8 we may choose any $n^{-1/3} \ll \varepsilon \ll 1$. Choosing $\varepsilon^3 = \frac{1}{\log \log \log n}$, we have $r_0 \varepsilon^3 \rightarrow \infty$ and the last term in the above inequality dominates, and we have $\log(\mathbb{E}(Y)) \leq -C \log n$ for any constant C . In particular, choosing $C = 12j/7$, we have $\mathbb{E}(Y) \leq n^{-12j/7}$. By Markov's inequality, this implies that with probability at least $1 - n^{-12j/7}$ we have $Y = 0$ and therefore no further components of size r .

Combining the two cases, this tells us that with probability at least $1 - 3n^{-12j/7}$, $\mathcal{H}^k(n, p)$ only has one non-trivial component.

Finally note that $M = pN = \Theta(n^j \log n)$. Thus by Lemma 5 we conclude that with probability at least $1 - 3n^{-12j/7} \sqrt{M} = 1 - o(n^{-8j/7})$, $\mathcal{H}^k(n, M)$ also has only one non-trivial j -component.

We now take a union bound over all possible M , of which there are at most $\frac{2\omega}{\binom{n}{k-j}} \binom{n}{k} = O(\omega n^j)$, and deduce that the probability that there is ever a second non-trivial within this time period is at most

$$O(\omega n^j) n^{-8j/7} = O(\omega n^{-j/7}) = o(1)$$

as required.

5 Proof of Theorem 3

Theorem 3 follows almost immediately from Theorems 1 and 2. In order to apply Theorem 1 in the binomial model, we apply the standard trick of birth times: to each k -tuple we assign a number (the *birth time*) between 0 and 1 uniformly at random and independently of all other k -tuples. Then the hypergraph process $\{\mathcal{H}^k(n, M)\}_M$ can be obtained by adding edges in increasing order of birth time (with probability 1 no two edges have the same birth time), while the hypergraph obtained by taking all edges with birth time at most p is distributed as $\mathcal{H}^k(n, p)$.

Theorem 6 (with $s = 0$) tells us that if $c_n \rightarrow \infty$, then whp there are no isolated j -sets, and therefore Theorem 1 tells us that whp the hypergraph is j -connected. This proves part (2). Part (1) is simply an application of Theorem 2 with $s = 0$.

6 Concluding remark

In [7], it is determined for the case $j = 1$ that the hitting time for d -strong 1-connectedness, i.e. the time at which the hypergraph first has the property that deleting any set of less than d vertices still leaves a 1-connected hypergraph, is the same as the hitting time for having no vertices of degree less than d whp. It would be interesting to generalise this result to d -strong j -connectedness (removing fewer than d j -sets still leaves a j -connected hypergraph), which is presumably attained whp when every j -set has degree at least d . However, this would present significant additional difficulties, not least that Lemma 7 would no longer give the substructure which we require.

Acknowledgement

The authors would like to thank an anonymous referee for detailed feedback which greatly improved the quality of the manuscript.

References

- [1] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.
- [2] B. Bollobás. *Random Graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [3] B. Bollobás and A. Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.
- [4] O. Cooley, M. Kang, and C. Koch. The size of the giant component in random hypergraphs. submitted. Preprint: [arXiv:1501.07835](https://arxiv.org/abs/1501.07835)

- [5] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [6] M. Kahle and B. Pittel. Inside the critical window for cohomology of random k -complexes. *Random Structures Algorithms*, 48(1):102–124, 2016.
- [7] D. Poole. On the strength of connectedness of a random hypergraph. *Electron. J. Combin.*, 22(1): #P1.69, 16, 2015.