

The size of the giant component in random hypergraphs: a short proof

Oliver Cooley*

Mihyun Kang[†]

Christoph Koch[‡]

Institute of Discrete Mathematics
Graz University of Technology
Steyrergasse 30
8020 Graz, Austria

Department of Statistics
University of Oxford
24–29 St. Giles
Oxford OX1 3LB, U.K.

{cooley,kang}@math.tugraz.at

christoph.koch@stats.ox.ac.uk

Submitted: Mar 8, 2018; Accepted: Jun 10, 2019; Published: Jul 5, 2019

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

We consider connected components in k -uniform hypergraphs for the following notion of connectedness: given integers $k \geq 2$ and $1 \leq j \leq k - 1$, two j -sets (of vertices) lie in the same j -component if there is a sequence of edges from one to the other such that consecutive edges intersect in at least j vertices.

We prove that certain collections of j -sets constructed during a breadth-first search process on j -components in a random k -uniform hypergraph are reasonably regularly distributed with high probability. We use this property to provide a short proof of the asymptotic size of the giant j -component shortly after it appears.

Mathematics Subject Classifications: 05C65, 05C80

1 Introduction and main results

One of the most well-known results in the field of random graphs is the *phase transition* for the emergence of the giant component. Originally observed by Erdős and Rényi and strengthened by various researchers over the years, we may state the result as follows. We say that an event holds *with high probability*, often abbreviated to whp, if its probability tends to 1 as n tends to infinity.

Theorem 1 ([4, 9, 10, 11]). *Let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \rightarrow 0$ and $\varepsilon^3 n \rightarrow \infty$.*

(a) *If $p = \frac{1-\varepsilon}{n}$, then whp all components of $G(n, p)$ have $O(\varepsilon^{-2} \log(\varepsilon^3 n))$ vertices.*

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

[†]Supported by Austrian Science Fund (FWF), I3747 and W1230 II.

[‡]Supported by EPSRC Grant No. EP/N004833/1.

(b) If $p = \frac{1+\varepsilon}{n}$, then whp the largest component of $G(n, p)$ has size $(1 \pm o(1))2\varepsilon n$, while all other components have $O(\varepsilon^{-2} \log(\varepsilon^3 n))$ vertices.

Our focus in this paper is an extension of this result to k -uniform hypergraphs¹, for which we need to know what we mean by a connected component in a hypergraph.

Given integers $k \geq 2$ and $1 \leq j \leq k - 1$ and a k -uniform hypergraph \mathcal{H} , we say that two distinct j -sets (of vertices) J_1, J_2 are j -connected if there is a sequence of edges E_1, \dots, E_m in \mathcal{H} such that:

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

In other words, we can “walk” from J_1 to J_2 using edges which consecutively intersect in at least j vertices. Additionally we say that a j -set is j -connected to itself. Then j -connectedness is an equivalence relation, and a j -connected component (or simply j -component) is an equivalence class of this relation. (Equivalently, a j -component is a maximal set of pairwise j -connected j -sets.) The size of a j -component is the number of j -sets it contains.

This provides a whole family of definitions for connectedness. The case $j = 1$, also known as *vertex-connectedness*, is by far the most studied, but larger j , which we refer to as *high-order connectedness*, provides new and richer challenges.

Given integers n, k, j and a real number $p \in [0, 1]$, let $\mathcal{H}^k(n, p)$ denote the random k -uniform hypergraph with vertex set $[n] := \{1, \dots, n\}$ in which each k -tuple of vertices forms an edge with probability p independently of one another. Furthermore, we define

$$\hat{p}_g = \hat{p}_g(n, k, j) := \frac{1}{\binom{k}{j}-1} \frac{1}{\binom{n}{j}}.$$

The following result is a generalisation of the main result in [7].

Theorem 2. *Let $1 \leq j \leq k - 1$ and let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \rightarrow 0$, and $\varepsilon^3 n^j \rightarrow \infty$ and $\varepsilon^2 n^{1-\delta} \rightarrow \infty$ for some constant $\delta > 0$.*

- (a) *If $p = (1 - \varepsilon)\hat{p}_g$, then whp all j -components of $\mathcal{H}^k(n, p)$ have size at most $O(\varepsilon^{-2} \log n)$.*
- (b) *If $p = (1 + \varepsilon)\hat{p}_g$, then whp the size 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 size at most $o(\varepsilon n^j)$.*

Note that for $j = 1$ the second condition on ε becomes $\varepsilon \gg n^{-1/3}$, which is best possible, while for larger j the condition $\varepsilon \gg n^{-\frac{1-\delta}{2}}$, which is probably not best possible, takes over. We discuss the critical window in more detail in Section 5.

A weaker version of this result appeared in [7] as Theorem 2, where the assumption that $\varepsilon^3 n^j, \varepsilon^2 n^{1-\delta} \rightarrow \infty$ was replaced by the stronger condition $\varepsilon^3 n^{1-2\delta} \rightarrow \infty$, for some

¹A k -uniform hypergraph is an ordered pair formed by a vertex set and an edge set, where each edge contains precisely k distinct vertices.

constant $\delta > 0$. The case $k = 2$ and $j = 1$ is simply Theorem 1. The case $j = 1$ for any $k \geq 2$ was proved by Schmidt-Pruzan and Shamir [12].

The proof of (the weaker version of) Theorem 2 in [7] was based on a short proof of Theorem 1 due to Bollobás and Riordan [6]. The idea is to study an exploration process modelling the growth of components and analyse this process based on a branching process approximation. In the setting of hypergraphs substantial challenges arise when adapting this approach. In order to overcome these obstacles, the proof of Theorem 2 in [7] required significant and lengthy technical details. In this paper we show how much of the technical detail can be avoided and thereby a slightly stronger result can be obtained.

Throughout the paper we fix integers $k \geq 2$ and $1 \leq j \leq k - 1$.

The main contribution of [7] to the proof of Theorem 2 was a result stating that certain collections of j -sets are *smooth* in the sense that for any $0 \leq \ell \leq j - 1$, no ℓ -set is contained in “too many” j -sets of the collection, see Corollary 5 of this paper. Indeed, Corollary 5 followed immediately from a far stronger result in [7]. In this paper we show that we can significantly shorten the proof of Theorem 2 by avoiding this stronger result and proving Corollary 5 more directly.

Let us fix a j -set J_1 and explore the j -component containing J_1 via a breadth-first search process $\text{BFS} = \text{BFS}(J_1)$. More precisely, given (arbitrary) orderings σ_j and σ_k of the j -sets and k -sets respectively, in BFS the first generation consists only of the j -set J_1 . We produce generation $i + 1$ from generation i in the following way. For each j -set J of generation i in order according to σ_j , we query all previously unqueried k -sets containing J , in order according to σ_k . If such a k -set forms an edge, any previously undiscovered j -sets within it are added to generation $i + 1$ and are called *discovered*.

We denote the i -th generation of this process by $\partial(i)$. Note that $\partial(i)$ is a set of j -sets, which we may also view as a j -uniform hypergraph. Thus for $1 \leq \ell \leq j - 1$ and an ℓ -set L , we may define the *degree of L in $\partial(i)$* , denoted $d_L(\partial(i))$, to be the number of j -sets of $\partial(i)$ that contain L . Our goal is to prove that the degrees $d_L(\partial(i))$ behave “nicely”.

Next, we fix a constant δ satisfying $0 < \delta < 1/6$, and think of it as an arbitrarily small constant – in general our results become stronger for smaller δ (the bounds on the error probabilities become weaker, but are still exponentially small). Furthermore, we fix a real-valued function $\lambda = \lambda(n)$ such that

$$n^{-1/2+\delta/2}, n^{-j/3} \ll \lambda \ll \varepsilon \ll 1.$$

We use BFS to grow the component of some j -set J_1 until at the beginning of some round $i \in \mathbb{N}$ one of the following three stopping conditions is reached:

- (S1) the component of J_1 is fully explored (i.e. $\partial(i) = \emptyset$);
- (S2) the (partial) component $\partial(1) \cup \dots \cup \partial(i)$ has reached size at least λn^j ;
- (S3) the i -th generation $\partial(i)$ has reached size at least $\lambda^2 n^j$.

We denote the first round² in which any one of these stopping conditions is invoked by

$$i_1 = i_1(\lambda) := \min_{i \in \mathbb{N}} \{ (S1) \vee (S2) \vee (S3) \text{ holds in round } i \}. \quad (1)$$

²Note that i_1 well-defined since BFS always terminates in finite time.

The key tool in this paper is the following upper bound on the degrees of ℓ -sets in each of the first i_1 generations.

Theorem 3. *For any $0 \leq \ell \leq j - 1$, there exists a constant $C_\ell > 0$ such that for all $i \leq i_1$, with probability $1 - \exp(-\Theta(n^{\delta/2}))$ we have*

$$\Delta_\ell(\partial(i)) \leq C_\ell \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right). \quad (2)$$

Note that Theorem 3 becomes trivial in the case $j = 1$, which is the main reason why vertex-connectedness is so much easier to handle than high-order connectedness.

In Section 2, we show how Theorem 3 can be applied to prove the supercritical case of Theorem 2 in a more concise way than in [7]. The proof of Theorem 3 forms the main body (Section 4) of the paper. The strategy is to split the contribution to the degree of an ℓ -set L into two parts, called “jumps” and “pivots”. The contributions made by each of these parts are bounded separately, in Propositions 11 and 12 respectively.

2 Application of Theorem 3: a simple proof of Theorem 2

In this section we show how Theorem 3 can be used to prove Theorem 2 (b) (i.e. the hard case of Theorem 2) by following the strategy of [6] for the graph case: we aim to determine asymptotically the number L of j -sets in *large* j -components, i.e. containing at least λn^j many j -sets, using the second moment method. Then it is a simple matter to show that in fact, almost all of these j -sets lie in the same component using a sprinkling argument. Since the argument already appeared in [7], we will give only an outline here.

We first explore the j -component C_{J_1} of J_1 using BFS until one of the three stopping conditions (S1), (S2) and (S3) is fulfilled. We define the *partial component* $C_{J_1}(i) := \partial(1) \cup \dots \cup \partial(i)$ for each i .

We can approximate the search process by a Galton-Watson branching process \mathcal{T} starting with a single vertex (which represents a j -set), and in which the number of children of any vertex v is a random variable X_v , where

$$\frac{X_v}{\binom{k}{j} - 1} \sim \text{Bi} \left(\binom{n}{k-j}, p \right)$$

and where the X_v are independent of each other. It is clear that this is an upper coupling for the search process; the fact that the stopping conditions are invoked before the component grows too large ensures that whp it will be a good approximation for sufficiently many rounds (see Lemma 7).

We use this branching process approximation to prove that the probability of the event \mathcal{E} , that one of stopping conditions (S2) and (S3) is invoked, is approximately the same as the survival probability of the associated branching process \mathcal{T} , i.e. the probability that the branching process survives indefinitely.

Claim 4. $\mathbb{P}(\mathcal{E}) = (1 + o(1)) \frac{2\varepsilon}{\binom{k}{j} - 1}$.

This claim follows from standard branching process techniques; for completeness, we include a proof in Appendix B. Furthermore, conditional on (S3) being invoked, C_{J_1} will be large whp. Also if (S2) is invoked, then clearly C_{J_1} is large. This already shows that the first moment of L , the number of j -sets in large components, is approximately $\frac{2\varepsilon}{\binom{k}{j}-1} \binom{n}{j}$.

In order to determine the second moment of L , we consider a second exploration process (of a j -component C_{J_2}) starting from another j -set J_2 outside $C_{J_1}(i_1)$. (The contribution to the second moment from j -sets J_2 inside $C_{J_1}(i_1)$ is easily shown to be negligible.) In order to ensure independence, we first delete all the j -sets of $C_{J_1}(i_1)$ from the hypergraph, so any k -set containing a deleted j -set may no longer be queried. However, since we have not deleted many j -sets, \mathcal{T} is still a good approximation for this search process. Again, the probability that this search process becomes large is approximately $\frac{2\varepsilon}{\binom{k}{j}-1}$.

This almost completes the proof of Theorem 2 (b), but there is one more case to consider, namely that the search process from J_2 remained small, while in fact the component is large, but we did not see this because we had deleted some j -sets. We need to show that the probability of this event is very small, and in particular contributes negligibly to the second moment. We therefore need to know how many queries we have not made because of deleting $C_{J_1}(i_1)$.

The first observation is that, since most of the j -sets of $C_{J_1}(i_1)$ were fully explored, many potential such queries had already been made while exploring C_{J_1} . The only ones we might have missed are the ones containing a j -set of $\partial(i_1)$ and a j -set of C_{J_2} .

However, given two j -sets J, J' , one from each of these components, how many k -sets contain both of them? For $j = 1$ (so in particular for graphs) this is simply $\binom{n-2}{k-2}$, but for $j \geq 2$ the answer is fundamentally dependent on the size of the intersection $J \cap J'$. We therefore need the following corollary of Theorem 3.

Corollary 5. *There exists a real-valued function $\xi = \xi(n)$ satisfying $\xi / \log n \rightarrow \infty$ and $\xi = o(\lambda^2 n)$ and constants $\kappa_1, \dots, \kappa_{j-1} > 0$ (independent of n) such that, conditional on \mathcal{E} , with probability at least $1 - \exp(-\Theta(\xi^{1/4}))$, for all $1 \leq \ell \leq j - 1$ and ℓ -sets L we have*

$$d_L(\partial(i_1)) \leq \kappa_\ell (|\partial(i_1)|n^{-\ell} + \xi). \quad (3)$$

Now given $J \in C_{J_2}$, we can consider all subsets L of J and bound from above the number of j -sets of $\partial(i_1)$ that intersect J in L .

Thus we obtain an upper bound on the number of k -sets that we have not queried because of deleting $C_{J_1}(i_1)$. With some careful calculation, we observe that the expected number of edges among these k -sets is small, and applying Markov's inequality, the probability that there is at least one such edge is small enough that it is negligible.

It remains to show how Corollary 5 can be deduced from Theorem 3.

Proof of Corollary 5. Note that Theorem 3 is a generalisation of Corollary 5 in that it applies to *any* generation, and furthermore, is not conditional on the event \mathcal{E} .

Let us denote by \mathcal{A} the event that (3) (i.e. the conclusion of Corollary 5) holds and by \mathcal{B} the event that (2) (i.e. the conclusion of Theorem 3) holds for all $i \leq i_1$. Then we

have $\mathcal{B} \subset \mathcal{A}$ and therefore

$$\mathbb{P}(\mathcal{A}|\mathcal{E}) \geq 1 - \frac{\Pr(\overline{\mathcal{B}} \wedge \mathcal{E})}{\mathbb{P}(\mathcal{E})} \geq 1 - \frac{\Pr(\overline{\mathcal{B}})}{\mathbb{P}(\mathcal{E})}.$$

Hence, using Theorem 3 and Claim 4, we obtain

$$\mathbb{P}(\mathcal{A}|\mathcal{E}) \geq 1 - \frac{\exp(-\Theta(n^{\delta/2}))}{\Omega(\varepsilon)} = 1 - \exp(-\Theta(n^{\delta/2})),$$

where the final equality we used the fact that $\log(1/\varepsilon) = o(n^{\delta/2})$. □

This also completes the proof of Theorem 2 (b).

Remark 6. As mentioned earlier, Corollary 5 already appeared as Lemma 4 in [7]. The proof was long and complicated, and in fact a much stronger result was proved (in which the degrees of ℓ -sets are asymptotically determined). Using Theorem 3 allows us to avoid much of this work.

3 Preliminaries

Throughout the paper we omit floors and ceilings whenever these do not significantly affect the argument.

We first prove some basic properties of the search process BFS. Most of these results were already stated in [7, 8] in a similar fashion. Because the proofs are short, we include them for local reference. Now we define two abstract branching processes, which will form upper and lower couplings on the search process BFS. For these branching processes, the j -sets will be represented by vertices.

Let \mathcal{T}^* denote a Galton-Watson branching process starting with a single vertex, and in which the number of children of any vertex is distributed as a random variable X , where

$$\frac{X}{\binom{k}{j} - 1} \sim \text{Bi} \left(\binom{n}{k-j}, p \right).$$

It is clear that \mathcal{T}^* forms an upper coupling on BFS, since from any j -set we may query at most $\binom{n}{k-j}$ many k -sets, each such k -set forms an edge with probability p , and if it forms an edge, we discover at most $\binom{k}{j} - 1$ new j -sets. If we actually discover fewer j -sets in BFS, we can artificially add in some dummy j -sets, and equally make some additional dummy queries, to “fill up” to \mathcal{T}^* . We denote this coupling by $\text{BFS} \prec \mathcal{T}^*$.

For the lower coupling we define a similar branching process \mathcal{T}_* , where in this case the number of children of any vertex is distributed as a random variable Y , where

$$\frac{Y}{\binom{k}{j} - 1} \sim \text{Bi} \left((1 - \varepsilon_*) \binom{n}{k-j}, p \right)$$

with some $\varepsilon_* = \varepsilon_*(n) > 0$ satisfying $\lambda \ll \varepsilon_* \ll \varepsilon$, say $\varepsilon_* := \sqrt{\lambda\varepsilon}$. Note that \mathcal{T}_* does not always form a lower coupling for BFS, but early on in the search process, it is very

likely to. More specifically, as long as it is still true that from each j -set we make at least $(1 - \varepsilon_*)\binom{n}{k-j}$ queries to k -sets which contain $\binom{k}{j} - 1$ undiscovered j -sets, \mathcal{T}_* forms a lower coupling, which we denote by $\mathcal{T}_* \prec \text{BFS}$.

We first show that whp \mathcal{T}_* will remain a lower coupling during the period of the search process that we are interested in. We introduce a new stopping time: let i_2 be the first round i for which either (S1) or (S2) is satisfied. Note that we have $i_2 \geq i_1$.

Lemma 7. *With probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ for each $0 \leq \ell \leq j - 1$ and $1 \leq i \leq i_2$ we have*

$$\Delta_\ell(C_{J_1}(i)) = O(\lambda n^{j-\ell}) = o(\varepsilon_* n^{j-\ell}).$$

In particular, with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ for any $1 \leq i \leq i_2$ we have

$$\mathcal{T}_* \prec \text{BFS}.$$

Note that this result has a broadly similar flavour to Corollary 5 and Theorem 3, with the crucial difference that the degree bound applies to the whole component, not to a single generation. While Lemma 7 bears a strong resemblance to Lemma 12 in [8] and Lemmas 14 and 17 in [7], it is not immediately apparent that these directly imply Lemma 7. Since a formal proof of this implication would involve checking tedious technicalities, we instead provide a stand-alone proof for Lemma 7 in Appendix A for completeness.

We use Lemma 7 twice: firstly, we use it in the proof of Claim 4 in Appendix B (with $i \leq i_2$); secondly, we use it to deduce that after seeing a sufficiently large generation in BFS, the next one will typically not be smaller.

Lemma 8. *With probability at least $1 - \exp(-\Theta(n^{\delta/2}))$, for all $1 \leq i \leq i_1$ such that $|\partial(i)| \geq n$ we have*

$$|\partial(i+1)| \geq |\partial(i)|.$$

This lemma will be applied in Section 4 during the proofs of Propositions 11 and 12.

Proof. By the coupling $\mathcal{T}_* \prec \text{BFS}$ provided by Lemma 7, conditional on $|\partial(i)|$, we may couple $|\partial(i+1)|$ with a random variable Z_* such that $|\partial(i+1)| \geq \left(\binom{k}{j} - 1\right) Z_*$, where

$$Z_* \sim \text{Bi} \left((1 - \varepsilon_*) \binom{n}{k-j} |\partial(i)|, p \right).$$

Note that $\mathbb{E}(Z_*) = (1 - \varepsilon_*)(1 + \varepsilon) \left(\binom{k}{j} - 1\right)^{-1} |\partial(i)|$, and thus we obtain

$$\begin{aligned} \mathbb{P}(|\partial(i+1)| < |\partial(i)|) &\leq \mathbb{P} \left(Z_* \leq \left(\binom{k}{j} - 1\right)^{-1} |\partial(i)| \right) \\ &\leq \exp \left(- \frac{((1 - \varepsilon_*)(1 + \varepsilon) - 1)^2 \left(\binom{k}{j} - 1\right)^{-1} |\partial(i)|}{2(1 - \varepsilon_*)(1 + \varepsilon)} \right) \\ &\leq \exp(-\Theta(\varepsilon^2 n)) \leq \exp(-\Theta(n^{\delta/2})), \end{aligned}$$

where the penultimate inequality follows since $|\partial(i)| \geq n$. □

At various points in the proof of Theorem 3 the following result, which is a direct consequence of a Chernoff bound, will be more convenient than the standard Chernoff bound formulations.

Lemma 9. *For any $m = m(n) \in \mathbb{N}$ and $p = p(n) \in [0, 1]$ and constant $a > 0$ the following holds. Let X be a random variable distributed as $\text{Bi}(m, p)$. Then with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$,*

$$X \leq (1 + a)mp + 2n^\delta.$$

Proof. We split the proof into two cases. If $mp \leq n^\delta$, then X is dominated by a random variable $Y \sim \text{Bi}(m, p')$, where $p' := n^\delta/m$ (if $p' > 1$, then $m < n^\delta$ and the upper bound is trivial). Let $\mu := \mathbb{E}(Y) = n^\delta$. Now using a multiplicative Chernoff bound, which states that

$$\mathbb{P}(Y \geq (1 + \eta)\mu) \leq \left(\frac{e^\eta}{(1 + \eta)^{1+\eta}} \right)^\mu$$

we have

$$\mathbb{P}(X \geq 2n^\delta) \leq \mathbb{P}(Y \geq 2n^\delta) = \mathbb{P}(Y \geq 2\mu) \leq \left(\frac{e}{2^2} \right)^{n^\delta} = \exp(-\Theta(n^\delta)).$$

On the other hand, if $mp > n^\delta$, then by an additive Chernoff bound, we obtain

$$\mathbb{P}(X \geq (1 + a)mp) \leq \exp\left(-\frac{a^2mp}{2(a + 1/3)}\right) = \exp(-\Omega(n^\delta)),$$

and the statement follows. □

4 Proof of main result

We prove Theorem 3 for a set of recursively defined constants C_ℓ . (Note that we have made no attempt to optimise these constants.) First we define $c_\ell := \binom{k-\ell}{j-\ell} - 1$ for each $0 \leq \ell \leq j-1$, and set $w_0 = w_0(\ell) := \max\{0, j + \ell - k\}$ and $r_\ell := \frac{c_\ell}{c_0}$. Note that $r_\ell < 1$ for $\ell > 0$. Therefore we can fix some constant $\alpha > 0$ such that $r'_\ell := (1 + \alpha)(1 + \varepsilon)r_\ell < 1$ for all $1 \leq \ell \leq j-1$.

Now we set $C_0 := 1$ and iteratively for $1 \leq \ell \leq j-1$ we let

$$C'_\ell := \binom{k-\ell}{j-\ell} \max \left\{ \frac{(1 + \alpha)(1 + \varepsilon)(k-j)!}{c_0} \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} \frac{C_w}{(k-j-\ell+w)!}, 3 \right\} \quad (4)$$

$$C_\ell := \frac{C'_\ell + 2c_\ell + 1}{1 - r'_\ell}. \quad (5)$$

Let us observe that $C'_\ell = f(C_0, \dots, C_{\ell-1}; k, j, \ell)$, while $C_\ell = g(C'_\ell; k, j, \ell)$, so these constants are recursively well-defined.

We will prove Theorem 3 by induction on ℓ ; the base case $\ell = 0$ is trivial, since then $\Delta_\ell(\partial(i)) = |\partial(i)| = C_0 \frac{|\partial(i)|}{n^\ell}$ for any i . Therefore in the remainder of the proof we will assume that $\ell \geq 1$ and that Theorem 3 holds for all ℓ' with $0 \leq \ell' \leq \ell - 1$.

The proof of Theorem 3 relies on distinguishing two types of contribution to the degrees of ℓ -sets.

Definition 10. Let L be an ℓ -set with $1 \leq \ell \leq j - 1$ and let $i \geq 1$.

- (i) A *jump* to L (in round i) occurs when we query a k -set containing L from a j -set in $\partial(i - 1)$ which did not contain L and the k -set forms an edge of $\mathcal{H}^k(n, p)$. Such an edge contributes at most $\binom{k-\ell}{j-\ell}$ to $d_L(\partial(i))$.
- (ii) A *pivot* at L (in round i) occurs when we query any k -set from a j -set in $\partial(i - 1)$ containing L and it forms an edge of $\mathcal{H}^k(n, p)$. Such an edge contributes at most $\binom{k-\ell}{j-\ell} - 1$ to $d_L(\partial(i))$.

The following two propositions bound the contribution to the degree $d_L(\partial(i))$ made by jumps and pivots, respectively. We first state these propositions and then show how they combine to prove Theorem 2, before providing their proofs.

Proposition 11. For any $0 \leq \ell \leq j - 1$, any ℓ -set L and any $2 \leq i \leq i_1$, with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ the contribution made to $d_L(\partial(i))$ by jumps is

$$d_L^{(jp)}(\partial(i)) \leq C'_\ell \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right).$$

Proposition 12. For any $0 \leq \ell \leq j - 1$, any ℓ -set L and any $2 \leq i \leq i_1$, with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ the contribution made to $d_L(\partial(i))$ by pivots is

$$d_L^{(pv)}(\partial(i)) \leq (r'_\ell C_\ell + 2c_\ell + 1) \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right).$$

Proof of Theorem 3. Now assuming Propositions 11 and 12 hold, we note that for any $0 \leq \ell \leq j - 1$, any ℓ -set L and any $2 \leq i \leq i_1$, assuming the conclusions of Propositions 11 and 12 hold, we have

$$d_L(\partial(i)) = d_L^{(jp)}(\partial(i)) + d_L^{(pv)}(\partial(i)) \leq (r'_\ell C_\ell + 2c_\ell + 1 + C'_\ell) \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right) = C_\ell \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right)$$

since $r'_\ell C_\ell + 2c_\ell + 1 + C'_\ell \stackrel{(5)}{=} r'_\ell C_\ell + (1 - r'_\ell)C_\ell = C_\ell$.

Taking a union bound over all choices of ℓ , L and i (of which there are certainly at most jn^{2j} , and observing that $jn^{2j} \exp(-\Theta(n^{\delta/2})) = \exp(-\Theta(n^{\delta/2}))$), this completes the proof of Theorem 3. \square

4.1 Jumps: proof of Proposition 11

Fix some ℓ , L and i and let $m'_2 = m'_2(L, i)$ denote the number of queries to be made in generation i that would result in jumps to L if an edge is found. Given a j -set J , the number of k -sets containing L and J is 0 if $|J \cap L| < w_0$, and at most $\binom{n}{k-j-\ell+w}$ if

$|J \cap L| = w \geq w_0$. Thus we consider the number of j -sets in $\partial(i-1)$ which intersect L in w vertices with $w_0 \leq w \leq \ell-1$. This is at most $\binom{\ell}{w} \Delta_w(\partial(i-1))$. Consequently, we have

$$\begin{aligned} m'_2 &\leq \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} \Delta_w(\partial(i-1)) \binom{n}{k-j-\ell+w} \\ &\leq \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} C_w \left(\frac{|\partial(i-1)|}{n^w} + n^\delta \right) \frac{n^{k-j-\ell+w}}{(k-j-\ell+w)!} =: m_2, \end{aligned}$$

with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$, where we have used the induction hypothesis for the second inequality.

The number of such edges we discover is dominated by $\text{Bi}(m_2, p)$, and so by Lemma 9, with probability $1 - \exp(-\Theta(n^{\delta/2}))$ is at most

$$\left(1 + \frac{\alpha}{2}\right) m_2 p + 2n^\delta.$$

Furthermore, each such edge results in at most $\binom{k-\ell}{j-\ell}$ new j -sets containing L being added to $\partial(i)$. Thus we have

$$\begin{aligned} \frac{d_L^{(\text{jp})}(\partial(i))}{\binom{k-\ell}{j-\ell}} &\leq \left(1 + \frac{\alpha}{2}\right) \frac{1 + \varepsilon}{c_0 \binom{n}{k-j}} \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} C_w \left(\frac{|\partial(i-1)|}{n^w} + n^\delta \right) \frac{n^{k-j-\ell+w}}{(k-j-\ell+w)!} + 2n^\delta \\ &= \frac{\left(1 + \frac{\alpha}{2}\right) (1 + \varepsilon) (k-j)!}{c_0 (1 - O(1/n))} \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} \frac{C_w}{(k-j-\ell+w)!} \left(\frac{|\partial(i-1)|}{n^\ell} + n^{-\ell+w+\delta} \right) + 2n^\delta. \end{aligned}$$

We now observe that since $w \leq \ell-1$, the term in $n^{-\ell+w+\delta}$ is always $O(n^{-1+\delta})$, and therefore we may absorb all such terms into the n^δ term by increasing the constant slightly.

We would also like to replace $|\partial(i-1)|$ by $|\partial(i)|$. This is certainly possible for an upper bound if $|\partial(i-1)| \geq n$ by Lemma 8. However, if $|\partial(i-1)| \leq n$, we simply observe that $\frac{|\partial(i-1)|}{n^\ell} \leq n^{1-\ell} = o(n^\delta)$ (because we have $\ell \geq 1$), and in this case we have $\frac{d_L^{(\text{jp})}(\partial(i))}{\binom{k-\ell}{j-\ell}} \leq 3n^\delta$.

Thus in either case we have

$$\begin{aligned} \frac{d_L^{(\text{jp})}(\partial(i))}{\binom{k-\ell}{j-\ell}} &\leq \frac{(1 + \alpha)(1 + \varepsilon)(k-j)!}{c_0} \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} \frac{C_w}{(k-j-\ell+1)!} \left(\frac{|\partial(i)|}{n^\ell} \right) + 3n^\delta \\ &\stackrel{(4)}{\leq} \frac{C'_\ell}{\binom{k-\ell}{j-\ell}} \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right) \end{aligned}$$

as claimed. □

4.2 Pivots: proof of Proposition 12

Fix some ℓ , L and i and let $m'_3 = m'_3(L, i)$ denote the number of queries to be made in generation i that would result in pivots at L if an edge is found. The number of j -sets in

$\partial(i-1)$ which could lead to a pivot is at most $\Delta_\ell(\partial(i-1))$ and from each such j -set we make at most $\binom{n}{k-j}$ many queries. Thus, with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$, we obtain

$$m'_3 \leq \Delta_\ell(\partial(i-1)) \binom{n}{k-j} \leq C_\ell \left(\frac{|\partial(i-1)|}{n^\ell} + n^\delta \right) \binom{n}{k-j} =: m_3,$$

where we have used the induction hypothesis for the second inequality.

The number of such edges we discover is dominated by $\text{Bi}(m_3, p)$, and so by Lemma 9, with probability $1 - \exp(-\Theta(n^{\delta/2}))$ is at most

$$(1 + \alpha)m_3p + 2n^\delta.$$

Furthermore, each such edge contributes at most $\binom{k-\ell}{j-\ell} - 1 = c_\ell$ to $d_L(\partial(i))$ (the j -set from which we are currently querying is already in $\partial(i-1)$ rather than $\partial(i)$). Thus we have

$$\begin{aligned} \frac{d_L^{(\text{pv})}(\partial(i))}{c_\ell} &\leq (1 + \alpha) \frac{1 + \varepsilon}{c_0 \binom{n}{k-j}} C_\ell \left(\frac{|\partial(i-1)|}{n^\ell} + n^\delta \right) \binom{n}{k-j} + 2n^\delta \\ d_L^{(\text{pv})}(\partial(i)) &\leq (1 + O(1/n)) r'_\ell C_\ell \frac{|\partial(i-1)|}{n^\ell} + (r'_\ell C_\ell + 2c_\ell) n^\delta, \end{aligned}$$

where we recall that $r'_\ell = (1 + \alpha)(1 + \varepsilon)c_\ell/c_0$.

As before, if $|\partial(i-1)| \geq n$, we may assume $|\partial(i-1)| \leq |\partial(i)|$ by Lemma 8, while if $|\partial(i-1)| \leq n$, we have $\frac{|\partial(i-1)|}{n^\ell} = o(n^\delta)$. In either case we have

$$\begin{aligned} d_L^{(\text{pv})}(\partial(i)) &\leq (1 + O(1/n)) r'_\ell C_\ell \frac{|\partial(i)|}{n^\ell} + (r'_\ell C_\ell + 2c_\ell + 1) n^\delta \\ &\leq (r'_\ell C_\ell + 2c_\ell + 1) \left(\frac{|\partial(i)|}{n^\ell} + n^\delta \right) \end{aligned}$$

as claimed. □

5 Concluding remarks

There are several topics which have been studied extensively for random graphs but remain open questions in random hypergraphs, particularly for $j \geq 2$.

5.1 Critical window

With our improvements on Theorem 2 compared to [7], we have come one step closer to determining the width of the critical window for the emergence of a unique largest j -component in $\mathcal{H}^k(n, p)$. However, the lower bound on ε given by $\varepsilon^2 n^{1-\delta} \rightarrow \infty$ is probably still not best possible for $j \geq 2$. We conjecture the following:

Conjecture 13. Theorem 2 holds for all $\varepsilon = \varepsilon(n)$ satisfying $\varepsilon \rightarrow 0$ and $\varepsilon^3 n^j \rightarrow \infty$. Furthermore, if $p = (1 + cn^{-j/3})\hat{p}_g$ for some fixed $c \in \mathbb{R}$, then whp all j -components are of size $O(n^{2j/3})$ and there is more than one j -component of $\mathcal{H}^k(n, p)$ of size $\Theta(n^{2j/3})$.

Note that if $\varepsilon^3 n^j$ is constant, the bounds on the size of the largest component from the super-critical case ($\Theta(\varepsilon n^j)$) and the sub-critical case ($O(\varepsilon^{-2} \log n)$) match up to the $\log n$ term, suggesting that we have a smooth transition. In particular, this condition is also sufficient for the sprinkling argument in Section 2 to work.

5.2 Asymptotic normality of the giant

In the supercritical case, it would be interesting to determine the asymptotic distribution of the size of the giant component more precisely, as was done for the case $j = 1$ in [3, 5], where the size of the giant component was shown to tend to a normal distribution.

5.3 Nullity

One can also consider the structure of the components in the subcritical, critical, or supercritical regimes. For graphs it is well-known that whp all components in the subcritical regime, and whp all except the giant in the supercritical regime, are either trees or contain at most one cycle. Thus we aim to generalise the notion of a tree to j -connectedness. One possibility is via the nullity (with respect to j -connectedness) of a k -uniform hypergraph $\mathcal{H} = (V, E)$, which we define to be

$$\nu = \nu(\mathcal{H}; j, k) := |\mathcal{C}| + c_0 |E| - \left| \binom{V}{j} \right|,$$

where \mathcal{C} denotes the set of j -components of \mathcal{H} , and we recall that $c_0 = \binom{k}{j} - 1$. Then, for a collection $\mathcal{J} \subset \binom{V}{j}$ and an edge set $E \subset \binom{V}{k}$ we define the pair (\mathcal{J}, E) to be a (*hyper-*)*tree* if \mathcal{J} is a j -connected component in \mathcal{H} and $\nu = 0$. (Note that this is not the only reasonable definition of a tree in hypergraphs.)

In contrast to graphs, a j -component does not necessarily contain a spanning tree, but in [7] it was asked what the minimal nullity of a spanning structure in the giant j -component after the phase transition is. We might also ask about the nullities of other components in the subcritical and supercritical regimes, and whether these are small whp, as is the case for graphs as shown in [10, 11]. In fact, this approach also provides crucial insights on the component structure of random graphs within the critical window and has laid the foundation for Aldous' [2] remarkable scaling limit description of the largest components in terms of Brownian motion, see also [1].

References

- [1] L. Addario-Berry, N. Broutin, and C. Goldschmidt, *The continuum limit of critical random graphs*, Probability Theory and Related Fields **152** (2012), no. 3, 367–406.
- [2] D. Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, The Annals of Probability **25** (1997), no. 2, 812–854.

- [3] M. Behrisch, A. Coja-Oghlan, and M. Kang, *Local limit theorems for the giant component of random hypergraphs*, *Combin. Probab. Comput.* **23** (2014), no. 3, 331–366. MR 3189416
- [4] B. Bollobás, *The evolution of random graphs*, *Trans. Amer. Math. Soc.* **286** (1984), no. 1, 257–274. MR 756039 (85k:05090)
- [5] B. Bollobás and O. Riordan, *Asymptotic normality of the size of the giant component in a random hypergraph*, *Random Structures Algorithms* **41** (2012), no. 4, 441–450. MR 2993129
- [6] ———, *A simple branching process approach to the phase transition in $G_{n,p}$* , *Electron. J. Combin.* **19** (2012), no. 4, #P21. MR 3001658
- [7] O. Cooley, M. Kang, and C. Koch, *The size of the giant high-order component in random hypergraphs*, *Random Structures Algorithms* **53** (2018), no. 2, 238–288. MR 3845097
- [8] O. Cooley, M. Kang, and Y. Person, *Largest components in random hypergraphs*, *Combin. Probab. Comput.* **27** (2018), no. 5, 741–762. MR 3868008
- [9] P. Erdős and A. Rényi, *On the evolution of random graphs*, *Bull. Inst. Internat. Statist.* **38** (1961), 343–347. MR 0148055 (26 #5564)
- [10] S. Janson, D. Knuth, T. Łuczak, and B. Pittel, *The birth of the giant component*, *Random Structures Algorithms* **4** (1993), no. 3, 231–358. MR 1220220
- [11] T. Łuczak, *Component behavior near the critical point of the random graph process*, *Random Structures Algorithms* **1** (1990), no. 3, 287–310. MR 1099794 (92c:05139)
- [12] J. Schmidt-Pruzan and E. Shamir, *Component structure in the evolution of random hypergraphs*, *Combinatorica* **5** (1985), no. 1, 81–94. MR 803242 (86j:05106)

A Lower coupling

Our aim in this section is to prove Lemma 7. We first prove that with very high probability

$$|C_{J_1}(i_2)| \leq 3\lambda n^j. \quad (6)$$

For by the definition of i_2 we have $|\partial(i_2 - 1)| \leq |C_{J_1}(i_2 - 1)| < \lambda n^j$, and therefore the number of queries that we make while exploring generation i_2 is at most

$$\lambda n^j \binom{n}{k-j}.$$

Therefore the expected number of edges we discover while exploring generation i_1 is at most $\lambda n^j \binom{n}{k-j} p = \frac{1+\varepsilon}{c_0} \lambda n^j$, and by a Chernoff bound, with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ the number of edges we discover is at most $\frac{2\lambda n^j}{c_0}$. Each such edge gives rise to at most c_0 new j -sets and therefore with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ we have $|\partial(i_2)| \leq 2\lambda n^j$. Thus we obtain $|C_{J_1}(i_2)| < \lambda n^j + 2\lambda n^j = 3\lambda n^j$, as claimed.

Proof of Lemma 7. We now prove the first assertion of Lemma 7 by induction on ℓ . Note that by monotonicity we may assume $i = i_2$. The base case $\ell = 0$ follows immediately from (6), so assume $\ell \geq 1$ and that $\Delta_{\ell'}(C_{J_1}(i_2)) \leq S_{\ell'} \lambda n^{j-i}$ for each $0 \leq \ell' \leq \ell - 1$ and some constants $S_0, \dots, S_{\ell-1}$.

Fix an ℓ -set L . Let us consider how the degree of L in $C_{J_1}(i_2)$ might grow. We bound the contribution from jumps and pivots separately.

We first consider how many queries we may make from a j -set $J \in C_{J_1}(i_2)$ not containing L to a k -set containing L , i.e. the number of queries which might lead to a jump to L . Given J , the number of k -sets containing L and J is at most $\binom{n}{k-j-\ell+w}$, where $w = |J \cap L|$. Thus we consider the number of j -sets in $C_{J_1}(i_2)$ which intersect L in w vertices. This is at most $\binom{\ell}{w} \Delta_w(C_{J_1}(i_2))$. Thus the number of queries which might result in a jump to L is

$$\begin{aligned} \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} \Delta_w(C_{J_1}(i_2)) \binom{n}{k-j-\ell+w} &\leq \sum_{w=w_0}^{\ell-1} \binom{\ell}{w} S_w \lambda n^{k-\ell} \\ &\leq 2^\ell \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{k-\ell}. \end{aligned}$$

By Lemma 9 and the fact that $\lambda p = \omega(n^{j-k-1+\delta})$, the number of edges we find in this way is at most $2p \lambda n^{k-\ell} 2^\ell \max_{w=w_0, \dots, \ell-1} S_w$ with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$. Finally, each such edge contributes at most $\binom{k-\ell}{j-\ell} \leq 2^k$ to the degree of L , and so the contribution made by jumps is at most

$$2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda p n^{k-\ell} \leq 2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{j-\ell}$$

with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$.

On the other hand, let us consider the pivots at L as forming a set of *pivot processes* – if we discover a j -set J' from a j -set J via a pivot at L , then J and J' both contain L and are part of the same pivot process. Each j -set arising from a jump to L (and possibly also J_1 if this contains L) gives rise to such a pivot process at L . A pivot process is a search process on j -sets containing L in a k -uniform hypergraph. By removing L from each of the j -sets, it becomes a search process on $(j-\ell)$ -sets in a $(k-\ell)$ -uniform hypergraph. We note that the number of children of any $(j-\ell)$ -set in such a process is dominated by a random variable $X^{(\text{pv})}$ with

$$\frac{X^{(\text{pv})}}{\binom{k-\ell}{j-\ell} - 1} \sim \text{Bi} \left(\binom{n}{k-j}, p \right)$$

and we therefore define $\mathcal{T}^{(\text{pv})}$ to be an abstract branching process on vertices (which represent j -sets containing L) in which the number of children of each vertex has this distribution. The expected number of children is

$$(1 + \varepsilon) \frac{\binom{k-\ell}{j-\ell} - 1}{\binom{k}{j} - 1} < 1,$$

where the inequality follows since $\ell \geq 1$ and $\varepsilon = o(1)$. In other words, $\mathcal{T}^{(\text{pv})}$ is subcritical. We can then show that with probability $1 - \exp(-\Theta(n^{\delta/2}))$, if we start $x \geq n^\delta$ such processes, then their combined total size is $O(x)$.

More precisely, consider a set of $x \geq n^\delta$ independent copies of $\mathcal{T}^{(\text{pv})}$. If $x - 1$ is the maximum of $\binom{k-\ell}{j-\ell}$ times the number of jumps to L and n^δ , then the total number of vertices in these x processes dominates the number of pivots at L .

We imagine generating children via a sequence of Bernoulli queries with success probability p , each success giving rise to $\binom{k-\ell}{j-\ell} - 1$ children. In order for the processes to reach total size Cx for some constant C , we would need at least $\frac{(C-1)x}{\binom{k-\ell}{j-\ell} - 1}$ of the first $Cx \binom{n}{k-j}$ queries to be successful. But the probability of this can be bounded by a Chernoff bound:

$$\begin{aligned} \mathbb{P} \left(\text{Bi} \left(Cx \binom{n}{k-j}, p \right) \geq \frac{(C-1)x}{\binom{k-\ell}{j-\ell} - 1} \right) &\leq \exp \left(- \frac{x^2 \left(\frac{C-1}{c_\ell} - Cp \binom{n}{k-j} \right)^2}{2 \left(Cx \binom{n}{k-j} p + 1/3 \right)} \right) \\ &\leq \exp \left(-x \frac{\left(\frac{C-1}{c_\ell} - \frac{C(1+\varepsilon)}{c_0} \right)^2}{2 \left(\frac{(1+\varepsilon)C}{c_0} + \frac{1}{3} \right)} \right) \\ &\leq \exp(-\Theta(n^\delta)), \end{aligned}$$

where the last line follows for sufficiently large C because $c_\ell < c_0$ and $\varepsilon = o(1)$ and because $x \geq n^\delta$.

Now recall that with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$ the number of jumps to L is at most $\leq 2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{j-\ell}$, therefore for sufficiently large C , the total contribution to the degree of L made by pivots is at most

$$C \left(2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{j-\ell} + 1 \right),$$

since $\lambda n^{j-\ell} \geq \lambda n = \omega(n^\delta)$.

Consequently, the total number of j -sets containing L is at most

$$(C+1) \left(2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{j-\ell} + 1 \right) \leq (C+2) 2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\} \lambda n^{j-\ell}$$

with probability at least $1 - \exp(-\Theta(n^{\delta/2}))$. We obtain the inductive step by setting $S_\ell := (C+2) 2^{k+\ell+1} \max_{w=w_0, \dots, \ell-1} \{S_w\}$ and taking a union bound over all ℓ -sets L (since $\binom{n}{\ell} \exp(-\Theta(n^{\delta/2})) = \exp(-\Theta(n^{\delta/2}))$).

Finally, we prove that \mathcal{T}_* is a lower coupling. For given a j -set $J \in \partial(i)$, we bound the number of k -sets K containing another j -set $J' \in C_{J_1}(i)$, and which therefore would give fewer than $c_0 = \binom{k}{j} - 1$ new j -sets. We distinguish cases based on $\ell = |J \cap J'|$ and observe that the number of such k -sets is at most

$$\sum_{\ell=\max\{0, 2j-k\}}^{j-1} \Delta_\ell(C_{J_1}(i)) \binom{n}{k-2j+\ell} \leq \sum_{\ell=0}^{j-1} o(\varepsilon_* n^{j-\ell}) n^{k-2j+\ell} = o(\varepsilon_* n^{k-j}).$$

Thus the number of k -sets that can be queried from J and which contain no further discovered j -sets is

$$\binom{n-j}{k-j} - o(\varepsilon_* n^{k-j}) = (1 - O(\frac{1}{n}) - o(\varepsilon_*)) \binom{n}{k-j} \geq (1 - \varepsilon_*) \binom{n}{k-j}$$

as required. □

B Survival probability

Proof of Claim 4. Let \mathcal{A} denote the event that $\mathcal{T}_* \prec \text{BFS}$ for all $i \leq i_2$. We observe that $\mathbb{P}(\mathcal{A}) \geq 1 - \exp(-\Theta(n^{\delta/2}))$ by Lemma 7. Thus we have

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\geq \mathbb{P}(|C_J(i_2)| \geq \lambda n^j) \geq \mathbb{P}(|C_J(i_2)| \geq \lambda n^j \wedge \mathcal{A}) \\ &\geq \mathbb{P}(|\mathcal{T}_*| \geq \lambda n^j \wedge \mathcal{A}) \\ &\geq \mathbb{P}(\mathcal{T}_* \text{ survives}) - \mathbb{P}(\neg \mathcal{A}) \\ &\geq \mathbb{P}(\mathcal{T}_* \text{ survives}) - \exp(-\Theta(n^{\delta/2})). \end{aligned}$$

On the other hand, note that by some elementary calculation (see Appendix A in [7]) the following holds: if we condition on the process \mathcal{T}^* dying out, we obtain a subcritical Galton-Watson branching process $\tilde{\mathcal{T}}$ where the number of children of each individual is distributed as a random variable \tilde{X} , where

$$\frac{\tilde{X}}{\binom{k}{j} - 1} \sim \text{Bi} \left(\binom{n}{k-j}, \tilde{p} \right)$$

for some $\tilde{p} = \tilde{p}(n)$ satisfying $\left(\binom{k}{j} - 1 \right) \binom{n}{k-j} \tilde{p} = 1 - \varepsilon \pm o(\varepsilon)$. Furthermore, we have $\mathbb{E}(|\tilde{\mathcal{T}}|) = (1 \pm o(1))\varepsilon^{-1}$ and thus we obtain the upper bound

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\leq \mathbb{P}(\mathcal{T}^* \text{ survives}) + \mathbb{P}(|\mathcal{T}^*| \geq \lambda n^j \mid \mathcal{T}^* \text{ dies out}) \\ &\leq \mathbb{P}(\mathcal{T}^* \text{ survives}) + (1 \pm o(1))(\varepsilon \lambda n^j)^{-1}, \end{aligned}$$

by Markov's Inequality. Note that the last term is $o(\varepsilon)$.

We therefore need to calculate the survival probabilities of \mathcal{T}_* and \mathcal{T}^* . We treat both cases in parallel by setting $\mathcal{T} := \mathcal{T}_*$ or $\mathcal{T} := \mathcal{T}^*$, $\zeta := 0$ or $\zeta := \varepsilon_*$ and $\varepsilon' := \varepsilon - \zeta - \varepsilon\zeta = (1 - \zeta)(1 + \varepsilon) - 1$.

It is slightly more convenient to consider the event \mathcal{D} of the process \mathcal{T} dying out and calculate its probability $\mathbb{P}(\mathcal{D})$. The process dies out if every subprocess starting at a child of the root also dies out. Recall that $c_0 := \binom{k}{j} - 1$. Because of the recursive nature of the

tree, we have

$$\begin{aligned} \mathbb{P}(\mathcal{D}) &= \sum_{i=0}^{\infty} \mathbb{P}\left(\text{Bi}\left((1-\zeta)\binom{n}{k-j}, p\right) = i\right) \mathbb{P}(\mathcal{D})^{\binom{k}{j}-1} i \\ &= \sum_{i=0}^{\infty} \binom{(1-\zeta)\binom{n}{k-j}}{i} \mathbb{P}(\mathcal{D})^{ic_0} (1-p)^{(1-\zeta)\binom{n}{k-j}-i} \\ &= (p\mathbb{P}(\mathcal{D})_0^c + 1-p)^{(1-\zeta)\binom{n}{k-j}} = (1-p(1-\mathbb{P}(\mathcal{D})^{c_0}))^{(1-\zeta)\binom{n}{k-j}}. \end{aligned}$$

We set $x := 1 - \mathbb{P}(\mathcal{D})^{c_0}$ and $y := \left((1-\zeta)c_0\binom{n}{k-j}\right)^{-1}$ and note that $p = (1+\varepsilon)\hat{p}_g = (1+\varepsilon)c_0^{-1}\binom{n}{k-j}^{-1} = (1+\varepsilon')y$. Hence, we obtain

$$1-x = (1-(1+\varepsilon')yx)^{1/y},$$

and furthermore solving for ε' yields

$$\varepsilon' = \frac{1-xy-(1-x)^y}{xy} = \frac{\frac{y(1-y)}{2}x^2 + \frac{y(1-y)(2-y)}{6}x^3 + \dots}{xy}$$

implying

$$\varepsilon' = \frac{x}{2} + O(x^2).$$

In other words, since $\varepsilon' = (1 \pm o(1))\varepsilon$, we have

$$1 - \mathbb{P}(\mathcal{D}) = 1 - (1-x)^{1/c_0} = (1 \pm o(1))\frac{2\varepsilon}{c_0},$$

as claimed. □