A sharp threshold of propagation connectivity for mixed random hypergraphs

Guangyan Zhou\textsuperscript{1,}\textsuperscript{a}, Bin Wang\textsuperscript{b}, Ke Xu\textsuperscript{c}

\textsuperscript{a}Department of Mathematics, Beijing Technology and Business University, China
\textsuperscript{b}Key Laboratory of Random Complex Structures and Data Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
\textsuperscript{c}State Key Laboratory of Software Development Environments, Beihang University, China

Abstract

This paper studies the propagation connectivity of a random hypergraph \(G\) containing both 2-edges and 3-hyperedges. We find an exact threshold of the propagation connectivity of \(G\): If \(I_{\epsilon,r} < -1\), then \(G\) is not propagation connected with high probability; while if \(I_{\epsilon,r} > -1\), then \(G\) is propagation connected with high probability, where \(I_{\epsilon,r}\) is a constant dependent on the parameters of 2 and 3-edge probabilities.

Keywords: Propagation connectivity, Hypergraph, Phase transition, Markov additive process.

1 Introduction

The study of phase transition phenomena of constraint satisfaction problems is one of the most interesting topics at the intersection of mathematics, statistical physics and computer science. In mathematics, since the seminal work of Erdős and Rényi \cite{ErdosRenyi1959}, identifying the thresholds for different properties has been a major task in the theory of random graphs and hypergraphs. Throughout the years, various types of graphs have been studied by graph theorists, such as planar graphs, routing networks and computational graphs that are used in designing algorithms or simulations, etc. Graph theory has emerged as a primary tool for detecting numerous hidden structures in various information networks, including internet graphs, social networks, or more generally, any graph representing relations in massive datasets. Nowadays, more important properties of random graph have been found, and the picture of the evolution of the random graph is fairly complete.

Connectivity is perhaps the most basic property of graphs and hypergraphs, and is also a fundamental combinatorial problem. For random graphs, the definition of connectivity is quite natural. For random graph \(G(n, p)\), which has \(n\) vertices and each edge appears with probability \(p = \frac{c}{n}\) (\(c\) is a constant), there exist double phase transitions near \(c = 1\) where the size of the largest connected component changes twice - first from \(\Theta(\ln n)\) to \(\Theta(n^{2/3})\), and then from \(\Theta(n^{2/3})\) to \(\Theta(n)\) \cite{Bollobas1980}. For random hypergraphs,
the definition of connectivity may differs. The ‘standard’ concept of random hypergraph connectivity (where edges are replaced by triangles) has been studied [1, 2, 5]. Propagation connectivity is one kind of connectivity that has been studied in recent years, e.g. on general hypergraphs [3] and on 3-uniform hypergraphs [7, 6]. Intuitively, propagation connectivity is analogous to the growth of a network: starting from a vertex, or a small initial graph, adding a new node and a new edge at each time following certain growth rules. It can be explained in terms of a simple marking process (or branching process): at each step, if there exists an edge $e$ where all vertices have been marked except one vertex, then we mark this vertex. If there exists a marking process such that all the vertices can be marked, then this graph is propagation connected. Specifically, Coja-Oghlan et al. [6] obtained the following results on 3-uniform hypergraphs.

**Theorem 1.** [6] Suppose that $p = \frac{r}{n \ln n}$ for a constant $r > 0$.
(1) If $r < 0.16$, then $H(n, p)$ fails to be propagation connected w.h.p. \(^1\)
(2) If $r > 0.25$, then $H(n, p)$ is propagation connected w.h.p.

In this paper, we study the phase transition window of random graph whose edge density stays below the threshold mentioned above (i.e. $c < 1$) by adding 3-hyperedges into it. Specifically, in this paper we consider the propagation connectivity of a random hypergraph $G = G(n, p_2, p_3)$, which is composed of both 2-edges and 3-hyperedges on $n$ vertices. Each possible 2-edge on exactly two vertices is included in $G(n, p_2, p_3)$ with probability $p_2$, and each possible 3-hyperedge on exactly three vertices is included in $G(n, p_2, p_3)$ with probability $p_3$. Throughout this paper, we let

$$p_2 = \frac{1 - \epsilon}{n}, p_3 = \frac{r}{n \ln n},$$

where parameters $0 < \epsilon < 1, r > 0$ are constants.

Using probabilistic methods, we study the propagation process starting from a random vertex, and present our most valuable and technical contribution in Theorem 2 that there exists a sharp phase transition of propagation connectivity of hypergraph $G(n, p_2, p_3)$. A special case of Theorem 2 closes the gap left by Coja-Oghlan et al. [6] in Theorem 1.

The rest of the paper is organized as follows. In Section 2, we introduce the definition of propagation connectivity in hypergraph $G(n, p_2, p_3)$. In Section 3, we present our main results on the threshold of propagation connectivity of the random hypergraph $G(n, p_2, p_3)$. The proof of our main theorem is put in Section 4. In particular, the main technique in this paper lies heavily on Markov chain techniques which give relatively precise estimates of some crucial probabilities. Section 5 draws a conclusion.

### 2 Propagation connectivity

**Definition 1.** (Propagation connectivity). Let $H = (V, E)$ be a hypergraph on $n = |V|$ vertices, and the edge set $E$ contains both edges of length 2 and edges of length 3. If there exits a propagation sequence $e_1, e_2, ..., e_{n-1} \in E$, such that for any integer $1 \leq l \leq n-2$ we have $|e_{l+1} \cap \bigcup_{i=1}^{l} e_i| = |e_{l+1}| - 1$, then we say $H$ is propagation connected.

\(^1\)We say a sequence of events $\xi_n$ occurs with high probability (w.h.p.) if $\lim_{n \to \infty} \Pr[\xi_n] = 1.$
To explain this concept, we show that the propagation sequence of a hypergraph $G(n, p_2, p_3)$ is in fact a certain Markov additive process. During the propagation process, vertices will be labeled as active, inactive and unexplored. We denote by $V$ the vertex set of $G$, $Y_t$ the set of active vertices at time $t$, and $D_t$ the set of inactive vertices at time $t$. At the beginning, $Y_0 = \{v_0\}$, $D_0 = \emptyset$. At time $t$ ($t = 0, 1, 2...$), we pick an active vertex $v_t$ and check all unexplored vertices that can be connected to $v_t$. For a random unexplored vertex $u$, if there exists a 2-edge which connects $v_t$ and $u$, or there exists a 3-edge which connects $v_t$, $u$ and an inactive vertex $\omega \in D_t$, then $u$ becomes active and at the same time, $v_t$ becomes inactive:

$$V_{t+1} = Y_t \cup \{u : u \text{ becomes active}\} \setminus \{v_t\},$$
$$D_{t+1} = D_t \cup \{v_t\}.$$ Intuitively, each step increases exactly one vertex to $D_t$ which can be considered as the formal set of vertices propagation connected from $v_0$. Let $Y_t$ be the total number of vertices that were “active” at time $t$. Denote by $Z_t$ the total number of new vertices $u$ that becomes active at step $t$. Once the set $Y_t$ becomes empty (i.e. $Y_t = 0$), the process terminates. We denote $C_{v_0}$ the component whose vertices are propagation connected from $v_0$. It is easy to see that the first hitting time $T_{v_0} = \inf\{t : Y_t = 0\}$ is exactly the size of $C_{v_0}$.

The process above is in fact a Markov additive process, where the distribution of the increment $Z_t$ is only dependent on $Y_t$ and $t$, which means

$$\Pr[Z_t | Y_1 = y_1, ..., Y_t = y_t] = \Pr[Z_t | Y_t = y_t].$$

It is easy to see that

$$\begin{align*}
Y_0 &= 1, \\
Y_{t+1} &= Y_t + Z_t - 1.
\end{align*}$$

Note that at time $t$, given an active vertex $v_t$, a random unexplored vertex $\omega$ is not connected to $v_t$ through a 2-edge with probability $1 - p_2$, and is not connected to $v_t$ through a 3-edge with probability $(1 - p_3)^t$, since there are $t$ inactive vertices at time $t$. On the other hand, the total number of unexplored vertices is $n - t - Y_t$. Thus the conditional distribution of $Z_t$ on $Y_t$ follows a binomial distribution as

$$Z_t | Y_t \sim B[n - t - Y_t, p(t)], \text{ where } p(t) = 1 - (1 - p_2)(1 - p_3)^t. \quad (2.2)$$

For simplicity, we define three functions which will be used later.

$$\begin{align*}
\lambda(x) &= 1 - \epsilon + rx, \\
\lambda_1(x) &= 1 - \epsilon + \frac{r}{2}x, \\
\lambda_2(x) &= \frac{1 - \epsilon}{2} + \frac{r}{6}x. \quad (2.3)
\end{align*}$$

3 Main results

In the following we present the most important theorem of this paper, a sharp threshold of propagation connectivity of random hypergraph $G(n, p_2, p_3)$. 

3 Main results

In the following we present the most important theorem of this paper, a sharp threshold of propagation connectivity of random hypergraph $G(n, p_2, p_3)$. 

3 Main results

In the following we present the most important theorem of this paper, a sharp threshold of propagation connectivity of random hypergraph $G(n, p_2, p_3)$. 

3 Main results

In the following we present the most important theorem of this paper, a sharp threshold of propagation connectivity of random hypergraph $G(n, p_2, p_3)$.
Theorem 2. Let \( p_2 = \frac{1-\epsilon}{n} \), \( p_3 = \frac{r}{n \ln n} \), where \( 0 < \epsilon < 1, r > 0 \) are constants. The random hypergraph \( \mathbb{G}(n, p_2, p_3) \) has the following properties:

- If \( I_{\epsilon,r} < -1 \), then w.h.p. \( \mathbb{G} \) is not propagation connected;
- If \( I_{\epsilon,r} > -1 \), then w.h.p. \( \mathbb{G} \) is propagation connected,

where

\[
I_{\epsilon,r} = \int_0^{\frac{1}{r}} \left[ 1 - \lambda(\omega) + \ln \lambda(\omega) \right] d\omega = -\frac{1}{r} \left[ \epsilon - \frac{\epsilon^2}{2} + (1 - \epsilon) \ln(1 - \epsilon) \right].
\]

\( (3.4) \)

Remark 1. It is noteworthy that a special case of \( p_2 = 0 \) in Theorem 2 corresponds to the 3-uniform hypergraph, and our results not only match the results of Theorem 1 obtained by [6] but also close the gap of threshold for connectivity left by them.

4 Proof of Theorem 2

Below we explain the main idea behind the proof of our main results. In Section 4.1, we prepare two lemmas which will be used later in our proof. In Section 4.2, we show that w.h.p. there exists no propagation component of intermediate size between \( O(\ln n) \) and \( n - 1 \). In Section 4.3, we prove that if \( I_{\epsilon,r} < -1 \), then w.h.p. the largest propagation component size is at most \( O(\ln n) \), thus w.h.p. \( \mathbb{G} \) can not be propagation connected. Based on the result that there is no existence of propagation component with intermediate size, we know that once a propagation process survives after \( O(\ln n) \) steps, then it is possible that this process will continue to time \( n \), which means that the entire hypergraph will be propagation connected by this process. Moreover, in Section 4.3.3, we compare the results of [6] with a special case of our results, and show that our results actually close the gap left by them. Finally, in Section 4.4, we prove that if \( I_{\epsilon,r} > -1 \), then \( \mathbb{G} \) is w.h.p. propagation connected.

4.1 Some lemmas

We prepare two lemmas which will be used for the proof of our main results later.

Lemma 3. Suppose random variables \( X_1, X_2, \ldots, X_n \) are independently and identically distributed, denote two events as

\[
A = \{ X_1 \geq 0, X_1 + X_2 \geq 0, \ldots, X_1 + X_2 + \ldots + X_n \geq 0 \},
\]

\[
B = \{ X_1 + X_2 + \ldots + X_n \geq 0 \}.
\]

Then

\[
\frac{\Pr(B)}{n} \leq \Pr(A) \leq \Pr(B).
\]

Proof. The second inequality on the right hand side holds trivially. So we will only prove the inequality on the left side.

For any \( k \in \mathbb{N} \), let the subscript of \( X_k \) module \( n \), i.e. \( X_{n+k} = X_k \), so that all the subscripts are at most \( n \). Define \( S_k = X_1 + \ldots + X_k \). Consider the following events:

\[
A_k = \{ X_k \geq 0, X_k + X_{k+1} \geq 0, \ldots, X_k + X_{k+1} + \ldots + X_{k+n-1} \geq 0 \}, k = 1, \ldots, n.
\]
By symmetry, it is easy to see that
\[ \Pr(A) = \Pr(A_1) = \ldots = \Pr(A_n). \]

Suppose the event \( B \) happens, i.e. \( S_n \geq 0 \), and suppose \( S_t (0 \leq t \leq n-1) \) is the minimum one of the set \( \{ S_0 \equiv 0, S_1, \ldots, S_{n-1} \} \). Then for all \( k = 1, 2, \ldots, n \), if \( t + k \leq n - 1 \) we have \( S_{t+k} \geq S_t \); if \( t + k \geq n \), we have \( S_{t+k} = S_{t+k-n} + S_n \geq S_t \). Therefore it always holds that
\[
X_{t+1} + X_{t+2} + \ldots + X_{t+k} = S_{t+k} - S_t \geq 0,
\]
which means that the event \( A_{t+1} \) happens, thus
\[
\Pr(B) \leq \Pr(A_1) + \Pr(A_2) + \ldots + \Pr(A_n) = n\Pr(A).
\]

\qed

**Definition 2.** Let the distribution function of a random variable \( X \) be \( F(x) = \Pr[X \leq x] \), and denote \( \overline{F}(x) = \Pr[X > x] = 1 - F(x) \). We say that \( F \succ G \), if two distribution functions \( F \) and \( G \) satisfying \( F(t) \leq G(t) \) (or \( \overline{F}(t) > \overline{G}(t) \)) for all \( t \in R \).

Take the binomial distribution for example, if \( n_1 \geq n_2 \) then \( B(n_1, p_1) \succ B(n_2, p_1) \); if \( p_1 \geq p_2 \), then \( B(n_1, p_1) \succ B(n_1, p_2) \).

With this notation, we consider two Markov additive processes, and have the following results which will be important to our argument.

**Lemma 4.** Suppose two stochastic processes \( \{S_k\} \) and \( \{T_k\} \) satisfy that
\[
S_{k+1} = S_k + X_k - q, \quad T_{k+1} = T_k + Y_k - q,
\]
where \( S_0 = T_0 = b \in Z, q \) is a nonnegative integer, \( X_k|S_k \sim F_{k,S_k} \) and \( Y_k|T_k \sim G_{k,T_k} \). If the distribution functions \( F_{k,S_k} \) and \( G_{k,T_k} \) take values of non-negative integers, and satisfying

1. For any \( k \in N, z \in Z, l \in R \), it holds that \( G_{k,z}(l) \leq G_{k,z-1}(l+1) \);

2. There exists an integer \( M \geq b \) such that for any \( z \leq M - kq \), it holds that \( F_{k,z} \succ G_{k,z} \).

Then for any sequence \( \{l_k|l_k \leq M - kq\} \),
\[
\Pr[S_1 > l_1, S_2 > l_2, \ldots, S_n > l_n] \geq \Pr[T_1 > l_1, T_2 > l_2, \ldots, T_n > l_n].
\]

**Proof.** We prove this claim by induction on \( n \). For simplicity, denote
\[
A_n = \{S_1 > l_1, S_2 > l_2, \ldots, S_n > l_n\}; \quad B_n = \{T_1 > l_1, T_2 > l_2, \ldots, T_n > l_n\}.
\]

If \( n = 1 \), and \( X_0 \sim F_{0,b}, Y_0 \sim G_{0,b}, F_{0,b} \succ G_{0,b} \), then the claim holds since
\[
\Pr[A_1] = \Pr[S_1 \geq l_1] = \Pr[X_0 > l_1 + q - b] = \overline{F}_{0,b}(l_1 + q - b)
\]
The inequality above is because for $\tilde{m} = q$, the fact $l_{n+1} = M - (n+1)q$ implies that $z \leq M - nq$. Applying condition 2 gives $\mathcal{F}_{n,z}(l) \geq \mathcal{T}_{n,z}(l)$.
Similarly, we can estimate $\Pr[B_{n+1}]$.

\[
\Pr[B_{n+1}] = \Pr[T_1 > l_1, \ldots, T_n > l_n, T_{n+1} > l_{n+1}]
= \sum_{z > l_{n+1}} \Pr[B_{n,z}] \Pr[Y_n|l_n = z > l_{n+1} - z + q]
= \sum_{l_n < z \leq l_{n+1} + q} \Pr[B_{n,z}] \mathbb{G}_n(z)(l_{n+1} - z + q) + \sum_{z > l_{n+1} + q} \Pr[B_{n,z}]
= \sum_{l_n < z \leq l_{n+1} + q} \left( \Pr[B_{n,z}] - \mathbb{G}_n(z)(l_{n+1} - z + q) \right) + \sum_{l_n < z \leq l_{n+1} + q} \Pr[B_{n,z}]
= \sum_{l_n < z \leq l_{n+1} + q} \Pr[B_{n,z}] \mathbb{G}_n(z)(l_{n+1} - z - 1 + q)
= \Pr[B_{n,l_n}] \mathbb{G}_n(z)(l_{n+1} - l_n - 1 + q)
= \sum_{l_n < z \leq l_{n+1} + q} \Pr[B_{n,z}] \mathbb{G}_n(z)(l_{n+1} - z - 1 + q) - \mathbb{G}_n(z)(l_{n+1} - z + q).
\]

Noticing that $\mathbb{G}_n(z)(l_{n+1} - z - 1 + q) \geq 0$, and combining condition 1 we know that $\mathbb{G}_n(z)(l_{n+1} - z + q) \geq \mathbb{G}_n(z)(l_{n+1} - z + q)$. If $z = l_n$ or $l_n < z \leq l_{n+1} + q$, then in fact $z \leq M - nq$. By the induction hypothesis we know that $\Pr[\mathbb{H}_{n,z}] \geq \Pr[B_{n,z}]$, therefore

\[
\Pr[A_{n+1}] \geq \Pr[B_{n+1}].
\]

This completes our proof. \(\square\)

4.2 No components of intermediate size

**Lemma 5.** There exists a constant $K_0 = \max\{1, \lambda^{-1}(10)\}$, such that with probability approaching one as $n \to \infty$, there is no propagation connected components of size between $K_0 \ln n$ and $n - 1$.

**Proof.** Suppose the propagation component connected from vertex $v$ is $C_v$, which contains $|C_v|$ vertices. There are two facts for $C_v$: first, there is no 2-edge connecting one vertex from $C_v$ and one vertex from $V \setminus C_v$, and there is no 3-edge connecting two vertices from $C_v$ and one vertex from $V \setminus C_v$; second, since $C_v$ is propagation connected, then $C_v$ contains at least $|C_v| - 1$ edges.

Let $E_v$ denote the number of edges in $C_v$. The probability that the size of the component $C_v$ is $z$ can be upper bounded as follows.

\[
\Pr[|C_v| = z] \leq \binom{n}{z} (1 - p_2)^z (1 - p_3)^{\binom{z}{2}} (n - z)^{n - z} \Pr[E_v \geq z - 1]. \tag{4.5}
\]

The stirling’s formula $n! = \sqrt{2\pi n} \left( \frac{n}{e} \right)^n e^{\theta_n} \left( \frac{1}{12n+1} < \theta_n < \frac{1}{12n} \right)$ implies that for all $1 \leq z \leq n$,

\[
\binom{n}{z} \leq \frac{n^n}{z^n(n - z)^{n-z}} = \frac{n^z}{z^z} \left( 1 + \frac{z}{n - z} \right)^{n-z} \leq \left( \frac{ne}{z} \right)^z. \tag{4.6}
\]
Also it is straightforward that
\[
(1 - p_2)^z(n-z)(1 - p_3)^n \left(\frac{z}{2}\right)^{(n-z)} \leq \exp\left\{-p_2z(n-z) - p_3 \left(\frac{z}{2}\right)(n-z)\right\}. \tag{4.7}
\]

We will estimate (4.5) by three cases.

**Case 1:** \( \frac{n}{2} \leq |C_v| = z \leq n - 1 \). We have
\[
\ln \left(\frac{n}{z}\right) = \ln \left(\frac{n}{n-z}\right) \leq \ln \left(\frac{ne}{n-z}\right)^{n-z} \leq (n-z)(\ln n + 1).
\]
Taking the logarithm on both sides of (4.5) and notice that \( \Pr[E_v \geq z - 1] \leq 1 \), we get
\[
\ln \Pr[|C_v| = z] \leq (n-z)(\ln n + 1) - (n-z) \left(\frac{1-\epsilon}{n}z + \frac{r}{n\ln n} \frac{z^2}{2}\right)
\]
\[= -(n-z)\frac{rz^2}{2n\ln n}(1 + o(1)) \leq -3 \ln n,
\]
which yields
\[
\Pr[|C_v| = z] \leq n^{-3}.
\]

**Case 2:** \((\ln n)^3 \leq |C_v| = z = \frac{n}{2}\). Similar to case 1, we can apply (4.6) and (4.7) to get
\[
\ln \Pr[|C_v| = z] \leq z(\ln n - \ln z + 1) - z \left(1 - \epsilon\right) \frac{n-z}{n} + \frac{r}{2n\ln n} \frac{z(n-z)}{2}
\]
\[\leq \ln n - \frac{r}{2} \frac{(\ln n)^3(n-n/2)}{n\ln n} \leq -z \ln n \leq -3 \ln n,
\]
which yields
\[
\Pr[|C_v| = z] \leq n^{-3}.
\]

**Case 3:** \( K_0 \ln n \leq |C_v| = z \leq (\ln n)^3 \). We rewrite \( z = z = K \ln n \), where \( K \) is dependent on \( n \). We claim that the probability of the component size \( |C_v| \) being \( z \) is also upper bounded by \( n^{-3} \).

To begin with, suppose \( C_v \) contains \( E_2 \) 2-edges and \( E_3 \) 3-edges, then \( E_2 \) and \( E_3 \) independently follow the binomial distribution with \( E_2 \sim B \left(\frac{z}{2}, p_2\right) \) and \( E_3 \sim B \left(\frac{z}{3}, p_3\right) \). Let \( E_v = E_2 + E_3 \) be the total number of edges in \( C_v \). Obviously, For all \( s \geq 1 \), the generating function of \( E_v \) satisfies
\[
E[s^{E_v}] \leq (1 + p_2(s-1))^{z^2/2}(1 + p_3(s-1))^{z^3/6}, \tag{4.8}
\]
thus it holds for all \( s \geq 1 \) that
\[
\Pr[E_v \geq z - 1] \leq \frac{E[s^{E_v}]}{s^{z-1}} \leq \frac{1}{s^{z-1}} \left(1 + p_2(s-1))^{z^2/2}(1 + p_3(s-1))^{z^3/6}. \tag{4.9}
\]
Now we can upper bound the probability that $C_v$ contains at least $z - 1$ edges as follows.

\[
\ln \Pr[E_v \geq z - 1] \leq \frac{z^2}{2} \ln (1 + p_2(s - 1)) + \frac{z^3}{6} (1 + p_3(s - 1)) - (z - 1) \ln s \\
\leq \frac{z^2}{2} p_2(s - 1) + \frac{z^3}{6} p_3(s - 1) - (z - 1) \ln s \\
= (s - 1) \frac{z^2}{n} \lambda_2(K) - (z - 1) \ln s.
\]

Substituting $s = s_0 \equiv \frac{n}{\lambda_2(K)} \geq 1$ into the above inequality, we obtain

\[
\ln \Pr[E_v \geq z - 1] \leq z - (z - 1) \ln s_0 \leq z - (z - 1)(\ln n - \ln z - \ln \lambda_2(K)). \tag{4.10}
\]

At the same time, notice that

\[
\ln \left( \frac{n}{z} \right) \leq z(\ln n - \ln z) + z, \\
(1 - p_2)^{z(n-z)}(1 - p_3) \frac{z(n-z)}{n} \leq \exp \left[ -z \frac{n - z}{n} \lambda_1(K) \right] \leq -0.9 z \lambda_1(K).
\]

Putting these two inequalities and (4.10) together, we get

\[
\ln \Pr[|C_v| = z] \leq 2z + \ln n - \ln z + (z - 1) \ln \lambda_2(K) - 0.9 z \lambda_1(K) \\
\leq z(2 + \ln \lambda_2(K) - 0.9 \lambda_1(K)) + \ln n.
\]

Note that we require $K_0 = \max\{1, \lambda_1^{-1}(10)\}$, thus for all $z = K \ln n \geq K_0 \ln n$, it holds that $\lambda_1(K) \geq 10$ and $\ln \lambda_2(K) - 0.9 \lambda_1(K) \leq -6$. Now for any $K_0 \ln n \leq z \leq (\ln n)^3$ we have

\[
\ln \Pr[|C_v| = z] \leq -4z + \ln n \leq -3 \ln n,
\]

which yields

\[
\Pr[|C_v| = z] \leq n^{-3}.
\]

As a consequence, putting the three cases together, we conclude that

\[
\sum_{z=K_0 \ln n}^{n-1} \Pr[|C_v| = z] \leq n \cdot n^{-3} = n^{-2},
\]

which completes our proof.

\[\square\]

4.3 Subcritical regime: $I_{\epsilon, r} < -1$

4.3.1 Proof of the lower bound in Theorem\[2\]

To prove that $G$ is not propagation connected, it suffices to show that almost all propagation processes terminate before $O(\ln n)$ steps. Recall that $K_0 = \max\{1, \lambda_1^{-1}(10)\}$.
Proposition 6. Let \( v_0 \) be any given vertex, then
\[
\Pr[|C_{v_0}| \geq K_0 \ln n] \leq n^{I_{c,r} + o(1)},
\]
(4.11)
where \( C_{v_0} \) is the propagation component generated by \( v_0 \).

From Proposition 6 if we let \( I_{c,r} < -1 \), then
\[
\Pr[\exists \text{ a vertex } v, \text{ s.t. } |C_v| \geq K_0 \ln n] \leq n \cdot n^{I_{c,r} + o(1)} = o(1),
\]
which proves that w.h.p. \( G \) is not propagation connected. This completes the proof of the lower bound of Theorem 1.

As a preliminary step to prove Proposition 6, we need the following result that starting from any vertex \( v \), the propagation process at step \( T \) (where \( T \leq K_0 \ln n \)) has the following property.

Proposition 7. If \( T \leq K_0 \ln n \), then for any constant \( c > 0 \), there exists a constant \( K_1(c) \) so that any propagation process has the following property
\[
\Pr[|Y_T| + |D_T| > K_1(c) \ln n] < n^{-c}.
\]

Proof. Suppose we start from a vertex \( v \), which has the following Markov additive process:
\[
\begin{align*}
Y_0 &= 1, \\
Y_{t+1} &= Y_t + Z_t - 1,
\end{align*}
\]
where the conditional distribution of \( Z_t \) on \( Y_t \) is
\[
Z_t | Y_t \sim B[n - t - Y_t, p(t)], \quad p(t) = 1 - (1 - p_2)(1 - p_3)^t.
\]

We introduce another Markov additive process which corresponds to it:
\[
\begin{align*}
Y_{t}^{(1)} &= 1, \\
Y_{t+1}^{(1)} &= Y_t^{(1)} + Z_t^{(1)} - 1,
\end{align*}
\]
where
\[
Z_t^{(1)} | Y_t^{(1)} \sim B[n, p(t)].
\]

Apparently, it holds that \( B[n, p(t)] \succ B[n - t - Y_t, p(t)] \). Take a sequence \( l_1 = ... = l_{T-1} = 0, l_T = K_1(c) \ln n - T \), then by Lemma 4 we have
\[
\begin{align*}
\Pr[|Y_T| + |D_T| > K_1(c) \ln n] &= \Pr[Y_1 > 0, ..., Y_{T-1} > 0, Y_T > K_1(c) \ln n - T] \\
&\leq \Pr[Y_1^{(1)} > 0, ..., Y_{T-1}^{(1)} > 0, Y_T^{(1)} > K_1(c) \ln n - T] \\
&\leq \Pr[Y_T^{(1)} > K_1(c) \ln n - T] \\
&= \Pr[Z^{(1)} \equiv Z_1^{(1)} + Z_2^{(1)} + ... + Z_T^{(1)} > K_1(c) \ln n].
\end{align*}
\]
The generating function of \( Z^{(1)} \) is
\[
E[s^{Z^{(1)}}] = \prod_{t=1}^{T} (1 + p(t)(s - 1))^n \leq \exp \left[ n(s - 1) \sum_{t=1}^{T} p(t) \right] = \exp \left[ (s - 1)((1 - \epsilon)K_0 + cK_0^2/2) \ln n \right] = n^{(s-1)((1-\epsilon)K_0+cK_0^2/2)}.
\]

Now we have
\[
\Pr[Z^{(1)} > K_1(c) \ln n] \leq \frac{E[s^{Z^{(1)}}]}{s^{K_1(c)}} \leq \frac{n^{(s-1)((1-\epsilon)K_0+cK_0^2/2)}}{K_1(c)}.
\]

Take \( s = \epsilon \), and let \( K_1(c) = (e - 1)((1 - \epsilon)K_0 + cK_0^2/2) \) yields the desired result that
\[
\Pr[|Y_T| + |D_T| > K_1(c) \ln n] = \Pr[Z^{(1)} > K_1(c) \ln n] \leq n^{-c}.
\]

\[\square\]

### 4.3.2 Proof of Proposition 6

Note that \( K_0 = \max \{1, \lambda^{-1}_1(10)\} \geq \frac{\epsilon}{r} \), so if suffices to prove
\[
\Pr[|C_v| \geq \overline{\omega} \ln n] \leq n^{I_c + o(1)}, \text{ where } \overline{\omega} = \frac{\epsilon}{r}.
\]

Since
\[
\Pr[|C_v| \geq \overline{\omega} \ln n] = \Pr[Y_t > 0, t = 0, 1, \ldots, \overline{\omega} \ln n],
\]
we only need to estimate the probability that \( Y_t > 0 \) holds for all \( t \in [0, \overline{\omega} \ln n] \) if \( Y_0 = 1 \).

**Step 1.** Take a large positive integer \( L \), we divide the interval \( [0, \overline{\omega} \ln n] \) into \( L \) intervals: \( \Delta_k = [t_{k-1}, t_k] \equiv [\omega_{k-1} \ln n, \omega_k \ln n] \), where \( \omega_k = \frac{k}{L} \overline{\omega}, \ k = 1, 2, \ldots, L \). We now estimate the following probability on the interval \( \Delta_k \).
\[
\Pr[\Delta_k] \equiv \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t > 0, Y_{t_k} = y_k | Y_{t_{k-1}} = y_{k-1}],
\]
where \( \{y_1, \ldots, y_k\} \) is a random sequence taking nonnegative integers, and \( y_0 = 1 \) for beginning.

Consider the propagation process on the interval \( t \in \Delta_k = [t_{k-1}, t_k] \):
\[
\begin{cases}
Y_{t_{k-1}} = y_{k-1}, \\
Y_{t+1} = Y_t + Z_t - 1,
\end{cases}
\]
where the conditional distribution of \( Z_t \) is
\[
Z_t | Y_t \sim B[n - t - Y_t, p(t)].
\]

We introduce another process which corresponds to it:
\[
\begin{cases}
Y_{t_{k-1}}^{(2)} = y_{k-1}, \\
Y_{t+1}^{(2)} = Y_t^{(2)} + Z_t^{(2)} - 1,
\end{cases}
\]
where
\[ Z_t^{(2)}|Y_t^{(2)} \sim B \left[ n, \frac{\lambda(\omega_k)}{n} \right]. \]

Since \( p(t) \leq \lambda(\omega_k)/n \), thus
\[ B[n - t - Y_t, p(t)] < B \left[ n, \frac{\lambda(\omega_k)}{n} \right]. \]

Taking \( l_{tk+1} = \ldots = l_{tk} = 0, l_{tk} = y_k - 1 \), and applying lemma 4, we have
\[
\Pr[\Delta_k] = \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t > 0, Y_{tk} = y_k|Y_{tk-1} = y_k-1] \\
\leq \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t > 0, Y_{tk} \geq y_k|Y_{tk-1} = y_k-1] \\
\leq \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t^{(2)} > 0, Y_{tk}^{(2)} \geq y_k|Y_{tk-1}^{(2)} = y_k-1] \\
\leq \Pr[Y_{tk}^{(2)} \geq y_k|Y_{tk-1}^{(2)} = y_k-1]. \tag{4.12}
\]

Denote
\[ Z^{(2)} = \sum_{t \in [t_{k-1}, t_k-1]} Z_t^{(2)} \sim B \left[ \frac{\omega}{L} \ln n, \frac{\lambda(\omega_k)}{n} \right]. \]

The generating function of \( Z^{(2)} \) is
\[ \mathbb{E}[s^{Z^{(2)}}] = \left( 1 + \frac{\lambda(\omega_k)}{n} (s - 1) \right)^{\frac{\omega}{L} \ln n}. \]

Denote
\[ d_k = \frac{y_k - y_k - 1}{\omega \ln n/L}, s_k = \frac{1 + d_k}{\lambda(\omega_k)}, \text{ where } k = 1, 2, \ldots, L. \]

If \( d_k \geq \lambda(\omega_k) - 1 \), then \( s_k \geq 1 \). Substituting \( s = s_k \) into the generating function, then we can rewrite the inequality (4.12) as
\[
\ln \Pr[\Delta_k] \leq \ln \Pr[Y_{tk}^{(2)} \geq y_k|Y_{tk-1}^{(2)} = y_k-1] = \ln \Pr[Z^{(2)} \geq y_k - y_k - 1 + \frac{\omega}{L} \ln n] \\
\leq \ln \mathbb{E}[s_k^{Z^{(2)}}] \leq (s_k - 1) \lambda(\omega_k) \frac{\omega}{L} \ln n - \left( y_k - y_k - 1 + \frac{\omega}{L} \ln n \right) \ln s_k \\
= \left( 1 + d_k - \lambda(\omega_k) \right) \frac{\omega}{L} \ln n - \left( 1 + d_k \right) \ln(1 + d_k) - \ln \lambda(\omega_k) \frac{\omega}{L} \ln n.
\]

By the additive relations of sequence \( \{Y_{tk} = y_k \geq 1\} \), we know that \( \{d_k\} \) satisfies
\[ d_1 \geq 0, d_1 + d_2 \geq 0, \ldots, d_1 + \ldots + d_L \geq 0. \]

For simplicity, define a function of \( d_k \):
\[ M_k(d_k) = \begin{cases} 
1 + d_k - \lambda(\omega_k) - (1 + d_k) \ln(1 + d_k) - \ln \lambda(\omega_k), & \text{if } d_k \geq \lambda(\omega_k) - 1, \\
0, & \text{if } d_k < \lambda(\omega_k) - 1.
\end{cases} \]

12
Obviously, \( M_k(d_k) \) decreases with \( d_k \). Now we have
\[
\ln \Pr[\Delta_k] \leq M_k(d_k) \frac{\bar{\omega}}{L} \ln n.
\]

**Step 2.** Now we estimate the probability that \( Y_t > 0 \) holds for any \( t = 0, \ldots, \bar{\omega} \ln n \) and \( Y_t \) takes given values on our interval endpoints.

\[
H(y_1, \ldots, y_L) = \prod_{1 \leq k \leq L} \Pr[\forall t \in [t_k-1, t_k-1], Y_t > 0, Y_{t_k} = y_k|Y_{t_k-1} = y_{k-1}]
\]

Thus
\[
\ln H(y_1, \ldots, y_L) \leq \frac{\bar{\omega}}{L} \ln n \sum_{1 \leq k \leq L} M_k(d_k).
\]

Now we make a minor modification of the sequence \( \{d_1, \ldots, d_L\} \). Denote \( d'_k = \max\{\lambda(\omega_k) - 1, d_k\} \), then \( M_k(d'_k) = M_k(d_k) \) and \( d'_k \geq d_k \), also it holds that
\[
d'_1 \geq 0, d'_1 + d'_2 \geq 0, \ldots, d'_1 + \ldots + d'_L \geq 0.
\]

With this in mind, we can rewrite and decompose the sum of \( M_k(d_k) \) as
\[
\sum_{1 \leq k \leq L} M_k(d_k) = \sum_{1 \leq k \leq L} M_k(d'_k)
\]
\[
= \sum_{1 \leq k \leq L} \left[ 1 + d'_k - \lambda(\omega_k) - (1 + d'_k) [\ln(1 + d'_k) - \ln \lambda(\omega_k)] \right]
\]
\[
= \sum_{1 \leq k \leq L} \left[ 1 - \lambda(\omega_k) + \ln \lambda(\omega_k) \right] + \sum_{1 \leq k \leq L} d'_k \ln \lambda(\omega_k) - \sum_{1 \leq k \leq L} [(1 + d'_k) \ln(1 + d'_k) - d'_k].
\]

Since \( \ln \lambda(\omega_k) \leq 0 \) is a decreasing sequence, we rewrite the second sum above by the Abel transformation, and obtain
\[
\sum_{1 \leq k \leq L} d'_k \ln \lambda(\omega_k) = (d'_1 + \ldots + d'_L) \ln \lambda(\omega_L) + \sum_{1 \leq k \leq L-1} (d'_1 + \ldots + d'_k) \ln \lambda(\omega_k) - \ln \lambda(\omega_{k+1}) \]
\[
\leq 0.
\]

On the other hand, \( g(x) = (1 + x) \ln(1 + x) - x \) is a concave function with respect to \( x \), thus
\[
\sum_{1 \leq k \leq L} [(1 + d'_k) \ln(1 + d'_k) - d'_k] = \sum_{1 \leq k \leq L} g(d'_k) \geq \frac{1}{L} \sum_{1 \leq k \leq L} \left( d'_1 + \ldots + d'_L \right) \geq 0.
\]

As a consequence,
\[
\sum_{1 \leq k \leq L} M_k(d_k) \leq \sum_{1 \leq k \leq L} \left[ 1 - \lambda(\omega_k) + \ln \lambda(\omega_k) \right].
\]
Consider \( I_{\epsilon,r} = \int_0^\infty [1 - \lambda(\omega) + \ln \lambda(\omega)] d\omega \). Under the partition \( \Delta_k, k = 1, ..., L \), the upper Darboux sum of \( I_{\epsilon,r} \) is

\[
\frac{\infty}{L} \sum_{1 \leq k \leq L} [1 - \lambda(\omega) + \ln \lambda(\omega)].
\]

Note that the integrand above is increasing with \( \omega \), thus

\[
\frac{\infty}{L} \sum_{1 \leq k \leq L} [1 - \lambda(\omega) + \ln \lambda(\omega)] - \int_0^\infty [1 - \lambda(\omega) + \ln \lambda(\omega)] d\omega
\]

\[
\leq \left[(1 - \lambda(\infty) + \ln \lambda(\infty)) - (1 - \lambda(0) + \ln \lambda(0))\right] \cdot \max(\omega_k - \omega_{k-1})
\]

\[
= - (\epsilon + \ln(1 - \epsilon)) \frac{\infty}{L}.
\]

We are now ready to estimate \( H(y_1, ..., y_L) \).

\[
\ln H(y_1, ..., y_L) \leq \frac{\infty}{L} \ln n \sum_{1 \leq k \leq L} M_k(d_k)
\]

\[
\leq \frac{\infty}{L} \ln n \sum_{1 \leq k \leq L} [1 - \lambda(\omega) + \ln \lambda(\omega)]
\]

\[
\leq I_{\epsilon,r} \ln n - (\epsilon + \ln(1 - \epsilon)) \infty \ln n/L
\]

\[
= (I_{\epsilon,r} + C/L) \ln n,
\]

where \( C = -(\epsilon + \ln(1 - \epsilon)) \infty \) is a constant.

As a consequence,

\[
H(y_1, ..., y_L) \leq n^{I_{\epsilon,r}+C/L}.
\]

**Step 3.** Finally, returning to the estimate of \( \Pr[|C_{v_0}| \geq \infty \ln n] \).

\[
\Pr[|C_{v_0}| \geq \infty \ln n] = \Pr[Y_t > 0, \forall t = 0, 1, ..., \infty \ln n]
\]

\[
= \sum_{y_1 > 0, ..., y_L > 0} \Pr[Y_t > 0, t = 0, 1, ..., \infty \ln n; Y_{t_k} = y_k, k = 1, ..., L]
\]

\[
= \sum_{y_1 > 0, ..., y_L > 0} H(y_1, ..., y_L).
\]

From Proposition 7 and take \( c = 2 - I_{\epsilon,r} \), then

\[
\sum_{y_1 > 0, ..., y_{L} > K_1(c) \ln n} H(y_1, ..., y_L)
\]

\[
\leq \sum_{y_1 > 0, ..., y_{L} > K_1(c) \ln n} \Pr[Y_{t_k} = y_k, k = 1, ..., L] = \Pr[Y_{t_j} > K_1(c) \ln n]
\]

\[
\leq n^{-c} = n^{I_{\epsilon,r}-2}.
\]

Consequently,

\[
\Pr[|C_{v_0}| \geq \infty \ln n]
\]
\[\sum_{y_1 \leq K_1(c) \ln n, \ldots, y_L \leq K_1(c) \ln n} H(y_1, \ldots, y_L) = \sum_{j=1}^{L} \sum_{y_1 > 0, \ldots, y_j > 0} \sum_{y_{j+1} \leq K_1(c) \ln n, \ldots, y_L > 0} H(y_1, \ldots, y_L) \leq (K_1(c) \ln n)^L n^{I_{\epsilon,r} - 2} + L n^{I_{\epsilon,r} - 2} \]
\[\leq n^{I_{\epsilon,r} + \epsilon},\]

where the last inequality holds for any sufficiently small constant \(\epsilon = \epsilon(L) > 0\).

4.3.3 Special case of \(p_2 = 0\)

In our previous analysis, all the propagation processes start from exactly one random variable. If we consider a special case of \(p_2 = 0\), then the hypergraph \(G = G(n, p_2, p_3)\) is exactly the 3-uniform hypergraph. In this case, we have to start the propagation process with two vertices instead of just one. Similarly we denote by \(C_{\{v_1, v_2\}}\) the component generated by starting from \(v_1, v_2\), where the propagation process stays the same except for the initial set \(Y_0\):

\[
\begin{align*}
Y_0 & = 2, \\
Y_{t+1} & = Y_t + Z_t - 1.
\end{align*}
\]

Similar to Proposition 6, propagation processes which start from any two arbitrary vertices will terminate w.h.p. before \(O(\ln n)\) steps.

**Corollary 1.** Let \(v_1, v_2\) be any random vertices, then

\[\Pr[|C_{v_1,v_2}| \geq K_0 \ln n] \leq n^{I_{\epsilon,r} + o(1)}\]

The proof of Corollary 1 follows the technique of Proposition 6. Because very minor modifications are needed, we move the proof to the appendix.

From Corollary 1 if we let \(I_{\epsilon,r} < -2\), then

\[\Pr[\exists \text{ two vertices } v_1, v_2, \text{ s.t. } |C_{v_1,v_2}| \geq K_0 \ln n] = \binom{n}{2} n^{I_{\epsilon,r} + o(1)} = o(1),\]

which implies that if \(I_{\epsilon,r} < -2\), then w.h.p. \(G\) is not propagation connected.

Recall the results in Theorem 1 obtained by [6], and note that if we let \(p_2 = 0\) (i.e. \(\epsilon = 1\)), then the condition \(I_{\epsilon,r} < -2\) is equivalent to \(r < 0.25\). Now we can say that if \(r < 0.25\), then \(H(n, p)\) fails to be propagation connected w.h.p., which shows that the propagation connectivity of 3-uniform hypergraph has a sharp phase transition. Therefore our results close the gap left by [6].

4.4 Supercritical regime: \(I_{\epsilon,r} > -1\)

From previous results, we know that once the propagation process of a vertex survives up to \(O(\ln n)\) steps, it may be possible to survive until the end, i.e., the entire hypergraph \(G\) can be propagation connected. This leads us to the definition of “good” vertices.

**Definition 3.** We call a vertex \(v\) good, if the propagation process starting from \(v\) survives at least \(K_0 \ln n\) steps, i.e. \(|C_v| > K_0 \ln n\).
Proposition 8. For a random vertex $v$, we have

$$\Pr[|C_v| \geq K_0 \ln n] \geq n^{I_{1,r-o(1)}}.$$ 

Proof. To estimate the probability that $Y_t > 0$ holds for all $t \in [0, K_0 \ln n]$, we first consider some interval $t \in [a, b] \subset [0, K_0 \ln n]$. Suppose at the beginning $Y_a = y > 0$, and the propagation process is

$$
\begin{aligned}
\begin{cases}
Y_a = y, \\
Y_{t+1} = Y_t + Z_t - 1,
\end{cases}
\end{aligned}
$$

where

$$Z_t | Y_t \sim \mathcal{B}[n - t - Y_t, p(t)], \quad p(t) = 1 - (1 - p_2)(1 - p_3)^t.$$ 

We define another Markov process which corresponds to it as

$$
\begin{aligned}
\begin{cases}
Y^{(3)}_a = y, \\
Y^{(3)}_{t+1} = Y^{(3)}_t + Z^{(3)}_t - 1,
\end{cases}
\end{aligned}
$$

where

$$Z^{(3)}_t | Y^{(3)}_t \sim \mathcal{B}[n - b, p(a)].$$

By lemma 4, $\mathcal{B}[n - t - Y_t, p(t)] \succ \mathcal{B}[n - b, p(a)]$ holds for all $Y_t \leq b - t$, where we take $M = b - a \geq y, q = 1, l_a = l_{a+1} = ... = l_b = 0$. Therefore

$$
\begin{aligned}
\Pr[\forall t \in [a, b], Y_t > 0 | Y_a = y] & \geq \Pr[\forall t \in [a, b], Y^{(3)}_t > 0 | Y^{(3)}_a = y] \\
& \geq \Pr[Z^{(3)} \geq b - a],
\end{aligned}
$$

where $Z^{(3)} \equiv Y^{(3)}_b - Y^{(3)}_a + b - a = Z^{(3)}_a + ... + Z^{(3)}_{b-1}$. Thus

$$Z^{(3)} \sim \mathcal{B}[(n - b)(b - a), p(a)].$$

As a result,

$$\Pr[Z^{(3)} = b - a] = \binom{(n - b)(b - a)}{b - a} p(a)^{b-a} (1 - p(a))^{(n-b-1)(b-a)}.$$ 

Note that

$$\ln \binom{(n - b)(b - a)}{b - a} \geq \ln \frac{(n - b - a)(b - a)^{b-a}}{(b - a)!} > (b - a) \left[ \ln (n - b - 1) + 1 \right] - \ln (b - a) - 2.$$ 

Now we have

$$\ln \Pr[Z^{(3)} = b - a] \geq (b - a) \left[ \ln (n - b - 1) + 1 + \ln p(a) - (n - b - 1)p(a) \right] - \ln (b - a) - 2$$

$$= (b - a) \left[ 1 + \ln \left( (n - b - a)p(a) \right) - (n - b - a)p(a) \right] - \ln (b - a) - 2.$$
Similarly, we consider the partition of the interval $[0, \bar{\omega} \ln n]$: $\Delta_k = [\omega_{k-1} \ln n, \omega_k \ln n], k = 1, \ldots, L$.

Take $[a, b] = \Delta_k$, then $(n - b - 1)p(a) = (1 + o(1))\lambda(\omega_{k-1})$.

Since random variables $Z_a^{(3)} - 1, \ldots, Z_{b-1}^{(3)} - 1$ are identically distributed, by lemma 4,

$$\Pr[\forall t \in [a, b], Y_t > 0|Y_a = y] \geq \Pr[\forall t \in [a, b], Y_t^{(3)} > 0|Y_{a}^{(3)} = y] \geq \Pr[\forall k = 0, 1, \ldots, b - a - 1, Z_a^{(3)} - 1 + \ldots + Z_{a+k}^{(3)} - 1 \geq 0] \geq \frac{1}{b - a} \Pr[Z^{(3)} \geq b - a].$$

Therefore,

$$\Pr[\forall t \in [a, b], Y_t > 0|Y_a > 0] \geq \frac{1}{b - a} \Pr[Z^{(3)} \geq b - a] \geq (b - a) \left(1 + \ln((n - b - a)p(a)) - (n - b - a)p(a)\right) - 2\ln(b - a) - 2.$$

Consequently,

$$\ln \Pr[\forall t \in [0, \bar{\omega} \ln n], Y_t > 0|Y_0 = 1] = \sum_{\Delta_k} \ln \Pr[\forall t \in \Delta_k, Y_t > 0|Y_0 = 1] \geq \sum_{\Delta_k} (b - a) \left(1 + \ln((n - b - a)p(a)) - (n - b - a)p(a)\right) - 2\ln(b - a) - 2 \geq \ln n \sum_{\Delta_k} (\omega_k - \omega_{k-1}) \left[1 + \ln \lambda(\omega_{k-1}) - \lambda(\omega_k) + o(1)\right] - O(L \ln \ln n) = \ln n \sum_{\Delta_k} (\omega_k - \omega_{k-1}) \left[1 + \ln \lambda(\omega_{k-1}) - \lambda(\omega_k)\right] - o(\bar{\omega} \ln n) - O(L \ln \ln n),$$

where the term $o(1)$ is of order $O(1/\ln n)$.

Note that $\sum_{\Delta_k} (\omega_k - \omega_{k-1})(1 + \ln \lambda(\omega_{k-1}) - \lambda(\omega_k))$ is a lower Darboux sum of the integral $I_{\epsilon,r} = \int_0^{\bar{\omega}} [1 - \lambda(\omega) + \ln \lambda(\omega)] d\omega$ under the partition $\Delta_k = [\omega_{k-1} \ln n, \omega_k \ln n], k = 1, \ldots, L$. Also note that the integrand is increasing with $\omega$, thus

$$I_{\epsilon,r} - \sum_{\Delta_k} (\omega_k - \omega_{k-1})(1 + \ln \lambda(\omega_{k-1}) - \lambda(\omega_k)) \leq \left[(1 - \lambda(\omega) + \ln \lambda(\omega)) - (1 - \lambda(0) + \ln \lambda(0))\right] \cdot \max(\omega_k - \omega_{k-1}) = - \left(\epsilon + \ln(1 - \epsilon)\right) \frac{\bar{\omega}}{L}.$$

Now we have

$$\ln \Pr[\forall t \in [0, \bar{\omega} \ln n], Y_t > 0|Y_0 = 1] - I_{\epsilon,r} \ln n \geq - \left(\epsilon + \ln(1 - \epsilon)\right) \frac{\bar{\omega}}{L} \ln n - o(\bar{\omega} \ln n) - O(L \ln \ln n) = -o(\ln n).$$
Next we take another interval \([a', b'] \equiv [\omega \ln n, K_0 \ln n]\), then \((n - b' - 1)p(a') = 1 - o(1)\).

\[
\ln P_r[\forall t \in [a', b'], Y_t > 0 | Y_{a'} > 0] \\
\geq (b' - a') \left(1 + \ln \left((n - b' - a')p(a')\right) - (n - b' - a')p(a')\right) - 2 \ln (b' - a') - 2 \\
= (b' - a')o(1) - 2 \ln (b' - a') - 2 = -o(\ln n).
\]

Consequently,

\[
\ln P_r[\forall t \in [0, \omega \ln n], Y_t > 0 | Y_0 = 1] - I_{\epsilon,r} \ln n \geq -o(\ln n),
\]

which leads to

\[
P_r[|C_v| \geq K_0 \ln n] \geq n^{I_{\epsilon,r}-o(1)}.
\]

With Proposition 8 in mind, and note that \(I_{\epsilon,r} > -1\) is a constant, it is straightforward to get the following result.

**Proposition 9.** Let \(X\) be the number of good vertices. There exists a constant \(\epsilon > 0\), such that

\[
\mathbf{E}[X] > n^{\epsilon}.
\]

The following result guarantees the existence of “good” vertices, which implies that random hypergraph \(G\) can be propagation connected.

**Lemma 10.** Let \(X\) be the number of good vertices. Then

\[
P_r[X > 0] = 1 - o(1).
\]

**Proof.** First we estimate the probability that vertices \(v_0\) and \(v_1\) are both good. To start with, we consider the conditional probability \(P_r[v_1 \text{ is good} | v_0 \text{ is good}].\)

Assume that \(\{Y_t, D_t\}\) and \(\{Y'_t, D'_t\}\) represent two propagation processes starting from \(v_0\) and \(v_1\), respectively. Recall the constant \(K_0\) which is defined in Lemma 8, then during the first \(K_0 \ln n\) steps, there are two cases:

1. We denote by \(B\) the event that there exists some \(t \leq K_0 \ln n\) such that two sets \(Y_t' \cup D_t'\) and \(Y_{K_0 \ln n} \cup D_{K_0 \ln n}\) have common vertices, i.e.,

\[
(Y_t' \cup D_t') \cap (Y_{K_0 \ln n} \cup D_{K_0 \ln n}) \neq \emptyset.
\]

2. Denote by \(\overline{B}\) the event that, for any \(t \leq K_0 \ln n\), the two sets \(Y_t' \cup D_t'\) and \(Y_{K_0 \ln n} \cup D_{K_0 \ln n}\) have no common vertices, i.e.,

\[
(Y_t' \cup D_t') \cap (Y_{K_0 \ln n} \cup D_{K_0 \ln n}) = \emptyset.
\]
Firstly, we consider the case (1).

Denote $R = Y_{K_0 \ln n} \cup D_{K_0 \ln n}$ the set of vertices connected from $v_0$. Then by Proposition 7 for $c = 2 - I_{\epsilon, \rho}$ there exists a constant $K_1(c)$ such that $\Pr[|R| > K_1(c) \ln n] < n^{-c}$. Let $B_t$ be the event that the two propagation processes starting from $v_0$ and $v_1$ do not intersect in the first $t - 1$ steps, but intersect at the $t$-th step. In fact, event $B$ is the union of all $B_t$ ($t = 1, \ldots, K_0 \ln n$).

$$\Pr[v_1 \text{ is good}, B|v_0 \text{ is good}] \leq \Pr[B|v_0 \text{ is good}]$$

$$\leq \Pr[B, |R| > K_1(c) \ln n|v_0 \text{ is good}] + \Pr[B, |R| \leq K_1(c) \ln n, |v_0 \text{ is good}]$$

$$\leq \Pr[|R| > K_1(c) \ln n|v_0 \text{ is good}] + \Pr[|R| \leq K_1(c) \ln n, v_0 \text{ is good}]$$

$$\leq \frac{n^{I_{\epsilon, \rho} - 2}}{\Pr[v_0 \text{ is good}]} + \sum_{t=0}^{K_0 \ln n} \Pr[B_t| |R| \leq K_1(c) \ln n, v_0 \text{ is good}]$$

$$\leq n^{-2+\epsilon} + \sum_{t=0}^{K_0 \ln n} |R| \cdot p(t) \leq n^{-2+\epsilon} + \sum_{t=0}^{K_0 \ln n} O[\ln n/n]$$

$$= O((\ln n)^2/n), \quad (4.13)$$

where the constant terms implicit in $O$ is at most $K_0 \cdot K_1(c) \cdot \lambda(K_0)$. The second to last inequality is because that, conditioned on the information of the propagation process starting from $v_0$, the event $B_t$ is equivalent to that the process from $v_1$ connects vertices in $R$ at step $t$. Recall that the probability for a vertex to be connected by a process at time $t$ is $p(t)$, and there are $|R|$ vertices in $R$, thus the inequality follows naturally.

Secondly, we consider the case (2).

Since the two sets $Y_{K_0 \ln n} \cup D_{K_0 \ln n}$ and $Y'_{K_0 \ln n} \cup D'_{K_0 \ln n}$ have no intersections, then the first $K_0 \ln n$ steps of the propagation process which starts from vertex $v_1$ contains no vertices from the set $Y_{K_0 \ln n} \cup D_{K_0 \ln n}$. In fact in this case, $v_1$ has the following Markov additive process:

\[
\begin{cases}
Y_0' = 1, \\
Y_{t+1}' = Y_t' + Z_t' - 1,
\end{cases}
\]

where

\[
Z_t'|Y_t' \sim B[n - t - |R| - Y_t', p(t)], p(t) = 1 - (1 - p_2)(1 - p_3)^t. \quad (4.14)
\]

Note that $B[n - t - z, p(t)] \succ B[n - t - |R| - z, p(t)]$ holds for all $z$. Applying lemma 4 if $t = K_0 \ln n$, then we have

$$\Pr[Y_t' \geq 0, Y_1' > 0, \ldots, Y_{t-1}' > 0] \leq \Pr[Y_1 \geq 0, Y_1 > 0, \ldots, Y_{t-1} > 0]. \quad (4.15)$$

This means that, the probability of the propagation process starting from $v_1$ survives up to time $K_0 \ln n$ is less than that of $v_0$, i.e.,

$$\Pr[v_1 \text{ is good} | B|v_0 \text{ is good}] \leq \Pr[v_1 \text{ is good} | B, v_0 \text{ is good}] \leq \Pr[v_1 \text{ is good}] = \Pr[v_0 \text{ is good}]. \quad (4.16)$$
Therefore, we combine case (1) and case (2) to get

\[ Pr[v_1 \text{ is good}|v_0 \text{ is good}] = Pr[v_1 \text{ is good}, B|v_0 \text{ is good}] + Pr[v_1 \text{ is good}, \overline{B}|v_0 \text{ is good}] \leq O\left(\frac{(\ln n)^2}{n}\right) + Pr[v_0 \text{ is good}] = (1 + o(1))Pr[v_0 \text{ is good}]. \quad (4.17) \]

Recall that \( Pr[v_0 \text{ is good}] > n^{-1+\epsilon} \), as a result,

\[
E[X^2] = n(n-1)Pr[v_1 \text{ is good}, v_0 \text{ is good}] + nPr[v_0 \text{ is good}]
= Pr[v_0 \text{ is good}]\left(n + n(n-1)Pr[v_1 \text{ is good}|v_0 \text{ is good}]\right)
= Pr[v_0 \text{ is good}](n + (1 + o(1))n(n-1)Pr[v_0 \text{ is good}])
= (1 + o(1))n^2Pr[v_0 \text{ is good}]^2 = (1 + o(1))E[X]^2. \quad (4.18)
\]

Finally, by the Cauchy inequality, we have

\[ Pr[X > 0] \geq \frac{E[X]^2}{E[X^2]} = 1 + o(1). \]

\[ \square \]

5 Conclusion

In this paper, we study the propagation connectivity of a special random hypergraph \( G(n, p_2, p_3) \) which is composed of edges of length 2 and hyperedges of length 3. Using probabilistic arguments, we prove that there is a sharp phase transition of propagation connectivity of \( G(n, p_2, p_3) \) that if \( I_{\epsilon,r} < -1 \) (\( I_{\epsilon,r} \) is a constant defined by the parameters \( \epsilon \) and \( r \)), then w.h.p. there is no vertex that propagation connects all the vertices in \( G \), moreover, every propagation connected component in \( G \) contains no more than \( O(\ln n) \) vertices; while if \( I_{\epsilon,r} > -1 \), then w.h.p. there exist “good” vertices which can surprisingly connect all vertices in the entire graph once they can propagation connect more than \( O(\ln n) \) vertices.

In view of the possible relevant connection with the discrete, realistic and complex networks, we believe the above results would be useful to provide a theoretical foundation for our understanding of various complex networks that permeate this information age.

References

[1] Behrisch M., Coja-Oghlan A. and Kang M. (2014). Local limit theorems for the giant component of random hypergraphs. Combinatorics, Probability and Computing. 23(03) 331-366.

[2] Behrisch M., Coja-Oghlan A. and Kang M. (2010). The order of the giant component of random hypergraphs. Random Structures and Algorithms. 36(2) 149-184.
Appendix. Proof of Corollary 1

The proof of Corollary 1 is a minor modification of Proposition 6.

Note that \( K_0 = \max\{1, \lambda^{-1}(10)\} \geq \frac{\epsilon}{r} \), so we only need to prove

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \overline{\omega} \ln n] \leq n^{I_{s,\omega}+o(1)}, \text{ where } \overline{\omega} = \frac{\epsilon}{r}.
\]

Since

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \overline{\omega} \ln n] = \Pr[Y_t > 0, t = 0, 1, ..., \overline{\omega} \ln n],
\]

we only need to estimate the probability that \( Y_t > 0 \) holds for all \( t \in [0, \overline{\omega} \ln n] \) if \( Y_0 = m \).

**Step 1.** Take a large positive integer \( L \), we divide the interval \( [0, \overline{\omega} \ln n] \) into \( L \) intervals: \( \Delta_k = [t_{k-1}, t_k] \equiv [\omega_{k-1} \ln n, \omega_k \ln n] \), where \( \omega_k = \frac{k}{L+1} \), \( k = 1, 2, ..., L \). We now estimate the following probability on the interval \( \Delta_k \).

\[
\Pr[|\Delta_k|] \equiv \Pr[\forall t \in [t_{k-1}, t_k-1], Y_t > 0, Y_{t+k} = y_k | Y_{t+k-1} = y_{k-1}],
\]

where \( \{y_1, ..., y_k\} \) is a random sequence taking nonnegative integers, and \( y_0 = 2 \) for beginning.

Consider the propagation process on the interval \( t \in \Delta_k = [t_{k-1}, t_k] \):

\[
\begin{cases}
Y_{t+k-1} = y_{k-1}, \\
Y_{t+1} = Y_t + Z_t - 1,
\end{cases}
\]

Appendix. Proof of Corollary 1

The proof of Corollary 1 is a minor modification of Proposition 6.

Note that \( K_0 = \max\{1, \lambda^{-1}(10)\} \geq \frac{\epsilon}{r} \), so we only need to prove

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \overline{\omega} \ln n] \leq n^{I_{s,\omega}+o(1)}, \text{ where } \overline{\omega} = \frac{\epsilon}{r}.
\]

Since

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \overline{\omega} \ln n] = \Pr[Y_t > 0, t = 0, 1, ..., \overline{\omega} \ln n],
\]

we only need to estimate the probability that \( Y_t > 0 \) holds for all \( t \in [0, \overline{\omega} \ln n] \) if \( Y_0 = m \).

**Step 1.** Take a large positive integer \( L \), we divide the interval \( [0, \overline{\omega} \ln n] \) into \( L \) intervals: \( \Delta_k = [t_{k-1}, t_k] \equiv [\omega_{k-1} \ln n, \omega_k \ln n] \), where \( \omega_k = \frac{k}{L+1} \), \( k = 1, 2, ..., L \). We now estimate the following probability on the interval \( \Delta_k \).

\[
\Pr[|\Delta_k|] \equiv \Pr[\forall t \in [t_{k-1}, t_k-1], Y_t > 0, Y_{t+k} = y_k | Y_{t+k-1} = y_{k-1}],
\]

where \( \{y_1, ..., y_k\} \) is a random sequence taking nonnegative integers, and \( y_0 = 2 \) for beginning.

Consider the propagation process on the interval \( t \in \Delta_k = [t_{k-1}, t_k] \):

\[
\begin{cases}
Y_{t+k-1} = y_{k-1}, \\
Y_{t+1} = Y_t + Z_t - 1,
\end{cases}
\]
where
\[ Z_t | Y_t \sim B[n - t - Y_t, p(t)], \] where \( p(t) = 1 - (1 - p_2)(1 - p_3)^t \).

Define another process which corresponds to it as:
\[
\begin{cases}
Y_{t_{k-1}}^{(2)} = y_{k-1}, \\
Y_{t+1}^{(2)} = Y_t^{(2)} + Z_t^{(2)} - 1,
\end{cases}
\]
where (recall (2.3))
\[ Z_t^{(2)} | Y_t^{(2)} \sim B \left[ n, \frac{\lambda(\omega_k)}{n} \right]. \]

Since \( p(t) \leq \lambda(\omega_k)/n \), thus
\[ B[n - t - Y_t, p(t)] < B \left[ n, \frac{\lambda(\omega_k)}{n} \right]. \]

Taking \( t_{k-1} + 1 = \ldots = t_{k-1} = 0, t_{k+1} = y_k - 1 \), and applying lemma 4 we have
\[
\Pr[\Delta_k] = \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t > 0, Y_{t_k} = y_k | Y_{t_{k-1}} = y_{k-1}]
\leq \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t > 0, Y_{t_k} \geq y_k | Y_{t_{k-1}} = y_{k-1}]
\leq \Pr[\forall t \in [t_{k-1}, t_k - 1], Y_t^{(2)} > 0, Y_{t_k}^{(2)} \geq y_k | Y_{t_{k-1}}^{(2)} = y_{k-1}]
\leq \Pr[Y_{t_k}^{(2)} \geq y_k | Y_{t_{k-1}}^{(2)} = y_{k-1}]. \tag{5.19}
\]

Denote
\[ Z^{(2)} = \sum_{t \in [t_{k-1}, t_k - 1]} Z_t^{(2)} \sim B \left( \frac{\omega}{L} n \ln n, \frac{\lambda(\omega_k)}{n} \right). \]

The generating function of \( Z^{(2)} \) is
\[ \mathbb{E}[s^{Z^{(2)}}] = \left( 1 + \frac{\lambda(\omega_k)}{n} (s - 1) \right)^{\frac{\omega}{L} n \ln n}. \]

Denote
\[ d_k = \frac{y_k - y_{k-1}}{\omega \ln n / L}, s_k = \frac{1 + d_k}{\lambda(\omega_k)} \quad \text{for } k = 2, \ldots, L. \]

and when \( k = 1 \) denote \( d_1 = \frac{y_1 - y_0 + 1}{\omega \ln n / L} = \frac{y_1 - 1}{\omega \ln n / L}, s_1 = \frac{1 + d_1}{\lambda(\omega_1)}. \)

If \( d_k \geq \lambda(\omega_k) - 1 \), then \( s_k \geq 1 \). Substitute \( s = s_k \) into the generating function, now we can rewrite the inequality (5.19) as (when \( k \geq 2 \))
\[
\ln \Pr[\Delta_k] \leq \ln \Pr[Y_{t_k}^{(2)} \geq y_k | Y_{t_{k-1}}^{(2)} = y_{k-1}] = \ln \Pr[Z^{(2)} \geq y_k - y_{k-1} + \frac{\omega}{L} \ln n]
\leq \ln \frac{\mathbb{E}[s^{Z^{(2)}}]}{s_k^{y_k - y_{k-1} + \frac{\omega}{L} \ln n}} \leq (s_k - 1) \lambda(\omega_k) \frac{\omega}{L} \ln n - \left( y_k - y_{k-1} + \frac{\omega}{L} \ln n \right) \ln s_k
= \left( 1 + d_k - \lambda(\omega_k) \right) \frac{\omega}{L} \ln n - (1 + d_k) \left[ \ln(1 + d_k) - \ln \lambda(\omega_k) \right] \frac{\omega}{L} \ln n.
\]

22
and (when \(k = 1\))

\[
\ln \Pr[\Delta_1] \leq (s_1 - 1)\lambda(\omega_1) \frac{\omega}{L} \ln n - \left( y_1 - y_0 + \frac{\omega}{L} \ln n \right) \ln s_1 \\
= (s_1 - 1)\lambda(\omega_1) \frac{\omega}{L} \ln n - \left( y_1 - y_0 + (m - 1) + \frac{\omega}{L} \ln n \right) \ln s_1 + \ln s_1 \\
= (1 + d_1 - \lambda(\omega_1)) \frac{\omega}{L} \ln n - (1 + d_1) \left[ \ln(1 + d_1) - \ln \lambda(\omega_1) \right] \frac{\omega}{L} \ln n + \ln s_1.
\]

Since \(\lambda(\omega_1) = 1 - \epsilon + r\omega_1 = 1 - \epsilon + r\frac{\omega}{L} \geq \frac{1}{L}\). And we assume that \(y_k \leq K_1(c) \ln n\) when we use the inequality for \(\ln \Pr[\Delta_k]\), which implies that \(d_1 \leq \frac{L K_1(c) \omega}{\epsilon} = \frac{L K_1(c) \epsilon}{\omega}\). So \(s_1 = \frac{1 + d_1}{\lambda(\omega_1)} \leq L^2 \times \text{const}\), and

\[
(m - 1) \ln s_1 \leq 2(m - 1) \ln L + \text{const} = O(\ln L).
\]

By the additive relations of sequence \(\{Y_{t_k} = y_k \geq 1\}\), we know that \(\{d_k\}\) satisfies

\[
d_1 \geq 0, d_1 + d_2 \geq 0, \ldots, d_1 + \ldots + d_L \geq 0.
\]

For simplicity, define a function of \(d_k\):

\[
M_k(d_k) = \begin{cases} 1 + d_k - \lambda(\omega_k) - (1 + d_k)(\ln(1 + d_k) - \ln \lambda(\omega_k)), & \text{if } d_k \geq \lambda(\omega_k) - 1, \\ 0, & \text{if } d_k < \lambda(\omega_k) - 1. \end{cases}
\]

Obviously, \(M_k(d_k)\) decreases with \(d_k\). Now we have

\[
\ln \Pr[\Delta_k] \leq M_k(d_k) \frac{\omega}{L} \ln n + O(\ln L)1_{k=1}.
\]

**Step 2.** Since step 2 is exactly the same with Proposition 6 we omit it to avoid duplication.

**Step 3.** Finally, returning to the estimate of \(\Pr[|C_{\{v_1,v_2\}}| \geq \omega \ln n]\).

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \omega \ln n] = \Pr[Y_t > 0, \forall t = 0, 1, \ldots, \omega \ln n] \\
= \sum_{y_1 > 0, \ldots, y_L > 0} \Pr[Y_t > 0, t = 0, 1, \ldots, \omega \ln n; Y_{t_k} = y_k, k = 1, \ldots, L] \\
= \sum_{y_1 > 0, \ldots, y_L > 0} H(y_1, \ldots, y_L).
\]

Recall Proposition 7 and take \(c = 2 - I_{\epsilon, r}\), then

\[
\sum_{y_1 > 0, \ldots, y_L > 0} H(y_1, \ldots, y_L) \\
\leq \sum_{y_1 > 0, \ldots, y_L > 0} \Pr[Y_{t_k} = y_k, k = 1, \ldots, L] = \Pr[Y_t > K_1(c) \ln n] \\
\leq n^{-c} = n^{I_{\epsilon, r} - 2}.
\]

Consequently,

\[
\Pr[|C_{\{v_1,v_2\}}| \geq \omega \ln n]
\]
\[
\begin{align*}
&= \sum_{y_1 \leq K_1(c) \ln n, \ldots, y_L \leq K_1(c) \ln n} H(y_1, \ldots, y_L) + \sum_{j=1}^{L} \sum_{y_1 > 0, \ldots, y_j > 0, y_j > K(c) \ln n, \ldots, y_L > 0} H(y_1, \ldots, y_L) \\
&\leq L^{O(1)}(K_1(c) \ln n)^L n^{I_{c,r} \ln n} + L n^{I_{c,r} - 2} \\
&\leq n^{I_{c,r} + \varepsilon},
\end{align*}
\]

where the last inequality holds for any sufficiently small constant \( \varepsilon = \varepsilon(L) > 0 \).