ISSN 1342-2812

Research Reports on Mathematical and Computing Sciences

Propagation Connectivity of Random Hypergraphs

A. Coja-Oghlan, M. Onsjö, and O. Watanabe

April 2010, C–271

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES C: Computer Science

Propagation Connectivity of Random Hypergraphs

Amin Coja-Oghlan[*], Mikael Onsjö[†], and Osamu Watanabe[†]

Tokyo Tech. Dept. MSC, Research Report C-271, April 2010

Abstract

We study the concept of *propagation connectivity* on random 3-uniform hypergraphs. This concept is inspired by a simple linear time algorithm for solving instances of certain constraint satisfaction problems. We derive upper and lower bounds for the propagation connectivity threshold, and point out some algorithmic implications.
Key words: random hypergraphs, constraint satisfaction problems, efficient algorithms.

1 Introduction and results

1.1 The propagation connectivity threshold

There are several natural ways to define connectivity for 3-uniform hypergraphs $H = (V, E)$. For instance, a standard concept is to consider H connected if the graph obtained by replacing each edge e by a triangle is connected (recall that in a 3-uniform hypergraph each edge is a set of three vertices).

In this paper we study a different concept that we call *propagation connectivity*.

Definition 1. Let $H = (V, E)$ be a 3-uniform hypergraph on $n = |V|$ vertices. We call a sequence $e_1, \ldots, e_{n-2} \in E$ a *propagation sequence* if for any $1 \leq l < n-2$ we have $|e_{l+1} \cap \bigcup_{i=1}^{l} e_l| = 2$. If H has a propagation sequence, then we say that H is *propagation connected*.

This definition is motivated by a simple algorithm for a certain kind of constraint satisfaction problem. For the time being, let us focus on the concrete example of a system of linear equations over a finite field with three variables per equation. We can associate a hypergraph H with this system by thinking of the variables as vertices and of the equations as hyperedges. If we are given a propagation sequence $e_1, \ldots, e_{n-2}$ for H, then we can find a solution to the system of equations in linear time (if there is one). Namely, suppose that the variables of e_1 are x, y, z. We can easily 'guess' the correct values of x, y (i.e., we can try all possible assignments because the field is finite). Then the value of z is implied. Now, assume inductively that we have obtained the values of the variables occurring in the first l edges/equations $e_1, \ldots, e_l$

[*]University of Warwick, Mathematics and Computer Science, Zeeman building, Coventry CV4 7AL, UK, a.coja-oghlan@warwick.ac.uk. Supported by EP/G039070/1 and DIMAP.

[†]Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, Meguro-ku Ookayama 2-12-1 W8-25, {mikael,watanabe}@is.titech.ac.jp. Supported in part by the JSPS Global COE program CompView and by Grants-in-Aid for Scientific Research on the Priority Area Dex-SMI from MEXT

already. Then e_{l+1} contains precisely one additional variable (by the definition of propagation sequence), whose value we can thus infer directly. Thus, after passing through the entire sequence $e_1, \dots, e_{n-2}$, we have determined the values of all n variables. If this solves the linear system, we are done. Conversely, if we find that no assignment to the first two variables x, y leads to a solution, then it is safe to conclude that no solution exists.

The contribution of this paper is close upper and lower bounds on the edge probability that the propagation connectivity holds in random hypergraphs. More precisely, we consider the following random hypergraph model $\mathcal{H}(n,p)$: the vertex set of the random hypergraph is $V = [n] = \{1, \dots, n\}$, and each of the $\binom{n}{3}$ possible edges is present with probability $0 \leq p \leq 1$ independently. We write $H : \mathcal{H}(n,p)$ to indicate that H is a random hypergraph chosen from this distribution. Moreover, we say that the random hypergraph has some property *with high probability* (w.h.p.) if the probability that the property holds converges to one as $n \to \infty$.

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

Determining the threshold for 'standard' connectivity (where each hyperedge is replaced by a triangle) is easy. The result is a hardly surprising $p \sim 2n^{-2} \ln n$, and the proof is via a simple coupon collecting argument. By contrast, analyzing propagation connectivity is quite non-trivial. Our proof is based on a kind of large deviations analysis of a time-dependent random walk. A precise solution of this problem might close the gap left by Theorem 1 for showing a propagation connectivity threshold (if it indeed exists).

1.2 Computing a propagation sequence

For a propagation connected hypergraph H one can determine a propagation sequence in polynomial time via a generalized breadth first search procedure. However, the running time of this algorithm is superlinear (in contrast to BFS on graphs). The following theorem shows that there is an algorithm with linear expected running time. (The proof is given in the appendix.)

Theorem 2. There is a randomized algorithm A that satisfies the following. Fix $c > 0.25$ and let $p = c/(n \ln n)$. Then A applied to $H : \mathcal{H}(n,p)$ outputs a propagation sequence w.h.p. in linear expected time.

As an application, we show how Theorem 2 yields an algorithm for deciding a class of random constraint satisfaction problems. A *CSP instance with domain* $[k] = \{1, \dots, k\}$ consists of a hypergraph $H = (V, E)$ with $V = [n]$ and a family $(f_e)_{e \in E}$ of maps $f_e : [k] \times [k] \times [k] \to \{0, 1\}$. Moreover, a *solution* is a map $\sigma : V \to [k]$ such that for any triple $1 \leq x < y < z \leq n$ of vertices with $e = \{x, y, z\} \in E$ we have $f_e(\sigma(x), \sigma(y), \sigma(z)) = 1$. Thus, intuitively the hypergraph H describes the interactions of the variables V, and for any edge e the map f_e characterizes the values that can be assigned to the variables in e so as to satisfy the constraint that e represents.

Furthermore, we say that a CSP instance is *propagating* if for any $x, y \in [k]$, any $i \in \{1,2,3\}$, and any edge $e \in E$ there is precisely one value $z_i \in [k]$ such that $f_e(z_1, x, y) = f_e(x, z_2, y) = f_e(x, y, z_3) = 1$. In other words, once we assign two variable in a constraint e, there is precisely one way to assign the third variable so as to satisfy e. Clearly, systems of linear equations over a finite field provide an example of propagating problems, but there are many others.

By combining Theorem 2 with the simple propagation procedure outlined after Definition 1, we obtain the following result.

Corollary 3. Fix $c > 0.25$ and $k \geq 2$ and let $p = c/(n \ln n)$. Moreover, assume that P is a probability distribution over propagating CSP instance with domain $[k]$ such that the distribution of the random hypergraph underlying the problem instance coincides with the distribution $\mathcal{H}(n,p)$. There is an algorithm with linear expected running time that decides whether a random CSP instance chosen from the distribution P has a solution w.h.p.

There are a variety of probability distribution over CSPs that satisfy the assumptions of Corollary 3. Examples include uniformly random systems of linear equations, which at the density assumed in Corollary 3 do not have solutions w.h.p. Thus, for these problems running the algorithm in Corollary 3 will provide a *succinct proof* that no solution exists w.h.p. On the other hand, distributions that do admit solutions w.h.p. include systems of linear equations with a 'planted' solution, for which the algorithm will find a solution in linear time w.h.p.

1.3 Related work

The 'standard' concept of random hypergraph connectivity (where edges are replaced by triangles) has been studied, e.g., in [BCK07, CMV07], particularly with respect to the emergence and size of the giant component. These results generalize what was known for random graphs (see [JLR00] for a comprehensive summary). A further related random hypergraph concept is that of a core. This concept is related to local search algorithms such as the 'pure literal rule' for the satisfiability. Contributions on these subjects include [DN05, Mol05].

Berke and Onsjö [BO09] approached the propagation connectivity threshold for random 3-uniform hypergraphs. They established a lower bound of $p = \Omega(1/n(\log n)^2)$ and an upper bound of $p = O(1/n(\log n)^{0.4})$. As Theorem 1 shows, the correct order of magnitude is $p = \Theta(1/(n \ln n))$.

With respect to the application to random constraint satisfaction problems, it is clear that the case of linear equations over finite fields can be solved in polynomial (albeit superlinear) time by Gaussian elimination. However, if the underlying hypergraph comes with a propagation sequence, then the problem can be solved in linear time as indicated. While linear equations provide an example of propagating constraint satisfaction problems, there exist NP-hard examples, too [CM04].

1.4 Preliminaries and notation

We will use the following Chernoff bound on the tails of a binomially distributed random variable X with mean μ (e.g., [JLR00, p. 21]): letting $\varphi(x) = (1+x)\ln(1+x) - x$, we have

for any $t > 0$

$$\begin{array}{rcl} \Pr[\, X \leq \mu - t\,] & \leq & \exp\bigl(-\mu \cdot \varphi(-t/\mu)\bigr), \quad \text{and} \\ \Pr[\, X \geq \mu + t\,] & \leq & \exp\bigl(-\mu \cdot \varphi(t/\mu)\bigr), \end{array} \tag{1}$$

We will also use the following Stirling bounds, see, e.g., [MU05, Lemma 7.3]: for any n, we have

$$\sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n \;\leq\; n! \;\leq\; 2\sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n.$$

For simplifying our notations we omit specifying ceiling or floor functions so long as they can be determined from the context.

2 The propagation process

In this section we show how the propagation connectivity problem can be modeled by a stochastic process, which we call the *propagation process*. We start out by describing this process for a fixed hypergraph $H = (V, E)$ with vertex set $V = \{1, \ldots, n\}$. Let (v_1, v_2) be a pair of distinct vertices, which we refer to as the *initial pair*. In the course of the prcoess, vertices are either *active*, *neutral*, or *dead*. Initially v_1 is dead, v_2 is active, and all other vertices are neutral; formally, we let

$$\mathcal{D}_0^{(v_1,v_2)}[H] = \{v_1\}, \quad \mathcal{A}_0^{(v_1,v_2)}[H] = \{v_2\}.$$

Once there is no active vertex left, the process stops. Otherwise at each time $t \geq 1$, the least active vertex u is chosen (recall that $V = [n]$ is an ordered set). All neutral vertices v for which there is a dead vertex w such that $\{u, v, w\} \in E$ are declared active, and then u is declared dead. In symbols, we let $u = \min \mathcal{A}_{t-1}^{(v_1,v_2)}[H]$ and

$$\begin{array}{rcl} \mathcal{D}_t^{(v_1,v_2)}[H] & = & \mathcal{D}_{t-1}^{(v_1,v_2)}[H] \cup \{u\}, \\ \mathcal{A}_t^{(v_1,v_2)}[H] & = & \Bigl(\mathcal{A}_{t-1}^{(v_1,v_2)}[H] \setminus \{u\}\Bigr) \cup \Bigl\{v \notin \mathcal{D}_{t-1}^{(v_1,v_2)}[H] : \exists w \in \mathcal{D}_{t-1}^{(v_1,v_2)}[H] : \{u, v, w\} \in E\Bigr\}. \end{array}$$

Thus, at time t the total number of dead vertices equals $t + 1$. Let $T^{(v_1,v_2)}[H]$ be the time when the process stops. To avoid case distinctions, we consider vertices dead (or active, or neutral) at times $t > T^{(v_1,v_2)}[H]$ if they had the corresponding predicate at time $T^{(v_1,v_2)}[H]$. Observe that for a fixed hypergraph H, the process is entirely deterministic.

The process is related to the propagation connectivity problem as follows. Assume that vertex v was declared active at time $t \geq 2$. Then H has an edge e_t that contains v and two vertices from $\mathcal{D}_t^{(v_1,v_2)}[H]$. Proceeding inductively, we obtain a sequence $e_2, \ldots, e_t$ such that $v_1, v_2 \in e_2$ and $|e_{l+1} \cap \bigcup_{i=2}^{l}| \geq 2$ for all $2 \leq l < t$. Hence, if all vertices are declared dead eventually, i.e., if $T^{(v_1,v_2)}[H] = n - 1$, then we obtain a propagation sequence. Conversely, if there is a propagation sequence $e_2, \ldots, e_{n-1}$ such that $v_1, v_2 \in e_2$, then the propagation process will not stop before time $n - 1$. Thus, we have the following.

Fact 1. *H is propagation connected iff there is a pair (v_1, v_2) such that $T^{(v_1,v_2)}[H] = n - 1$.*

To prove Theorem 1, we are going to study the propagation process on a random hypergraph $H : \mathcal{H}(n, p)$. In this case we omit the reference to H, i.e., we just write $\mathcal{D}_t^{(v_1,v_2)}$ etc. It will be

convenient to use the terminology of stochastic processes. In particular, for $t \geq 0$ we let $\mathcal{F}_t^{(v_1,v_2)}$ signify the coarsest σ-algebra on $\mathcal{H}(n,p)$ in which all events $\{v \in \mathcal{D}_s^{(v_1,v_2)}\}$ and $\{v \in \mathcal{A}_s^{(v_1,v_2)}\}$ for $s \leq t$ and $v \in V$ are measurable. Then $(\mathcal{F}_t^{(v_1,v_2)})_{t\geq 0}$ is a filtration. We will also use the concept of conditional probabilities with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ (see [D05]). To remind the reader, for an event A and a (fixed) hypergraph H_0 the conditional probability is

$$\Pr\left[A|\mathcal{F}_t^{(v_1,v_2)}\right](H_0) \quad = \quad \frac{\Pr\left[A \text{ occurs and } \mathcal{D}_s^{(v_1,v_2)} = \mathcal{D}_s^{(v_1,v_2)}[H_0],\, \mathcal{A}_s^{(v_1,v_2)} = \mathcal{A}^{(v_1,v_2)}[H_0] \text{ for all } s \leq t\right]}{\Pr\left[\mathcal{D}_s^{(v_1,v_2)} = \mathcal{D}_s^{(v_1,v_2)}[H_0],\, \mathcal{A}_s^{(v_1,v_2)} = \mathcal{A}^{(v_1,v_2)}[H_0] \text{ for all } s \leq t\right]}.$$

In words, $\Pr\left[A|\mathcal{F}_t^{(v_1,v_2)}\right](H_0)$ is the probability of the event A in a random hypergraph $H : \mathcal{H}(n,p)$ given that the first t steps of the propagation process on H work out the same as in H_0. As per standard practice, where the argument H_0 is omitted, it is understood that the corresponding statement holds for all H_0.

For any $t \geq 1$ the first t steps of the propagation process on the random hypergraph $H : \mathcal{H}(n,p)$ *only* depend on the presence (or absence) of edges that contain at least two vertices that have been declared dead by time t, i.e., from the set $\mathcal{D}_t^{(v_1,v_2)}$. This means that the presence of edges e with $|e \cap \mathcal{D}_t^{(v_1,v_2)}| < 2$ is stochastically independent of the first t steps.

Fact 2. Given $\mathcal{F}_t$, for all triples $e = \{u,v,w\}$ such that $\left|e \cap \mathcal{D}_t^{(v_1,v_2)}\right| < 2$, the edge e is present in $H : \mathcal{H}(n,p)$ with probability p independently. In symbols, for any set

$$\mathcal{E} \subset \left\{e \in \binom{V}{3} : \left|e \cap \mathcal{D}_t^{(v_1,v_2)}\right| < 2\right\}$$

we have $\Pr\left[\mathcal{E} \subset E(H)|\mathcal{F}_t^{(v_1,v_2)}\right] \;=\; p^{|\mathcal{E}|}$.

The above propagation process is similar in spirit to the branching process approach for the giant component problem in random graphs/digraphs [Kar90]. The difference between our proofs and the standard argument is that we need to investigate whether *there exists* a pair (v_1,v_2) such that $T^{(v_1,v_2)} \geq n-1$ (cf. Fact 1). Since there are a total of $\binom{n}{2}$ initial pairs to choose from, this means that we need to study *unlikely* trajectories of the propagation process (that occur with probability merely about $1/\binom{n}{2}$).

By contrast, for the giant component problem the corresponding process has to be studied only from a *random* start vertex, a problem which relatively easily reduces to the typical behavior of a standard Galton-Watson branching process. Alternatively, the problem can be tackled via a whole arsenal of different techniques, ranging from differential equations to random walks. Unfortunately, the fact that here we need to study an 'exceptional' event puts these standard arguments out of business.

To get started, we point out that the hypergraph distribution $H : \mathcal{H}(n,p)$ is invariant w.r.t. permutations of the vertices. Therefore, the distribution of the propagation process is the same for any initial pair. For the sake of concreteness we will refer to $(v_1,v_2) = (1,2)$. For this

initial pair we will omit the superscript (v_1, v_2) from the notation. Moreover, we let $A_t = |\mathcal{A}_t|$ be the number of active vertices at time t (from the initial pair $(1, 2)$). Then $A_0 = 1$ by construction. For any $t \geq 1$, we define a further random variable X_t via

$$X_t \;=\; A_t - A_{t-1} + 1. \tag{2}$$

That is, X_t is the number of vertices that got declared active at time t.

Fact 3. If $1 \leq t \leq T$, then given $\mathcal{F}_{t-1}$, the random variable X_t is binomially distributed $\mathrm{Bin}(n - t - A_{t-1}, 1 - (1-p)^t)$.

Proof. The number of neutral vertices at time $t-1$ equals $n - A_{t-1} - |\mathcal{D}_{t-1}| = n - A_{t-1} - t$. Suppose that v is neutral at time $t-1$ and let $u = \min \mathcal{A}_{t-1}$. Then v becomes active at time t iff there is $w \in \mathcal{D}_{t-1}$ such that $\{u, v, w\} \in E$. By Fact 2 each of these t edges is present in H with probability p independently. Hence, the probability that all of them are absent is $1 - (1-p)^t$. □

To outline the proof of Theorem 1, let us interpret the propagation process in terms of a time-dependent random walk. The process continues up to time t iff $A_s > 0$ for all $1 \leq s \leq t$. Due to (2), this is true iff $\sum_{q=1}^{s}(X_q - 1) \geq 0$ for all $1 \leq s \leq t$. Thus, if we think of the random variables $X_s - 1$ as the steps of a random walk, then the propagation process continues to time t iff the random walk stays non-negative at all times $s \leq t$. As Fact 3 shows, this random walk is time-dependent.

In the regime $p = \Theta(1/(n \ln n))$ that we are interested in, and for times $s \ll \ln n$, the random walk has a negative drift. More precisely, for $s \ll \ln n$ Fact 3 implies that the expectation of $X_s - 1$ is $(1 + o(1))nps - 1 < 0$. Therefore, standard results on random walks show that the probability that the random walk will continue to time, say, $\ln n$ is $o(1)$. If, however, the process happens to survive up to time $t = 1/(np) = \Theta(\ln n)$ for a fixed $\epsilon > 0$, then Fact 3 shows that the 'drift' of $X_t - 1$ becomes positive and thus the process is likely to continue up to time $n - 1$.

The previous paragraph shows that the probability that one specific initial pair leads to a propagation sequence is $o(1)$. But this does *not* imply that the random hypergraph $H : \mathcal{H}(n, p)$ is not propagation connected w.h.p., because there is a total $\binom{n}{2}$ initial pairs to choose from. This observation suggests that in order find the threshold for propagation connectivity we need to determine for what p the random walk continues to time $1/(np)$ with probability $1/\binom{n}{2}$. In Section 3 we will derive a lower bound on this value of p. The more challenging problem is to obtain an upper bound, which we address in Section 4.

3 The lower bound

In this section we prove the first part of Theorem 1, i.e., we show that the random hypergraph $H : \mathcal{H}(n, p)$ is *not* propagation connected w.h.p. if $p < 0.16/(n \ln n)$. To this end, we will derive that the probability that the initial pair $(1, 2)$ leads to a propagation sequence is $o(n^{-2})$. By symmetry and the union bound, this implies that w.h.p. *no* initial pair (v_1, v_2) does. We start

by reducing the problem of estimating the probability that $(1,2)$ yields a propagation sequence to an exercise in calculus.

Lemma 4. Let $p = c/(n \ln n)$ for a fixed $c > 0$ and assume that $0 < d \leq 2/c$ is such that $d(cd/2 + \ln(2/cd) - 1) > 2$. Let $t_0 = d \ln n$. Then $\Pr[T > t_0] = o(n^{-2})$.

Proof. Let $\{\widetilde{X}_t\}_{t \geq 1}$ be a family of mutually independent random variables such that $\widetilde{X}_t$ has distribution $\mathrm{Bin}(nt, p)$. Let $t \geq 1$. By construction, for each vertex $v \in \mathcal{A}_t \setminus \mathcal{A}_{t-1}$ that becomes active at time t, there is an edge $\{u, v, w\}$ in $H : \mathcal{H}(n,p)$ such that $u = \min \mathcal{A}_{t-1}$ and $w \in \mathcal{D}_{t-1}$. In particular, the number $X_t = |\mathcal{A}_t \setminus \mathcal{A}_{t-1}| + 1$ of newly active vertices v is bounded by the number of such edges $\{u, v, w\}$. By Fact 2, given $\mathcal{F}_{t-1}$, each such edge is present in H with probability p independently. As $|\mathcal{D}_{t-1}| = t$ and because the number of neutral vertices v to choose from is bounded by n, this shows that $X_t | \mathcal{F}_{t-1}$ is stochastically dominated by the binomial variable $\widetilde{X}_t = \mathrm{Bin}(nt, p)$.

If the stopping time T exceeds some specific time t_0, then $A_t \geq 1$ for all $t \in [t_0]$. Hence, (2) implies $\sum_{1 \leq t \leq t_0} X_t \geq t_0$. Because each X_t is dominated by $\widetilde{X}_t$, we can bound the probability of this event by

$$\begin{aligned} \Pr[T \geq t_0] &\leq \Pr\left[\sum_{1 \leq t \leq t_0} X_t \geq t_0\right] \leq \Pr\left[\sum_{1 \leq t \leq t_0} \widetilde{X}_t \geq t_0\right] \\ &= \Pr\left[\mathrm{Bin}\left(n \cdot \sum_{1 \leq t \leq t_0} t, p\right) \geq t_0\right] = \Pr\left[\mathrm{Bin}\left(n \cdot \frac{t_0(t_0+1)}{2}, p\right) \geq t_0\right]. \quad (3) \end{aligned}$$

Let μ_0 denote the expectation of this last binomial distribution. Then

$$\mu_0 = n \cdot \frac{t_0(t_0+1)}{2} \cdot p = \frac{cd^2}{2}\left(1 + \frac{2}{d \ln n}\right) \ln n \sim \frac{cd^2}{2} \ln n. \quad (4)$$

We are going to verify that our assumption on c, d implies that the r.h.s. of (3) is $o(n^{-2})$. Since we assume $d \leq 2/c$, we have $\frac{cd^2}{2} \ln n = \mu_0 \leq t_0 = d \ln n$. Therefore, we can bound the probability (3) via Chernoff (1) as follows:

$$\begin{aligned} \Pr\left[\mathrm{Bin}\left(n \cdot \frac{t_0(t_0+1)}{2}, p\right) \geq t_0\right] &\leq \Pr\left[\mathrm{Bin}\left(\frac{nt_0(t_0+1)}{2}, p\right) \geq \mu_0 + (t_0 - \mu_0)\right] \\ &\leq \exp\left(-\mu_0 \cdot \varphi\left(\frac{t_0}{\mu_0} - 1\right)\right) = n^{-\frac{\mu_0}{\ln n} \cdot \varphi\left(\frac{t_0}{\mu_0} - 1\right)}. \end{aligned}$$

Thus, we just need to verify that

$$\frac{\mu_0}{\ln n} \cdot \varphi\left(\frac{t_0}{\mu_0} - 1\right) > 2. \quad (5)$$

Using the approximation (4), we obtain

$$\begin{aligned} \frac{\mu_0}{\ln n} \cdot \varphi\left(\frac{t_0}{\mu_0} - 1\right) &= \frac{\mu_0}{\ln n} \cdot \left(\frac{t_0}{\mu_0} \ln \frac{t_0}{\mu_0} - \frac{t_0}{\mu_0} + 1\right) \\ &\sim \frac{cd^2}{2}\left(\frac{2}{cd} \ln \frac{2}{cd} - \frac{2}{cd} + 1\right) = d\left(\frac{cd}{2} + \ln \frac{2}{cd} - 1\right). \end{aligned}$$

Thus, our assumption on c, d implies (5). □

Proof of Theorem 1, part (1). Let $c = 0.16$ and $p = c/(n \ln n)$. Letting $f(d) = d(cd/2 + \ln(2/cd) - 1)$, we see that $\max_{0<d<2/c} f(d) > 2$. Hence, Lemma 4 entails that for $c < 0.16$, we have $\Pr[T > t_0] = o(n^{-2})$ for a certain $t_0 = O(\ln n)$. By the union bound, this implies that w.h.p. there is no pair (v_1, v_2) such that $T^{(v_1,v_2)} = n-1$, whence $H : \mathcal{H}(n, p)$ is not propagation connected w.h.p. by Fact 1. □

4 The upper bound

In this section we sketch the proof of part (2) of our main theorem, that is, an upper bound for p such that $H : \mathcal{H}(n, p)$ is propagation connected w.h.p. The detail proofs of three propositions stated here will be given in the following subsections. As we saw in Section 2, the propagation process can be viewed as a time-dependent random walk. At first, the drift of this random walk is negative, but after a certain time the drift turns positive. The following proposition reflects this fact by showing that once the process has survived up to a certain time, it will likely continue to time $n - 1$. In the following, we use $\nu = \lceil (\ln n)^3 \rceil$.

Proposition 5. Suppose that $c > 0$ is a constant and let $p = c/(n \ln n)$. Then w.h.p. there is no pair (u, v) such that $\nu \le T^{(u,v)} < n - 1$.

In the light of Proposition 5, we call a pair of vertices (u, v) *good* if $T^{(u,v)} \ge \nu$. Let N be the number of good pairs of $H : \mathcal{H}(n, p)$. Then by Proposition 5 in order to prove that $H : \mathcal{H}(n, p)$ is propagation connected w.h.p., we just need to establish that $N > 0$ w.h.p. We first estimate the *expected* number of good pairs.

Proposition 6. For any fixed $c > 0.25$ there is a number $\delta = \delta(c) > 0$ such that for $p = c/(n \ln n)$ we have $\mathrm{E}[N] \ge \Omega(n^\delta)$.

Then by the following proposition, we relate the above result on the expectation of N to showing that $N > 0$ w.h.p. The proof of this proposition is based on a second moment argument.

Proposition 7. Assume that $\delta, c > 0$ are constants such that for $p = c/(n \ln n)$ we have $\mathrm{E}[N] \ge \Omega(n^\delta)$. Then in fact $N \ge \Omega(n^\delta) > 0$ w.h.p.

Now the second part of Theorem 1 is a direct consequence of Propositions 5–7.

4.1 Proof of Proposition 5

Fix any constant $c > 0$, and let $p = c/(n \ln n)$. Also fix any sufficiently large n. Recall that $\nu = \lceil (\ln n)^3 \rceil$.

Consider a random hypergraph $H : \mathcal{H}(n, p)$. For any $L \subseteq V$, we say that L is *closed* if there is no edge in H having two vertices in L and one vertex in $V \setminus L$. Below we estimate the probability that H has *some* $Z \subseteq V$ satisfying

$$Z \text{ is closed } \wedge \ \nu < |Z| < n. \tag{6}$$

Note that, starting from some vertex pair (u,v), if $\nu \le T^{(u,v)}[H] < n-1$, then the set $\mathcal{D}_t^{(u,v)}[H]$ of dead vertices at time $t = T^{(u,v)}[H]$ has exactly $t+1$ vertices and satisfies (6). Thus, the proposition is proved by showing this probability is $o(1)$.

Consider any z, $\nu + 1 \le z \le n-1$. Then we have

$$\begin{aligned}
&\Pr[\,\exists Z\,[\,Z \text{ is closed} \wedge |Z| = z\,]\,] \\
&\le \binom{n}{z}(1-p)^{\binom{z}{2}(n-z)} \le n \cdot \frac{n^n}{z^z \cdot (n-z)^{n-z}} \cdot \exp\left(-p \cdot \frac{z(z-1)(n-z)}{2}\right) \\
&\le \exp(\ln n + z\ln n - z\ln z) \cdot \left(1 + \frac{z}{n-z}\right)^{n-z} \cdot \exp\left(-p \cdot \frac{z(z-1)(n-z)}{2}\right) \\
&\le \exp(\ln n + z\ln n - z\ln z) \cdot \exp(z) \cdot \exp\left(-pz^2 n/8\right) \\
&\le \exp\left(z \cdot \left(\frac{\ln n}{z} + \ln n - \ln z + 1 - \frac{pzn}{8}\right)\right) \\
&= \exp\left(z \cdot \left(\frac{1}{(\ln n)^2} + \ln n - \ln z + 1 - \frac{c(\ln n)^2}{8}\right)\right) \le \exp(-z) = n^{-(\ln n)^2}.
\end{aligned}$$

Now by the union bound, the target probability is bounded by $n \cdot n^{-(\ln n)^2} = o(1)$.

4.2 Proof of Proposition 6

As indicated in Section 2, we basically need to analyze the probability that the random walk described by the variables $X_t = A_t - A_{t-1} + 1$ remains positive. From now on, we fix a number $c > 0.25$ and let $p = c/(n\ln n)$ for n sufficiently large. We will keep the notation from Section 2.

For a time t and a number $g \ge 1$, we let $\mathrm{AT}(t,g)$ denote the event that $A_t \ge g$. That is, the process does not stop before time t, and at this time there are at least g active vertices. As we saw in Section 2, the 'drift' of the time-dependent random walk described by the variables X_t is negative for small $t \ll \ln n$. The following lemma will help us get over the first few steps of the process. Intuitively, it shows that with a decent probability the process will not only survive up to time $\gamma \ln n$, but also amass a small excess of $\gamma \ln n$ active vertices for a small $\gamma > 0$.

Lemma 8. For any $\delta > 0$, there is $\gamma_0 = \gamma_0(c,\delta) > 0$ such that for all $0 < \gamma < \gamma_0$, the event $\mathrm{AT}(\lceil\gamma\ln n\rceil, \lceil\gamma\ln n\rceil)$ holds with probability at least $n^{-\delta}$.

Proof. As $\lim_{\gamma\to 0} 2\gamma\ln(c) - c\gamma^2/2 + 2\gamma\ln(\gamma/2) = 0$, for any $\delta > 0$, there is $\gamma_0 > 0$ such that for all $0 < \gamma < \gamma_0$, we have $2\gamma\ln(c) - c\gamma^2/2 + 2\gamma\ln(\gamma/2) > -\delta$. Assume that γ, $0 < \gamma < \gamma_0$, is sufficiently small so that this is the case. Let $t_1 = \lceil\gamma\ln n\rceil$ and $t_0 = \lfloor t_1/2\rfloor$. Then

$$\Pr\left[\mathrm{AT}(\lceil\gamma\ln n\rceil, \lceil\gamma\ln n\rceil)\right] \ge \Pr\left[\bigwedge_{1\le t\le t_0} X_t = 1 \wedge \bigwedge_{t_0 < t \le t_1} X_t = 3\right].$$

(For if $X_t > 0$ for all $t \in [t_1]$, then the process won't stop before time t_1, i.e., $T \ge t_1$. Moreover, the number of active vertices at time t_1 equals $\sum_{t=1}^{t_1}(X_t - 1) = 2(t_1 - t_0) \ge \gamma\ln n$.)

For $0 \le t \le t_1$, we let $\mathcal{E}_t$ signify the event that $X_s = 1$ for all $1 \le s \le \min\{t, t_0\}$ and $X_s = 3$ for all $t_0 < s \le t$. Then our objective is to lower bound $\Pr[\mathcal{E}_{t_1}]$.

If we condition on the event $\mathcal{E}_{t-1}$ for some $t \in [t_1]$, then the number of neutral vertices at time t works out to be $n-(t+1)-A_t \ \ge \ n-2t_1-2 \ = \ n-O(\ln n)$. Furthermore, Fact 3 entails that X_t given $\mathcal{E}_{t-1}$ is binomially distributed $\mathrm{Bin}(n-t-A_{t-1}, 1-(1-p)^t)$. Consequently,

$$\begin{aligned}\Pr[X_t = 1|\mathcal{E}_{t-1}] &\ge (n-O(\ln n))(1-(1-p)^t)(1-p)^{tn} \sim \frac{ct}{\ln n}\cdot\exp(-ct/\ln n), \quad \text{and}\\ \Pr[X_t = 3|\mathcal{E}_{t-1}] &\ge \binom{n-O(\ln n)}{3}(1-(1-p)^t)^3(1-p)^{tn} \sim \frac{(ct)^3}{(\ln n)^3}\cdot\exp(-ct/\ln n).\end{aligned}$$

Therefore, a small computation shows that

$$\begin{aligned}\Pr[\mathcal{E}_{t_1}] &= \prod_{1\le t\le t_0}\Pr[X_t = 1|\mathcal{E}_{t-1}]\cdot\prod_{t_0<t\le t_1}\Pr[X_t = 3|\mathcal{E}_{t-1}]\\ &\ge c^{3t_1-2t_0}n^{-c\gamma^2/2}\left(\frac{\gamma}{2}\right)^{3\gamma\ln(n)/2}\cdot\frac{\exp\left(\sum_{t=1}^{t_0}\ln t\right)}{(\ln n)^{t_0}} \ \ge \ \Omega\left(n^{2\gamma\ln(c)-c\gamma^2/2+2\gamma\ln(\gamma/2)}\right).\end{aligned}$$

Since we have chosen γ so that $2\gamma\ln(c) - c\gamma^2/2 + 2\gamma\ln(\gamma/2) > -\delta$, the assertion follows. □

Lemma 8 shows that with a decent probability the first few steps of the process will yield a good number of active vertices. The following lemma studies the continuation of the process up to the time $c^{-1}\ln n$ where the 'drift' of the random walk turns positive.

Lemma 9. There exists $\delta > 0$ such that $\Pr[T \ge \lceil(c^{-1}+\delta)\ln n\rceil] \ \ge \ n^{\delta-2}$.

Proof. Since $c > 0.25$, we can choose $\delta > 0$ so that $4c(1-\delta) > 1$. Let γ_0 be the number promised by Lemma 8. Moreover, choose $0 < \gamma < \gamma_0$ sufficiently small so that $1 + 4c\gamma - \ln(1-c\gamma) < 4c(1-\delta)$. We may also assume that $\lceil(c^{-1}+\delta)\ln n\rceil \le \lceil\gamma\ln n\rceil\cdot\left(\lfloor(c\gamma)^{-1}\rfloor+1\right)$.

Let $g = \lceil\gamma\ln n\rceil$ and $s_0 = \lfloor(c\gamma)^{-1}\rfloor$. Then our goal is to estimate the probability that the propagation process lasts at least $(s_0+1)g$ steps. To this end, we partition this period into s_0+1 chunks of size g. That is, for each $s \in [s_0]$, we define $Y_s \ = \ \sum_{sg<t\le(s+1)g} X_t$. We are going to lower bound the probability of the event

$$\mathrm{AT}(g,g) \wedge (Y_1 \ge g) \wedge \cdots \wedge (Y_{s_0} \ge g). \tag{7}$$

If this event occurs, then $T \ge g(s_0+1)$. To see this, we show by induction that for each $1 \le s \le s_0$ at time $t = sg$ there are at least g active vertices. For $s = 1$ this follows directly from the definition for $\mathrm{AT}(g,g)$. Proceeding inductively, we note that the following period up to time $(s+1)g$ will generate g new active vertices, because $Y_{s+1} \ge g$. This ensures that at time $(s+1)g$ there are at least g active vertices as well.

Thus, in order to establish the proposition, we just need to prove that the event (7) holds with probability $n^{\delta-2}$. Lemma 8 shows that $\Pr[\mathrm{AT}(g,g)] \ge n^{-\delta}$. In addition, we are going to estimate probability that $Y_s \ge g$ given $\mathrm{AT}(g,g) \wedge (Y_1 \ge g) \wedge \cdots \wedge (Y_{s-1} \ge g)$ for any $s \in [s_0]$. In doing so we may assume that $A_{sg} \le 2c^{-1}\ln n$, because otherwise the process will continue to time $2c^{-1}\ln n > (c^{-1}+\delta)\ln n$ with certainty. Hence, we may assume that there are always

more than $n' = (n - 2c^{-1}\ln n) = n(1-o(1))$ neutral vertices. On the other hand, at times $sg < t \le (s+1)g$ there are at least sg dead vertices. Thus, Fact 3 implies that

$$\begin{aligned}
&\Pr\big[\, A_{sg} \ge 2c^{-1}\ln n \vee Y_s \ge g \,|\, \mathrm{AT}(g,g) \wedge (Y_1 \ge g) \wedge \cdots \wedge (Y_{s-1} \ge g)\,\big] \\
&\quad\ge\ \Pr\left[\sum_{sg<t\le(s+1)g} X_t \ge g \,\middle|\, \big(A_{sg} < 2c^{-1}\ln n\big) \wedge \mathrm{AT}(g,g) \wedge (Y_1 \ge g) \wedge \cdots \wedge (Y_{s-1} \ge g)\right] \\
&\quad\ge\ \Pr\big[\,\mathrm{Bin}(gn', 1-(1-p)^{sg}) \ge g\,\big] \ \ge\ \binom{g^2n's}{g} p^g (1-p)^{g^2n's-g}.
\end{aligned}$$

Let $\mu_s = g^2n'sp$ and $x_s = g/\mu_s$. Applying Stirling's formula, we obtain

$$\begin{aligned}
\binom{g^2n's}{g} p^g(1-p)^{g^2n's-g} &\ge c' \cdot \sqrt{\frac{m}{g(m-g)}} \cdot \frac{m^m}{g^g\cdot(m-g)^{m-g}} \cdot p^g(1-p)^{m-g} \\
&\ge \frac{c'}{\sqrt{g}} \cdot \left(\frac{pm}{g}\right)^g \cdot \left(\frac{m-pm}{m-g}\right)^{m-g} \\
&= \exp\left(\ln c' - \frac{\ln g}{2} - g\ln x_s + (m-g)\ln\left(1-\frac{\mu_s-g}{m-g}\right)\right) \\
&\ge \exp\left(\ln c' - \frac{\ln g}{2} - g\ln x_s - (\mu_s - g) - \frac{(\mu_s-g)^2}{m-g}\right) \\
&\ge \exp\left(-g\ln x_s + g - \mu_s - O(\ln\ln n)\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\Pr[\,(7)\,] &= \Pr[\,\mathrm{AT}(g,g)\,]\cdot\Pr[\,(Y_1 \ge g)\wedge\cdots\wedge(Y_{s_0}\ge g)\,|\,\mathrm{AT}(g,g)\,] \\
&\ge n^{-\delta}\cdot\prod_{1\le s\le s_0}\exp\big(-g\ln x_s + g - \mu_s - c''\ln\ln n\big) \\
&= n^{-\delta}\cdot\exp\left(\sum_{1\le s\le s_0}\big(-g\ln x_s + g - \mu_s - c''\ln\ln n\big)\right) \qquad (8)
\end{aligned}$$

Approximating the sum in the exponent by an integral, we see that

$$\begin{aligned}
\sum_{1\le s\le s_0}\big(-g\ln x_s + g - \mu_s - c''\ln\ln n\big) &\ge -\frac{\ln n}{2c}\cdot\big(1+o(1)+3c\gamma-\ln(1-c\gamma)\big) \\
&> -2\ln n + 2\delta\ln n,
\end{aligned}$$

where the last step is due to our choice of γ and δ. Finally, combining this estimate with (8) yields $\Pr[\,(7)\,] \ge n^{-\delta}\cdot n^{-1/2c_{\mathrm{pos}}+2\delta} = n^{\delta-2}$, as desired. □

The basic idea in the above proof was to study the behavior of the random walk by partitioning the time up to about $c^{-1}\ln n$ in short periods of length $g = \lceil\gamma\ln n\rceil$ with a small $\gamma > 0$. What we estimated was the probability that for each of these periods the *total* number of newly generated active vertices is at least g, without taking into account how these g vertices are distributed over the period. Alternatively, one could lower bound the probability

that the process survives up to time $c^{-1}\ln n$ by the probability that the process generates at least one active vertex *at each individual step*. However, this argument gives a significantly weaker result. Intuitively, this means that typically the process will generate a little bit of 'leeway' for itself by aggregating a certain excess of active vertices.

Once the process 'survives' up to time $c^{-1}\ln n$, we are on firm ground, because then the 'drift' of the underlying random walk becomes positive. This observation yields the following corollary to Lemma 9, which in turn implies Proposition 6.

Corollary 10. There is $\delta > 0$ such that $\Pr[(1,2) \text{ is good}] = \Omega(n^{\delta-2})$.

Proof. Let δ be as in Proposition 6 and set $\theta = \lceil (c^{-1}+\delta)\ln n \rceil$. We condition on the event that the propagation process for $(1,2)$ continues for at least θ steps, i.e., $A_\theta \neq 0$. Then at time θ there is a set of A_θ of active vertices, and a set of $\theta+1$ of dead vertices. Let

$$\tau \;=\; \min\left\{t > \theta : A_t = 0 \text{ or } (t+1) + A_t \geq (\ln n)^3\right\}.$$

In order to prove the proposition, we need to show that

$$\Pr[\, A_\tau \neq 0 \,|\, A_\theta \neq 0] \;=\; \Omega(1). \tag{9}$$

This implies the assertion, because Lemma 9 shows that $\Pr[A_\theta \neq 0] \geq n^{\delta-2}$.

In order to prove (9), we are going to approximate the propagation process for times $\theta < t < \tau$ by a Galton-Watson branching process with successor rate greater than one. This is possible because for $\theta < t < \tau$, the number of neutral vertices at time t is at least $n-(t+1)-A_t \geq n-(\ln n)^3$. Therefore, the number X_t of new active vertices at time t has a binomial distribution $\mathrm{Bin}(n-(t+1)-A_t, 1-(1-p)^t)$ (see Fact 3). Its expectation bounded away from one. That is, for all t, $\theta < t < \tau$, we have

$$\mathrm{E}[\, X_t | \mathcal{F}_{t-1}\,] \;\sim\; t(n-t-A_t)p \;\geq\; (1-o(1))\theta np \;=\; 1+\delta-o(1).$$

To set up the analogy with the branching process, let $\{\widetilde{X}_s\}_{s\geq 1}$ be a family of mutually independent random variables with distribution $\mathrm{Bin}(\theta(n-(\ln n)^3), p)$ with mean $\mathrm{E}(\widetilde{X}_s) \geq 1+\delta-o(1) > 1$. Let $A'_0 = A_\theta > 0$ and let $A'_s = A'_{s-1} + \widetilde{X}_s - 1$ for all $s \geq 1$ be the branching process corresponding to the sequence $(\widetilde{X}_s)_{s\geq 1}$. Furthermore, let τ' be the least $s \geq 1$ such that $A'_s = 0$ if there is such an s, and set $\tau' = \infty$ otherwise. Because $n - A_t - t \geq n - (\ln n)^3$, the random variable X_t dominates $\widetilde{X}_{t-\theta}$ for all $\theta < t < \tau$. Therefore,

$$\Pr[A_\tau = 0 | A_\theta > 0] \;\leq\; \Pr\left[\tau' \leq \tau - \theta\right] \;\leq\; \Pr\left[\tau' < \infty\right]. \tag{10}$$

Finally, as the random variables of $\{\widetilde{X}_s\}_{s\geq 1}$ are i.i.d. with expectation greater than one, the theory of branching processes (e.g., [Fe50, p. 297]) shows that $\Pr[\tau' < \infty] \leq 1-\alpha$ for some number $\alpha = \alpha(\delta) > 0$ that depends on δ only. Thus, (10) implies (9). □

4.3 Proof of Proposition 7

Recall that a pair (x, y) of (distinct) vertices is good if $T^{(u,v)} \geq \nu$ $(= \lceil (\ln n)^3 \rceil)$, in other words, the process from (x, y) continues at least to time ν.

To bound the probability that there exists *some* good initial pair, we study the propagation process first from $(1, 2)$ and, given its outcome, the process from either $\pi = (3, 4)$ or from $\pi = (1, 3)$. Thus, let $\mathcal{A}_t = \mathcal{A}_t^{(1,2)}$, $\mathcal{D}_t = \mathcal{D}_t^{(1,2)}$, $T = T^{(1,2)}$, $A_t = |\mathcal{A}_t|$, X_t be the quantities that characterize the process from $(1, 2)$ as in Section 3. Then $(1, 2)$ is good iff $T \geq \nu$. Let $(\mathcal{F}_t)_{t \geq 0}$ be the filtration corresponding to this process. (Recall that $\mathcal{A}_t$, $\mathcal{F}_t$, etc. are defined even for $t > T$.)

In addition, we consider random sets/variables $\mathcal{A}'_t = \mathcal{A}_t^{\pi}$, $\mathcal{D}'_t = \mathcal{D}_t^{\pi}$, $T' = T^{\pi}$, $A'_t = |\mathcal{A}'_t|$, $\widetilde{X}_t = X_t^{\pi}$ associated to the process commencing from π (= either $(3, 4)$ or $(1, 3)$). Let

$$C_t \;=\; \begin{cases} 1, & \text{if } |(\mathcal{A}'_t \cup \mathcal{D}'_t) \cap (\mathcal{A}_\nu \cup \mathcal{D}_\nu)| \geq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathcal{F}'_0 = \mathcal{F}_\nu$. Moreover, for $t \geq 1$, let $\mathcal{F}'_t$ be the coarsest σ-algebra such that $\mathcal{F}'_t \supset \mathcal{F}_\nu$ and such that all events $\{v \in \mathcal{A}'_s\}$ for $s \leq t$ and $v \in V$ are $\mathcal{F}'_t$-measurable. Intuitively, $\mathcal{F}'_t$ captures the propagation process from $(1, 2)$ up to time $\min\{\nu, T\}$ and the process from $(3, 4)$ (or $(1, 3)$) up to time t. In analogy to Fact 2, we have the following.

Fact 4. Given $\mathcal{F}'_t$, for all triples $e = \{u, v, w\}$ such that $\max\{|e \cap \mathcal{D}_\nu|, |e \cap \mathcal{D}'_t|\} < 2$, the edge e is present in $H : \mathcal{H}(n, p)$ with probability p independently.

Fact 5. Given $\mathcal{F}'_{t-1}$, random variable $(1 - C_{t-1})\widetilde{X}_t$ is stochastically dominated by

$$\mathrm{Bin}(n - t - A'_{t-1}, 1 - (1 - p)^t).$$

Proof. If $C_{t-1} = 1$ or $t > T'$, then the statement is trivially true. Thus, we may condition on $C_{t-1} = 0$ and $t \leq T'$. Let $a = \min \mathcal{A}'_{t-1}$ be the active vertex chosen at time t, and let $d \in \mathcal{D}'_{t-1}$ be any dead vertex. Since $C_{t-1} = 0$, $\mathcal{A}_\nu \cup \mathcal{D}_\nu$ contains at most one of a, d. Let $b \notin \mathcal{A}'_{t-1} \cup \mathcal{D}'_{t-1}$ be another vertex and set $e = \{a, b, d\}$. We are going to show that given $\mathcal{F}'_{t-1}$, the edge e is present with probability at most p independently. We consider several cases; note that $|e \cap \mathcal{D}_{t-1}| < 2$.

- **Case 1:** $(a, d \notin \mathcal{A}_\nu \cup \mathcal{D}_\nu)$ In this case $|e \cap \mathcal{D}_\nu| \leq 1$, and thus Fact 4 shows that e is present with probability p.
- **Case 2:** $(b \notin \mathcal{A}_\nu \cup \mathcal{D}_\nu)$ As $C_{t-1} = 0$ and $a, d \in \mathcal{A}'_{t-1} \cup \mathcal{D}'_{t-1}$, at most one of a, d is in $\mathcal{A}_\nu \cup \mathcal{D}_\nu$. Thus, $|e \cap \mathcal{D}_\nu| < 2$, and therefore e is present with probability p by Fact 4.
- **Case 3:** $(a, b \in \mathcal{A}_\nu \cup \mathcal{D}_\nu)$ We have $d \notin \mathcal{A}_\nu \cup \mathcal{D}_\nu$, because otherwise $\mathcal{A}'_{t-1} \cup \mathcal{D}'_{t-1}$ would contain two vertices from $\mathcal{A}_\nu \cup \mathcal{D}_\nu$ and thus $C_{t-1} = 1$. If in the $(1, 2)$-process both a, b are dead at time ν, then e is not present, because otherwise d would have been included in $\mathcal{A}_\nu \cup \mathcal{D}_\nu$ as well. If in the $(1, 2)$-process at least one of a, b is in $\mathcal{A}_\nu$, then $|e \cap \mathcal{D}_\nu| < 2$ and thus the probability that e is present equals p by Fact 4.
- **Case 4:** $(a, d \in \mathcal{A}_\nu \cup \mathcal{D}_\nu)$ Identical to Case 3.

We have shown that for each of the $n - t - A'_{t-1}$ vertices $b \notin \mathcal{A}'_{t-1} \cup \mathcal{D}'_{t-1}$ and each of the t vertices $d \in \mathcal{D}'_{t-1}$ the edge $e = \{a, b, d\}$ is present with probability at most p. Hence, for any b the probability that at least one such edge is present is bounded by $1 - (1-p)^t$. □

Lemma 11. Let H_0 be any hypergraph such that $A_\nu[H_0] \le \nu^2$.

- If $\pi = (3,4)$, then $\Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) \le \nu^8 n^{-2}$.
- If $\pi = (1,3)$, then $\Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) \le \nu^4 n^{-1}$.

Proof. If $C_\nu = 1$, then at least two vertices in $\mathcal{A}_\nu \cup \mathcal{D}_\nu$ belong to $\mathcal{A}'_\nu \cup \mathcal{D}'_\nu$.

Consider the first assertion for the case $\pi = (3,4)$. For a pair $1 \le s \le s' \le \nu$, we let $\mathcal{E}(s,s')$ be the event that $(\mathcal{A}'_s \setminus \mathcal{A}'_{s-1}) \cap (\mathcal{A}_\nu \cup \mathcal{D}_\nu) \neq \emptyset$, $C_{s'-1} = 0$, and $C_{s'} = 1$. In other words, $\mathcal{E}(s,s')$ is the event that $s \le s'$ are the first times when vertices from $\mathcal{A}_\nu \cup \mathcal{D}_\nu$ become active in the $(3,4)$-process.

Let $a = \min \mathcal{A}'_s$ and $a' = \min \mathcal{A}'_{s'}$. Note that the event $\mathcal{E}(s,s')$ implies the existence of edges $e = \{a, b, d\}$ and $e' = \{a', b', c'\}$ for some $d \in \mathcal{D}'_s$, $d' \in \mathcal{D}'_{s'}$ and $b, b' \in \mathcal{A}_\nu \cup \mathcal{D}_\nu$ in a random $H : \mathcal{H}(n,p)$ consistent with H_0. For these edges, by construction, we have $\left|e \cap \mathcal{D}'_{s-1}\right| < 2$ and $\left|e' \cap \mathcal{D}'_{s'-1}\right| < 2$; also $|e' \cap \mathcal{D}_\nu| < 2$ holds. Moreover, if $|e \cap \mathcal{D}_\nu| \ge 2$, then $b' \in \mathcal{D}_\nu$ and $d' = b \in \mathcal{D}_\nu$, and in this case e' is not present in H because otherwise we would have $C_{s'-1} = 1$. Thus, by Fact 4, the probability that the random H contains e (resp., e') is bounded as

$$\Pr\left[e \in H | \mathcal{F}'_{s-1}\right](H_0) \;=\; p, \quad \text{and} \quad \Pr\left[e' \in H | \mathcal{F}'_{s'-1}\right](H_0) \;\le\; p. \tag{11}$$

Note that the total number of possible ways to choose each of d, d' is bounded by $\nu+1$ (because $s \le s' \le \nu$ and $|\mathcal{D}'_\nu| = \nu + 1$) and that there are $|\mathcal{A}_\nu \cup \mathcal{D}_\nu| \le \nu^2 + \nu + 1$ ways to choose each of b, b' (because $A_\nu \le \nu^2$ by assumption). Thus, from (11) and $p = O(1/(n \ln n))$, we obtain

$$\Pr\left[\mathcal{E}(s,s') | \mathcal{F}'_0\right](H_0) \;\le\; (\nu^2 + \nu + 1)^2 (\nu+1)^2 p^2 = o(\nu^6/n^2)$$

Hence, by the union bound, it holds that

$$\Pr\left[C_\nu = 1 | \mathcal{F}'_0\right](H_0) \;\le\; \Pr\left[\exists s < s' \in [\nu]\ [\mathcal{E}(s,s') \text{ occurs}] \mid \mathcal{F}'_0\right](H_0) \;\le\; \nu^2 \cdot o(\nu^6/n^2) \le \nu^8/n^2.$$

This proves the first assertion.

Next consider the case $\pi = (1,3)$. For $1 \le s' \le \nu$, let $\mathcal{E}(s')$ be the event that $C_{s'-1} = 0$ and $C_{s'} = 1$. Let $a' = \min \mathcal{A}_{s'}$; then $a' \notin \mathcal{A}_\nu \cup \mathcal{D}_\nu$, because $C_{s'-1} = 0$. If $\mathcal{E}(s')$ occurs, then $e' = \{a', b', d'\}$ is present in the random graph for some $d' \in \mathcal{D}'_{s'-1}$ and $b' \in \mathcal{A}_\nu \cup \mathcal{D}_\nu$. Then we have $\left|e' \cap \mathcal{D}'_{s'-1}\right| < 2$. Moreover, if $|e' \cap \mathcal{D}_\nu| = 2$, then $d' = 1$ and $b' \in \mathcal{D}_\nu$, because $C_{s'-1} = 0$; but then e' is not present, because otherwise the process from $(1,2)$ would have included a' into $\mathcal{A}_\nu \cup \mathcal{D}_\nu$. Furthermore, if $|e' \cap \mathcal{D}_\nu| < 2$, then by Fact 4, we have $\Pr\left[e' \in H | \mathcal{F}'_{s-1}\right](H_0) = p$. Thus, in any case, it holds that

$$\Pr\left[e' \in H | \mathcal{F}'_{s-1}\right](H_0) \;\le\; p. \tag{12}$$

Now because the number of ways to choose each of d' is bounded by $\nu + 1$, and as there are $|\mathcal{A}_\nu \cup \mathcal{D}_\nu|$ choices for b, b', from $p = O(1/(n \ln n))$ and (12), we have

$$\Pr\left[\mathcal{E}(s') | \mathcal{F}'_0\right](H_0) \;\le\; \Pr\left[\mathcal{E}(s') | \mathcal{F}'_0\right](H_0) \;\le\; (\nu^2 + \nu + 1)(\nu+1)p \;=\; o(\nu^3/n).$$

Finally, taking the union bound over $s' \leq \nu$, we get

$$\Pr\left[C_\nu = 1 | \mathcal{F}'_0\right](H_0) \;\leq\; \Pr\left[\exists s' \in [\nu]\,[\mathcal{E}(s') \text{ occurs}]\,|\mathcal{F}'_0\right](H_0) \;\leq\; \nu \cdot o(\nu^3/n) \;\leq\; \nu^4/n,$$

as desired. □

Lemma 12. We have $\Pr\left[A_\nu > \nu^2\right] = o(n^{-4})$.

Proof. We have $\mathcal{A}_\nu \leq \sum_{t=1}^{\nu} X_t$. Furthermore, by Fact 3 the variable X_t given $\mathcal{F}_{t-1}$ is dominated by

$$\widetilde{X}_t \stackrel{\text{def}}{=} \text{Bin}(n, 1-(1-p)^t),$$

where $\widetilde{X}_1, \widetilde{X}_2, \ldots$ are mutually independent.

As a consequence,

$$\Pr\left[\mathcal{A}_\nu > \nu^2\right] \;\leq\; \Pr\left[\sum_{t=1}^{\nu} \widetilde{X}_t > \nu^2\right]. \tag{13}$$

Furthermore,

$$\begin{aligned} \mathrm{E}\left[\sum_{t=1}^{\nu} \widetilde{X}_t\right] &= \sum_{t=1}^{\nu} n(1-(1-p)^t) \;\leq\; \nu n(1-(1-p)^\nu) \\ &\leq (1+o(1))\nu^2 np \;=\; o(\nu^2). \end{aligned} \tag{14}$$

Combining (13) and (14) with the Chernoff bound (1), the bound of the lemma is shown. □

Lemma 13. Let $\pi \in \{(1,3),(3,4)\}$. Let H_0 be any hypergraph such that $T[H_0] \geq \nu$ and $A_\nu[H_0] \leq \nu^2$. Then we have

$$\Pr\left[T' \geq \nu \wedge C_\nu = 0 \,|\mathcal{F}'_0\right](H_0) \;\leq\; \Pr\left[T^{(3,4)} \geq \nu\right].$$

Proof. Since T' is the least time t such that $A'_t = 0$, we have $T' \geq \nu$ iff $A'_t > 0$ for all $t \in [\nu]$. Recall that $A'_t = A'_{t-1} + X'_t - 1$; hence, $T' \geq \nu$ implies that $X'_t \geq 1$ for all $t \in [\nu]$. On the other hand, $C_\nu = 0$ implies $C_{t-1} = 0$ for all $t \in [\nu]$. Thus, $T' \geq \nu \wedge C_\nu = 0$ implies $(1-C_{t-1})X'_t \geq 1$ for all $t \in [\nu]$. In other words, letting

$$A''_t \;=\; A''_{t-1} + (1-C_{t-1})X'_t - 1,$$

we see that

$$\Pr\left[T' \geq \nu \wedge C_\nu = 0 \,|\,\mathcal{F}'_0\right](H_0) \;\leq\; \Pr\left[\forall t \in [\nu]\;\left[A''_t > 0\right]\,|\,\mathcal{F}'_0\right](H_0).$$

Thus, in order to prove the lemma, it suffices to show that

$$\Pr\left[\forall t \in [\nu]\;\left[A''_t > 0\right]\,|\,\mathcal{F}'_0\right](H_0) \;\leq\; \Pr\left[\forall t \in [\nu]\;[A_t > 0]\right] \;=\; \Pr\left[T \geq \nu\right] \tag{15}$$

since $\Pr[T \geq \nu] = \Pr\left[T^{(3,4)} \geq \nu\right]$ by symmetry.

For simplifying expressions, we let μ denote the probability measure $\Pr[\,\cdot\,|\,\mathcal{F}'_0](H_0)$. Now we are going to prove that

$$\mu\left[\,A''_t \geq a\,|\,\forall s \in [t-1]\ \left[A''_s > 0\right]\,\right] \ \leq\ \Pr\left[\,A_t \geq a\,|\,\forall s \in [t-1]\ [A_s > 0]\,\right] \tag{16}$$

holds for any t, $0 \leq t \leq \nu$, and for any integer $a \geq 1$.

Fix any $a \geq 1$. We proceed by induction on t. As $A''_0 = A_0 = 1$, (16) is trivial for $t = 0$. Consider any $t \geq 1$. Note that given $A''_{t-1} = b > 1$ we have $A''_t \geq a$ iff $(1 - C_{t-1})X'_t \geq a - b + 1$. Thus, it holds that

$$\begin{aligned}
&\mu\left[\,A''_t \geq a\,|\,\forall s \in [t-1]\ \left[A''_s > 0\right]\,\right] \\
&\quad= \sum_{1 \leq b \leq n} \mu\left[\,A''_t \geq a\,|\,A''_{t-1} = b\,\right] \cdot \mu\left[\,A''_{t-1} = b\,|\,\forall s \in [t-2]\ \left[A''_s > 0\right]\,\right] \\
&\quad= \sum_{1 \leq b \leq n} \mu\left[(1 - C_{t-1})X'_t = a - b + 1\,|\,A''_{t-1} = b\,\right] \cdot \mu\left[\,A''_{t-1} = b\,|\,\forall s \in [t-2]\ \left[A''_s > 0\right]\,\right].
\end{aligned} \tag{17}$$

Here (and in the following) we omit unrelated conditions and write, e.g., $\mu\left[A''_t \geq a | A''_{t-1} = b\right]$ instead of $\mu\left[A''_t \geq a | A''_{t-1} = b \wedge \forall s \in [t-2]\ [A''_s > 0]\right]$.

By Fact 5 the variable $(1 - C_{t-1})X'_t$ given $\mathcal{A}''_t = b$ is stochastically dominated by a binomial distribution $\mathrm{Bin}(n - t - b, 1 - (1-p)^t)$. By comparison, Fact 3 shows that given $A_{t-1} = b$ the variable X_t has a binomial distribution $\mathrm{Bin}(n - t - b, 1 - (1-p)^t)$. Hence, X_t given $A_{t-1} = b$ dominates $(1 - C_{t-1})X'_t$ given $A''_{t-1} = b$. Therefore, (17) yields

$$\begin{aligned}
&\mu\left[\,A''_t \geq a\,|\,\forall s \in [t-1]\ \left[A''_s > 0\right]\,\right] \\
&\quad\leq \sum_{1 \leq b \leq n} \Pr\left[\,X_t \geq a - b + 1\,|\,A_{t-1} = b\,\right] \cdot \mu\left[\,A''_{t-1} = b\,|\,\forall s \in [t-2]\ \left[A''_s > 0\right]\,\right] \\
&\quad= \sum_{1 \leq b \leq n} \Pr\left[\,A_t \geq a\,|\,A_{t-1} = b\,\right] \cdot \mu\left[\,A''_{t-1} = b\,|\,\forall s \in [t-2]\, A''_s > 0\,\right].
\end{aligned} \tag{18}$$

By induction, A_{t-1} given $A_s > 0$ for all $s \in [t-2]$ dominates A''_{t-1} given $A''_s > 0$ for all $s \in [t-2]$. Furthermore, the function $b \mapsto \Pr[A_t \geq a | A_{t-1} = b]$ is monotonically increasing. Combining these two facts with (18), we obtain

$$\begin{aligned}
&\mu\left[\,A''_t \geq a\,|\,\forall s \in [t-1]\ \left[A''_s > 0\right]\,\right] \\
&\quad\leq \sum_{1 \leq b \leq n} \Pr\left[\,A_t \geq a\,|\,A_{t-1} = b\,\right] \cdot \Pr\left[\,A_{t-1} = b\,|\,\forall s \in [t-2]\ [A_s > 0]\,\right] \\
&\quad= \Pr\left[\,A_t \geq a\,|\,\forall s \in [t-1]\ [A_s > 0]\,\right].
\end{aligned}$$

This proves (16) for all $t \in [\nu]$. Then using the fact that $A''_0 = A_0 = 1$, we can show

$$\Pr\left[\,\forall s \in [t]\ \left[A''_s > 0\right]\ |\,\mathcal{F}'_0\,\right](H_0) \ \leq\ \Pr\left[\,\forall s \in [t]\ [A_s > 0]\,\right],$$

for any $t \in [\nu]$. In particular, our goal (15) is obtained as its special case. □

Now we assume that $\delta > 0$ is the one promised by Proposition 6 for any constant $c > 0.25$ and $p = c/(n \ln n)$.

Lemma 14. We have

$$\Pr\left[\,\text{both } (1,2) \text{ and } (3,4) \text{ are good}\,\right] \;\leq\; (1+o(1))\left(\Pr\left[(1,2) \text{ is good}\right]\right)^2.$$

Proof. Note first that

$$\begin{aligned}&\Pr\left[\,(1,2),\,(3,4) \text{ are good}\,\right]\\ &\quad=\;\Pr\left[\,(3,4) \text{ is good} \,\middle|\, (1,2) \text{ is good} \wedge A_\nu \leq \nu^2\,\right]\cdot \Pr\left[\,(1,2) \text{ is good}\,\right] + \Pr\left[\,A_\nu > \nu^2\,\right].\end{aligned}$$

Since $\Pr\left[\mathcal{A}_\nu > \nu^2\right] = o(n^{-4})$ (from Lemma 12) and $\Pr\left[(3,4) \text{ is good}\right] = \Omega(n^{\delta-2})$ (from Corollary 10 and by symmetry), we see that $\Pr\left[\mathcal{A}_\nu > \nu^2\right] = o\left(\left(\Pr\left[(1,2) \text{ is good}\right]\right)^2\right)$. Thus, we just need to prove

$$\Pr\left[\,(3,4) \text{ is good} \,\middle|\, (1,2) \text{ is good} \wedge A_\nu \leq \nu^2\,\right] \;\leq\; (1+o(1))\Pr\left[\,(3,4) \text{ is good}\,\right]. \tag{19}$$

Let $\pi = (3,4)$. In order to establish (19), it suffices to prove that for any hypergraph H_0 such that $T\left[H_0\right] \geq \nu$ (i.e., $(1,2)$ is good in H_0) and $A_\nu \leq \nu^2$, we have

$$\Pr\left[\,T' \geq \nu \,\middle|\, \mathcal{F}_0'\,\right](H_0) \;\leq\; (1+o(1))\cdot \Pr\left[\,T^{(3,4)} \geq \nu\,\right]. \tag{20}$$

Fix any such H_0. Then

$$\Pr\left[\,T' \geq \nu \,\middle|\, \mathcal{F}_0'\,\right](H_0) \;\leq\; \Pr\left[\,C_\nu = 1 | \mathcal{F}_0'\,\right](H_0) + \Pr\left[\,T' \geq \nu \wedge C_\nu = 0 \,\middle|\, \mathcal{F}_0'\,\right](H_0). \tag{21}$$

Since we are assuming that $\Pr\left[T^{(3,4)} \geq \nu\right] = \Pr\left[(3,4) \text{ is good}\right] = \Omega\left(n^{\delta-2}\right)$, Lemma 11 yields that

$$\Pr\left[\,C_\nu = 1 | \mathcal{F}_0'\,\right](H_0) \;\leq\; \nu^8 n^{-2} \;=\; o\left(\Pr\left[T^{(3,4)} \geq \nu\right]\right). \tag{22}$$

In addition, Lemma 13 shows that

$$\Pr\left[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}_0'\right](H_0) \;\leq\; \Pr\left[T^{(3,4)} \geq \nu\right] \tag{23}$$

Finally, combining (21), (22), and (23), we obtain (20), thereby completing the proof. □

Lemma 15. We have

$$\Pr\left[\,\text{both } (1,2) \text{ and } (1,3) \text{ are good}\,\right] \;\leq\; o(n)\left(\Pr\left[(1,2) \text{ is good}\right]\right)^2.$$

Proof. We use a similar argument as in the proof of Lemma 14. As in the that proof, we have

$$\begin{aligned}&\Pr\left[(1,2),\,(1,3) \text{ are good}\right]\\ &\quad=\;\Pr\left[\,(1,3) \text{ is good} \,\middle|\, (1,2) \text{ is good} \wedge A_\nu \leq \nu^2\right]\cdot \Pr\left[\,(1,2) \text{ is good}\,\right] + \Pr\left[A_\nu > \nu^2\right].\end{aligned} \tag{24}$$

Now from Lemma 12, it suffices to show that

$$\Pr\left[\,(1,3) \text{ is good} \,\middle|\, (1,2) \text{ is good} \wedge A_\nu \leq \nu^2\,\right] \;\leq\; (1+o(1))n\cdot \Pr\left[\,(1,3) \text{ is good}\,\right]. \tag{25}$$

Let $\pi = (1,3)$. To prove (25), we are going to show that for any hypergraph H_0 such that $T[H_0] \geq \nu$ (i.e., $(1,2)$ is good) and $A_\nu \leq \nu^2$, we have

$$\Pr\left[T' \geq \nu | \mathcal{F}_0'\right](H_0) \;\leq\; (1+o(1)) \cdot \Pr\left[T^{(3,4)} \geq \nu\right]. \tag{26}$$

For any such H_0, we have

$$\Pr\left[T' \geq \nu \,|\mathcal{F}_0'\right](H_0) \;\leq\; \Pr\left[C_\nu = 1 \,|\, \mathcal{F}_0'\right](H_0) + \Pr\left[T' \geq \nu \wedge C_\nu = 0 \,|\, \mathcal{F}_0'\right](H_0). \tag{27}$$

Since we are assuming that $\Pr\left[T^{(1,3)} \geq \nu\right] = \Pr\left[(1,3) \text{ is good}\right] = \Omega\left(n^{\delta-2}\right)$, Lemma 11 yields

$$\Pr\left[C_\nu = 1 | \mathcal{F}_0'\right](H_0) \;\leq\; \nu^4 n^{-1} = o\left(n \cdot \Pr\left[T^{(1,3)} \geq \nu\right]\right). \tag{28}$$

In addition, Lemma 13 shows that

$$\Pr\left[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}_0'\right](H_0) \;\leq\; \Pr\left[T^{(3,4)} \geq \nu\right]. \tag{29}$$

Finally, combining (27), (28) and (29), we obtain (26), thereby completing the proof. □

Proof of Proposition 7. Let N be the number of good pairs. We are going to show that $\mathrm{E}\left[N^2\right] \sim (\mathrm{E}[N])^2$. More specifically, we analyze $\mathrm{E}[N(N-1)]$. Let W be the set of all pairs $(x,y) \in V^2$ such that $x \neq y$. We use (x,y) and (x',y') to denote distinct elements of W. Then

$$\mathrm{E}[N(N-1)] \;=\; \sum_{(x,y),(x',y') \in W} \Pr[\,\text{both } (x,y),\, (x',y') \text{ are good}\,]. \tag{30}$$

We split the sum into two cases where

- the summands (x,y), (x',y') are pairwise distinct, that is, $|\{x,y,x',y'\}| = 4$, and
- the summands (x,y), (x',y') satisfy $|\{x,y,x',y'\}| = 3$.

There are $n(n-1)(n-2)(n-3)$ possibilities for (x,y,x',y') in the first case, and $2n(n-1)(n-2)$ possibilities in the second case. Since the hypergraph distribution $H : \mathcal{H}(n,p)$ is symmetric with respect to permutations of the vertices, in the first case we have

$$\Pr[\,\text{both } (x,y),\, (x',y') \text{ are good}\,] \;=\; \Pr[\,\text{both } (1,2),\, (3,4) \text{ are good}\,],$$

and in the second case we get

$$\Pr[\,\text{both } (x,y),\, (x',y') \text{ are good}\,] \;=\; \Pr[\,\text{both } (1,2),\, (1,3) \text{ are good}\,].$$

Hence, we can rephrase (30) as

$$\begin{aligned}\mathrm{E}[N(N-1)] \;&=\; n(n-1)(n-2)(n-3) \cdot \Pr[\,\text{both } (1,2),\, (3,4) \text{ are good}\,] \\ &\quad +2n(n-1)(n-2) \Pr[\,\text{both } (1,2),\, (1,3) \text{ are good}\,].\end{aligned}$$

Then invoking Lemmas 14 and 15, we thus obtain

$$\begin{aligned}\mathrm{E}[N(N-1)] \;&\leq\; (1+o(1))n(n-1)(n-2)(n-3)\left(\Pr[\,(1,2) \text{ is good}\,]\right)^2 \\ &\qquad +o\left(n^2(n-1)(n-2)\left(\Pr[\,(1,2) \text{ is good}\,]\right)^2\right) \\ &\leq\; (1+o(1))\left(n(n-1)\Pr[\,(1,2) \text{ is good}\,]\right)^2 \;=\; (1+o(1))\left(\mathrm{E}[N]\right)^2.\end{aligned}$$

As a consequence, we get $\mathrm{Var}[N] = \mathrm{E}[N^2] - (\mathrm{E}[N])^2 = o\big((\mathrm{E}[N])^2\big)$. Therefore, for any $\gamma > 0$, Chebyshev's inequality shows that

$$\Pr[N < (1-\gamma)\mathrm{E}[N]] \;\le\; \frac{\mathrm{Var}[N]}{(\gamma \mathrm{E}[N])^2} \;=\; \frac{o(1)}{\gamma^2} = o(1).$$

From this, the proposition follows. □

5 Computing a Propagation Sequence

5.1 Algorithm and an Outline of its Analysis

We show the algorithm A claimed in Theorem 2. For any constant $\epsilon > 0$ and for any $p > (0.25+\epsilon)/(n \ln n)$, our algorithm A finds a propagation sequence of a given hypergraph $H : \mathcal{H}(n,p)$ w.h.p.; furthermore, its expected running time is linear in the number of edges of H.

We first describe our algorithm and give an outline of its analysis. The detail analysis will be given in the next subsection. Throughout this section, we fix $p = c/(n \ln n)$ for any constant $c > 0.25$; this lose no generality for proving the theorem because for any random hypergraph $H' : \mathcal{H}(n,p')$ with $p' > p$, we may consider a graph H following $H : \mathcal{H}(n,p)$ by removing each edge of H with probability $1 - p/p'$, and on this H, the algorithm finds a propagation sequence w.h.p., which can be used as a propagation sequence of H'.

Figure 5.1 states the outline of the algorithm A. The algorithm's execution is divided into three steps. At the first step, a random hash table `HashTable` is constructed. This table is used, for any given pair of vertices x and y, to check whether it is *positive*, namely, there exists an edge in E containing x and y. (If positive, we also need to enumerate all such edges; but this information can be added at the corresponding entry of each positive pair by a linked list.) The second step is based on the propagation process we introduced in section 2. In the second step, the algorithm searches for a *successful* initial pair such that the propagation process *succeeds*, that is, it continues until to time $n-1$. Let `D`, `A`, and `N` be variables for the current set of dead, active, and neutral vertices respectively. Note that we only need to start from a pair of vertices that appear in some edge of E, which we call an *initial edge*. When a successful initial edge is found, then the algorithm proceeds to the third step to compute an propagation sequence starting from this edge. For this, we simply need to collect all edges recorded as path edge candidates during the successful propagation process.

Clearly this algorithm finds a propagation sequence if the given graph is propagation connected. As guaranteed by Theorem 1 (2), a random $H = (V,E) : \mathcal{H}(n,p)$ is propagation connected w.h.p., and hence, the algorithm on H succeeds to find its propagation sequence w.h.p. Thus, our task is to show that the expected running time of the algorithm is $O(|E|)$.

We explain some important points for showing the linear expected running time. In the following consider any random hypergraph $H : \mathcal{H}(n,p)$. Let $R(H)$ denote the running time of the algorithm on H. Note that $R(H)$ is a random variable depending on both $H : \mathcal{H}(n,p)$ and the algorithm's randomness used for constructing `HashTable`. Note that the number of edges

$|E|$ itself is a random variable. What we will show is precisely that

$$\exists c_{\mathrm{alg}},\, \forall m \;\left[\; \mathrm{E}\big[\, R(H) \,\big|\, |E| = m \,\big] \;\le\; c_{\mathrm{alg}} \max\left(m,\, \frac{cn^2}{\ln n}\right) \;\right] \tag{31}$$

holds. Note that $|E|$ is very well concentrated in $\Theta(cn^2/\ln n)$. Thus, this technical goal (31) is sufficient for the theorem. In the following, we will simply use m to denote the number of edges of a given H.

Let us go through the algorithm to check what is necessary to show (31). Note first that the time bound for the third step of the algorithm can be subsumed by the one for the second step. Thus, we consider only the first and the second steps of the algorithm.

The first step of the algorithm is for preparing an appropriate hash table for checking whether a given pair of vertices is positive. Note that there are $3m$ positive pairs. Thus, by using a standard pair-wise independent random hash function family (see, e.g., [MU05, Theorem 13.11]), we can construct a 'perfect' random hash table with $O(m)$ entries in $O(m)$ time *on average*. Here by 'perfect' we mean a hash table with which each query can be answered in constant time.

Now consider the second step, the main step of the algorithm. For any $e \in E$, we consider the running time R_e for simulating the propagation process starting from (a pair of vertices of) e. Let A_e and C_e denote the number of vertices that get active and that of edges that are examined at $(*)$ of the algorithm during the process starting from e. Note that # of all pairs $(\mathtt{x}, \mathtt{y})$ that are examined during the process is bounded by $(A_e)^2$. Hence, it is easy to see that R_e is bounded by $O(\max(A_e^2, C_e))$. This yields a trivial bound $R_e = O(\max(m, n^2))$ since we have $A_e \le n$ and $C_e \le m$. We would like to show that it is much smaller *on average*; more specifically, $\mathrm{E}[R_e] = O(m)$ if the process succeeds from e and $\mathrm{E}[R_e] = O(1)$ otherwise.

Recall that if an initial edge (and the initial pair that it defines) is 'good' and the propagation process from the edge lasts more than ν steps, then most likely it succeeds to reach to time $n-1$ and a propagation sequence is obtained. On the other hand, the probability of hitting a good initial pair is small, and in fact, the process stops much earlier for most of the 'bad' initial pairs. We use this to show that $\mathrm{E}[R_e] = O(1)$ if the process does not succeed.

For this analysis, we consider the following four cases determined by initial edge e: (i) the process (from e) terminates in two steps, (ii) the process (from e) terminates in $\lfloor\sqrt{\ln n/c(n)}\rfloor$ steps, (iii) the process (from e) succeeds, and (iv) the other case. Let $E_1, \ldots, E_4$ denote the set of initial edges for which each of these four cases occurs respectively.

Clearly, $R_e = O(1)$ for $e \in E_1$. On the other hand, we can show that the probability that $e \notin E_1$ is bounded by $O(c/\ln n)$. This shows that $\mathrm{E}[R_e]$ for $e \in E_2$ is $O(1)$. We then consider initial edges $e \in E_4$. It is possible that the process lasts more than $(\ln n)^3$ steps; but from Lemma 5, such probability is quite small, and this case can be ignored in our average-case running time analysis. Thus, we may assume that the process terminates in $(\ln n)^3$ steps, which can be simulated in $O((\ln n)^6)$ time. On the other hand, it is also possible to show that the probability that $e \notin E_2$ is bounded by $o(1/(\ln n)^d)$ for any $d > 0$. This shows that $\mathrm{E}[R_e]$ for $e \in E_4$ is also $O(1)$.

Finally, consider the case $e \in E_3$. That is, the case where the propagation process succeeds from e. Note that the algorithm immediately proceeds to step (3) as soon as any successful

initial edge is found; thus, we need to consider the time for simulating *one* successful process. From our trivial bound, we have $R_e = O(\max(m, n^2))$, which is a bit large to bound by m (since $m = O(cn^2/\ln n)$ with high probability). In order to reduce this time, we split our simulation of the propagation process into two stages. The first one is until time $(n/\sqrt{\ln n})$. Since $|\mathtt{D}| \leq n/\sqrt{\ln n}$ during this stage, the simulation of this stage can be done in time $O(\max(m, n^2/\ln n))$. In the second stage, i.e., after the $(n/\sqrt{\ln n})$th step of the process, we switch our strategy of obtaining edges examined at $(*)$; for any active vertex $\mathtt{x}$, instead of searching for edges containing a pair of vertices $\mathtt{x}, \mathtt{y}$ for each dead vertex $\mathtt{y}$, we use another table and examine all edges of E containing $\mathtt{x}$. Note that the number of edge candidates (for each active vertex) is *on average* $cn/\ln n$; in fact, the case that there exists some active vertex that appears in more than $2cn/\ln n$ edges occurs with very small probability, and it can be ignored. This proves that the second stage can be simulated in time $O(cn^2/\ln n)$. Hence, altogether the process from e can be simulated in time $O(\max(m, cn^2/\ln n))$ on average.

5.2 Detail running time analysis

We give detail running time analysis on the second step of the algorithm. We fix n to any sufficiently large number, and consider the execution of our algorithm on a random input $H : \mathcal{H}(n, p)$. We will use notations E, A_e, C_e, and $E_1, \ldots, E_4$ as defined in the previous subsection.

We introduce some notation. Let $o_{\mathrm{e}}(1)$ denote any positive function that is bounded by $1/e(n)$ with some subexponentially growing function $e(n)$. Note that the algorithm's running time is bounded (even in the worst-case) by some polynomial in n. Thus, for discussing the average case running time, we may ignore any event that occurs with probability $o_{\mathrm{e}}(1)$. We summarize such events as follows. (We omit their proofs because Fact 8 is from Lemma 5, and the other facts are easy to show by using the Chernoff bound.)

Fact 6. There exist constants c_1 and c_2 such that

$$\Pr\left[c_1 \frac{cn^2}{\ln n} \leq |E| \leq c_2 \frac{cn^2}{\ln n} \right] \;=\; 1 - o_{\mathrm{e}}(1).$$

Fact 7. For any vertex $v \in V$, let N_v denote the number of edges containing v. Then we have

$$\Pr\left[\exists v \in V \left[N_v \geq \frac{cn}{\ln n} \right] \right] \;=\; o_{\mathrm{e}}(1).$$

Fact 8.

$$\Pr\left[\exists e \in E \left[C_e > (\ln n)^3 \text{ and the propagation process from } e \text{ fails} \right] \right] \;=\; o_{\mathrm{e}}(1).$$

Now we prove our technical goal, namely (31). Consider any m such that $c_1 cn^2/\ln n \leq m \leq c_2 cn^2/\ln n$ holds for the constants c_1, c_2 of Fact 6. We estimate the expectation of $T(F)$ under the condition that $|E| = m$.

As explained in the previous subsection, for the set of $3m$ positive pairs, a perfect hash table of size $O(m)$ can be constructed in $O(m)$ time on average. Thus, we may assume that the

expected time for the step (1) of the algorithm to create a perfect hash table for `HashTable` is $O(m)$. Thus, for proving (31), it suffices to show that the expected running time of the step (2) of the algorithm is $O(m)$ assuming that `HashTable` is a perfect hash table and it can be used to check whether a given pair of vertices is positive or not in constant time.

Recall that T_e is the time for simulating the propagation process starting from edge e. Thus, our goal is to show that $\mathrm{E}\left[\sum_{e\in E} T_e\right]$ is $O(\max(m, cn^2/\ln n))$. As explained in the previous subsection, we estimate $\mathrm{E}\left[\sum_{e\in E} T_e\right]$ considering the four cases in the following way[1]:

$$\begin{aligned}\mathrm{E}\left[\sum_{e\in E} T_e\right] &= \sum_{e\in E}\sum_{i=1,2,4} \mathrm{E}\left[T_e | e\in E_i\right]\cdot \Pr\left[e\in E_i\right] \\ &\quad + \mathrm{E}\left[T_{e_0}\,\middle|\, \text{the process succeeds from } e_0\right]\end{aligned}$$

where by symmetry we may choose e_0 as any edge in E. Below we give lemmas showing the desired bounds for these four cases.

First consider the case that the exploration succeeds from e_0. In this case, by using Fact 6, it is easy to show the following bound.

Lemma 16. Suppose that the propagation process succeeds from edge e_0. Then we have $\mathrm{E}\left[T_{e_0}\right] = O(cn^2/\ln n)$, where T_{e_0} is the time for simulating this process following the implementation explained in the previous subsection.

Finally, we show the following bounds for the other three failed cases.

Lemma 17. Let E_i be either E_1, E_2, or E_4. Then for any $e \in E$, we have

$$\mathrm{E}\left[T_e | e\in E_i\right]\cdot \Pr\left[e\in E_i\right] \;=\; O(1). \tag{32}$$

Proof. Let e denote any edge in E. First we estimate probabilities $\Pr\left[e\in E_i\right]$ for $i\in\{1,2,4\}$ as follows.

Claim 1.(1) $\Pr\left[e\in E_2\right] \;\le\; \Pr\left[e\notin E_1\right] \;<\; 3\ln n/c$.
(2) $\Pr\left[e\in E_4\right] \;\le\; \Pr\left[e\notin E_2\right] \;=\; o(1/(\ln n)^d)$ for any $d\ge 1$.

We prove the lemma by using these bounds. Since the process from e fails, we may assume that $C_e < n$ from Fact 8. Hence, we have $A_e \ge C_e/2$, which implies that $T_e = O((A_e)^2)$. Then the case $e\in E_1$ is clear because we have a trivial bound $\mathrm{E}\left[(A_e)^2 | e\in E_1\right] = O(1)$. Consider the case $e\in E_2$. Again by definition we have $(A_e)^2 = \ln n/c$. Hence, from (1) of the above claim, we have

$$\mathrm{E}[(A_e)^2 | e\in E_2]\cdot \Pr[e\in E_2] \;\le\; \frac{\ln n}{c}\cdot\frac{3c}{\ln n} \;=\; 3.$$

Finally, consider the case $e\in E_4$, that is, the case $e\notin E_1\cup E_2\cup E_3$. Since e is a failed initial edge, we may assume from Fact 8 that $A_e \le (\ln n)^3$. Thus, by using (2) of the above claim, we have

$$\mathrm{E}[(A_e)^2 | e\in E_4]\cdot \Pr[e\in E_4] \;\le\; (\ln n)^6\cdot o\left(\frac{1}{(\ln n)^6}\right) \;=\; o(1).$$

[1]For simplifying expressions/statements, we treat E as a nonrandom variable in the following analysis. Modifying this analysis to the one with precise expressions/statements is easy, and it is left to the reader.

This prove the lemma. □

Proof of Proof of Claim 1. (1) Let $e = \{u, v, w\}$, and let u and v be the dead and the active vertices at the 1st step of the process. Then clearly at least one active vertex (i.e., w) is found, and the process goes to the second step. In this situation, there are three possibilities to continue to the third step: (i) yet another vertex gets active at the 1st step, (ii) some vertex gets active by $\{x, z\}$ at the 2nd step, and (iii) some vertex gets active by $\{y, z\}$ at the 2nd step. The probability that each case occurs is bounded by $\Pr[\mathrm{Bin}(n, p) \geq 1]$, which is bounded by

$$\Pr[\mathrm{Bin}(n,p) \geq 1] \;=\; 1-(1-p)^n \;\leq\; np \;=\; \frac{c}{\ln n}.$$

This implies the bound of the lemma.

(2) The argument is similar to the one for proving the part (1) of the main theorem. Let $t_0 = \lfloor\sqrt{\ln n/c}\rfloor + 1$. We note that if $|A_e| \geq t_0$, then at least t_0 edges are found by checking at most $nt_0(t_0+1)/2$ triples. This probability is at most $\Pr[\mathrm{Bin}(nt_0(t_0+1)/2, p) \geq t_0] \leq \Pr\left[\mathrm{Bin}(nt_0^2, p) \geq t_0\right]$. This last probability is bounded by

$$\Pr\left[\,\mathrm{Bin}\left(nt_0^2, p\right) \geq t_0\,\right] \;\leq\; \Pr\left[\,t_0 \leq \mathrm{Bin}\left(\frac{n\ln n}{c}, p\right) < c'\ln n\,\right] + \Pr\left[\,\mathrm{Bin}\left(\frac{n\ln n}{c}, p\right) \geq c'\ln n\,\right]$$

for any $c' \geq 1$. Here by taking c' sufficiently large, we can show that the second term of the above bound is $o(n^{-1})$ by the Chernoff bound. On the other hand, the first term is bounded by

$$\begin{aligned}\Pr\left[\,t_0 \leq \mathrm{Bin}\left(\frac{n\ln n}{c}, p\right) < c'\ln n\,\right] &\leq (c'\ln n)\cdot\Pr\left[\,\mathrm{Bin}\left(\frac{n\ln n}{c}, p\right) = t_0\,\right]\\ &\leq (c'\ln n)\cdot\binom{(n\ln n)/c}{t_0}p^{t_0}(1-p)^{t_0},\end{aligned}$$

which can be shown $o((\ln n)^{-d})$ for any $d \geq 1$. This proves the bound of the claim. □

References

[BCK07] M. Behrisch, A. Coja-Oghlan, M. Kang, Local limit theorems for the giant component of random hypergraphs, in *Proc. 11th International Workshop RANDOM* (APPROX+RANDOM'07), Lecture Notes in Computer Sciecne 4627, 341–352, 2007.

[BO09] R. Berke and M. Onsjö, Propagation connectivity of random hyptergraphs, in *Proc. 5th Symposium on Stochastic Algorithms, Foundations and Applications* (SAGA'09), Lecture Notes in Computer Science 5792, 117–126, 2009.

[CMV07] A. Coja-Oghlan, C. Moore, V. Sanwalani, Counting connected graphs and hypergraphs via the probabilistic method, *Random Structure and Algorithms* 31, 288–329, 2007.

[CM04] H. Connamacher and M. Molloy, The exact satisfiability threshold for a potentially intractable random constraint satisfaction problem, in *Proc. 45th Annual Symposium on Foundations of Computer Science* (FOCS'04), IEEE, 590–599, 2004.

[D05] R. Durrett, *Probability and examples, 3rd edition*, 2005.

[DN05] R.W.R. Darling and J.R. Norris, Structure of large random hypergraphs, *Ann. App. Probability* 15(1A), 125–152, 2005.

[Fe50] W. Feller, *An introduction to probability theory and its applications*, Wiley, 1950.

[JLR00] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley, 2000.

[MU05] M. Mitzenmacher and E. Upfal, *Probability and Computing, Randomized Algorithms and Probabilistic Analysis*, Cambridge Univ. Press, 2005.

[Kar90] R.M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* 1, 73–93, 1990.

[Mol05] M. Molloy, Cores in random hypergraphs and Boolean formulas, *Random Structures and Algorithms* 27(1), 124–135, 2005.

algorithm A (for computing a propagation sequence);
given $H = (V, E)$ following $\mathcal{H}(n, p)$, where $V = [n]$ and we denote $E = \{e_1, \ldots, e_m\}$;
(1) prepare a random hash table `HashTable` with $O(m)$ entries so that one can search edges containing a pair of vertices using the pair as a key;
(2) **for each** initial edge `e` $\in E$ that has not been examined **do** {
  let u, v, w be vertices of the edge `e`;
  `D` $\leftarrow \{u\}$; `A` $\leftarrow \{v\}$; `N` $\leftarrow V -$ `D` $\cup$ `A`;
  // start the propagation process from edge `e`
  while `A` $\neq \emptyset$ and `N` $\neq \emptyset$ **do** {
    `x` $\leftarrow$ any one element of `A`;
    for each `y` $\in$ `D` **do** {
      use `HashTable` to search edges containing `x` and `y`;
      for each one of the obtained edges $(*)$ **do** {
        `z` $\leftarrow$ the third vertex of the edge;
        if `z` $\in$ `N` **then** {
          save this edge as a candidate path edge;
          `A` $\leftarrow$ `A` $\cup \{$`z`$\}$; `N` $\leftarrow$ `N` $- \{$`z`$\}$;
        } }
    }
    `A` $\leftarrow$ `A` $- \{$`x`$\}$; `D` $\leftarrow$ `D` $\cup \{$`x`$\}$;
  }
  if `N` $= \emptyset$ (i.e., the process succeeds) **then goto** (3);
  reset the list of candidate path edges;
} // end of the while loop
report failure;
(3) compute the propagation sequence starting from the 'successful' initial edge `e` by following all candidate path edges, and output it;
end-procedure.

Figure 1: Outline of Algorithm A