

Propagation Connectivity of Random Hypergraphs*

Amin Coja-Oghlan[†]

University of Warwick
Mathematics and Computer Science
Zeeman building, Coventry CV4 7AL, UK
a.coja-oghlan@warwick.ac.uk

Mikael Onsjö[‡]

Tokyo Institute of Technology
Department of Mathematical and Computing Sciences
Meguro-ku Ookayama 2-12-1 W8-25
mikael@is.titech.ac.jp

Osamu Watanabe[§]

Tokyo Institute of Technology
Department of Mathematical and Computing Sciences
Meguro-ku Ookayama 2-12-1 W8-25
watanabe@is.titech.ac.jp

Submitted: Aug 18, 2011; Accepted: Jan 2, 2012; Published: Jan 16, 2012

Abstract

We study the concept of *propagation connectivity* on random 3-uniform hypergraphs. This concept is inspired by a simple propagation algorithm for solving instances of certain constraint satisfaction problems. We derive upper and lower bounds for the propagation connectivity threshold. Our proof is based on a kind of large deviations analysis of a time-dependent random walk. Based on the analysis, we also give an upper bound on the expected running time of the simple propagation algorithm.

*This is a journal version of an extended abstract presented at RANDOM 2010 (Lecture Notes in Computer Sci. 6302: 490–503, 2010).

[†]Supported in part by EPSRC grant EP/G039070/2 and DIMAP.

[‡]Supported in part by the “CompView” Global COE of JSPS.

[§]Supported in part by KAKENHI No. 22300003 of MEXT.

1 Introduction and results

1.1 The propagation connectivity

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 $e_1, \dots, e_{n-2} \in E$ a propagation sequence if for any $1 \leq l < n - 2$ we have $|e_{l+1} \cap \bigcup_{i=1}^l e_i| = 2$. 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, \dots, 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, \dots, e_l 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 $n \rightarrow \infty$.

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.

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.

Theorem 2 *There is a randomized algorithm A that satisfies the following. For any given hypergraph, A finds a propagation sequence if it exists. For any parameter p , the expected running time of A applied to $H : \mathcal{H}(n, p)$ is linear in the number of edges of H .*

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 3-uniform hypergraph $H = (V, E)$ with $V = [n]$ and a family $(f_e)_{e \in E}$ of maps $f_e : [k] \times [k] \times [k] \rightarrow \{0, 1\}$. Moreover, a *solution* is a map $\sigma : V \rightarrow [k]$ 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 super-linear) 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]; for such NP-hard propagating constraint satisfaction problems, as far as the authors know, no better way has been known for solving them other than the propagation algorithm and its variations.

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 $t > 0$

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

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

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

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 $V = \{1, \dots, n\}$. Let (v_1, v_2) be a pair of distinct vertices, which we refer to as the *initial pair*. In the course of the

process, 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

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

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 $t \geq 1$. 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_1, \dots, e_t such that $v_1, v_2 \in e_1$ and $|e_{l+1} \cap \bigcup_{i=1}^l e_i| = 2$ for all $1 \leq l < t$. 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_1, \dots, e_{n-1} such that $v_1, v_2 \in e_1$, 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. Also for discussing conditional probabilities on a random hypergraph $H : \mathcal{H}(n, p)$, we will use some convenient terminology of stochastic processes. Let A be any event on a hypergraph for which we would like to discuss the probability that A occurs in $H : \mathcal{H}(n, p)$ given that the first t steps of the propagation process on H satisfies some condition; this can be formally defined as the probability that A occurs in $H : \mathcal{H}(n, p)$ given that the first t steps of the propagation process on H work out the same as in H_0 , for a certain fixed hypergraph H_0 . This can be expressed both succinctly and formally by using the notion of filtration. 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 then 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

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

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 are 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 $\mathcal{F}_t^{(v_1, v_2)}$, let $\mathcal{E}_t^{(v_1, v_2)}$ denote the set of triples $e = \{u, v, w\}$ such that $|e \cap \mathcal{D}_t^{(v_1, v_2)}| < 2$. Then for all $e \in \mathcal{E}_t^{(v_1, v_2)}$, the event that e is present as an hyperedge in $H : \mathcal{H}(n, p)$ is mutually independent and it occurs with probability p . That is, for any set $\mathcal{E} \subset \mathcal{E}_t^{(v_1, v_2)}$, 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 the probability that the propagation sequence is obtained, i.e., $T^{(v_1, v_2)} \geq n - 1$ (cf. Fact 1), from each *fixed* initial pair is relatively small, and we need to investigate whether *there exists* a pair (v_1, v_2) such that $T^{(v_1, v_2)} \geq n - 1$ holds. Thus, 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 $\text{Bin}(n - t - A_{t-1}, 1 - (1 - p)^t)$.*

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 $1 - (1 - p)^t$. \square

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 $t = (1 + \epsilon)/(np) = \Theta(\ln n)$ 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 w.h.p. not propagation connected, because there is a total $\binom{n}{2}$ initial pairs to choose from. This observation suggests that in order to 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 w.h.p. *not* propagation connected 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 $\{\tilde{X}_t\}_{t \geq 1}$ be a family of mutually independent random variables such that \tilde{X}_t has distribution $\text{Bin}(nt, p)$. 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 $\tilde{X}_t = \text{Bin}(nt, p)$.

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 \tilde{X}_t , we can bound the probability of this event by

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

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:

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

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 w.h.p. not propagation connected by Fact 1. \square

4 The upper bound

In this section we outline 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 proofs of the 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 let $\nu = \lceil (\ln n)^3 \rceil$.

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 $\nu \leq T^{(u, v)} < n - 1$.*

In the light of Proposition 5, we call a pair of vertices (u, v) *good* if $T^{(u, v)} \geq \nu$. 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 $\mathbb{E}[N] \geq \Omega(n^\delta)$.*

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 $\mathbb{E}[N] \geq \Omega(n^\delta)$. Then in fact $N \geq \Omega(n^\delta) > 0$ 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

$$Z \text{ is closed} \wedge \nu < |Z| < n. \quad (6)$$

Note that, starting from some vertex pair (u, v) , if $\nu \leq T^{(u,v)}[H] < n - 1$, 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 $z, \nu + 1 \leq z \leq n - 1$. Then we have

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

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 $g \geq 1$, we let $\text{AT}(t, g)$ denote the event that $A_t \geq g$. 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 $t \ll \ln n$. 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 an 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 $\text{AT}(\lceil \gamma \ln n \rceil, \lceil \gamma \ln n \rceil)$ holds with probability at least $n^{-\delta}$.*

Proof. As $\lim_{\gamma \rightarrow 0} 2\gamma \ln(c) - c\gamma^2/2 + 2\gamma \ln(\gamma/2) = 0$, 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

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

(For if $X_t > 0$ for all $t \in [t_1]$, then the process won't stop before time t_1 , i.e., $T \geq t_1$. Moreover, the number of active vertices at time t_1 equals $\sum_{t=1}^{t_1} (X_t - 1) = 2(t_1 - t_0) \geq \gamma \ln n$.) For $0 \leq t \leq t_1$, we let \mathcal{E}_t signify the event that $X_s = 1$ for all $1 \leq s \leq \min\{t, t_0\}$ and $X_s = 3$ for all $t_0 < s \leq t$. 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 $n - (t + 1) - A_t \geq n - 2t_1 - 2 = n - O(\ln n)$. Furthermore, Fact 3 entails that X_t given \mathcal{E}_{t-1} is binomially distributed $\text{Bin}(n - t - A_{t-1}, 1 - (1 - p)^t)$. Consequently,

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

Therefore,

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

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

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 $\Pr[T \geq \lceil (c^{-1} + \delta) \ln n \rceil] \geq n^{\delta-2}$.*

Proof. Since $c > 0.25$, we can choose $\delta > 0$ so that $4c(1 - \delta) > 1$ and $c\delta > 0$. 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 $\lceil (c^{-1} + \delta) \ln n \rceil \leq \lceil \gamma \ln n \rceil \cdot (\lfloor (c\gamma)^{-1} \rfloor + 1)$.

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 $Y_s = \sum_{sg < t \leq (s+1)g} X_t$. We are going to lower bound the probability of the event

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

If this event occurs, then $T \geq g(s_0 + 1)$. To see this, we show by induction that for each $1 \leq s \leq s_0$ at time $t = sg$ there are at least g active vertices. For $s = 1$ this follows directly from the definition for $\text{AT}(g, g)$. Proceeding inductively, we note that the following period up to time $(s + 1)g$ will generate g new active vertices, because $Y_{s+1} \geq g$. 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 $\Pr[\text{AT}(g, g)] \geq n^{-\delta}$. In addition, we are going to estimate the probability that $Y_s \geq g$ given $\text{AT}(g, g) \wedge (Y_1 \geq g) \wedge \cdots \wedge (Y_{s-1} \geq g)$ for any $s \in [s_0]$. In doing so we may assume that $A_{sg} \leq 2c^{-1} \ln n$, 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 $sg < t \leq (s + 1)g$ there are at least sg dead vertices. Thus, Fact 3 implies that

$$\begin{aligned} & \Pr[A_{sg} \geq 2c^{-1} \ln n \vee Y_s \geq g \mid \text{AT}(g, g) \wedge (Y_1 \geq g) \wedge \cdots \wedge (Y_{s-1} \geq g)] \\ & \geq \Pr \left[\sum_{sg < t \leq (s+1)g} X_t \geq g \mid (A_{sg} < 2c^{-1} \ln n) \wedge \text{AT}(g, g) \wedge (Y_1 \geq g) \wedge \cdots \wedge (Y_{s-1} \geq g) \right] \\ & \geq \Pr[\text{Bin}(gn', 1 - (1-p)^{sg}) \geq g] \geq \binom{gn'}{g} (1 - (1-p)^{sg})^g ((1-p)^{sg})^{gn'-g} \\ & \geq \binom{gn'}{g} (psg)^g (1 - psg)^g (1-p)^{sg(gn'-g)}. \end{aligned}$$

Here for the last bound, we used the fact that $(1-p)^{sg} \leq 1 - psg + (psg)^2$ holds for sufficiently large n .

Let $\mu_s = n'psg^2$ and $x_s = g/\mu_s (= (n'psg)^{-1})$. Applying Stirling's formula, we obtain

$$\begin{aligned} & \binom{gn'}{g} (1-p)^{sg(gn'-g)} (psg)^g (1-psg)^g \\ & \geq c' \cdot \sqrt{\frac{gn'}{g(gn'-g)}} \cdot \frac{(gn')^{gn'}}{g^g \cdot (gn'-g)^{gn'-g}} \cdot (1-p)^{sg(gn'-g)} (psg)^g (1-psg)^g \\ & \geq \frac{c'}{\sqrt{g}} \cdot \left(\frac{(gn')(psg)}{g}\right)^g \cdot \left(\frac{gn'}{gn'-g}\right)^{gn'-g} \cdot (1-p)^{sg(gn'-g)} \cdot (1-psg)^g \\ & = \exp \left(\ln c' - \frac{\ln g}{2} - g \ln x_s \right. \\ & \quad \left. + (gn'-g) \ln \left(\frac{gn'}{gn'-g}\right) + (gn'-g)sg \ln(1-p) + g \ln(1-psg) \right). \quad (8) \end{aligned}$$

Here by using the bounds $\ln(1-x) \geq -x - x^2$ (for sufficiently small x) and $\ln(1+x) \geq -x - x^2/2$, we have

$$\begin{aligned}
(8) &\geq \exp\left(\ln c' - \frac{\ln g}{2} - g \ln x_s + \frac{(gn' - g)g}{gn' - g} \right. \\
&\quad \left. - \frac{(gn' - g)g^2}{2(gn' - g)^2} - (gn' - g)sgp - (gn' - g)sgp^2 - g(psg + (psg)^2)\right) \\
&\geq \exp\left(\ln c' - \frac{\ln g}{2} - g \ln x_s + g - o(1) - n'psg^2 - n'p^2sg^2 - psg^2 - (psg)^2g\right) \\
&\geq \exp(-g \ln x_s + g - \mu_s - O(\ln \ln n)).
\end{aligned}$$

Hence,

$$\begin{aligned}
\Pr[(7)] &= \Pr[\text{AT}(g, g)] \cdot \Pr[(Y_1 \geq g) \wedge \dots \wedge (Y_{s_0} \geq g) \mid \text{AT}(g, g)] \\
&\geq n^{-\delta} \cdot \prod_{1 \leq s \leq s_0} \exp(-g \ln x_s + g - \mu_s - c'' \ln \ln n) \\
&= n^{-\delta} \cdot \exp\left(\sum_{1 \leq s \leq s_0} (-g \ln x_s + g - \mu_s - c'' \ln \ln n)\right) \tag{9}
\end{aligned}$$

Now by estimating the sum in the exponent, we see that

$$\begin{aligned}
\sum_{1 \leq s \leq s_0} (-g \ln x_s + g - \mu_s - c'' \ln \ln n) &= \sum_{1 \leq s \leq s_0} (-g \ln g + g \ln \mu_s + g - \mu_s - c'' \ln \ln n) \\
&= s_0(-g \ln g + g - c'' \ln \ln n + g \ln(n'g^2p)) + g \sum_{1 \leq s \leq s_0} \ln s - (n'g^2p) \sum_{1 \leq s \leq s_0} s \\
&\geq s_0(-g \ln g + g - c'' \ln \ln n + g \ln(n'g^2p)) + gs_0 \ln\left(\frac{s_0}{e}\right) - (n'g^2p) \cdot \frac{s_0(s_0 + 1)}{2} \\
&\geq s_0g \ln\left(\frac{e \cdot n'g^2p \cdot s_0}{g \cdot e}\right) - s_0c'' \ln \ln n - \frac{n'(g(s_0 + 1))^2p}{2} \\
&= s_0g \ln(n'gs_0p) - s_0c'' \ln \ln n - \frac{n'(g(s_0 + 1))^2p}{2}. \tag{10}
\end{aligned}$$

Here from our choice of $s_0 = \lfloor (c\gamma)^{-1} \rfloor$ (and noting that $1/c\gamma \geq s_0 \geq (1/c\gamma)(1 - c\gamma)$), we can bound further by

$$\begin{aligned}
(10) &\geq \left(\frac{1}{c\gamma}\right)(1 - c\gamma) \cdot \gamma \ln n \cdot \left(\frac{1}{c\gamma}\right)(1 - c\gamma) \\
&\quad \times \ln\left(n(1 - o(1)) \cdot (\gamma \ln n) \cdot \left(\frac{1}{c\gamma}\right)(1 - c\gamma) \cdot \frac{c}{n \ln n}\right) \\
&\quad - \left(\frac{1}{c\gamma}\right)c'' \ln \ln n - \frac{n}{2} \cdot (\gamma \ln n)^2 \left(1 + \frac{1}{\gamma \ln n}\right)^2 \cdot \left(\frac{1}{c\gamma}\right)^2 (1 + c\gamma)^2 \cdot \frac{c}{n \ln n}
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{\ln n}{2c} \cdot 2(1 - c\gamma) \cdot \ln((1 - o(1))(1 - c\gamma)) \\
&\quad - \frac{\ln n}{2c} \cdot \frac{2c'' \ln \ln n}{\gamma \ln n} - \frac{\ln n}{2c} \cdot \left(1 + \frac{1}{\gamma \ln n}\right)^2 (1 + c\gamma)^2 \\
&= \frac{\ln n}{2c} \cdot 2(1 - c\gamma)(o(1) + \ln(1 - c\gamma)) \\
&\quad - \frac{\ln n}{2c} \cdot (1 - o(1))o(1) - \frac{\ln n}{2c} \cdot (1 - o(1))(1 + c\gamma)^2 \\
&\geq -\frac{\ln n}{2c} \cdot (1 + o(1) + 3c\gamma - \ln(1 - c\gamma)) > -2 \ln n + 2\delta \ln n,
\end{aligned}$$

where the last step is due to our choice of γ and δ . Finally, combining this estimate with (9) yields $\Pr[(7)] \geq n^{-\delta} \cdot n^{-2} = n^{\delta-2}$, as desired. \square

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 A_θ of active vertices, and a set of $\theta + 1$ of dead vertices. Let

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

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

$$\Pr[A_\tau \neq 0 \mid A_\theta \neq 0] = \Omega(1). \tag{11}$$

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

In order to prove (11), 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 $\text{Bin}(n - (t + 1) - A_t, 1 - (1 - p)^t)$ (see Fact 3). Its expectation is bounded away from one. That is, for all t , $\theta < t < \tau$, we have

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

To set up the analogy with the branching process, let $\{\tilde{X}_s\}_{s \geq 1}$ be a family of mutually independent random variables with distribution $\text{Bin}(\theta(n - (\ln n)^3), p)$ with mean $\mathbb{E}(\tilde{X}_s) \geq 1 + c\delta - o(1)$, which is larger than 1 for sufficiently large n because we chose δ so that $c\delta > 0$ holds. Let $A'_0 = A_\theta > 0$ and let $A'_s = A'_{s-1} + \tilde{X}_s - 1$ for all $s \geq 1$ be the branching process corresponding to the sequence $(\tilde{X}_s)_{s \geq 1}$. 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 $\tilde{X}_{t-\theta}$ for all $\theta < t < \tau$. Therefore,

$$\Pr[A_\tau = 0 | A_\theta > 0] \leq \Pr[\tau' \leq \tau - \theta] \leq \Pr[\tau' < \infty]. \quad (12)$$

Finally, as the random variables of $\{\tilde{X}_s\}_{s \geq 1}$ 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, (12) implies (11). \square

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 $\mathcal{A}'_t = \mathcal{A}_t^\pi$, $\mathcal{D}'_t = \mathcal{D}_t^\pi$, $T' = T^\pi$, $A'_t = |\mathcal{A}'_t|$, $X'_t = X_t^\pi$ 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} , the random variable $(1 - C_{t-1})X'_t$ is stochastically dominated by

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

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}_ν , 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 $1 - (1 - p)^t$. \square

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

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

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 $1 \leq s \leq s' \leq \nu$, 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 $s \leq s'$ 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 $e' = \{a', b', d'\}$ 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 $|e \cap \mathcal{D}'_{s-1}| < 2$ and $|e' \cap \mathcal{D}'_{s'-1}| < 2$; also $|e \cap \mathcal{D}_\nu| < 2$ holds. Moreover, if $|e' \cap \mathcal{D}_\nu| \geq 2$, then e' is not present in H ; because otherwise, $e' = \{a', b', d'\}$ must be chosen during the process from $(1, 2)$ and we would have $a', d' \in \mathcal{A}_\nu \cup \mathcal{D}_\nu$ implying $C_{s'-1} = 1$. Thus, by Fact 4, the probability that the random H contains e (resp., e') is bounded as

$$\Pr[e \in H | \mathcal{F}'_{s-1}](H_0) = p, \quad \text{and} \quad \Pr[e' \in H | \mathcal{F}'_{s'-1}](H_0) \leq p. \quad (13)$$

Note that the total number of possible ways to choose each of d, d' is bounded by $\nu + 1$ (because $s \leq s' \leq \nu$ and $|\mathcal{D}'_\nu| = \nu + 1$) and that there are $|\mathcal{A}_\nu \cup \mathcal{D}_\nu| \leq \nu^2 + \nu + 1$ ways to choose each of b, b' (because $A_\nu \leq \nu^2$ by assumption). Thus, from (13) and $p = O(1/(n \ln n))$, we obtain

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

Hence, by the union bound, it holds that

$$\begin{aligned} \Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) &\leq \Pr[\exists s < s' \in [\nu] [\mathcal{E}(s, s') \text{ occurs}] | \mathcal{F}'_0](H_0) \\ &\leq \nu^2 \cdot o(\nu^6/n^2) \leq \nu^8/n^2. \end{aligned}$$

This proves the first assertion.

Next consider the case $\pi = (1, 3)$. For $1 \leq s' \leq \nu$, 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 $|e' \cap \mathcal{D}'_{s'-1}| < 2$. 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 $\Pr[e' \in H | \mathcal{F}'_{s-1}](H_0) = p$. Thus, in any case, it holds that

$$\Pr[e' \in H | \mathcal{F}'_{s-1}](H_0) \leq p. \tag{14}$$

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 (14), we have

$$\begin{aligned} \Pr[\mathcal{E}(s') | \mathcal{F}'_0](H_0) &\leq \Pr[\mathcal{E}(s') | \mathcal{F}'_0](H_0) \\ &\leq (\nu^2 + \nu + 1)(\nu + 1)p = o(\nu^3/n). \end{aligned}$$

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

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

as desired. □

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

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

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

where $\tilde{X}_1, \tilde{X}_2, \dots$ are mutually independent.

As a consequence,

$$\Pr[\mathcal{A}_\nu > \nu^2] \leq \Pr\left[\sum_{t=1}^\nu \tilde{X}_t > \nu^2\right]. \tag{15}$$

Furthermore,

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^{\nu} \tilde{X}_t \right] &= \sum_{t=1}^{\nu} n(1 - (1-p)^t) \leq \nu n(1 - (1-p)^\nu) \\ &\leq (1 + o(1))\nu^2 np = o(\nu^2). \end{aligned} \tag{16}$$

Combining (15) and (16) with the Chernoff bound (1), the bound of the lemma is shown. \square

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*

$$\Pr[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}'_0](H_0) \leq \Pr[T \geq \nu] \quad (= \Pr[T^{(1,3)} \geq \nu] = \Pr[T^{(3,4)} \geq \nu]).$$

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

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

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

$$\Pr[\forall t \in [\nu] [A''_t > 0] | \mathcal{F}'_0](H_0) \leq \Pr[\forall t \in [\nu] [A_t > 0]] = \Pr[T \geq \nu]. \tag{17}$$

Here note that $\Pr[T \geq \nu] = \Pr[T^{(1,2)} \geq \nu] = \Pr[T^{(3,4)} \geq \nu]$ 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

$$\mu[A''_t \geq a | \forall s \in [t-1] [A''_s > 0]] \leq \Pr[A_t \geq a | \forall s \in [t-1] [A_s > 0]] \tag{18}$$

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 $A''_0 = A_0 = 1$, (18) 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

$$\begin{aligned} &\mu[A''_t \geq a | \forall s \in [t-1] [A''_s > 0]] \\ &= \sum_{1 \leq b \leq n} \mu[A''_t \geq a | A''_{t-1} = b] \cdot \mu[A''_{t-1} = b | \forall s \in [t-2] [A''_s > 0]] \\ &= \sum_{1 \leq b \leq n} \mu[(1 - C_{t-1})X'_t \geq a - b + 1 | A''_{t-1} = b] \\ &\quad \times \mu[A''_{t-1} = b | \forall s \in [t-2] [A''_s > 0]]. \end{aligned} \tag{19}$$

Here and in the following, we omit unrelated conditions and write, for example, $\mu [A_t'' \geq a | A_{t-1}'' = b]$ instead of $\mu [A_t'' \geq a | A_{t-1}'' = b \wedge \forall s \in [t-2] [A_s'' > 0]]$.

By Fact 5 the variable $(1 - C_{t-1})X_t'$ given $A_t'' = b$ is stochastically dominated by a binomial distribution $\text{Bin}(n - t - b, 1 - (1 - p)^t)$. By comparison, Fact 3 shows that given $A_{t-1} = b$ the variable X_t has a binomial distribution $\text{Bin}(n - t - b, 1 - (1 - p)^t)$. Hence, X_t given $A_{t-1} = b$ dominates $(1 - C_{t-1})X_t'$ given $A_{t-1}'' = b$. Therefore, (19) yields

$$\begin{aligned} & \mu [A_t'' \geq a | \forall s \in [t-1] [A_s'' > 0]] \\ & \leq \sum_{1 \leq b \leq n} \Pr [X_t \geq a - b + 1 | A_{t-1} = b] \cdot \mu [A_{t-1}'' = b | \forall s \in [t-2] [A_s'' > 0]] \\ & = \sum_{1 \leq b \leq n} \Pr [A_t \geq a | A_{t-1} = b] \cdot \mu [A_{t-1}'' = b | \forall s \in [t-2] [A_s'' > 0]]. \end{aligned} \quad (20)$$

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 (20), we obtain

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

This proves (18) for all $t \in [\nu]$. Then using the fact that $A_0'' = A_0 = 1$, we can show

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

for any $t \in [\nu]$. In particular, our goal (17) is obtained as its special case. \square

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*

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

Proof. Note first that

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

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

$$\Pr [(3, 4) \text{ is good} | (1, 2) \text{ is good} \wedge A_\nu \leq \nu^2] \leq (1 + o(1)) \Pr [(3, 4) \text{ is good}]. \quad (21)$$

Let $\pi = (3, 4)$. In order to establish (21), it suffices to prove that for any hypergraph H_0 such that $T[H_0] \geq \nu$ (i.e., $(1, 2)$ is good in H_0) and $A_\nu \leq \nu^2$, we have

$$\Pr[T' \geq \nu | \mathcal{F}'_0](H_0) \leq (1 + o(1)) \cdot \Pr[T^{(3,4)} \geq \nu]. \quad (22)$$

Fix any such H_0 . Then

$$\Pr[T' \geq \nu | \mathcal{F}'_0](H_0) \leq \Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) + \Pr[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}'_0](H_0). \quad (23)$$

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

$$\Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) \leq \nu^8 n^{-2} = o(\Pr[T^{(3,4)} \geq \nu]). \quad (24)$$

In addition, Lemma 13 shows that

$$\Pr[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}'_0](H_0) \leq \Pr[T^{(3,4)} \geq \nu] \quad (25)$$

Finally, combining (23), (24), and (25), we obtain (22), thereby completing the proof. \square

Lemma 15 *We have*

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

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

$$\begin{aligned} & \Pr[(1, 2), (1, 3) \text{ are good}] \\ &= \Pr[(1, 3) \text{ is good} | (1, 2) \text{ is good} \wedge A_\nu \leq \nu^2] \cdot \Pr[(1, 2) \text{ is good}] \\ & \quad + \Pr[A_\nu > \nu^2]. \end{aligned} \quad (26)$$

Now from Lemma 12, it suffices to show that

$$\Pr[(1, 3) \text{ is good} | (1, 2) \text{ is good} \wedge A_\nu \leq \nu^2] \leq o(n) \cdot \Pr[(1, 3) \text{ is good}]. \quad (27)$$

Let $\pi = (1, 3)$. To prove (27), 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

$$\Pr[T' \geq \nu | \mathcal{F}'_0](H_0) \leq o(n) \cdot \Pr[T^{(1,3)} \geq \nu]. \quad (28)$$

For any such H_0 , we have

$$\Pr[T' \geq \nu | \mathcal{F}'_0](H_0) \leq \Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) + \Pr[T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}'_0](H_0). \quad (29)$$

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

$$\Pr[C_\nu = 1 | \mathcal{F}'_0](H_0) \leq \nu^4 n^{-1} = o(n \cdot \Pr[T^{(1,3)} \geq \nu]). \quad (30)$$

In addition, Lemma 13 shows that

$$\Pr [T' \geq \nu \wedge C_\nu = 0 | \mathcal{F}'_0] (H_0) \leq \Pr [T^{(1,3)} \geq \nu]. \quad (31)$$

Finally, combining (29), (30) and (31), we obtain (28), thereby completing the proof. \square

Proof of Proposition 7. Let N be the number of good pairs. We are going to show that $E[N^2] \sim (E[N])^2$. More specifically, we analyze $E[N(N-1)]$. 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

$$E[N(N-1)] = \sum_{(x,y),(x',y') \in W} \Pr [\text{both } (x, y), (x', y') \text{ are good}]. \quad (32)$$

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 (32) as

$$E[N(N-1)] = n(n-1)(n-2)(n-3) \cdot \Pr [\text{both } (1, 2), (3, 4) \text{ are good}] + 2n(n-1)(n-2) \Pr [\text{both } (1, 2), (1, 3) \text{ are good}].$$

Then invoking Lemmas 14 and 15, we thus obtain

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

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

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

From this, the proposition follows. \square

5 Computing a Propagation Sequence

An algorithm A with the properties claimed in Theorem 2 is outlined in Figure 1. This algorithm behaves as stated for any probability parameter p ; for convenience we write $p = c(n)/(n \ln n)$. We will assume $c(n) \geq \ln n/n$, as otherwise the corresponding graph will have less than n edges with very high probability and the statements of the theorem follow trivially.

Algorithm A (for computing a propagation sequence);
given $H = (V, E)$ following $\mathcal{H}(n, p)$, where $V = [n]$ and $E = \{e_1, \dots, e_m\}$;
(1) prepare a random hash table **HashTable** with $O(m)$ entries such that one may search for edges containing any pair of vertices using the pair as a key;
(2) **for each** initial edge $e \in E$ that has not been examined **do** {
 let the candidate propagation sequence be (e) ;
 let u, v, w be the vertices of 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 ;
 (a) **for each** $y \in D$ **do** {
 use **HashTable** to search for edges containing x and y ;
 (b) **for each** such edge with x and y **do** {
 $z \leftarrow$ the third vertex of the edge;
 if $z \in N$ **then** {
 append this edge to the candidate propagation sequence;
 $A \leftarrow A \cup \{z\}$; $N \leftarrow N - \{z\}$;
 } }
 }
 $A \leftarrow A - \{x\}$; $D \leftarrow D \cup \{x\}$;
 } // end of the while loop
 if $N = \emptyset$ (i.e., the process succeeds) **then**
 output the candidate propagation sequence and terminate;
 } // end of the for loop
report failure;

Figure 1: Outline of Algorithm A

The algorithm is divided into two steps. At the first step (i.e., step (1) of Figure 1) a random hash table **HashTable** is constructed. For any given pair of vertices x and y , the hash table yields a list (e.g. by reference to a linked list) of edges in E that contain both x and y . If this list is not empty, the pair x, y is called *positive*. The second step (i.e., step (2) of of Figure 1) is based on the propagation process we introduced in

section 2. In this step the algorithm searches for an initial pair, that is called *good*, such the propagation process *succeeds*, i.e., such that $n - 1$ edges are obtained in order with each (after the first) covering exactly one previously uncovered vertex. Pairs that are not good are called *bad*. 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 pairs of vertices that appear in some edge of E , which we call an *initial edge*. As soon as a good initial edge (an edge containing a good initial vertex pair) is found, the algorithm outputs the candidate propagation sequence that by manner of its construction at this point must be a proper propagation sequence as in Definition 1. To facilitate the analysis we shall assume that the order in which edges are tried in step (2) is chosen randomly by the algorithm.

Clearly this algorithm finds a propagation sequence if the given graph is propagation connected. Thus our task is to show that the expected running time of the algorithm is linear in the number of edges. In the following consider a random hypergraph $H : \mathcal{H}(n, p)$; let E and m denote its edge set and $|E|$ respectively. Note that m is a binomially distributed random variable, sharply concentrated around its expectation $E[m] = \binom{n}{3}p$. We shall assume without further remark from here on that $m \propto n^3p$.

Let us go through the algorithm to check what is necessary to show the desired time bound. The first step of the algorithm is to prepare an appropriate hash table for checking whether a given pair of vertices is positive. By using a standard pairwise 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.

Thus it remains to show that the second step, the main step of the algorithm, can be executed in time $O(m)$ on average. Notice that in the propagation process of step (2), the worst case time needed to propagate to some i vertices is trivially $O(i^3)$, as each vertex pair is used at most once and each time at most i new vertices are processed. Furthermore, the worst case time of the entire algorithm is $O(n^6)$ as there are less than n^2 initial vertex pairs to try.

Before proceeding however, let us point out that for the sake of brevity and readability, the exposition of algorithm A in Figure 1 is slightly wrong on one point and should be refined. Namely, if an initial edge in step (2) is good, the process will eventually deplete the set, N , of neutral vertices; but when N becomes small, the number of pairs x, y the algorithm must try to find an appropriate z , may be too large.

Modification of the algorithm: Let $n' = n - |N|$ be the number of active and dead vertices at some point in stage (2) of the algorithm. If $n' > (\ln n)^3$, then the way to search for a new vertex z is changed as follows: Fix any active vertex x and scan the edges containing x for an edge $\{x, y, z\}$ where $z \in N$ and $y \in D$.

Proposition 16 *If an initial vertex pair is good, the time to complete step (2) in algorithm A with the modification described, is $O(m)$ on average.*

Proof. After reaching $(\ln n)^3$ vertices in expected $O((\ln n)^9)$ time, the algorithm investigates active vertices and corresponding edges. No vertex is investigated more than once

as it is moved from active to dead immediately after. Since an edge contains exactly three vertices, no edge is investigated more than three times. Thus the total time is $O((\ln n)^9 + 3m) = O(m)$. \square

Since the algorithm terminates as soon as it has found (and processed) a good pair, it remains now to show that the time spent on *bad* initial pairs before this, is not too large on average. From the proof of Proposition 5, it is easy to show that the probability there is some bad initial vertex pair from which we reach $(\ln n)^3$ vertices, is smaller than $O(n^{-6})$. Since the worst case time of algorithm A is trivially $O(n^6)$ we may therefore assume without further remark that any initial pair that propagates to $(\ln n)^3$ vertices will propagate to the entire graph (the total expected time for the other case is at most a constant and therefore negligible). We now come to the key points of our analysis of the expected running time for step (2). First we consider a parameter range in which we know the graph is almost surely propagation connected.

Proposition 17 *Let $c(n) > 0.3$. The time spent on step (2) in algorithm is $O(m)$ on average.*

Proof. By the proof of Proposition 7 we have that with probability $> 1 - n^{-\epsilon}$ the graph has at least n^δ good pairs for some positive constants δ and ϵ . This is true for $c(n) > 0.3$. If there are indeed more than n^δ good pairs, and the pairs are investigated in random order, then we expect to hit on a good pair in $O(n^{2-\delta})$ tries. Since the propagation from bad pairs must finish before reaching $(\ln n)^3$ vertices, the expected time is trivially $O((\ln n)^9 n^{2-\delta}) = O(m)$.

With probability less than $n^{-\epsilon}$ it happens that there is not a sufficient number of good pairs in the graph. But then the expected time spent on bad pairs is still $O((\ln n)^9 n^{2-\epsilon}) = O(m)$. \square

Finally we have the case where there may or may not exist a propagation sequence, but the graph at any rate has not too many edges. In this region, any given vertex pair is in fact *very bad* with high probability, as will be seen.

Proposition 18 *Let $c(n) \leq 0.3$. The time spent on step (2) in algorithm is $O(m)$ on average.*

Proof. Consider the argument for the lower bound in Section 3, but this time choose a constant stopping time, $t_0 = 100$. The probability that a given vertex pair will propagate past the corresponding number of vertices is therefore bounded as

$$\begin{aligned} \Pr \left[\text{Bin} \left(n \frac{t_0(t_0 + 1)}{2}, p \right) \geq t_0 \right] &\leq \exp(-\mu_0 \varphi(t_0/\mu_0 - 1)) \\ &\leq \exp \left(-\frac{t_0}{2} \ln \ln n \right) = (\ln n)^{-50} \end{aligned}$$

for sufficiently large n . We have the following three cases, each of which has expected linear time.

- (i) The process stops short of 100 vertices: Expected time is $O(100^3 m) = O(m)$.
- (ii) The process stops between 100 and $(\ln n)^3$ vertices: Expected time is $O(m(\ln n)^{9-50}) = O(m)$.
- (iii) The process propagates to the entire graph: Expected time $O(m)$ as given by Proposition 16.

Note again that we ignored the possibility that the process stops after $(\ln n)^3$ but short of the whole graph, as this case is exceedingly unlikely and the expected time no more than a constant. \square

Proof of Theorem 2. The Propositions 17 and 18 together with the previous argument about step (1), show that the expected time of the algorithm is linear in m . If the algorithm fails, it must have tried all possible initial pairs and found them bad, hence there is no propagation sequence. On the other hand, any sequence outputted by the algorithm is trivially a propagation sequence.

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 Science 4627, 341–352, 2007.
- [BO09] R. Berke and M. Onsjö, Propagation connectivity of random hypergraphs, 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.