Connectedness of graphs generated by a random d-process

A. Ruciński and N.C. Wormald

Abstract

Suppose that a graph process begins with n isolated vertices, to which edges are added randomly one by one so that the maximum degree of the induced graph is always at most d. In a previous article, the authors showed that as $n \to \infty$, with probability tending to 1, the result of this process is a d-regular graph. This graph is shown to be connected with probability asymptotic to 1.

AMS subject classifications: 05C80, 60K99.

1 Introduction

A random d-process begins with n isolated vertices, to which edges are added randomly one by one so that the maximum degree of the induced graph is always at most d. Multiple edges are always forbidden. The process effectively stops when no more edges can be legally added, which must happen after at most $N = \lfloor dn/2 \rfloor$ steps. Let $G_0, \ldots, G_N$ be the (random) graphs obtained in the course of such a process, where, if the process stops with less than N edges, the sequence is padded out with copies of the last graph. In [4] the authors showed that with probability tending to 1 as $n \to \infty$, the result of this process is a d-regular graph (except for one vertex of degree $d-1$ when dn is odd). More precisely, the following result was proved.

[0]Research supported by the Australian Research Council and by KBN grant # 2 P03A 032 16

Theorem 1 *For $d \geq 1$, in a random d-process*

$$\lim_{n\to\infty} \mathbf{P}(|E(G_N)| = \lfloor \tfrac{1}{2}dn \rfloor) = 1.$$

In this paper we consider the connectedness of a random d-process.

Theorem 2 *For $d \geq 3$,*

$$\lim_{n\to\infty} \mathbf{P}(G_N \textit{ is connected }) = 1.$$

In fact, we show that with probability asymptotic to 1, the graph G_i becomes connected already for $i = \lfloor dn/2 - \log^{1/8} n \rfloor$.

For $d = 1$ (and $n \geq 3$) G_N is always disconnected, while for $d = 2$, $\mathbf{P}(G_N \text{ is connected}) = \mathbf{P}(G_N \text{ is Hamiltonian})$, and this was shown in [6] to be $\Theta(1/\sqrt{n})$.

One is easily led to conjecture the result of Theorem 2 from simulation results, although for Theorem 1 this is not the case, the reason being that the convergence in Theorem 1 is only at the rate of $1/\log n$ (approximately — the true rate is not known exactly). We briefly mention here some results obtained from simulating a random 3-process. The events in Theorem 1 and Theorem 2 seem to be highly correlated; no doubt the easiest way for a 3-process to produce a disconnected G_N is for it to have two components, one 3-regular and the other a triangle. (No more edges can then be added by the process.) One of our main interests is in comparing the *regular* graphs produced by d-processes with other models of random regular graphs, such as the uniform model. To this end, we carried out simulation of random 3-processes on even numbers of vertices to estimate the probability that G_N is disconnected, conditional on it being 3-regular.

This was done with two computer programs, one of which repeatedly simulated the random 3-process and recorded the number of times G_N was both disconnected and 3-regular, and the other recording how many times it was 3-regular. It was done this way because the former was essentially slower, but savings could be made by noting that each single run did not have to go to the end but could be terminated as soon as the graph became connected. To make errors fairly uniform, the disconnectedness testing went for each number n of vertices until 10,000 successes. We denote by i_d the number of iterations, or runs of the process, required to achieve this. Consequently we get the probability of success $(10000/i_d)$ with expected error of only a couple of percent.

The second program was much faster and so could afford to test for regularity until 10^6 successes, that is, regular final graphs. The number of iterations for this is denoted by i_r. Finally we estimate $\mathbf{P}(\mathcal{D}|\mathcal{R})$, where $\mathcal{D}$ is the event that G_N is disconnected and $\mathcal{R}$ that it is regular. The programs took a few days to run on a DEC Alpha station. The results are in the following table.

n	i_d	i_r	$\mathbf{P}(\mathcal{D}\|\mathcal{R})$
20	24342275	1613430	0.00066281
30	61121366	1535365	0.00025120
40	110628317	1492160	0.00013488
50	169726426	1461720	0.000086122
60	237254000	1439774	0.000060685
70	314481451	1423049	0.000045251
80	408520743	1407122	0.000034444

Table 1: Simulation estimates for the probability that a random 3-process produces a disconnected graph, given that it produces a 3-regular graph.

Setting $p = \mathbf{P}(\mathcal{D}|\mathcal{R})$, and graphing $-\log p$ against $\log n$ gives an almost perfect straight line with slope approximately 2 and $(-\log p)$-intercept about 1.4. So we conjecture the probability of G_N being disconnected, given that it is 3-regular, is about cn^{-2} where $c \approx e^{-1.4} \approx 0.25$.

It is interesting to compare this with the uniform model of random 3-regular graphs, where the probability of disconnectedness is asymptotic to $\frac{2}{27}n^{-2}$ as $n \to \infty$. This can be calculated by showing that the major contribution comes from graphs with a component isomorphic to K_4.

Indeed, the natural line of attack for the connectivity question, which has been successful for uniform regular models as well as for ordinary random graphs (see [2]), relies on ruling out the presence of components of order $s = 1, 2, \ldots, n/2$. Due to Theorem 1, this has been checked already for the random d-process for $s = 1, 2, \ldots, d$. For any fixed s it is tedious, but possible to rule out components of order s. We include calculations for $d = 3$ and $s = 4$ in Section 2. However, for larger s — say around $\sqrt{n}$ — it seems to be a very difficult approach.

In the proof of Theorem 1, a key role was played by isolated vertices. Their disappearance early in the process was crucial there. Let I_i denote the number of isolated vertices in G_i. The analysis of the behaviour of I_i, carried out in

Section 3, will be crucial for the present proof too. In the evolution of the ordinary random graph (which is the same as a d-process except that there is no upper bound on vertex degree), connectedness occurs, with high probability, at the same moment as the last isolated vertex disappears (see [2] or [1]). We are unable to verify if this phenomenon holds also for d-processes.

Besides I_i, another important random variable of the process is U_i — the number of *unsaturated* vertices in G_i; that is, vertices with current degree less than d. These are the vertices which are still 'in the game', contrary to the vertices of degree d, which remain 'dead' till the end of the process.

We classify connected components occurring in the course of the d-process according to the number of unsaturated vertices in them. This is called the *active size* of a component. Components of active size 0 will be called *dead.* Once a dead component is created, it remains such till the very end, and, if its order is less than n, the final graph G_N is bound to be disconnected. This 'irreversibility of death' is a new feature as compared to the ordinary model of random graphs, where at any time any component can be linked to another one. It contributes to the higher level of difficulty one encounters when studying the connectedness of a random d-process.

A component is said to be *in danger* if its active size is 1 or 2, but it is not an isolated vertex K_1 or an isolated edge K_2. A component in danger may become dead, or contained in a dead one, in a single step of the process (unless its active size is 2 and the two vertices of degree less than d are already adjacent, but it turns out that such components cause no problems). Let D_i be the number of components in danger belonging to the graph G_i. In Section 4 we carefully analyse the behaviour of D_i making almost sure that the number of components in danger is kept low throughout the process and thus no edge hits and kills any of them.

The fact that components in danger almost never become dead is to a large extent caused by the presence of giant components which 'attract small fry', catching them before they 'die'; that is, catching any components in danger before a single edge can have both ends in them. For technical convenience, any component of G_i with active size at least $\frac{2}{5}U_i$ will be called a *giant.* The existence of giant components is established with some help from a supermartingale concentration inequality in Section 5. The proof of Theorem 2 is completed in Section 6.

As most of the interesting things (from our perspective in this paper) happen toward the end of the process, we quite often prefer to measure the remaining (residual) time; that is, $N-i$ rather than i. The same happens with a runner who measures distance from the start when near the start, but distance to go when near the end. In order to smoothly shift between the two modes of counting, for any i we define $r(i) = N - i$. Thus at time i there are at most $r(i)$ steps to go. This remaining time will be referred to as *r-time*. That is, at time i, the r-time is $r(i)$. Similarly let $i(r) = N - r$. If symbols i and r appear in the same expression or sentence, it will be always assumed that $r = r(i)$, and thus $i = i(r)$. Also, phrases such as 'before time x' have the obvious meaning ('up to time $\lfloor x \rfloor$') for x a non-integer.

We now state a simple fact which bounds the number of unsaturated vertices U_i in terms of the r-time r. Note that $dn - 2i = 2r$ if dn is even and $dn - 2i = 2r + 1$ otherwise.

Fact 1. *For all $d \geq 1$ and all $i = 0, \ldots, N$,*

$$\frac{2r}{d} \leq U_i \leq dn - 2i - (d-1)I_i \leq dn - 2i \leq 2r + 1 \ .$$

Proof. The proof is immediate by suitably bounding from below and from above the quantity

$$\sum_{v \in [n]} (d - \deg_{G_i}(v)) = dn - 2i \ . \quad \blacksquare$$

As a corollary, observe that the (conditional) probability of inserting the $(i+1)$-st edge of the d-process into a given place (a pair of unsaturated vertices not being an edge of G_i) is at most

$$\frac{1}{\binom{U_i}{2} - (d-1)U_i/2} = O(r^{-2}) \ .$$

Throughout we have $d \geq 3$ fixed, and n always denotes the number of vertices. We will be using the phrase *asymptotically almost surely*, abbreviated to a.a.s., to indicate the fact that the probability of the event in question converges to 1 as $n \to \infty$. When saying that *a.a.s.* a random variable X is $o(a_n)$ we mean that for some function $\omega(n)$ which goes to infinity with n, *a.a.s.* $X \leq a_n/\omega$. For our purposes here it is sufficient to set $\omega(n) = \log\log\log n$.

2 Dead End

Here we examine the connectivity question, using an approach successful for ordinary random graphs (see [2]). In this, we try to rule out the presence of components of order $s = 1, 2, \ldots, n/2$, by showing that the implied separation of the graph does not occur. Let $\mathcal{A}_s$ be the event that there is a component of order s in the final graph G_N. It follows from Theorem 1 that $\mathbf{P}(\bigcup_{s=1}^{d} \mathcal{A}_s) = o(1)$. We will now show calculations leading to bounding the term $\mathbf{P}(\mathcal{A}_4)$ in the smallest case when $d = 3$. After this experience we will be happy to give up this approach, though it is a gentle introduction to computing probabilities with this model of a random graph process.

Let P_4 be the probability that the set of vertices $T = \{1, 2, 3, 4\}$ forms a component of G_N. By symmetry,

$$\mathbf{P}(\mathcal{A}_4) \leq \binom{n}{4} P_4 \, . \tag{2.1}$$

Observe that the only way the vertices of T can form a component is when T spans the complete graph K_4. Let $\pi = (\pi_1, \ldots, \pi_6)$ be a permutation of the six edges in T, and let $i_1 < i_2 < \cdots < i_6$ be the times of appearance of these edges in the 3-process in the order determined by π. Furthermore, let $\mathcal{E}_i$ be the event that the i-th edge of the process will be π_j for $i = i_j$, $j = 1, \ldots, 6$, and that it will have both endpoints outside the set T otherwise. Note that here $N = \lfloor 3n/2 \rfloor$, and set $r_j = r(i_j)$, $j = 1, \ldots, 6$, and

$$P_4^{\pi}(r_1, \ldots, r_6) = \mathbf{P}\left(\bigcap_{i=1}^{N} \mathcal{E}_i\right) \, .$$

Then, by the chain formula,

$$P_4^{\pi}(r_1, \ldots, r_6) = \mathbf{P}(\mathcal{E}_1)\mathbf{P}(\mathcal{E}_2|\mathcal{E}_1)\cdots\mathbf{P}(\mathcal{E}_N|\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{N-1}) \, .$$

In order to estimate $\mathbf{P}(\mathcal{E}_i|\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{i-1})$, let $\mathcal{H}_i = (G_0, \ldots, G_{i-1})$ be the history of the process up to time $i-1$. Setting $\mathcal{E} = \mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{i-1}$ for brevity, we then have

$$\mathbf{P}(\mathcal{E}_i|\mathcal{E}) = \sum_{\mathbf{h}_i \in \mathcal{E}} \mathbf{P}(\mathcal{E}_i|\mathcal{H}_i = \mathbf{h}_i)\mathbf{P}(\mathcal{H}_i = \mathbf{h}_i|\mathcal{E})$$

and

$$\mathbf{P}(\mathcal{E}_i|\mathcal{H}_i = \mathbf{h}_i) = \mathbf{P}(\mathcal{E}_i|G_{i-1} = g_{i-1}),$$

where the lower case letters represent concrete values of random variables designated by upper case counterparts, thus $\mathbf{h}_i = (g_0, \ldots, g_{i-1})$. We know already that for $i = i_1, \ldots, i_6$,

$$\mathbf{P}(\mathcal{E}_i | G_{i-1} = g_{i-1}) = O(1/r^2) \ .$$

In order to handle the remaining cases, we need to know how many vertices of T have been saturated so far during the process. Let s_k, $k = 1, 2, 3, 4$, be the element of the set $\{r_1, \ldots, r_6\}$ which is the r-time of saturation of the k-th vertex of T. The values of s_k depend on the permutation π, but we always have $s_1, s_2 \in \{r_3, r_4, r_5\}$ and $s_3 = s_4 = r_6$. Let us set $s_0 = N$ for convenience. Then, for $i(s_k) < i < i(s_{k+1})$, $k = 0, 1, 2$, as $\mathbf{h}_i \in \mathcal{E}_i$ implies that g_{i-1} has no edges from T to outside T, we have

$$\mathbf{P}(\mathcal{E}_i | G_{i-1} = g_{i-1}) < 1 - \frac{(4-k)(u_i - 4 + k)}{\binom{u_i}{2}} < 1 - \frac{4-k}{r} + O(r^{-2}) \ .$$

Hence

$$\begin{aligned}
(r_1 r_2 \ldots r_6)^2 P_4^\pi(r_1, \ldots, r_6) &= O(1) \exp\left(- \sum_{s_0 > r > s_1} \frac{4}{r} - \sum_{s_1 > r > s_2} \frac{3}{r} - \sum_{s_2 > r > s_3} \frac{2}{r} \right) \\
&= O(1) \exp\{-4 \log n + \log s_1 + \log s_2 + 2 \log s_3\} \\
&= O(1) \frac{s_1 s_2 s_3^2}{n^4} \ .
\end{aligned}$$

Our goal, in view of (2.1), is to prove that $n^4 P_4 = o(1)$, and we just have shown that

$$n^4 P_4 = O(1) \sum_\pi \sum_{r_1 > \cdots > r_6} \frac{s_1 s_2}{(r_1 \cdots r_5)^2} \ .$$

It can be verified that for any π, and thus for any particular values of s_1 and s_2, the above quantity converges to 0 provided that the range of r_1 can be bounded from below by any function of n tending with n to ∞. Consider, for example, the case that $s_1 = r_3$ and $s_2 = r_5$. Then

$$\sum_{r_1 > \cdots > r_6} \frac{1}{r_1^2 r_2^2 r_3 r_4^2 r_5} \leq \sum_{r_1 > \cdots > r_4} \frac{1}{r_1^2 r_2^2 r_3 r_4} \leq \sum_{r_1 > r_2} \frac{1}{r_1^2 r_2} < \sum_{r_1} \frac{\log r_1}{r_1^2} \ .$$

Can we actually restrict the range of r_1? Some extra knowledge is needed here. With some sweat it was proved in [4] (and it will be sketched in the next section) that *a.a.s.* the last isolated vertex disappears by r-time around $\log n/\log\log n$. Before r-time r_1 all vertices in T are still isolated and we obtain the desired lower bound on r_1.

Nevertheless this straightforward approach to proving the connectedness of G_N via estimating $\mathbf{P}(\mathcal{A}_s)$ seems very hard to push through for higher values of s, in particular for $s = n^c$, say $s = \sqrt{n}$.

Hence, we give up this approach, turning our attention to analysing the behaviour of I_i throughout the d-process, as the early disappearence of isolated vertices will play a crucial role in our proof, as it did in [4] where we proved Theorem 1.

3 Isolated vertices

In this section we take $d \geq 2$. We begin with the expected number of vertices which remain isolated throughout a certain time interval. It was proved in [4] that the conditional probability that a given vertex v is isolated at time i_2, given that it was isolated at time i_1, is at most $O(r(i_2)/r(i_1))$. Applying this for $i_1 = 0$, we find that the expected number of isolated vertices at time i is $O(r)$, which is in any case a deterministic bound on this number. On the other hand, if we know that at some time i, *a.a.s.* there are $o(r)$ isolates, then we can use this property to conclude that at some later r-time $r \to \infty$ they will disappear *a.a.s.* It will turn out that *a.a.s.* I_i becomes $o(r)$ at any r-time $o(n)$. By Fact 1, this implies that I_i will be dropping significantly as compared to U_i, the number of unsaturated vertices. We collect together all the required facts about isolates (and isolated edges) in the next lemma.

Lemma 1 *The following are true* a.a.s. *(for some constant $c < 1$ in (i)).*

(a) For all r in the range $r(0.1n) \geq r \geq n^{9/10}$, we have $I_i < cU_i$.

(b) For all r in the range $n^{9/10} \geq r \geq \log n/\log\log n$, we have $I_i < U_i/\log^{0.99} n$.

(c) For all r satisfying $\log n/\log\log n \geq r$, there are no isolated vertices.

(d) For all r satisfying $\log^{1.3} n \geq r$, there are no isolated edges.

Proof. First consider (a). Define the functions $b = b(x)$ and $q = q(x)$ as in [4] by

$$b'(x) = \frac{-2b}{d - 2x - (d-1)b}\,, \qquad b(0) = 1 \tag{3.1}$$

and

$$q = \frac{2b}{d - 2x} \tag{3.2}$$

for $0 \leq x < d/2$. As was shown in [4, Section 2], $q'(x)$ is negative on the interval $(0, d/2)$, and so there is some $\epsilon > 0$ such that $q(x) < 2/d - \epsilon$ for, say, $0.1 < x < d/2$. Then [4, Lemma 3.2] implies that *a.a.s.*

$$\begin{aligned} I_i &\leq nb\left(\frac{i}{n}\right) + o\left(n^{9/10}\right) \\ &= rq\left(\frac{i}{n}\right) + o\left(n^{9/10}\right) \end{aligned} \tag{3.3}$$

for all $i = 0, \ldots, dn/2 - \lfloor n^{9/10} \rfloor$. Thus we obtain that for some $\epsilon' > 0$, *a.a.s.*

$$I_i \leq \frac{2r(i)}{d}(1 - \epsilon')$$

from time $0.1n$ to r-time $n^{9/10}$. On the other hand, $U_i \geq 2r(i)/d$ by Fact 1. This proves (a).

We turn to the proof of (b). From [4, equation (2.3)] we have

$$q(x) \sim \frac{-2}{(d-1)\log(d/2 - x)} \tag{3.4}$$

as $x \to d/2$. Thus from (3.3) we obtain that *a.a.s.* with $r_0 = \lfloor n^{10/11} \rfloor$,

$$I_{i(r_0)} \leq \frac{11r_0 + o(r_0)}{\log n} < \frac{12r_0}{\log n}. \tag{3.5}$$

Define the function

$$s = s(i) = \left\lfloor \frac{r(i)}{\log^{.99} n} \right\rfloor.$$

Choose $m < 12r_0/\log n$. For $i \geq dn/2 - n^{9/10}$, let P_i be the conditional probability that $I_i \geq s(i)$ given that $I_{i(r_0)} = m$.

At time i, the probability that the next edge is incident with some vertex from a given set of s isolates is at least

$$\frac{(U_i - s)s + \binom{s}{2}}{\binom{U_i}{2}} \geq \frac{s}{r} - O(\frac{s^2}{r^2}) \ .$$

Thus, for a given set of $s(i)$ isolates at time $i_0 = i(r_0)$, the probability that all vertices in this set remain isolated until time i is at most

$$\begin{aligned} \prod_{j=i_0}^{i-1} \left(1 - \frac{s}{r(j)} + O\left(\frac{s^2}{r(j)^2}\right)\right) &\leq \exp\left\{-s \sum_{k=r(i)+1}^{r_0} \frac{1}{k} + O\left(\frac{s}{k^2}\right)\right\} \\ &= \left(\frac{O(r(i))}{r_0}\right)^s . \end{aligned}$$

Hence, multiplying by $\binom{m}{s}$,

$$P_i = \left(\frac{O(mr(i))}{sr_0}\right)^s = \left(\frac{O(1)}{\log^{.01} n}\right)^{r(i)/\log^{.99} n} .$$

Thus if $r_1 = \lfloor \log n / \log\log n \rfloor$,

$$\begin{aligned} \sum_{r=r_1}^{\lfloor n^{9/10} \rfloor} P_{i(r)} &= \sum_{r=r_1}^{\lfloor \log^2 n \rfloor} P_{i(r)} + \sum_{r=\lfloor \log^2 n \rfloor}^{\lfloor n^{9/10} \rfloor} P_{i(r)} \\ &= \sum_{r=r_1}^{\lfloor \log^2 n \rfloor} \left(\frac{O(1)}{\log^{0.01} n}\right)^{r_1/\log n^{.99}} + \sum_{r=\lfloor \log^2 n \rfloor}^{\lfloor n^{9/10} \rfloor} \left(\frac{O(1)}{\log^{0.01} n}\right)^{\log^{1.01} n} \\ &= o(1). \end{aligned}$$

Since this bound is uniform over $m < 12r_0/\log n$, by (3.5) *a.a.s.* $I_i \leq s(i)$ for all i in the range $i(r_1) \geq i \geq i(n^{9/10})$. This gives (b).

For (c), define P_i as above in (b), but with $s = 1$. Then as before we obtain

$$P_{i(r_1)} = \left(\frac{O(mr_1)}{r_0}\right) = O\left(\frac{1}{\log\log n}\right) .$$

This shows that $I_{i(r_1)} = 0$ *a.a.s.*

Finally, we prove (d). We will focus on three r-times:

$$r_1 = \log^3 n, \qquad r_2 = \log^{1.9} n, \qquad r_3 = \log^{1.3} n.$$

By part (b), there are *a.a.s.* $O(\log^{2.01} n)$ isolated vertices in $G_{i(r_1)}$, and the number of isolated edges is at most $\frac{1}{2}U_i = O(\log^3 n)$ by Fact 1. The isolated edges present at r-time r_2 may have one of two origins: either they are survivors from r-time r_1 or they were created since. The expected number of surviving isolated edges from r-time r_1 until r-time r_2 is, by calculations similar to those in (b) (see also [4]), $O(r_1(r_2/r_1)^2) = O(\log^{0.8} n)$, while the expected number of isolated edges created between r-times r_1 and r_2 is, again using (b),

$$O(r_1 - r_2) \max_{r_1 \geq r \geq r_2} \binom{I_i(r)}{2} \frac{1}{r(i)^2} = O(r_1/(\log n)^{1.98}) = O((\log n)^{1.02}) \ .$$

Thus, by Markov's inequality *a.a.s.* there are at most $(\log n)^{1.03}$ isolated edges at r-time r_2. Similar arguments show that *a.a.s.* none of these will survive until r-time r_3 (since $(\log n)^{1.03}(r_3/r_2)^2 = o(1)$). On the other hand, there are *a.a.s.* too few isolated vertices at r-time r_2 to create any isolated edges before r-time $\log n/\log\log n$ (since $r_2/(\log n)^{1.98} = o(1)$, again using (b)), nor at any time afterwards (when, by (c), *a.a.s.* there are no isolated vertices at all). ∎

4 Components in danger

In this section we take $d \geq 3$. Recall that a component is said to be in danger if its active size is 1 or 2 but it is different from K_1 and K_2. In the previous section we verified that isolated copies of K_1 and K_2 (which are themselves dangerous rather than in danger, because components in danger can quickly be created from them) will vanish after r-time $\log n/\log\log n$. We shall now take a firm grasp of the components in danger, showing that throughout the process there are not too many of them. As an eventual consequence, it will follow that *a.a.s.* there is no dead component formed before the end of the process; that is, G_N is connected.

The following section will examine the components of very large active size. In each of these two sections, the results about the process after r-time $\log^3 n$ depend on the results in the other section before r-time $\log^3 n$, so a logical development would reorder them, dealing first with early time, and then with later time in the process. However, we do not follow this path because the arguments concerning one type of component (in danger or giant) before and after r-time $\log^3 n$ tend to be similar and better kept together. The effect of this is that the proof of Lemma 3 depends on Corollary 4 in the later section.

Let D_i be the number of components in danger in graph G_i, and let $\mathcal{D}_1$ and $\mathcal{D}_2$ be the events that $D_i < \omega \log n$ for all $i = 0, \ldots, N$, and that $D_i < \omega \log\log n$ for all $i > i(\log^3 n)$, respectively. Here and in the rest of the paper, ω stands for $\log\log\log n$. The following lemma implies that the event $\mathcal{D}_1$ holds *a.a.s.*

Lemma 2 *The expected number of times that the number of components in danger rises, or an edge is inserted into a component in danger, between time i_1 and time i_2 is $O(\log(r(i_1)/r(i_2)))$.*

Proof. The number of components in danger can only rise when an edge is inserted into a component of active size 3 or 4. So, let J_i be the indicator variable of the event that the $(i+1)$-st edge of the d-process is inserted into a component of active size at most 4 (and at least 2, of course). By Fact 1, there are at most $U_i/2 \le r + 1/2$ such components in G_i, and each has at most 6 places for an edge. Thus

$$\mathbf{P}(J_i = 1) = \sum_{g_i} \mathbf{P}(J_i = 1 | G_i = g_i)\mathbf{P}(G_i = g_i) = O\left(\frac{1}{r(i)}\right)$$

using Fact 1 again. The lemma follows on summing this bound over $i_1 \le i \le i_2 - 1$. ∎

Corollary 1 *The event $\mathcal{D}_1$ holds* a.a.s.

Proof. Since $D_0 = 0$, then, for each i, D_i is not greater than the number D of times that the number of components in danger rises during the entire process. By Lemma 2 with $i_1 = 0$ and $i_2 = N - 1$, $\mathbf{E}(D) = O(\log n)$, and the corollary follows by Markov's inequality. ∎

Corollary 2 *There are* a.a.s. *no dead components in G_i for any $i < i(\log^3 n)$.*

Proof. We define $\mathcal{A}_i$ to be the event that the $(i+1)$-st edge of the d-process creates a dead component. (This may only happen if it has both ends in components in danger.) Then with $r_1 = dn/2$ and $r_2 = \log^3 n$,

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{i=i_1}^{i_2-1} \mathcal{A}_i\right) &\leq \sum \mathbf{P}\left(\mathcal{A}_i | D_i < \omega \log n\right) + \mathbf{P}(\neg \mathcal{D}_1) \\
&= O((\omega \log n)^2) \sum_{r=r_2}^{r_1-1} \frac{1}{r^2} \quad + o(1) \\
&= O(\omega^2 / \log n) + o(1) = o(1)\ . \quad \blacksquare
\end{aligned}$$

Thus, we have gained a brief respite until r-time $\log^3 n$. The following lemma pushes this up a little further. Its proof depends on Corollary 4 from the next section.

Lemma 3 *The event $\mathcal{D}_2$ holds* a.a.s.

Before proving this, let us extend the 'dead-component-free zone' of the d-process using Lemma 3. The choice of $\log^{1/8} n$ below is arbitrary and certainly not best possible, but easily sufficient for our task.

Corollary 3 *There are* a.a.s. *no dead components in G_i for any $i < i(\log^{1/8} n)$.*

Proof. In view of Corollary 2, we may focus on the range of r-time between $\log^3 n$ and $\log^{1/8} n$ and repeat the calculations of Corollary 2 with $\mathcal{D}_1$ replaced by $\mathcal{D}_2$ (here we apply Lemma 3) and the $\log n$ term replaced with $\log\log n$. $\blacksquare$

The strategy for our proof of Theorem 2 can now be described as follows: along with proving Lemma 3 and hence Corollary 3, we will show that *a.a.s.* at r-time $\log^{1/8} n$ all unsaturated vertices belong to the same component, and thus the graph G_i, $i = i(\log^{1/8} n)$, is connected.

Proof of Lemma 3. Let $r_1 = \log^6 n$ and $r_2 = \log^3 n$, and set $i_1 = i(r_1)$ and $i_2 = i(r_2)$. By Lemma 2, *a.a.s.* D_i rises at most $\omega \log\log n$ times after r-time r_1. Hence we are done if we show that the event $\mathcal{B}$, that all components in danger at time i_1 will cease to be in danger by time i_2, holds *a.a.s.* The presence of the giants will help to achieve this task. We will bound $\mathbf{P}(\mathcal{B})$ by the probability that all components in danger at time i_1 will join a giant sometime before time i_2. Note that a giant at time i may no longer be a giant (and also may be contained in a larger component) at time $i+1$.

For any i, denote by $\mathcal{D}_{1,i}$ the event $D_i < \omega(\log n)$, and by $\mathcal{L}_{1,i}$ the event that for all j in the range $i_1 \le j \le i$ there is a giant in G_j. We have $\mathcal{D}_{1,i_1}$ *a.a.s.* by Corollary 1, and Corollary 4 gives $\mathcal{L}_{1,i_2}$ *a.a.s.* (this is called $\mathcal{L}_1$ in Section 5). Thus it is enough to show

$$\mathbf{P}(\neg\mathcal{B}|G_{i_1} = g_{i_1}) = o(1) \tag{4.1}$$

for any $g_{i_1} \in \mathcal{D}_{1,i_1} \cap \mathcal{L}_{1,i_2}$. The following argument is in this conditional space. Noting that the proofs of Lemma 2 and Corollary 2 are valid conditional on $G_{i_1} = g_{i_1}$, *a.a.s.* no component which was in danger at r-time r_1 becomes dead before r-time r_2. We will show that the probability that at least one component in danger at time i_1 does not become dead and does not join a giant by time i_2, is $o(1)$, thereby establishing (4.1) as required.

Consider a fixed component C which is in danger at time i_1, and define $\mathcal{E}_i$ to be the event that no edge from time i_1 up to time i either makes the component containing C dead or joins it to a giant. Since we have assumed that $\mathbf{P}(\neg\bigcap_{i=i_1}^{i_2-1}\mathcal{L}_{1,i}) = o(1)$, it suffices to show that

$$\mathbf{P}\left(\bigcap_{i=i_1+1}^{i_2}\mathcal{E}_i \cap \bigcap_{i=i_1}^{i_2-1}\mathcal{L}_{1,i}\right) = o\left(\frac{1}{\omega\log n}\right).$$

Note that the probability of an active non-giant component joining to a given giant in one step is at least

$$\frac{\frac{2}{5}U_i}{\binom{U_i}{2}} \ge \frac{2}{5r(i)}$$

by Fact 1. Thus we have

$$\begin{aligned}
&\mathbf{P}\left(\bigcap_{i=i_1+1}^{i_2}\mathcal{E}_i \cap \bigcap_{i=i_1}^{i_2-1}\mathcal{L}_{1,i}\right)\\
\le\;& \mathbf{P}(\mathcal{E}_{i_1+1}|\mathcal{L}_{1,i_1})\mathbf{P}(\mathcal{E}_{i_1+2}|\mathcal{E}_{i_1+1}\cap\mathcal{L}_{1,i_1+1})\mathbf{P}(\mathcal{E}_{i_1+3}|\mathcal{E}_{i_1+2}\cap\mathcal{L}_{1,i_1+2})\ldots\\
\le\;& \prod_{i=i_1}^{i_2-1}\left(1-\frac{2}{5r(i)}\right)\\
=\;& O\left((\log^3 n/\log^6 n)^{2/5}\right)\\
=\;& o\left(\frac{1}{\omega\log n}\right) \quad \blacksquare.
\end{aligned}$$

5 The giants

In this section we will derive some crucial facts concerning the presence of large active components during the d-process (see Corollary 4). Recall that any component of G_i with active size at least $\frac{2}{5}U_i$ is called a giant.

Let L_i denote the greatest active size of a component in G_i. We now introduce the following events:

$\mathcal{L}_0$ is the event that $L_i > c'U_i$ from r-time $\frac{1}{6}n$ until r-time $\log^6 n$, where $c' = 0.99\min\{0.5 - c/2, 0.19\}$, with c defined as in Lemma 1 (a).

$\mathcal{L}_1$ is the event that $L_i > \frac{2}{5}U_i$ from r-time $\log^6 n$ until r-time $\log^3 n$.

$\mathcal{L}_2$ is the event that $L_i > \frac{2}{5}U_i$ from r-time $\log^3 n$ until r-time $\log^{1/8} n$.

It turns out that showing that these events hold *a.a.s.* is easier once we show that a large component exists early in the d-process. Twice along the way we will use the following result from [7]. The proof is not given since it is exactly the same as that of Azuma's inequality (see [3, Lemma 4.1] or [5, Theorem 3]).

Lemma 4 *Let $Y_0, Y_1, \dots$ be a supermartingale with respect to a sequence of σ-algebras $\{\Phi_i\}$ with Φ_0 empty, and suppose $Y_0 = 0$ and $|Y_{k+1} - Y_k| \le c$ for $k \ge 0$ always. Then for all $\alpha > 0$,*

$$\mathbf{P}(Y_i \ge \alpha c) \le \exp(-\alpha^2/2i). \quad \blacksquare$$

Now we will establish the existence of a large component in an early phase of the d-process.

Lemma 5 *A.a.s. at some time i_1, $\frac{1}{10}n < i_1 < N - \frac{1}{6}n$, there exists a component with active size at least $\frac{1}{5}U_{i_1}$.*

Proof. Let X_i denote the number of components in G_i. If $L_i < \frac{1}{5}U_i$, then the conditional expectation of $X_{i+1} - X_i$ given G_i is at most $-\frac{4}{5}$. This is because the next edge can be selected by choosing first one unsaturated vertex u, and then another v, and then repeating both selections if uv is already an edge. No matter where u is, it has at least $\frac{4}{5}U_i$ unsaturated vertices outside its component, and so the probability that v lies in some other component is at least $\frac{4}{5}$. In this case the selections are not repeated. Alternatively, we may argue that the probability that the number of components will drop by 1 is

$$\mathbf{P}(X_{i+1} - X_i = -1) \geq \frac{U_i(\frac{4}{5}U_i)/2}{\binom{U_i}{2}} > \frac{4}{5} \; .$$

If this trend was maintained for at least $i_2 = N - \frac{1}{6}n$ initial steps of the d-process then $\mathbf{E}(X_{i_2})$ would drop below $n - \frac{4}{5}i_2 < 0$ — an obvious contradiction. To formalise this heuristic, let us say that the d-process is *successful* at time $i > 0$ if $L_j \geq \lfloor \frac{1}{5}U_j \rfloor$ for some $0 \leq j < i$. Define

$$Y_i = X_i + \frac{4}{5}i - n$$

provided the d-process is not successful at time i, and $Y_i = Y_{i-1}$ otherwise.

We will apply Lemma 4 to Y_i. For a process which is not successful at time i, it follows from the above calculations that $\mathbf{E}(Y_{i+1}|G_i) \leq Y_i$. On the other hand, if the process is successful then $\mathbf{E}(Y_{i+1}|G_i) = Y_i$ because $Y_{i+1} = Y_i$. Hence the sequence $\{Y_k\}$ is a supermartingale with respect to the sequence of σ-algebras $\sigma(\mathcal{H}_{k+1})$ (the history of the process).

Applying Lemma 4 with $c = 1$ and $i = i_2$, we deduce that

$$\mathbf{P}(Y_{i_2} > n^{2/3}) = o(1).$$

But for any process not successful at time i_2,

$$Y_{i_2} = X_{i_2} + \frac{4}{5}i_2 - n > \frac{1}{15}n \; .$$

Thus the process is *a.a.s.* successful at time i_2. Finally, it is easy to check that the inequality $L_i > \frac{1}{5}U_i$ yields $i > \frac{1}{10}n$. ∎

Lemma 5 ensures that a 'large' active component exists at some time early in the process. The key, rather technical, lemma ensuring that this property is *a.a.s.* preserved through time is the following.

Lemma 6 *Suppose that for some $\delta > 0$ and some $i_1 = i_1(n)$ we have $L_{i_1} > \delta U_{i_1}$. Suppose further that for some $i_2 \geq i_1$ with $r(i_2) \to \infty$ and some $\lambda > 0$,* a.a.s. *for all times $i_1 \leq i \leq i_2$ we have $D_i = o(r(i))$ and $I_i/U_i < 1 - 2\lambda$. Then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{L_i \geq \min\{(\lambda - \epsilon)U_i, L_{i_1} - 2\sqrt{r(i_1)}\} \textit{ for all } i_1 \leq i \leq i_2\right\} = 1.$$

Before proving Lemma 6, we show how useful it is.

Corollary 4 *The events $\mathcal{L}_0$, $\mathcal{L}_1$ and $\mathcal{L}_2$ all hold* a.a.s.

Proof of corollary. Set $\delta = 0.19$ and i_1 as in Lemma 5. Lemma 1 and Corollary 1 verify the truth of the hypotheses of Lemma 6 for any time $i_2 \leq i(\log^6 n)$, with $\lambda = \frac{1}{2} - \frac{c}{2}$, where c is defined in Lemma 1 (a). Applying Lemma 6 with i_1 as in Lemma 5 and $i_2 = i(\lfloor \log^6 n \rfloor)$, we deduce $\mathcal{L}_0$ *a.a.s.* since $c' < \min\{\lambda, \delta\}$.

If, in an application of Lemma 6, $r(i) = o(r(i_1))$ and $L_{i_1} = \Theta(r(i_1))$, then $U_i = o(L_{i_1} - 2\sqrt{r(i_1)})$ and consequently in the conclusion of the lemma $L_i \geq (\lambda - \epsilon)U_i$. If, moreover, $I_i = o(U_i)$ then one can choose $\lambda = 1/2 - \epsilon$ yielding that, say, $L_i > \frac{2}{5}U_i$ *a.a.s.* Thus, by Lemma 1, Corollary 1 and the fact that $\mathcal{L}_0$ holds *a.a.s.*, Lemma 6 with $r(i_1) = n^{9/10}$, $r(i_2) = \log^3 n$ and $\lambda = \frac{4}{9}$ yields the event $\mathcal{L}_1$ *a.a.s.* In turn, this together with Lemma 1, Corollaries 1 and 2, and another application of Lemma 6 with $r(i_1) = \log^4 n$, $r(i_2) = \log^{1/8} n$ and $\lambda = \frac{4}{9}$ implies the event $\mathcal{L}_2$ *a.a.s.* ∎

Note that the above proof uses results from the previous section only pertaining to time up to r-time $\log^3 n$, in particular not Lemma 3. In this way our proof of Lemma 3, which uses Corollary 4, is not circular and is now complete.

The remainder of this section is devoted to proving Lemma 6. The heart of this lemma is the following claim.

Claim. *Assume $i_1 \leq i_3$, $i_3 + \left\lfloor \sqrt{r(i_3)} \right\rfloor \leq i_2$ and $\epsilon < \frac{1}{2}\lambda$. Define $\mathcal{E}_i$ to be the event that $D_j = o(r(j))$ and $\frac{I_j}{U_j} < 1 - 2\lambda$ for all j with $i_1 \leq j \leq i$. Then*

$$\mathbf{P}(L_{i_3 + \lfloor \sqrt{r(i_3)} \rfloor} < L_{i_3} | \mathcal{E}_{i_3} \cap \{\epsilon U_{i_3} < L_{i_3} < (\lambda - \tfrac{1}{2}\epsilon)U_{i_3}\}) = O(\exp(-r(i_3)^{1/3})).$$

We postpone the proof of this until after showing that the lemma follows from the claim.

Proof of Lemma 6. Note that from the outset we can assume that $\epsilon = \delta < \frac{1}{2}\lambda$. Using r to denote $r(i)$, we have by the assumptions of the lemma and by the Claim

$$\begin{aligned}
&\mathbf{P}\left(\exists_{i_1 \leq i \leq i_2 - \lfloor \sqrt{r} \rfloor} \max\{L_{i+\lfloor \sqrt{r} \rfloor}, \epsilon U_i\} < L_i < (\lambda - \tfrac{1}{2}\epsilon)U_i\right) \\
&\leq \mathbf{P}\left(\exists_{i_1 \leq i \leq i_2 - \lfloor \sqrt{r} \rfloor} \left\{\max\{L_{i+\lfloor \sqrt{r} \rfloor}, \epsilon U_i\} < L_i < (\lambda - \tfrac{1}{2}\epsilon)U_i\right\} \cap \mathcal{E}_i\right) + o(1)
\end{aligned}$$

$$
\begin{aligned}
&\leq o(1) + \sum_{i=i_1}^{i_2-\lfloor\sqrt{r}\rfloor} \mathbf{P}\left(L_{i+\lfloor\sqrt{r}\rfloor} < L_i | \mathcal{E}_i \cap \{\epsilon U_i < L_i < (\lambda - \tfrac{1}{2}\epsilon)U_i\}\right) \\
&= o(1) + \sum_{i=i_1}^{i_2-\lfloor\sqrt{r}\rfloor} O(\exp(-r(i)^{1/3})) \\
&= o(1).
\end{aligned}
$$

Thus we can assume that *a.a.s.*

$$
\epsilon U_i < L_i < (\lambda - \tfrac{1}{2}\epsilon)U_i \Rightarrow L_{i+\lfloor\sqrt{r(i)}\rfloor} \geq L_i \tag{5.1}
$$

for each such i.

Define $k_1 = i_1$, and for $j \geq 1$ define $k_{j+1} = k_j + \left\lfloor\sqrt{r(k_j)}\right\rfloor$. We first show by induction on j that *a.a.s.*

$$
L_{k_j} \geq \min\left\{(\lambda - \epsilon)U_{k_j} + 2\sqrt{r(k_j)}, L_{i_1}\right\} \tag{5.2}
$$

for $k_j \leq i_2$. Since L_i can fall by at most 2 in each step, and $k_{j+1} - k_j \leq \sqrt{r(k_j)} \leq \sqrt{r(i_1)}$, this implies the lemma.

To prove (5.2), observe that it is trivially true for $j = 1$, and assume that it is true for some $j > 1$ with $k_{j+1} \leq i_2$. We distinguish two cases. First, assume $L_{k_j} \geq (\lambda - \frac{1}{2}\epsilon)U_{k_j}$. Then again since L_i falls by at most 2 in each step, and since U_i is nonincreasing,

$$
\begin{aligned}
L_{k_{j+1}} &\geq (\lambda - \tfrac{1}{2}\epsilon)U_{k_j} - 2\sqrt{r(k_j)} \\
&\geq (\lambda - \epsilon)U_{k_j} + 2\sqrt{r(k_j)} \\
&\geq (\lambda - \epsilon)U_{k_{j+1}} + 2\sqrt{r(k_{j+1})}.
\end{aligned}
$$

On the other hand, if $L_{k_j} < (\lambda - \frac{1}{2}\epsilon)U_{k_j}$ then since $\lambda - \epsilon > \epsilon$, and since from the hypotheses of the lemma $L_{i_1} > \delta U_{i_1} = \epsilon U_{i_1} > \epsilon U_{k_j}$, we have from (5.2) that $L_{k_j} \geq \epsilon U_{k_j}$. Hence by (5.1), $L_{k_{j+1}} \geq L_{k_j}$. We conclude that (5.2) holds for k_{j+1}. ∎

Proof of Claim. Define $i_4 = i_3 + \left\lfloor\sqrt{r(i_3)}\right\rfloor$. In this proof we work entirely in the conditional space defined by the conditional probability appearing in the claim. Let C_i denote the component of largest active size in G_i (chosen by any canonical method if there is more than one largest) and consider the edge e added at time $i+1$. Define

$Z_{i+1} = -2$ if both endvertices of e belong to C_i,

$Z_{i+1} = -1$ if e joins C_i to a component in danger,

$Z_{i+1} = 0$ if e joins C_i to an isolate or is not incident with C_i,

$Z_{i+1} = 1$ otherwise, in which case e joins C_i to either an isolated edge or to a component different from C_i of active size at least 3.

Note that

$$\begin{aligned}
\mathbf{P}(Z_{i+1} = -2 | G_i) &\geq \frac{\binom{L_i}{2}}{\binom{U_i}{2} - O(U_i)}, \\
\mathbf{P}(Z_{i+1} = -1 | G_i) &= \frac{L_i D_i}{\binom{U_i}{2} - O(U_i)}, \\
\mathbf{P}(Z_{i+1} = 1 | G_i) &= \frac{L_i(U_i - L_i - I_i - D_i)}{\binom{U_i}{2} - O(U_i)}.
\end{aligned}$$

Hence

$$\mathbf{E}(Z_{i+1} | G_i) = \frac{L_i(U_i - 2L_i - I_i - 2D_i + 1)}{\binom{U_i}{2} - O(U_i)} > 4\frac{L_i}{U_i}\left(\lambda - \frac{L_i}{U_i} + o(1)\right) \tag{5.3}$$

as by the (conditional) assumption $\mathcal{E}_{i_3}$ of the Claim, since $r(i_4) \sim r(i_3)$, we have $I_i/U_i < 1 - 2\lambda + o(1)$ and $D_i = o(r(i))$. Clearly, $L_{i+1} - L_i \geq Z_{i+1}$.

We define a process to be *successful* at time j if $L_j \geq U_j(\lambda - \frac{1}{4}\epsilon)$ for some $i_3 \leq j \leq i$. Also define $\hat{Z}_i = Z_i$ for a process which is not successful at time i, and $\hat{Z}_i = 1$ otherwise. Then defining $Y_i = \sum_{k=i_3}^{i-1}(\hat{Z}_k - \frac{1}{2}\epsilon^2)$, we have

$$L_i \geq L_{i_3} + Y_i + (i - i_3)\frac{1}{2}\epsilon^2 \tag{5.4}$$

for $i = i_3, \ldots, i_4$ and n sufficiently large, where this is true in the case of successful processes because L_i can drop by at most 2 in each step and $i_4 - i_3 = o(U_{i_3})$.

We will apply Lemma 4 to $-Y_{i_3}, -Y_{i_3+1}, \ldots$. Noting the conditional assumption in the claim that $\epsilon U_{i_3} < L_{i_3}$, and recalling $\epsilon < \frac{1}{2}\lambda$, we get from (5.3), for a process which is not successful at time i,

$$\mathbf{E}(Z_{i+1}|\mathcal{H}_i) \geq \epsilon(\lambda - \tfrac{1}{4}\epsilon) - o(1) > \tfrac{1}{2}\epsilon^2$$

for n sufficiently large. From this it follows that $\mathbf{E}(\hat{Z}_{i+1}|\mathcal{H}_i) > \frac{1}{2}\epsilon^2$ in all cases, successful or not. Hence $-Y_{i_3}, -Y_{i_3+1}, \ldots, -Y_{i_4}$ is a supermartingale with respect to the sequence of σ-algebras $\{\sigma(\mathcal{H}_k)\}$, $k \geq i_3$, and so by Lemma 4,

$$\mathbf{P}\left(-Y_{i_4} \geq \frac{1}{2}\epsilon^2\sqrt{r_3}\right) = O(\exp(-r_3^{1/3})) \ ,$$

where, as usual, $r_3 = r(i_3)$. The claim follows from this and (5.4).

6 Proof of Theorem 2

Let us first summarise the knowledge about the d-process we have gained so far. Among other things, we proved that *a.a.s.* all isolated vertices and edges disappear by the r-time $\log n/\log\log n$ (Lemma 1), no dead component has been created by the r-time $\log^{1/8} n$ (Corollary 3), and there is at least one giant during the r-time interval from $\log^3 n$ down to $\log^{1/8} n$ (Corollary 4).

We will show that *a.a.s.* at r-time $\log^{1/8} n$ all unsaturated vertices belong to giants. This comes from examining two phases.

The first phase of the proof of Theorem 2 is from r-time $r_1 = \log n/\log\log n$ until r-time $r_2 = \log^{1/7} n$. We will show that all components of active size at least 3 at r-time r_1 will *a.a.s.* at some time during this interval either join a giant (that is, become contained in a giant) or be in danger (that is, become contained in a component in danger). The analysis is similar to the proof of (4.1).

This time we condition on the existence of at least one giant in the entire interval from r-time r_1 to r_2.

If at time i a component C is not in danger, then the probability that the next edge joins C to a giant is at least $3 \times \frac{2}{5}U_i/\binom{U_i}{2} \geq 6/5r(i)$ by Fact 1. Arguing as for (4.1), it follows that the probability of a fixed component, not in danger at r-time r_1, not joining a giant and not becoming in danger by r-time r_2 is

$$O(1)\left(\frac{r_2}{r_1}\right)^{6/5} = o\left(\frac{1}{r_1}\right).$$

As there are $O(r_1)$ components at r-time r_1, they all join a giant or become in danger by r-time r_2 *a.a.s.* Thus, if T denotes the set of components outside the giants at r-time r_1 which either are in danger or will be in danger sometime by r-time r_2, we have that *a.a.s.* all unsaturated vertices at r-time r_2, not belonging to a giant, were in components in T at r-time r_1. We next bound the number of these components, and consequently, the number of these vertices.

Let F_i denote twice the number of components in danger at time i of active size 1 plus the number of components in danger at time i of active size 2. Note that F_i can only increase either by D_i increasing or by an edge being inserted into a component in danger. By Lemma 2 and Markov's inequality this happens *a.a.s.* at most $O(\omega \log\log n)$ times after r-time r_1. On the other hand, by $\mathcal{D}_2$, *a.a.s.* $F_{r_1} = O(\omega \log\log n)$. Hence, since F_r is non-negative, *a.a.s.* F_r decreases at most $O(\omega \log\log n)$ times for $r_1 \geq r \geq r_2$. A component which is not in danger can only become in danger either by gaining an edge and increasing D_i, or by joining to a component in danger of active size 1 and creating a component in danger of active size 2, thereby decreasing F_i. Hence, this happens $O(\omega \log\log n)$, and so by $\mathcal{D}_2$, $|T| = O(\omega \log\log n)$.

Any vertex of a component in T not in danger which remains unsaturated until r-time r_2 must have been unsaturated when that component was in danger, at which time it had active size at most 2. Hence, denoting by S the set of unsaturated vertices outside giant components at r-time r_2, we have that *a.a.s.* $|S| \leq 2|T| = O(\omega \log\log n)$.

The second phase we consider is from r-time r_2 to r-time $r_3 = 2\log^{1/8} n$. This is similar to but simpler than the analysis of $\mathcal{D}_2$ in Section 4. We condition on $|S| = O(\omega \log\log n)$. The probability that some edge in this phase has both ends in S is

$$O((\omega \log\log n)^2) \sum_{r=r_3}^{r_2} \frac{1}{r^2} = o(1).$$

Also, the probability that some fixed vertex in S is incident with no edge added in this phase is at most

$$\prod_{r=r_3}^{r_2} \left(1 - \frac{1-o(1)}{r}\right) = o\left(\frac{1}{|S|}\right).$$

Thus *a.a.s.* every vertex in S is incident with some edge added in this phase, and hence *a.a.s.* every vertex in S gets joined to a giant during this phase.

We can now conclude that *a.a.s.* at r-time $r_3 = 2\log^{1/8} n$, all unsaturated vertices not in a giant at r-time r_2 have been joined to a component which *was giant at the time of joining.*

In fact it can be shown that a giant *a.a.s.* maintains its identity; that is, that for every i between $i(\log^3 n)$ and $i(\log^{1/8} n)$ there is a giant C_i, and, moreover, this can be chosen so that $C_i \subseteq C_{i+1}$. However, rather than relying on this type of argument, it is simpler to finish by showing that *a.a.s.* all vertices which at any r-time from r_2 to r_3 were in any giant component, lie together in the same giant component at some time after r_3 but before $r_3/2$. Then at this time the graph is *a.a.s.* already connected and Theorem 2 follows.

We can condition on the existence of a giant at every step, and so a giant can only become non-giant by shrinking in size by 2 in the presence of another component of active size asymptotically at least $\frac{2}{5}U_i$. Clearly if there are two such large components (which are both giants, but perhaps not quite simultaneously but only on consecutive steps) at any time up to r-time $r_3/2$, then the probability they do not join in any of the next $\log^{1/10} n$ steps is $O\left((1-\epsilon)^{\log^{1/10} n}\right) = o(\log^{-3} n)$ for any $\epsilon < 8/25$. Hence *a.a.s.* every time from r-time r_2 to r-time $r_3/2$ that a giant becomes non-giant, it joins within the next $\log^{1/10} n$ steps to another huge component to create a giant again. Whenever such a merge happens, there can only be one giant, and so by induction, *a.a.s.* inside every time-period of length $\log^{1/10} n$ from r-time r_2 to $r_3/2$, there will be a time when there is a unique giant which contains all vertices in all components which were in or joined to any giant component since r-time r_2. This is all that was required. ∎

References

[1] B. Bollobás, *Random graphs*, Academic Press, London, 1985.

[2] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* **6** (1959), 290–297.

[3] C. McDiarmid, On the method of bounded differences, *Surveys in Combinatorics 1989*, J. Siemons ed., Cambridge University Press, Cambridge (1989) pp. 148–188.

[4] A. Ruciński and N.C. Wormald, Random graph processes with degree restrictions, *Combin. Probab. Comput.* **1** (1992), 169–180.

[5] E. Shamir and J. Spencer, Sharp concentration of the chromatic number on random graphs $G_{n,p}$, *Combinatorica* **7** (1987), 121–129.

[6] A. Telcs and N.C. Wormald, Hamiltonicity of random 2-processes (in preparation).

[7] N.C. Wormald, Differential equations for random processes and random graphs, *Ann. Appl. Probab.* **5** (1995), 1217–1235.

Department of Discrete Mathematics
Faculty of Mathematics and Computer Science
Adam Mickiewicz University
email: rucinski@amu.edu.pl

Department of Mathematics and Statistics
University of Melbourne
email: nick@ms.unimelb.edu.au