Combinatorics, Probability and Computing (2013) 22, 133–145.
doi:10.1017/S0963548312000442

Explosive Percolation in Erdős–Rényi-Like Random Graph Processes

KONSTANTINOS PANAGIOTOU[1], RETO SPÖHEL[2]†,
ANGELIKA STEGER[3] and HENNING THOMAS[3]‡

[1]University of Munich, Mathematical Institute, Theresienstr. 39, 80333 Munich, Germany
(e-mail: kpanagio@math.lmu.de)
[2]Max Planck Institute for Informatics, 66123 Saarbrücken, Germany
(e-mail: rspoehel@mpi-inf.mpg.de)
[3]Institute of Theoretical Computer Science, ETH Zurich, 8092 Zurich, Switzerland
(e-mail: {steger,hthomas}@inf.ethz.ch)

Received 24 April 2012; revised 28 August 2012; first published online 3 October 2012

The study of the phase transition of random graph processes, and recently in particular Achlioptas processes, has attracted much attention. Achlioptas, D'Souza and Spencer (*Science*, 2009) gave strong numerical evidence that a variety of edge-selection rules in Achlioptas processes exhibit a discontinuous phase transition. However, Riordan and Warnke (*Science*, 2011) recently showed that all these processes have a continuous phase transition.

In this work we prove discontinuous phase transitions for three random graph processes: all three start with the empty graph on n vertices and, depending on the process, we connect in every step (i) one vertex chosen randomly from all vertices and one chosen randomly from a restricted set of vertices, (ii) two components chosen randomly from the set of all components, or (iii) a randomly chosen vertex and a randomly chosen component.

AMS 2010 *Mathematics subject classification*: Primary 03C90; 03-02

1. Introduction

In their seminal 1960 paper [10], Erdős and Rényi analysed the size of the largest component in the random graph model $G_{n,m}$, a graph drawn uniformly at random from all graphs on n vertices with m edges. For any graph G, let $L_1(G)$ denote the size of its largest component. We say that an event occurs *asymptotically almost surely* (a.a.s.) if it occurs with probability $1 - o(1)$ as n tends to infinity. Moreover, we omit floors and ceilings whenever they are not essential.

† Supported by a fellowship of the Swiss National Science Foundation.
‡ Supported by the Swiss National Science Foundation, grant 200021-120284.

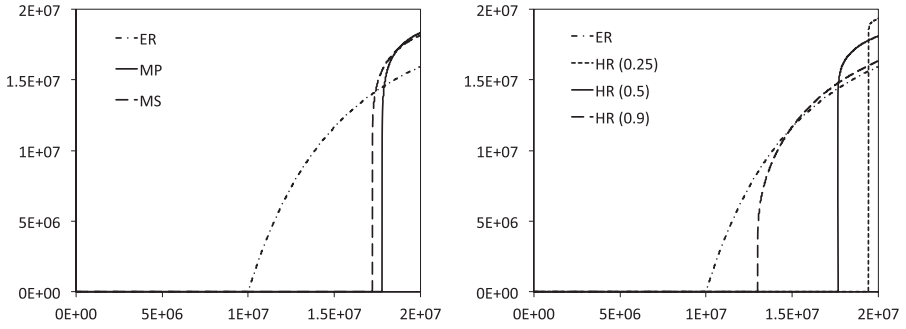

Figure 1. Evolution of the largest component over the first n edges of the Erdős–Rényi process, the min-product and min-sum rule and the half-restricted process with parameters 0.25, 0.5 and 0.9 on 20 million vertices.

Theorem 1.1 ([10]). *For any constant $c > 0$ the following holds.*

- *If $c < 0.5$, then a.a.s. $L_1(G_{n,cn}) = O(\log n)$.*
- *If $c > 0.5$, then a.a.s. $L_1(G_{n,cn}) = \Omega(n)$, and all other components have $O(\log n)$ vertices.*

This result can be viewed from a random graph *process* perspective. Starting with the empty graph on the vertex set $[n] := \{1, 2, \ldots, n\}$, we add a single edge chosen uniformly at random from all non-edges in every step. It is not hard to see that the graph after inserting m edges is distributed as $G_{n,m}$. In this context Theorem 1.1 states that, asymptotically, a linear-sized component (a so-called 'giant component') appears around the time when we have inserted $n/2$ edges. We say that the Erdős–Rényi process has a *phase transition* at $n/2$.

This phase transition has been studied in great detail (for a survey see Chapter 3 in [12]). It is known that $L_1(G_{n,cn})$ a.a.s. satisfies $L_1(G_{n,cn}) = (f(c) + o(1))n$ for some continuous function $f(c)$ with $f(c) = 0$ for every $c < 0.5$ and $\lim_{c \to 0.5^+} f(c) = 0$. Thus, the phase transition in the Erdős–Rényi process is continuous[1] (see Figure 1).

A class of variants of the Erdős–Rényi process which has gained much attention over the last decade consists of the so-called *Achlioptas processes*. These processes start with the empty graph on the vertex set $[n]$. In every step a pair of edges is chosen uniformly at random from all pairs of non-edges in the current graph, and a fixed edge-selection rule selects exactly one of them to be inserted into the graph while the other is put back into the pool of non-edges. Most papers concerned the question of whether one can delay or accelerate the appearance of the giant component in these processes: see, *e.g.*, [3] and [17].

In a paper in *Science* from 2009 [1], Achlioptas, D'Souza and Spencer provided strong numerical evidence that the min-product rule (select the edge that minimizes the product of the component sizes of the endpoints) and min-sum rule (select the edge that minimizes the respective sum) exhibit discontinuous phase transitions (see Figure 1), in contrast to a variety of closely related edge-selection rules, in particular the ones analysed in [17]. A discontinuous phase transition essentially means that for some constant $d > 0$ the size of the largest component

[1] In the literature such a phase transition is also called *second-order* while discontinuous phase transitions are called *first-order*.

a.a.s. 'jumps' from $o(n)$ to dn within $o(n)$ edge insertions, that is, at the phase transition a constant fraction of the vertices is accumulated into a single giant component within a *sublinear* number of steps. This phenomenon is also called *explosive percolation*. Since it is of great interest, in particular in physics, a series of papers has been devoted to understanding the phase transition of the min-product and min-sum rule (see, *e.g.*, [4], [5], [8] and [14]), most of the arguments not being rigorous but supported by computer simulations.

Countering the numerical evidence, it was claimed in [6] that the transition is actually continuous. In a recent *Science* paper [16] (see also [15]) Riordan and Warnke indeed confirmed this claim with a rigorous proof. In fact, their argument shows continuous phase transitions for an even larger class of processes. Loosely speaking, a random graph process $(G(t))_{t\geqslant 0}$ is called an *ℓ-vertex rule* if $G(0)$ is empty, and $G(t)$ is obtained from $G(t-1)$ by drawing a set V_t of ℓ vertices uniformly at random and adding one or more edges within V_t. Informally, this states that graph processes that operate locally in every step (*i.e.*, on a vertex set of constant size) cannot exhibit explosive percolation. Note that Achlioptas processes are essentially 4-vertex rules.

In this paper we study the characteristics of random graph processes that *can* cause a discontinuous phase transition. Before stating our results in the forthcoming sections, let us give an informal picture of processes that exhibit explosive percolation. Note that a process which inserts one edge per step, can only have a discontinuous transition if at some point in time the components with size $\omega(1)$ and $o(n)$ occupy a constant fraction of the vertices (see, *e.g.*, [15]). Employing the terminology of [11], these components form a type of 'powder keg' that 'explodes' at the phase transition. Intuitively speaking, such a powder keg is formed if the process keeps the sizes of the largest components close together and prevents a single component from growing too large before the phase transition occurs. To some extent, the min-product and min-sum rule try to approximate this effect by favouring the construction of smaller components. However, as discussed above, this does not lead to a discontinuous phase transition. In this paper we consider different approaches to forming a powder-keg.

2. Our results

We introduce three variants of the Erdős–Rényi process and prove discontinuous phase transitions for all of them.

2.1. The half-restricted process

First, we introduce a graph process which we call the *half-restricted process*. The idea is to connect two vertices in every step, but to restrict one of them to be within smaller components.

The half-restricted process has a parameter $0 < \beta \leqslant 1$ and starts with the empty graph $\mathcal{H}_\beta(0) = \mathcal{H}_{n,\beta}(0)$ on the vertex set $[n]$. In every step $t \geqslant 1$ we draw an *unrestricted* vertex $u \in [n]$ uniformly at random and, independently, a *restricted* vertex v which is drawn uniformly at random from a restricted vertex set $R_\beta(\mathcal{H}_\beta(t-1))$ defined as follows. For every graph G we denote by $R_\beta(G)$ the $\lfloor \beta n \rfloor$ vertices in smallest components. Precisely, let $v_1, v_2, \ldots, v_n$ be the vertices of G sorted in increasing order according to the sizes of the components they are contained in (where vertices with the same component size are sorted according to their labels). Then $R_\beta(G) := \{v_1, v_2, \ldots, v_{\lfloor \beta n \rfloor}\}$. We obtain $\mathcal{H}_\beta(t)$ from $\mathcal{H}_\beta(t-1)$ by inserting an edge between

u and v if $u \neq v$ and the edge is not already present (otherwise we do nothing). Note that for $\beta < 1$ the half-restricted process is not an ℓ-vertex rule.

For a half-restricted process let α_t be the random variable that denotes the maximum size of all components that the restricted vertex can be drawn from in step t. Hence, α_t denotes the size of the component of $v_{\lfloor \beta n \rfloor}$ for an ordering $v_1, v_2, \ldots, v_n$ of the vertices according to their component sizes in $\mathcal{H}_\beta(t-1)$. Clearly, α_t is increasing in t. For every positive integer k let T_k denote the last step in which α_t is still below k, that is,

$$T_k := \max\{t : \alpha_t < k\}.$$

We show that for any parameter $\beta < 1$ the half-restricted process exhibits explosive percolation. Thus, even though Figure 1 suggests that the phase transitions of the min-product (or min-sum) rule and that of the half-restricted process behave similarly, their mathematical structure is fundamentally different. More precisely, we show that around the step in the half-restricted process when the number of vertices in components of constant size drops below βn, a giant component of size almost $(1-\beta)n$ is created within a sublinear number of steps.

Theorem 2.1. *Let $0 < \beta < 1$. For every $K = K(n)$ with $K \geqslant (\ln n)^{1.02}$, every $C = C(n)$ with $C = \omega(1)$ and $C = o(\ln K)$, and every $\varepsilon > 0$ there exists a constant $c = c(\beta, \varepsilon)$ such that a.a.s.*

(i) $L_1(\mathcal{H}_\beta(T_C)) \leqslant K$, *and*
(ii) $L_1(\mathcal{H}_\beta(T_C + n/c\sqrt{C})) \geqslant (1-\varepsilon)(1-\beta)n$.

Note that for $K = (\ln n)^{1.02}$ and $C = (\ln \ln n)^{0.99}$ the theorem states that, a.a.s., the number of steps from the first appearance of a component of size $(\ln n)^{1.02}$ to the appearance of a component of size $(1-\varepsilon)(1-\beta)n$ is $O(n/(\ln \ln n)^{0.49}) = o(n)$.

2.2. The component process

In this subsection we introduce a graph process which we call the *component process*. We start with the empty graph $\mathcal{C}(0)$. In the tth step (for every integer $t \geqslant 1$) we obtain the graph $\mathcal{C}(t)$ from $\mathcal{C}(t-1)$ by drawing a pair of components uniformly at random from the set of all components in $\mathcal{C}(t-1)$, and inserting an arbitrary edge between them. In this way, after $n-1$ steps we obtain a tree $\mathcal{C}(n-1)$ which connects all n vertices, and the process then stops. This process is closely related to what is called, in the physics literature, Smoluchowski's coagulation equation with constant kernel: see, *e.g.*, [2], [7] and [13].

Theorem 2.2. *For every $\varepsilon > 0$ there exists a constant $K = K(\varepsilon) > 0$ such that a.a.s. $L_1(\mathcal{C}((1-\varepsilon)n)) \leqslant K \ln n$.*

Observe that the graph of the component process is connected after $n-1$ steps. Hence, Theorem 2.2 implies that the component process exhibits a discontinuous phase transition.

2.3. The mixed process

We also consider a mixture of the Erdős–Rényi and the component processes, which we call the *mixed process*. In this process, we start with the empty graph $\mathcal{M}(0)$ and in the tth step (for every integer $t \geqslant 1$) we obtain the graph $\mathcal{M}(t)$ from $\mathcal{M}(t-1)$ as follows. We draw a vertex v

uniformly at random from the set of all vertices and a component uniformly at random from the set of all components in $\mathcal{M}(t-1)$ except the one that v is contained in. We then insert an arbitrary edge between them.

Theorem 2.3. *For every $\varepsilon > 0$ there exists a constant $K = K(\varepsilon) > 0$ such that a.a.s. $L_1(\mathcal{M}((1-\varepsilon)n)) \leqslant K \ln n$.*

Observe that the graph of the mixed process is connected after $n-1$ steps. Hence, Theorem 2.3 implies that the mixed process exhibits a discontinuous phase transition.

3. Proof for the half-restricted process

In this section we prove Theorem 2.1. In our proof we need a technical lemma introduced in the following.

Let N be a positive integer, and for every $0 \leqslant i < N$ let $X_i \sim \mathrm{Geom}(\frac{N-i}{N})$ be a geometrically distributed random variable with parameter $\frac{N-i}{N}$ and set $X(a,b) := \sum_{i=a}^{b} X_i$ for every $0 \leqslant a < b < N$. All subsequent statements and arguments are about sums of geometrically distributed random variables, but we will use that they have the following combinatorial interpretation in a coupon collector scenario with N coupons. (In a coupon collector scenario, we have a number of different coupons and repeatedly draw one uniformly at random with replacement. We are interested in how often we have to draw until we have seen every coupon.) Observe that X_i is distributed like the number of coupons we draw while holding exactly i different coupons, waiting for the $(i+1)$st, and thus $X(a,b)$ can be viewed as the number of coupons we draw while holding at least a and at most b different coupons.

Note that

$$\mathbb{E}[X(a,b)] = \sum_{i=a}^{b} \frac{N}{N-i} = N \sum_{i=N-b}^{N-a} \frac{1}{i} = N(H_{N-a} - H_{N-b-1}), \tag{3.1}$$

where $H_n = \sum_{i=1}^{n} 1/i$ denotes the nth harmonic number for every $n \geqslant 1$. It is well known that

$$H_n = (1 + o(1)) \ln n. \tag{3.2}$$

Lemma 3.1. *Let $k = k(N) = \omega(1)$ and $s = s(N)$ with $s = o(N \ln k)$. Then, for N large enough,*

$$\mathbb{P}[X(N-k, N-2) \leqslant s] \leqslant \mathrm{e}^{-k^{0.99}}.$$

Proof. First note that by using (3.1), (3.2) and $k = \omega(1)$ we have

$$\mathbb{E}[X(N-k, N-2)] = N(H_k - H_1) = (1 + o(1))N \ln k.$$

Thus, $s = o(\mathbb{E}[X(N-k, N-2)])$. Consider a coupon collector scenario with N different coupons of which we have already seen $N-k$. For each of the remaining k coupons let Y_j (where $1 \leqslant j \leqslant k$) denote the indicator random variable for the event that it is not drawn within the next s trials. By the comments preceding this lemma, the probability of the event $X(N-k, N-2) \leqslant s$

equals the probability that at most one of these coupons is not drawn within the s trials. Hence, for $Y := \sum_{j=1}^{k} Y_j$ we observe that

$$\mathbb{P}[X(N-k, N-2) \leqslant s] = \mathbb{P}[Y \leqslant 1]. \tag{3.3}$$

Using the identity $1 - x \geqslant \mathrm{e}^{-2x}$, which holds for all $0 \leqslant x \leqslant 1/2$, we have for every $1 \leqslant j \leqslant k$ that

$$\mathbb{E}[Y_j] = \mathbb{P}[Y_j = 1] = \left(1 - \frac{1}{N}\right)^s \geqslant \mathrm{e}^{-\frac{2s}{N}},$$

and thus, for N large enough,

$$\mathbb{E}[Y] \geqslant k\mathrm{e}^{-\frac{2s}{N}} = k^{1-\frac{2s}{N\ln k}} \geqslant 8k^{0.99},$$

where we used $s = o(N \ln k)$ in the last step. One can check that the random variables $Y_1, \ldots, Y_k$ are negatively associated (see, *e.g.*, Chapter 3 in [9]). Hence, we can apply Chernoff bounds to Y and obtain for N large enough that

$$\mathbb{P}[Y \leqslant 1] \leqslant \mathbb{P}\left[Y \leqslant \left(1 - \frac{1}{2}\right)\mathbb{E}[Y]\right] \leqslant \mathrm{e}^{-\mathbb{E}[Y]/8} \leqslant \mathrm{e}^{-k^{0.99}},$$

which together with (3.3) finishes the proof. □

We now turn to the proof of Theorem 2.1.

Proof for Theorem 2.1. We fix $\beta < 1, K = K(n)$ with $K \geqslant (\ln n)^{1.02}$, $C = C(n)$ with $C = \omega(1)$ and $C = o(\ln K)$, and $\varepsilon > 0$. To simplify notation we write $\mathcal{H}(t)$ instead of $\mathcal{H}_\beta(t)$.

(i) We first address (i). We need to show that at the step when the restricted vertex can be in a component of size C for the first time, there is a.a.s. no component of size larger than K. The main idea is that a large component can only form if the unrestricted vertex is drawn from this component so often that this is unlikely to happen within T_C steps.

We first note that it is not hard to show that $T_C \leqslant 4n$ asymptotically almost surely (see, *e.g.*, Lemma 3 and the remark following its proof in [15]). Let $\mathcal{E}^*$ denote this event.

Note that up to step T_C two components of size at least C can never be merged by an edge since the restricted vertex in every step is drawn from vertices in components of size less than C. Hence, we can easily keep track of the components of size at least C and call them *chunks*. Let $A_1, A_2, \ldots$ denote all chunks in order of appearance during the process, where a chunk keeps its label if merged with another component, and the new component is *not* inserted into the list. Clearly, there can be at most n/C chunks. For every $1 \leqslant i \leqslant n/C$ we denote by $\mathcal{E}_i$ the event that chunk A_i has size larger than K in $\mathcal{H}(T_C)$. We will show that $\mathbb{P}[\mathcal{E}_i \cap \mathcal{E}^*] \leqslant 1/n$ for every $1 \leqslant i \leqslant n$. By applying a union bound this implies

$$\begin{aligned}\mathbb{P}[\mathcal{H}(T_C) \text{ contains a component of size } > K] &= \mathbb{P}\left[\bigcup_{i=1}^{n/C} \mathcal{E}_i \cap \mathcal{E}^*\right] + \mathbb{P}[\bar{\mathcal{E}}^*] \\ &\leqslant \frac{n}{C} \cdot \frac{1}{n} + o(1) = o(1). \end{aligned} \tag{3.4}$$

It remains to bound $\mathbb{P}[\mathcal{E}_i \cap \mathcal{E}^*]$ for every $1 \leqslant i \leqslant n/C$. Let $1 \leqslant i \leqslant n/C$ be fixed for the remainder of the proof. Clearly, a chunk has size at most $2C$ when it appears. Moreover, since for every $t \leqslant T_C$ the restricted vertex is drawn from vertices in components of size smaller than C, the chunk A_i can grow by at most C in every step. Hence, the chunk has size at most $(j+1) \cdot C$ before a vertex from the chunk is drawn for the jth time. Hence, a vertex from chunk A_i needs to be drawn in at least $K/C - 1$ steps after its appearance for $\mathcal{E}_i$ to occur. Let $X_2, X_3, \ldots, X_{K/C}$ denote the number of steps between steps in which we draw a vertex from A_i. That is, X_2 is the number of steps from the appearance of chunk A_i until a vertex from A_i is drawn for the first time, X_3 is the time from that step until a vertex from A_i is drawn for the second time, and so on. Furthermore, let $X := \sum_{j=2}^{K/C} X_j$. Then,

$$\mathbb{P}[\mathcal{E}_i \cap \mathcal{E}^*] \leqslant \mathbb{P}[\{X \leqslant T_C\} \cap \mathcal{E}^*]. \tag{3.5}$$

Recall that for time steps $t \leqslant T_C$ only the unrestricted vertex in every step can be in A_i. In the period $t \leqslant T_C$ we thus have that X_j conditioned on $X_1, X_2, \ldots, X_{j-1}$ is geometrically distributed with parameter at most jC/n. Therefore we can bound $\mathbb{P}[\{X \leqslant T_C\} \cap \mathcal{E}^*]$ as follows. Let Y_j be an independent random variable with $Y_j \sim \mathrm{Geom}(jC/n)$, and set $Y := \sum_{j=2}^{K/C} Y_j$. Then

$$\mathbb{P}[\{X \leqslant T_C\} \cap \mathcal{E}^*] \leqslant \mathbb{P}[\{Y \leqslant T_C\} \cap \mathcal{E}^*] \leqslant \mathbb{P}[Y \leqslant 4n]. \tag{3.6}$$

Here, we use Lemma 3.1 with $N = n/C$, $k = K/C$ and $s = 4n$. Concerning the prerequisites of the lemma, we have $k = \omega(1)$ since $C = o(K)$. Furthermore, we have $s = o(N \ln k)$ since $C = o(\ln K)$, and thus

$$N \ln k = \frac{n}{C} \cdot \ln\left(\frac{K}{C}\right) = (1 - o(1))\frac{n \ln K}{C} = \omega(n).$$

Hence, Lemma 3.1 gives us for large enough n that

$$\mathbb{P}[Y \leqslant 4n] \leqslant \mathrm{e}^{-(K/C)^{0.99}}.$$

Since $K \geqslant (\ln n)^{1.02}$ and $C = o(\ln K)$, we have

$$\mathbb{P}[Y \leqslant T_C] \leqslant \mathrm{e}^{-\ln n} = \frac{1}{n}.$$

Using (3.4), (3.5) and (3.6) this settles the proof of (i), and it remains to prove (ii), *i.e.*, we have to show that for an appropriate constant $c = c(\beta, \varepsilon)$, which we will specify later, we have that $\mathcal{H}(T_C + n/c\sqrt{C})$ contains a component of size $(1-\varepsilon)(1-\beta)n$ with high probability.

(ii) We set $a := n/(2c\sqrt{C})$ and split the proof into two parts. In the first part (the first a additional steps after T_C) we collect a suitable amount of vertices in components of size at least C, and in the second part (the remaining a steps) we actually build a giant component on these vertices.

Consider steps $T_C + 1$ to $T_C + a$ in the graph process. For every $t \geqslant 1$ let $U(t)$ denote the set of vertices in components of size at least C in $\mathcal{H}(t)$. Note that by definition of T_C we have $|U(T_C + 1)| \geqslant (1-\beta)n$. We now show that with high probability we have

$$|U(T_C + a)| \geqslant (1-\beta)n + \frac{1-\beta}{8c\sqrt{C}}n. \tag{3.7}$$

Clearly, this holds if $|U(T_C + a)| \geqslant (1 - \beta/2)n$ and we thus assume $|U(T_C + i)| \leqslant (1 - \beta/2)n$ for every $1 \leqslant i \leqslant a$ in the remainder.

For every $1 \leqslant i \leqslant a$ let $X_i := |U(T_C + i)| - |U(T_C + i - 1)|$ denote the number of vertices added to the components of size at least C in the ith additional step. Furthermore, let $X := \sum_{i=1}^{a} X_i$. We prove a lower bound on the probability that X_i contributes at least 1 vertex. Clearly, $X_i \geqslant 1$ if the unrestricted vertex is drawn from components of size at least C, which by definition of T_C happens with probability at least $1 - \beta$, and if the restricted vertex is drawn from components of size smaller than C, which happens with probability at least

$$\frac{(\beta/2)n}{\beta n} = 1/2$$

since we assume $|U(T_C + i)| \leqslant (1 - \beta/2)n$. Hence, for every $1 \leqslant i \leqslant a$ we have

$$\mathbb{P}[X_i \geqslant 1] \geqslant (1 - \beta)/2$$

and thus also

$$\mathbb{E}[X] \geqslant a(1 - \beta)/2 = (1 - \beta)n/(4c\sqrt{C}).$$

Using Chernoff bounds we obtain

$$\mathbb{P}\left[X < \frac{1}{2} \cdot \frac{(1 - \beta)n}{4c\sqrt{C}}\right] \leqslant \mathrm{e}^{-\Theta(\frac{n}{\sqrt{C}})} = o(1),$$

which establishes (3.7).

We now look at the second half of the additional steps, *i.e.*, steps $T_C + a + 1$ to $T_C + 2a$, and show that a.a.s. in these steps a sufficiently large component is created within $U := U(T_C + a)$. Note that U is a fixed set of vertices which does not change from step to step.

Let $\mathcal{E}$ denote the event that $\mathcal{H}(T_C + 2a)$ has no component of size $(1 - \varepsilon)(1 - \beta)n$ conditioned on (3.7). We now show that $\mathbb{P}[\mathcal{E}] = o(1)$, which clearly finishes the proof. Observe that $\mathcal{E}$ can only occur if, for every $a + 1 \leqslant i \leqslant 2a$, we have that the largest component in $\mathcal{H}(T_C + i)$ has size less than $(1 - \varepsilon)(1 - \beta)n$. In the following we bound the probability of this occurring.

We call a step *successful* if it connects two components in U. Since every component in U has size at least C, we have that n/C successful steps will connect all components in U such that U forms one giant component of size at least $(1 - \beta)n$. We now compute the probability of having a successful step conditioned on not having a component of size $(1 - \varepsilon)(1 - \beta)n$. For a successful step, the restricted vertex needs to be in U and the unrestricted vertex needs to be drawn from a different component in U. This happens with probability at least

$$\frac{|U| - (1 - \beta)n}{\beta n} \cdot \frac{|U| - (1 - \varepsilon)(1 - \beta)n}{n} \geqslant \frac{1 - \beta}{8\beta c\sqrt{C}} \cdot \varepsilon(1 - \beta) =: p, \tag{3.8}$$

where we used (3.7) to bound the first factor. Thus, setting $c := (1 - \beta)\sqrt{\varepsilon/(32\beta)}$, the number S of successful steps satisfies

$$\mathbb{E}[S] \geqslant ap = \frac{n}{2c\sqrt{C}} \cdot \frac{\varepsilon(1 - \beta)^2}{8\beta c\sqrt{C}} \geqslant \frac{2n}{C},$$

and by Chernoff bounds

$$\mathbb{P}\left[S \leqslant \frac{n}{C}\right] \leqslant \mathbb{P}\left[S \leqslant \frac{1}{2}\mathbb{E}[S]\right] \leqslant \mathrm{e}^{-\mathbb{E}[S]/8} = \mathrm{e}^{-\Omega(\frac{n}{C})} = o(1),$$

Hence, we a.a.s. have at least n/C successful steps conditioned on that we do not have a component of size $(1-\varepsilon)(1-\beta)n$. Together with the considerations before (3.8) this implies that $\mathbb{P}[\mathcal{E}] = o(1)$. □

4. Proof for the component process

In this section we prove Theorem 2.2. Our proof relies on the following lemma, which bounds the probability that the component process contains a component of size exactly k after exactly t steps.

Lemma 4.1. *For every $\varepsilon > 0$, $k \geqslant 1$ and $1 \leqslant t \leqslant (1-\varepsilon)n$, the probability that $\mathcal{C}(t)$ contains a component of size exactly k is for n large enough at most*

$$n\mathrm{e}^{-(k-1)\varepsilon+13k^2/(\varepsilon n)}.$$

Proof. Let $\varepsilon > 0$, let $k \geqslant 1$ be fixed and let $1 \leqslant t \leqslant (1-\varepsilon)n$. We use a union bound argument. We first count the number of possible ways to create a component of size k within t steps. First, there are $\binom{n}{k} \leqslant \frac{n^k}{k!}$ ways to choose the vertices of the component. Let us fix such a set S of k vertices. We now estimate the number of ways that a component on S can be created. Recall that at any time of the process the graph generated by the component process is a forest. In particular, if S is a component after t steps then S forms a tree, and there are exactly $k-1$ edges within S. In the beginning, S consists of k separate singleton components and there are $\binom{k}{2}$ ways to connect two of them. In general, after inserting i edges within S, there are $k-i$ components in S and thus $\binom{k-i}{2}$ ways to connect two of them in the component process. Hence, the number of ways to create a component on S is

$$\binom{k}{2}\binom{k-1}{2}\cdots\binom{2}{2} = \frac{k!(k-1)!}{2^{k-1}}.$$

It now remains to choose the steps $1 \leqslant t_1 < t_2 < \cdots < t_{k-1} \leqslant t$ in which we insert an edge in S. Observe that this can be done in $\binom{t}{k-1} \leqslant \frac{t^{k-1}}{(k-1)!}$ ways. Altogether, the number of ways to create a component of size k is at most

$$\frac{n^k}{k!} \cdot \frac{k!(k-1)!}{2^{k-1}} \cdot \frac{t^{k-1}}{(k-1)!} = n^k 2^{-(k-1)} t^{k-1}. \tag{4.1}$$

Having fixed S and the order and steps in which we connect components in S, let us estimate the probability of the event that S forms a component in this way. Set $t_0 = 0$ and $t_k = t+1$. For every $1 \leqslant i \leqslant k-1$ the probability of choosing the fixed component pair in step t_i is $1/\binom{n-t_i+1}{2}$, since there are $n - t_i + 1$ components to choose from in step t_i. Moreover, in every step $j \notin \{t_1, \ldots, t_{k-1}\}$ two components outside S need to be chosen, which for every $t_i < j < t_{i+1}$ happens with probability

$$\begin{aligned}\binom{n-j+1-(k-i)}{2} \Big/ \binom{n-j+1}{2} &\leqslant (1-(k-i)/(n-j+1))^2 \\ &\leqslant \exp(-2(k-i)/(n-j+1)).\end{aligned}$$

Hence, the probability that between steps t_i and t_{i+1} only components outside S are chosen is bounded by

$$\exp\bigl(-2(k-i)\cdot(H_{n-t_i}-H_{n-t_{i+1}+1})\bigr).$$

Altogether, we obtain that the probability of creating a component on S in a fixed way is at most

$$\begin{aligned}\left(\prod_{i=1}^{k-1}\frac{1}{\binom{n-t_i}{2}}\right)\cdot\left(\prod_{i=0}^{k-1}\mathrm{e}^{-2(k-i)\cdot(H_{n-t_i}-H_{n-t_{i+1}+1})}\right)\\ \leqslant 2^{k-1}\exp\Bigg(-2\sum_{i=1}^{k-1}\ln(n-t_i-1)\\ -2\sum_{i=0}^{k-1}(k-i)(H_{n-t_i-1}-H_{n-t_{i+1}-2})+\frac{6k^2}{n-t_k-1}\Bigg),\end{aligned}\tag{4.2}$$

where the term $6k^2/(n-t_k-1)$ accounts for replacing $H_{n-t_{i+1}+1}$ by $H_{n-t_{i+1}-2}$. It is easy to see that

$$H_a-H_b\geqslant\ln\left(\frac{a}{b+1}\right)\qquad\text{for all } a,b\in\mathbb{N}, a>b.$$

Thus, using $t_0=0$ we deduce for the second sum in (4.2) that

$$\begin{aligned}-2\sum_{i=0}^{k-1}(k-i)(H_{n-t_i-1}-H_{n-t_{i+1}-2})\\ \leqslant -2\sum_{i=0}^{k-1}(k-i)\cdot(\ln(n-t_i-1)-\ln(n-t_{i+1}-1))\\ =-2k\ln(n-1)+2\sum_{i=1}^{k-1}\ln(n-t_i-1)+2\ln(n-t_k-1).\end{aligned}\tag{4.3}$$

Using $n-t_k-1=n-t-2\geqslant\varepsilon n/2$ for n large enough, we thus obtain an upper bound for (4.2) of

$$2^{k-1}\exp\left(-2k\ln(n-1)+\frac{12k^2}{\varepsilon n}+2\ln(\varepsilon n/2)\right)\leqslant 2^{k-1}n^{-2(k-1)}\mathrm{e}^{13k^2/(\varepsilon n)}.\tag{4.4}$$

Putting together (4.1) and (4.4) and using $t\leqslant(1-\varepsilon)n$, we obtain that the probability that $\mathcal{C}(t)$ contains a component of size exactly k is at most

$$n^k 2^{-(k-1)}t^{k-1}\cdot 2^{k-1}n^{-2(k-1)}\mathrm{e}^{13k^2/(\varepsilon n)}\leqslant n\mathrm{e}^{-(k-1)\varepsilon+13k^2/(\varepsilon n)}.$$

□

Proof of Theorem 2.2. We use a union bound argument. Let $\varepsilon>0$ and $K=3/\varepsilon$. Clearly, the component process cannot contain a component of size larger than $K\ln n$ after exactly $(1-\varepsilon)n$ steps if it does not contain a component of size between $K\ln n+1$ and $2K\ln n$ at any time $t\in\{1,2,\dots,(1-\varepsilon)n\}$. For every $k\geqslant 1$ and $1\leqslant t\leqslant(1-\varepsilon)n$, let $\mathcal{E}(k,t)$ denote the event that the component process contains a component of size exactly k after exactly t steps. Then we

have by Lemma 4.1 that

$$\mathbb{P}\left[\bigcup_{\substack{K\ln n+1\leqslant k\leqslant 2K\ln n\\ 1\leqslant t\leqslant(1-\varepsilon)n}}\mathcal{E}(k,t)\right]\leqslant\sum_{k=K\ln n+1}^{2K\ln n}\sum_{t=1}^{(1-\varepsilon)n}n\mathrm{e}^{-(k-1)\varepsilon+6k^2/(\varepsilon n)}$$
$$\leqslant(1-\varepsilon)K\ln n\cdot n^{2-K\varepsilon+o(1)}=o(1)\,.\qquad\square$$

5. Proof for the mixed process

In this section we prove Theorem 2.3. The proof proceeds similarly to the proof of the component process (see Section 4). It relies on a lemma which bounds the probability that the mixed process contains a component of size exactly k after exactly t steps.

Lemma 5.1. *For every $\varepsilon>0$, $k\geqslant 1$ and $1\leqslant t\leqslant(1-\varepsilon)n$ the probability that $\mathcal{M}(t)$ contains a component of size exactly k is for n large enough at most*

$$n\mathrm{e}^{1-(k-1)\varepsilon^2/2+13k^2/(\varepsilon n)}.$$

Proof. Let $\varepsilon>0$, let $k\geqslant 1$ be fixed and let $1\leqslant t\leqslant(1-\varepsilon)n$. Similar to the component process (see Lemma 4.1) we apply a union bound argument. We first count the number of possible ways to create a component of size k within t steps. As for the component process there are $\binom{n}{k}\leqslant\frac{n^k}{k!}$ ways to choose the vertex set S of such a component, and $\binom{t}{k-1}\leqslant\frac{t^{k-1}}{(k-1)!}$ ways to choose the steps $1\leqslant t_1<t_2<\cdots<t_{k-1}\leqslant t$ in which we connect two components in S. (Note that the mixed process also satisfies that for every $i\geqslant 0$ the graph $\mathcal{M}(i)$ is a forest.)

Observe that just before step t_i the set S consists of $k-(i-1)$ components, two of which are connected in step t_i. Now, for the choice of the vertex in step t_i we have k possibilities and for the choice of the component we have $k-(i-1)-1=k-i$ possibilities (all components in S are fine except the one that the chosen vertex is contained in). Hence, after fixing the set S and the steps $t_1,\ldots,t_{k-1}$, there are

$$k(k-1)\cdot k(k-2)\cdot\ldots\cdot k\cdot 1=k^{k-1}(k-1)!$$

ways to create a component on S. Using $k!\geqslant(k/\mathrm{e})^k$, this yields altogether at most

$$\frac{n^k}{k!}k^{k-1}(k-1)!\frac{t^{k-1}}{(k-1)!}\leqslant(n\mathrm{e})^k\mathrm{e}^{(k-1)\ln t}\tag{5.1}$$

ways to create a component of size exactly k in exactly t steps.

Having fixed S and the steps and way in which we connect components in S, let us estimate the probability of this event. Set $t_0=0$ and $t_k=t+1$. First observe that before step j the graph has exactly $n-j+1$ components, and the component drawn in that step is drawn from a set of $n-j$ components, namely all $n-j+1$ components of the graph except the one that contains the vertex drawn in that step. Thus, for every $1\leqslant i\leqslant k-1$ the probability of choosing the fixed vertex and component in step t_i is $1/n\cdot 1/(n-t_i)$. Moreover, in every step $t_i<j<t_{i+1}$ we need to choose a vertex outside S together with a component outside S. This happens with probability

$(1-\frac{k}{n})\cdot(1-\frac{k-i}{n-j})$. Altogether, we obtain (similarly to (4.2)) that the probability of creating a component on S in a fixed way is at most

$$\begin{aligned}\prod_{i=1}^{k-1}\left(\frac{1}{n}\cdot\frac{1}{n-t_i}\right)\cdot\prod_{i=0}^{k-1}\prod_{j=t_i+1}^{t_{i+1}-1}\left(1-\frac{k}{n}\right)\left(1-\frac{k-i}{n-j}\right)\\ \leqslant n^{-(k-1)}\exp\Bigg(-\sum_{i=1}^{k-1}\ln(n-t_i)-\frac{k}{n}(t-(k-1))\\ -\sum_{i=0}^{k-1}(k-i)(H_{n-t_i-1}-H_{n-t_{i+1}-2})+\frac{2k^2}{n-t_k-1}\Bigg).\end{aligned}\tag{5.2}$$

Estimating the last sum in the exponent similar to (4.2) in the previous section (see (4.3) – the only difference is a missing factor of 2), we obtain that (5.2) is at most

$$\begin{aligned}&n^{-(k-1)}\exp\left(-\frac{k}{n}(t-(k-1))-k\ln(n-1)+\ln(\varepsilon n/2)+\frac{4k^2}{\varepsilon n}\right)\\ &\leqslant n^{-(k-1)}\mathrm{e}^{-kt/n-(k-1)\ln n+6k^2/(\varepsilon n)}.\end{aligned}\tag{5.3}$$

Combining (5.1) and (5.3), and using $t\leqslant(1-\varepsilon)n$ and $1-x\leqslant\mathrm{e}^{-x-x^2/2}$, we obtain that the probability that $\mathcal{M}(t)$ contains a component of size exactly k is at most

$$\begin{aligned}&(n\mathrm{e})^k\mathrm{e}^{(k-1)\ln t}\cdot n^{-(k-1)}\mathrm{e}^{-kt/n-(k-1)\ln(n)+6k^2/(\varepsilon n)}\\ &=n\mathrm{e}^{k+(k-1)\ln(1-(n-t)/n)-kt/n+6k^2/(\varepsilon n)}\\ &\leqslant n\mathrm{e}^{k-(k-1)(1+\frac{(n-t)^2}{2n^2})+6k^2/(\varepsilon n)}\\ &\leqslant n\mathrm{e}^{1-(k-1)\varepsilon^2/2+6k^2/(\varepsilon n)}.\end{aligned}$$

□

Proof of Theorem 2.3. The proof is similar to that of Theorem 2.2 if we set $K=5/\varepsilon^2$ instead of $3/\varepsilon$. □

Acknowledgements

The authors would like to thank the anonymous referee for helpful comments which improved the presentation of this paper.

References

[1] Achlioptas, D., D'Souza, R. M. and Spencer, J. (2009) Explosive percolation in random networks. *Science* **323** 1453–1455.

[2] Aldous, D. J. (1999) Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli* **5** 3–48.

[3] Bohman, T. and Frieze, A. (2001) Avoiding a giant component. *Random Struct. Alg.* **19** 75–85.

[4] Cho, Y. S., Kahng, B. and Kim, D. (2010) Cluster aggregation model for discontinuous percolation transitions. *Phys. Rev. E* **81** 030103.

[5] Cho, Y. S., Kim, J. S., Park, J., Kahng, B. and Kim, D. (2009) Percolation transitions in scale-free networks under the Achlioptas process. *Phys. Rev. Lett.* **103** 135702.

[6] da Costa, R. A., Dorogovtsev, S. N., Goltsev, A. V. and Mendes, J. F. F. (2010) Explosive percolation transition is actually continuous. *Phys. Rev. Lett.* **105** 255701.
[7] Deaconu, M. and Tanré, E. (2000) Smoluchowski's coagulation equation: probabilistic interpretation of solutions for constant, additive and multiplicative kernels. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie IV* **29** 549–580.
[8] D'Souza, R. M. and Mitzenmacher, M. (2010) Local cluster aggregation models of explosive percolation. *Phys. Rev. Lett.* **104** 195702.
[9] Dubhashi, D. and Panconesi, A. (2009) *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press.
[10] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** 17–61.
[11] Friedman, E. J. and Landsberg, A. S. (2009) Construction and analysis of random networks with explosive percolation. *Phys. Rev. Lett.* **103** 255701.
[12] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley-Interscience.
[13] Kreer, M. and Penrose, O. (1994) Proof of dynamical scaling in Smoluchowski's coagulation equation with constant kernel. *J. Statist. Phys.* **75** 389–407.
[14] Radicchi, F. and Fortunato, S. (2010) Explosive percolation: A numerical analysis. *Phys. Rev. E* **81** 036110.
[15] Riordan, O. and Warnke, L. (2012) Achlioptas process phase transitions are continuous. *Ann. Appl. Probab.* **22** 1450–1464.
[16] Riordan, O. and Warnke, L. (2011) Explosive percolation is continuous. *Science* **333** 322–324.
[17] Spencer, J. and Wormald, N. (2007) Birth control for giants. *Combinatorica* **27** 587–628.