On the Connectivity Threshold of Achlioptas Processes

Mihyun Kang* Konstantinos Panagiotou†

July 28, 2014

Abstract

In this paper we study the connectivity threshold of Achlioptas processes. It is well known that the classical Erdős-Rényi random graph with n vertices becomes connected whp (with high probability, i.e., with probability tending to one as $n \to \infty$) when the number of edges is asymptotically $\frac{1}{2} n \log n$. Our first result asserts that the connectivity threshold of the well-studied Bohman-Frieze process, which is known to delay the phase transition, coincides asymptotically with that of the Erdős-Rényi random graph. Moreover, we describe an Achlioptas process that pushes backward the threshold for being connected (only $\frac{1}{4} n \log n$ edges, i.e., asymptotically half of what is required in the Erdős-Rényi process, are sufficient), but which simultaneously retains the property of delaying the phase transition.

1 Introduction

In the classical and well-studied Erdős-Rényi random graph process we begin with a graph that contains n isolated vertices and add edges randomly one at a time. We denote this process with $\mathsf{ER} = (\mathsf{ER}_n(t))_{t \geq 0}$ for short, where $\mathsf{ER}_n(t)$ is the graph that is obtained after having added t random edges. Suppose that t is parametrized as τn, where $\tau \geq 0$. A major discovery of Erdős and Rényi in their seminal paper [5] was the identification of a *phase transition* with respect to the component structure: with probability tending to one as $n \to \infty$ (with high probability, whp for short), if $\tau > 1/2$, there is a unique component that contains linearly in n many vertices, called the *giant component*, while if $\tau < 1/2$, every component contains $O(\log n)$ vertices. The value $\tau_{\mathsf{ER}} = 1/2$ is called the *critical point* of the phase transition in the ER process.

Since the seminal work of Erdős and Rényi, many different modifications of the ER process have been proposed. Aiming at studying processes that exhibit different characteristics, Dimitris Achlioptas suggested exploiting the principle of many choices. A (generic) *Achlioptas process* starts with a graph on n vertices and no edges. In each subsequent step *two* potential edges are drawn uniformly at random, and one of them is selected according to a given rule and added to the graph. An important question that initiated a series of studies was whether there is a rule that shifts the position of the phase transition or, more generally, that substantially changes the distribution of the component sizes. For example, Bohman and

*Technische Universität Graz. kang@math.tugraz.at. Supported in part by the Deutsche Forschungsgemeinschaft (KA 2748/3-1).

†Ludwig-Maximilians-Universität München. kpanagio@math.lmu.de. Supported in part by the Deutsche Forschungsgemeinschaft (PA 2080/0-1).

Frieze [1] considered the following rule, which is now known as the Bohman-Frieze process or the BF process for short: add the first edge if it joins two vertices that are isolated in the current graph, and otherwise add the second edge. They showed that their rule indeed delays the appearance of the giant component, that is, the critical point τ_{BF} of the phase transition in the BF process is strictly larger than $1/2$. Spencer and Wormald [10] and Bohman and Kravitz [2] proved that the critical point τ_{BF} can be expressed as the blow-up point of a function that describes the susceptibility, i.e., the average size of the component containing a randomly chosen vertex. The finer behaviour of the phase transition of the BF process was investigated by Janson and Spencer [7]. The phase transition of many other Achlioptas processes was studied in [10] and in more generality by Riordan and Warnke [9].

In this paper we study several Achlioptas processes after the phase transition, in particular, we consider the property of being connected. For the ER process the threshold for being connected is when the number of edges is around $\frac{1}{2}n \log n$ (see e.g. [3, 6]), which is the same as the threshold of the (non-)existence of isolated vertices. Our main interest is in studying the effect of specific rules to the connectivity transition of the underlying random graph process. We show that for the BF process, see Theorem 4.2, the threshold for being connected coincides asymptotically with that of the ER process, the reason being surprisingly that whp the number of isolated vertices in the BF process is asymptotically the same as that in the ER process.

In Section 3 we introduce a second process that is a simple modification of the BF process: it starts with a graph with n vertices and no edge, in each step two potential edges are chosen uniformly at random, and the first edge is added to the graph only if *at least one* of its endvertices is isolated. We call this the KP process. We show that the KP process exhibits two at first sight contradictory characteristics: while, similarly to the BF process, it delays the phase transition (see Theorem 3.1), it simultaneously needs whp asymptotically only *half* as many edges to create a connected graph (see Theorem 3.2). In other words, the KP process pushes *backward* the threshold for being connected, while it pushes *forward* the critical point of the phase transition.

2 Preliminaries

2.1 The Erdős-Rényi Process

For technical convenience we alter slightly the definition of the classical Erdős-Rényi random graph process $(\mathsf{ER}_n(t))_{t\geq 0}$. The graph $\mathsf{ER}_n(0)$ contains n vertices and no edges. We obtain $\mathsf{ER}_n(t)$ by adding an edge that contains two uniformly random vertices to $\mathsf{ER}_n(t-1)$. Note that $\mathsf{ER}_n(t)$ may contain loops and multiple edges; we allow this here and in the rest of the paper, and the asymptotic results are not affected by this modification.

It is well known that the Erdős-Rényi process exhibits a phase transition at time $t \approx n/2$, see e.g. [3, 6].

Theorem 2.1. *The following statements are true whp.*

- *If $t < (1-\varepsilon)n/2$, then all components of $\mathsf{ER}_n(t)$ contain $O(\log n)$ vertices.*
- *If $t > (1+\varepsilon)n/2$, then $\mathsf{ER}_n(t)$ contains a component with $\Omega(n)$ vertices.*

The property of being connected is also well studied. The following result states that at $t = (1/2 + o(1))n \log n$ the graph becomes whp connected [3, 6].

Theorem 2.2. *Let $\varepsilon > 0$ and set $t_0(n) = \frac{1}{2} n \log n$. Then*

$$\Pr[\mathsf{ER}_n(t) \text{ is connected}] = \begin{cases} 1 - o(1), & \text{if } t > (1+\varepsilon)t_0(n), \\ o(1), & \text{if } t < (1-\varepsilon)t_0(n). \end{cases}$$

We say that a property of graphs $\mathcal{Q}$ is convex, if $A \subseteq B \subseteq C$ and $A, C \in \mathcal{Q}$ imply $B \in \mathcal{Q}$. In our proofs it will be convenient to switch between $\mathsf{ER}_n(t)$ and the classical random graph $G_{n,p}$, where each edge is included independently with probability p. The following statement allows us to do so for convex properties.

Proposition 2.3 ([6])**.** *Let $\mathcal{Q}$ be a convex property of graphs with n vertices. Let $p = t/\binom{n}{2}$. If $\Pr[G_{n,p} \in \mathcal{Q}] = 1 - o(1)$, then also $\Pr[\mathsf{ER}_n(t) \in \mathcal{Q}] = 1 - o(1)$.*

3 The KP Process

We consider the following random graph process $\mathsf{KP}_n = (\mathsf{KP}_n(t))_{t \geq 0}$. The graph $\mathsf{KP}_n(0)$ contains n vertices and no edges. In each time step t, two random edges $e_1(t)$ and $e_2(t)$ (where, as in the ER process, the endpoints of those edges are selected uniformly at random) are presented. If $e_1(t)$ contains an isolated vertex, then $\mathsf{KP}_n(t) = \mathsf{KP}_n(t-1) \cup \{e_1(t)\}$. Otherwise $\mathsf{KP}_n(t) = \mathsf{KP}_n(t-1) \cup \{e_2(t)\}$. In Section 3.1 we show the following result regarding the phase transition of the KP process.

Theorem 3.1. *There exists a constant $\tau_c > 1/2$ such that for any $\varepsilon > 0$ the following statements are true whp.*

- *If $t < (1-\varepsilon)\tau_c n$, then all components of $\mathsf{KP}_n(t)$ contain $O(\log n)$ ver- tices.*
- *If $t > (1+\varepsilon)\tau_c n$, then $\mathsf{KP}_n(t)$ contains a component with $\Omega(n)$ vertices.*

If we compare this statement to the behaviour of the classical random graph process, c.f. Theorem 2.1, we see that the emergence of the giant connected component is delayed. However, our next result shows that the property of being connected is *accelerated* in $\mathsf{KP}_n(t)$. The proof can be found in Section 3.2.

Theorem 3.2. *Let $\varepsilon > 0$ and set $t_0(n) = \frac{1}{4} n \log n$. Then whp*

$$\Pr[\mathsf{KP}_n(t) \text{ is connected}] = \begin{cases} 1 - o(1), & \text{if } t > (1+\varepsilon)t_0(n), \\ o(1), & \text{if } t < (1-\varepsilon)t_0(n). \end{cases}$$

3.1 Proof of Theorem 3.1

Spencer and Wormald [10] studied a wide class of Achlioptas processes defined by so-called bounded-size rules. Let K be a fixed constant. A bounded-size Achlioptas process $\mathsf{A}_n = (\mathsf{A}_n(t))_{t \geq 0}$ starts with a graph with vertex set $[n] := \{1, \ldots, n\}$ and no edges. In each subsequent time step t the endpoints v_1, w_1, v_2, w_2 of two edges $e_1(t) = \{v_1, w_1\}$ and $e_2(t) = \{v_2, w_2\}$ are chosen uniformly and independently at random from $[n]$. Then, exactly one of these two edges is included in the resulting graph, where the choice depends only on the sizes of the components containing v_1, w_1, v_2, w_2, and all components of size larger than K are treated the same. The Erdős-Rényi process is when $K = 0$, while the KP process and the BF process are examples of bounded-size Achlioptas processes with $K = 1$.

The main topic of study in [10] is the phase transition in bounded-size Achlioptas processes. The analysis reveals that the crucial parameter is the so-called *susceptibility*, which is defined as the expected number of vertices in the component containing a randomly chosen vertex. In particular, if we denote by $C(v)$ the size of the component containing the vertex v, and if $C_1, C_2, \ldots, C_\ell$ denote the sizes of the components of a graph G, then it is straightforward to establish that the susceptibility $S(G)$ is given by

$$S(G) = \frac{1}{n} \sum_{v \in [n]} C(v) = \frac{1}{n} \sum_{i=1}^{\ell} C_i^2.$$

Spencer and Wormald proved the following results (this is a simplified version of Theorem 1.1. in [10]).

Theorem 3.3. *Let A_n be a bounded-size Achlioptas process. Then there exist a constant $\tau_c > 0$ and functions $x(\tau)$ and $s(\tau)$ such that for any $\varepsilon > 0$ the following is true whp.*

(1) The function $x(\tau) \in (0, 1)$ for all $\tau > 0$. For any $\tau > 0$, let $X(\tau n)$ denote the proportion of isolated vertices in $\mathsf{A}_n(\tau n)$. Then $X(\tau n) = x(\tau) + o(1)$.

(2) The function $s(\tau)$ is defined for all $\tau \in [0, \tau_c)$ and $\lim_{\tau \to \tau_c^-} s(\tau) = \infty$. For any $0 \le \tau < \tau_c$, let $S(\tau n)$ denote the susceptibility of $\mathsf{A}_n(\tau n)$. Then $S(\tau n) = s(\tau) + o(1)$.

(3) For all $t < (1 - \varepsilon)\tau_c n$, all components of $\mathsf{A}_n(t)$ contain $O(\log n)$ vertices, while for all $t > (1 + \varepsilon)\tau_c n$, $\mathsf{A}_n(t)$ contains a component with $\Omega(n)$ vertices.

To deduce Theorem 3.1 from Theorem 3.3, it suffices to show that $\tau_c > 1/2$ for the KP process. Theorem 3.3 (2) implies that there exists a deterministic function $s(\tau)$ such that whp the susceptibility $S(t)$ of $\mathsf{KP}_n(t)$ is concentrated around $s(\tau)$. Following the general principles of the differential equations method, see [10], one can show that $s(\tau)$ for the KP process is the solution of the differential equation

$$s'(\tau) = 2(1 - (1 - x(\tau))^2)s(\tau) + 2(1 - x(\tau))^2 s(\tau)^2 \tag{3.1}$$

with the initial condition $s(0) = 1$. Since all the work was done in [10] here we sketch only the rough ideas of how to derive (3.1) by studying the average evolution of $S(t)$ for the KP process by adding a single edge. First of all, let us suppose that the first edge $e_1(t+1) = \{v_1, w_1\}$ contains an isolated vertex in $\mathsf{KP}_n(t)$, i.e., at least one of v_1, w_1 is isolated. This happens with probability $1 - (1 - X(t))^2$. In this case, two components of sizes $C(v_1), C(w_1)$ of $\mathsf{KP}_n(t)$ are merged to form a new component of size $C(v_1) + C(w_1)$ in $\mathsf{KP}_n(t+1)$. Therefore, unless v_1 and w_1 are contained in the same component, we have

$$S(t+1) - S(t) = \frac{1}{n}\left((C(v_1) + C(w_1))^2 - C(v_1)^2 - C(w_1)^2\right) = \frac{2}{n}\, C(v_1)C(w_1).$$

Since at least one of $C(v_1)$ and $C(w_1)$, say $C(v_1)$, is, by assumption, equal to one and $C(w_1)$ is equal to C_i (including the possibility of C_i being equal to one) with probability $\frac{C_i}{n}$ for any $1 \le i \le \ell$, we have in this case $S(t+1) - S(t) = \frac{2}{n} \sum_i C_i \frac{C_i}{n} = 2S(t)/n$.

On the other hand, given $\mathsf{KP}_n(t)$, if both v_1, w_1 are not isolated, which happens with probability $(1 - X(t))^2$, then two components of sizes $C(v_2)$, $C(w_2)$ of $\mathsf{KP}_n(t)$ are merged to form a new component of size $C(v_2) + C(w_2)$ in $\mathsf{KP}_n(t + 1)$. Thus, unless again v_2 and w_2 belong to the same component, $S(t + 1) - S(t) = 2C(v_2)C(w_2)/n$. We note further that $C(v_2) = C_i$ with probability C_i/n and $C(w_2) = C_j$ with probability C_j/n for any $1 \leq i, j \leq \ell$. Thus $S(t + 1) - S(t) = \frac{2}{n} \sum_{i,j} C_i C_j \frac{C_i}{n} \frac{C_j}{n} = 2S(t)^2/n$.

Putting all together, we arrive at the bound

$$\begin{aligned}\mathbb{E}[S(t+1)-&S(t) \mid \mathsf{KP}_n(t)] \\ &= (1 - (1 - X(t))^2)\frac{2}{n}S(t) + (1 - X(t))^2 \frac{2}{n} S(t)^2 + o(n^{-1}),\end{aligned}$$

where the $o(n^{-1})$ term accounts for the event that two of the vertices v_1, w_1, v_2, w_2 are in the same component of $\mathsf{KP}_n(t)$. This already looks quite similar to (3.1), and the analysis in [10] makes this heuristic derivation rigorous, i.e., Theorem 3.3 (2) follows.

In order to prove $\tau_c > 1/2$ for the KP process, we observe from Theorem 3.3 (2) that $\lim_{\tau \to \tau_c^-} s(\tau) = \infty$, and so it is enough to show that $s(1/2)$ is finite. To this end, we shall use some properties of $s(\tau)$ that are specific to the KP process. Observe that $s(\tau) \geq 1$ and that $s(\tau)$ is increasing, so we can take τ_0 such that $s(\tau_0/2) = 3/2$. Then for all $\tau \geq \tau_0/2$, using (3.1)

$$\begin{aligned} s'(\tau) &\leq 2(1 - (1 - x(\tau))^2)\frac{2}{3}s(\tau)^2 + 2(1 - x(\tau))^2 s(\tau)^2 \\ &\leq 2(1 - x(\tau)/3)s(\tau)^2 \qquad\qquad [\text{since } 0 \leq x(\tau) \leq 1]. \end{aligned}$$

Let $x(1/2)/3 = \delta$. Then $0 < \delta < 1$ and for all $\tau_0/2 \leq \tau \leq 1/2$, since $x(\tau)$ is decreasing, we have

$$s'(\tau) \leq 2(1 - \delta)s(\tau)^2.$$

From this, together with the boundary condition $s(\tau_0/2) = 3/2$, it follows that for all $\tau_0/2 \leq \tau \leq 1/2$,

$$s(\tau) \leq \frac{1}{2/3 + (1 - \delta)\tau_0 - 2(1 - \delta)\tau},$$

and in particular,

$$s(1/2) \leq \frac{1}{2/3 - (1 - \delta)(1 - \tau_0)}. \tag{3.2}$$

Moreover, from (3.1) and the fact $s(\tau) \geq 1$ we have that $s'(\tau) \leq 2s(\tau)^2$. This, together with the initial condition that $s(0) = 1$ implies that $s(\tau) \leq \frac{1}{1-2\tau}$. As a consequence, we have $3/2 = s(\tau_0/2) \leq 1/(1 - \tau_0)$ and so, from (3.2) $s(1/2) \leq 3/2\delta < \infty$, as desired.

3.2 Proof of Theorem 3.2

We call a random graph process $(\mathsf{A}_n(t))_{t \geq 0}$ a k-edge process if $\mathsf{A}_n(t+1)$ is obtained by adding at most one out of k random edges to $\mathsf{A}_n(t)$. With this notation, $\mathsf{ER}_n(t)$ is a 1-edge process, and $\mathsf{BF}_n(t)$ and $\mathsf{KP}_n(t)$ are 2-edge processes. We start with a simple lower bound for the connectivity threshold of $\mathsf{A}_n(t)$.

Lemma 3.4. *Let $\varepsilon > 0$ and let $(\mathsf{A}_n(t))_{t\geq 0}$ be a k-edge process. If $t < (1-\varepsilon)\frac{1}{2k} n \log n$, then whp $\mathsf{A}_n(t)$ is disconnected.*

Proof. Let G_t be the graph where all $kt < (1-\varepsilon)\frac{1}{2} n \log n$ edges are added. Then, certainly $\mathsf{A}_n(t) \subset G_t$, implying that $\mathsf{A}_n(t)$ is disconnected whenever G_t is. Since G_t is distributed like $\mathsf{ER}_n(kt)$, by applying Theorem 2.2 the conclusion of the lemma follows. ∎

Note that this lemma (with $k = 2$) immediately implies the lower bound for t_0 in Theorem 3.2. The rest of this section is devoted to the proof of the upper bound for t_0. Note that it is enough to show that if $t > (1+\varepsilon)\frac{1}{4} n \log n$, whp $\mathsf{KP}_n(t)$ contains no components of size $1, \ldots, \lfloor n/2 \rfloor$. We shall first show that for $\varepsilon > 0$, if $t > (1/4+\varepsilon) n \log n$, then whp $\mathsf{KP}_n(t)$ has no isolated vertices. To this end, we compute an upper bound for the expected number of isolated vertices in $\mathsf{KP}_n(t)$.

Lemma 3.5. *Let $0 < \delta < 1/2$. Let $X_n(t) = X(t)$ denote the proportion of isolated vertices in $\mathsf{KP}_n(t)$. Then, there exists a constant $C = C(\delta) > 0$ such that for sufficiently large n*

$$\mathbb{E}\left[X(t)\right] \leq C\, e^{-(4-5\delta)t/n}.$$

Proof. Let us begin with computing some elementary probabilities. From the definition of the process it follows that

$$\Pr[e_1(t+1) \text{ contains two isolated vertices} \mid \mathsf{KP}_n(t)] = X(t)(X(t) - 1/n).$$

Moreover, conditional on the event that $e_1(t+1)$ contains two isolated vertices we have that $n(X(t+1) - X(t)) = -2$. Another circumstance that decreases the number of isolated vertices in $\mathsf{KP}_n(t+1)$ is the event that $e_1(t+1)$ connects an isolated vertex to a larger component in $\mathsf{KP}_n(t)$. The probability for this event is

$$\Pr[e_1(t+1) \text{ contains one isolated vertex} \mid \mathsf{KP}_n(t)] = 2X(t)(1 - X(t)).$$

Note that in this case we have that $n(X(t+1) - X(t)) = -1$. Finally, the probability of the remaining events is

$$\Pr[e_1(t+1) \text{ contains no isolated vertex} \mid \mathsf{KP}_n(t)] = (1 - X(t))^2.$$

If $e_1(t+1)$ contains no isolated vertex, then a random edge is added to $\mathsf{KP}_n(t)$. So, in this case $n(X(t+1) - X(t))$ equals the number of distinct endpoints of a random edge that are isolated vertices. By putting everything together we infer that

$$\begin{aligned}
&n\,\mathbb{E}\left[X(t+1) - X(t) \mid \mathsf{KP}_n(t)\right] \\
&= -\,2X(t)(X(t) - 1/n) - 2X(t)(1 - X(t)) - 2X(t)(1 - X(t))^2 \\
&= -\,(4 - 2/n)X(t) + 4X(t)^2 - 2X(t)^3.
\end{aligned}$$

With this relation in mind it can be shown that whp

$$X(\lfloor \tau n \rfloor) = x(\tau) + o(1),$$

for any $\tau \in [0, \infty)$, where x is the unique solution of the differential equation

$$x' = -4x + 4x^2 - 2x^3, \qquad x(0) = 1.$$

This task was performed in [10], and this is the function x in Theorem 3.3 (1). Since $X(t) \le 1$, note that we also have

$$\mathbb{E}\left[X(\lfloor \tau n \rfloor)\right] = x(\tau) + o(1).$$

Let τ_δ be the solution to $x(\tau) = \delta$. Since X is non-increasing, we infer that for any $t > \tau_\delta n$ and sufficiently large n

$$\begin{aligned} n\,\mathbb{E}\left[X(t+1) - X(t) \mid \mathsf{KP}_n(t)\right] &= -(4 - 2/n)X(t) + 4X(t)^2 - 2X(t)^3 \\ &\le -(4 - 5\delta)X(t). \end{aligned} \tag{3.3}$$

So, for all such t we get the bound

$$\mathbb{E}\left[X(t+1) \mid \mathsf{KP}_n(t)\right] \le \left(1 - \frac{4-5\delta}{n}\right) X(t) \le e^{-(4-5\delta)/n} X(t),$$

which implies that

$$\mathbb{E}\left[X(t + \tau_\delta n)\right] \le e^{-(4-5\delta)t/n}\,\mathbb{E}\left[X(\lfloor \tau_\delta n \rfloor)\right] = e^{-(4-5\delta)t/n}\,(\delta + o(1)).$$

The claim of the lemma then follows by replacing t with $t - \tau_\delta n$ in the previous calculation, and choosing, say, $C = 2\delta e^{4\tau_\delta}$. ∎

By applying the previous lemma with $t = (1+\varepsilon)\frac{1}{4}n\log n$ and $\delta = 2\varepsilon/5$ we infer that $\mathbb{E}\left[X(t)\right] = o(n^{-1})$, and Markov's inequality implies that $\mathsf{KP}_n(t)$ contains whp no isolated vertices. In order to complete the proof of Theorem 3.2, we show next that whp $\mathsf{KP}_n(t)$ contains no component of size $s = 2, \dots, \lfloor n/2 \rfloor$.

Lemma 3.6. *Let $\varepsilon > 0$ and set $t_+ = (1+\varepsilon)\frac{1}{4}n\log n$. Then, $\mathsf{KP}_n(t_+)$ contains whp no components with a number of vertices in $[2, \lfloor n/2 \rfloor]$.*

Proof. Let us construct an auxiliary graph sequence $\mathsf{G}_n(t)$ as follows. The graph $\mathsf{G}_n(0)$ contains n vertices and no edges. Moreover, set

$$\mathsf{G}_n(t+1) = \begin{cases} \mathsf{G}_n(t) \cup \{e_2(t+1)\}, & \text{if } e_1(t+1) \text{ contains no isolated vertex} \\ & \text{of } \mathsf{KP}_n(t), \\ \mathsf{G}_n(t), & \text{otherwise} \end{cases}.$$

In words, $\mathsf{G}_n(t)$ contains all edges in $\mathsf{KP}_n(t)$ that were included in a time step where the first edge contained no isolated vertex, i.e., it contains random edges. Since $\mathsf{G}_n(t) \subset \mathsf{KP}_n(t)$, it is sufficient to show that whp there are no components of size $s \in [2, \lfloor n/2 \rfloor]$ in $\mathsf{G}_n(t)$.

Note that $\mathsf{G}_n(t)$ is distributed like $\mathsf{ER}_n(t')$, where $t' = t - Z$, and Z is the number of edges in $\mathsf{KP}_n(t)$ that were not included in $\mathsf{G}_n(t)$. But $0 \le Z \le n$, since every such edge eliminates at least one isolated vertex. So, there is a coupling guaranteeing that

$$\mathsf{ER}_n(t-n) \subseteq \mathsf{G}_n(t) \subseteq \mathsf{ER}_n(t).$$

We will argue that $\mathsf{ER}_n(t^*)$, where $t^* = (1+\varepsilon/2)\frac{1}{4}n\log n < t_+ - n$ for sufficiently large n, contains whp no components with a number of vertices in $[2, \lfloor n/2 \rfloor]$, thus completing the proof. Actually, this property of ER_n is well known – see e.g. page 104 in [6] – but we are unaware of any explicit proof in the literature. We include one here for completeness.

Note that the property of containing no component of size $s \in [2, \lfloor n/2 \rfloor]$ is convex. Thus, by applying Proposition 2.3, we may perform all our calculations in the $G_{n,p}$ model of random graphs, where $p = t^*/\binom{n}{2} = \frac{(1+\varepsilon/2)\log n}{2(n-1)}$. The expected number of components of size s in $G_{n,p}$ is at most

$$\binom{n}{s}\, s^{s-2}\, p^{s-1}\, (1-p)^{s(n-s)}. \tag{3.4}$$

Indeed, the binomial coefficient accounts for the number of choices of the vertices in a component of size s. The term s^{s-2}, by Cayley's formula [4], is the number of ways to choose a (spanning) tree on the set of selected vertices. Finally, p^{s-1} is the probability that the edges of the tree are included in $G_{n,p}$, and $(1-p)^{s(n-s)}$ is the probability that no edge exists between the selected vertices and the rest of the graph.

Using the facts $\binom{n}{s} \le n^s/s!$ and $s! \ge (s/e)^s$ and the inequality $1 - x \le e^{-x}$ we infer that there exists a constant $c > 0$ (independent of n) such that (3.4) is at most

$$p_s = n\,(c\log n)^s\, n^{-s(1+\varepsilon/2)(1-s/n)/2}.$$

Note that p_2, p_3, p_4 are all $o(1)$. Moreover, for $5 \le s \le \lfloor n/2 \rfloor$ it can easily be verified that $p_s = o(n^{-1})$. Thus, the expected number of components of size $s \in [2, \lfloor n/2 \rfloor]$ in $G_{n,p}$ is $o(1)$, and the proof is completed. ∎

4 The BF process

In this section we consider the following random graph process $(\mathsf{BF}_n(t))_{t\ge 0}$, which was described by Bohman and Frieze [1]. The initial graph $\mathsf{BF}_n(0)$ contains n vertices and no edges. In each time step t, two uniform random edges $e_1(t)$ and $e_2(t)$ (where, as in the ER process, the endpoints of those edges are selected uniformly at random) are presented. If $e_1(t)$ contains two isolated vertices in $\mathsf{BF}_n(t-1)$, then $\mathsf{BF}_n(t) = \mathsf{BF}_n(t-1) \cup \{e_1(t)\}$. Otherwise $\mathsf{BF}_n(t) = \mathsf{BF}_n(t-1) \cup \{e_2(t)\}$.

The details of the phase transition in the Bohman-Frieze process were studied in several papers, see e.g. [1, 2, 10]. In particular, it was shown that this rule delays the emergence of the giant component.

Theorem 4.1. *There exists a constant $\tau_c > 1/2$ such that for any $\varepsilon > 0$ the following statements are true whp.*

- *If $t < (1-\varepsilon)\tau_c n$, then all components of $\mathsf{BF}_n(t)$ contain $O(\log n)$ ver- tices.*
- *If $t > (1+\varepsilon)\tau_c n$, then $\mathsf{BF}_n(t)$ contains a component with $\Omega(n)$ vertices.*

The main result in this section demonstrates that whp the time of the connectivity transition in the Bohman-Frieze process coincides asymptotically with the time of that in the Erdős-Rényi process.

Theorem 4.2. *Let $\varepsilon > 0$ and set $t_0(n) = \frac{1}{2} n \log n$. Then whp*

$$\Pr[\mathsf{BF}_n(t) \textit{ is connected}] = \begin{cases} 1 - o(1), & \textit{if } t > (1+\varepsilon)t_0(n), \\ o(1), & \textit{if } t < (1-\varepsilon)t_0(n). \end{cases}$$

By Theorem 4.1, the following lemma implies the upper bound of t_0 in Theorem 4.2.

Lemma 4.3. *Let $\varepsilon > 0$ and set $t_+ = (1+\varepsilon)\frac{1}{2}n \log n$. Then, $\mathsf{BF}_n(t_+)$ con- tains whp no components with a number of vertices in $[1, \lfloor n/2 \rfloor]$.*

Proof. First we prove that $\mathsf{BF}_n(t_+)$ contains whp no isolated vertices, by following the lines of the proof of Lemma 3.5. Here, (3.3) is replaced by

$$\begin{aligned} n\,\mathbb{E}\left[X(t+1) - X(t) \mid \mathsf{BF}_n(t)\right] &= -2X(t)(X(t) - 1/n) - 2(1 - X(t)^2)X(t) \\ &\leq -(2 - 5\delta)X(t), \end{aligned}$$

where the first term after the equality accounts for the case that the endpoints of $e_1(t+1)$ are distinct isolated vertices, and the second term accounts for the event that $e_2(t+1)$ is added to $\mathsf{BF}_n(t)$. It follows that for any $0 < \delta < 1/2$ there exits a constant $C' = C'(\delta) > 0$ such that $\mathbb{E}[X(t)] \leq C' e^{-(2-5\delta)t/n}$ for sufficiently large n. Markov's inequality then implies what is desired.

To prove the rest, i.e., $\mathsf{BF}_n(t_+)$ contains whp no components with a number of vertices in $[2, \lfloor n/2 \rfloor]$, we follow the lines of the proof of Lemma 3.6, where t^* is replaced by $(1 + \varepsilon/2)\frac{1}{2}n \log n$. ∎

The lower bound of t_0 in Theorem 4.2 is an immediate consequence of the following lemma.

Lemma 4.4. *Let $0 < \varepsilon < 1$ and set $t_- = (1-\varepsilon)\frac{1}{2}n \log n$. Then $\mathsf{BF}_n(t_-)$ contains whp at least one isolated vertex.*

Proof. Let $0 < \delta < \varepsilon$ and let $t^* = t^*(\delta)$ be the smallest t such that $\mathsf{BF}_n(t)$ has less than $n^{1-\delta}+2$ isolated vertices. We will split up the rounds $t > t^*$ in chunks of length n. In particular, the jth chunk contains all rounds $t_j < t \leq t_{j+1}$ with $t_j = t^* + jn$ and $0 \leq j \leq (1-\varepsilon)\log n/2$. Let $\mathcal{E}_j$ be the event that $X(t_j) \geq (e^{-2} - 1/\log^2 n)^j n^{-\delta}$. Then by assumption $\Pr[\mathcal{E}_0] = 1$ (and $X(t_0) \leq n^{-\delta} + 2/n$). We will show that

$$\Pr[\mathcal{E}_{j+1} \mid \mathcal{E}_j] \geq 1 - o(1/\log n)$$

uniformly for all $0 \leq j \leq (1-\varepsilon)\log n/2$; the assertion of the lemma follows immediately, since we obtain that whp the number of isolated vertices in $\mathsf{BF}_n(t_-)$ is

$$nX(t_-) \geq (1 - o(1)) \cdot n \cdot n^{-1+\varepsilon-\delta} = \omega(1).$$

Let us call an isolated vertex v in $\mathsf{BF}_n(t_j)$ *bad* if

a) v is contained in $e_2(t)$, for some $t_j < t \leq t_{j+1}$ or

b) v is contained in $e_1(t)$ together with some other isolated vertex in $\mathsf{BF}_n(t_j)$, for some $t_j < t \leq t_{j+1}$.

We call an isolated vertex v in $\mathsf{BF}_n(t_j)$ *good* otherwise. It follows that the number of good vertices in $\mathsf{BF}_n(t_j)$ is a lower bound for the number of isolated vertices in $\mathsf{BF}_n(t_{j+1})$.

We will first bound the number B of bad vertices. In the following calculations we always condition on $\mathcal{E}_j$, i.e., $X(t_j) \geq (e^{-2} - 1/\log^2 n)^j n^{-\delta}$. For an isolated vertex v in $\mathsf{BF}_n(t_j)$ the probability for a) is $1 - (1 - 1/n)^{2n} = 1 - e^{-2} + O(1/n)$ and for b) it is at most

$$n \cdot \frac{2}{n} \cdot X(t_j) \leq 2X(t_0) \leq 2n^{-\delta} + 4/n.$$

Thus $\mathbb{E}[B] = (1-e^{-2}+f_n)nX(t_j)$, where $|f_n| \le 3n^{-\delta}$ for sufficiently large n (and independent of j). Note that if for some $t_j < t \le t_{j+1}$ we change any of the vertices in $e_1(t)$ or $e_2(t)$, the value of B changes by at most two. Moreover, if $B \ge r$, this event can be certified by exposing at most r different edges from the set $\{e_1(t)\}_{t_j<t\le t_j+1} \cup \{e_2(t)\}_{t_j<t\le t_j+1}$. Thus, the combinatorial version of Talagrand's inequality applies, see e.g. [6, Theorem 2.29 and (2.43)] and we obtain that there is a constant $\gamma > 0$ such that for any $h \ge 0$

$$\Pr[|B - \mathbb{E}[B]| \ge h] \le \exp(-\gamma h^2/(\mathbb{E}[B] + h)).$$

Set $h = \frac{nX(t_j)}{2\log^2 n}$. The conditioning on $\mathcal{E}_j$ and $j \le (1-\varepsilon)\log n/2$ guarantees that $h^2/(\mathbb{E}[B]+h) = \omega(\log\log n)$. Thus, with probability at least $1 - o(1/\log n)$ we have for sufficiently large n that $B \le (1 - e^{-2} + 1/\log^2 n)nX(t_j)$. Consequently, with probability at least $1 - o(1/\log n)$ the fraction of good vertices in $\mathsf{BF}_n(t_{j+1})$ is at least

$$X(t_j) - (1 - e^{-2} + 1/\log^2 n)X(t_j) \ge (e^{-2} - 1/\log^2 n)X(t_j).$$

We arrive at the claimed bound $\Pr[\mathcal{E}_{j+1} \mid \mathcal{E}_j] \ge 1 - o(1/\log n)$. ▪

References

[1] T. Bohman and A. Frieze, Avoiding a giant component, *Random Structures & Algorithms*, 19(1):75–85, 2001.

[2] T. Bohman and D. Kravitz, Creating a giant component, *Combinatorics, Probability and Computing*, 15:489–511, 2006.

[3] B. Bollobás, *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, second edition, 2001.

[4] A. Cayley, A theorem on trees, *Quart. J. Math.*, 23:376–378, 1889.

[5] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960), 17–61.

[6] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[7] S. Janson and J. Spencer, Phase transitions for modified Erdős-Rényi processes, *Ark. Mat.*, 50(2):305–329, 2012.

[8] M. Kang, W. Perkins and J. Spencer, The Bohman-Frieze process near criticality, *Random Structures & Algorithms*, 43(2):221–250, 2013.

[9] O. Riordan and L. Warnke, Achlioptas process phase transitions are continuous, *Annals of Applied Probability*, 22:1450–1464, 2012.

[10] J. Spencer and N. Wormald, Birth control for giants, *Combinatorica*, 27:587–628, 2007.

[11] N. Wormald, The differential equation method for random graph processes and greedy algorithms, *Lectures on approximation and randomized algorithms*, pages 73–155, 1999.