Sets that are connected in two random graphs

Michael Molloy*

October 1, 2011

Abstract

We consider two random graphs G_1, G_2 from the $G_{n,p=c/n}$ model, both on the same vertex set. We ask whether there is a non-trivial set of vertices S, so that S induces a connected subgraph both in G_1 and in G_2. We determine the threshold for the appearance of such a subset, as well as the size of such a subset.

1 Introduction

The giant component - the linear-sized connected component in the random graph $G_{n,p=c/n}$ for $c > 1$ - was the first great discovery in random graph theory[3] and still remains as one of the most studied phenomena in the field. In this paper, we consider a very natural variation on the giant component which, somewhat surprisingly, doesn't appear to have been studied before: giant vertex sets that are connected in two different random graphs, simultaneously.

Consider two random graphs, G_1, G_2, each drawn from the model $G_{n,p=c/n}$ for some constant $c > 1$, and each on the same vertex set, $V = \{1, ..., n\}$. We say that $S \subseteq V$ is a *doubly connected set* of (G_1, G_2) if S induces a connected subgraph in G_1 and S induces a connected subgraph in G_2.

Clearly, a doubly connected set S must lie in the intersection of the giant components of G_1 and G_2. It does not take long to realize that intersection is w.h.p.[1] not doubly connected. Indeed, it will contain a linear number of vertices which were, eg. connected to the giant component of G_1 through a path of vertices that are not in the giant component of G_2. At first thought, it is not clear whether there will be a non-trivial doubly connected set at all, even for very large constant c. In this paper, we show that there is. We determine the threshold for the appearance of a non-trivial doubly connected set, and we determine its size.

Define

$$c^* = \min_{\xi > 0} \frac{\xi}{(1 - e^{-\xi})^2} = 2.4554...,$$

and for any $c > c^*$ define $\alpha = \alpha(c)$ to be the unique positive solution to

$$\alpha = (1 - e^{-\alpha c})^2.$$

---

*Dept of Computer Science, University of Toronto, molloy@cs.toronto.edu. Research supported by an NSERC Discovery Grant.

[1] We say that a property holds w.h.p. (with high probability) if the probability tends to 1 as n grows.

Substituting $\xi = \alpha c$, we obtain $c = \xi/(1-e^{-\xi})^2$ and so $\alpha(c)$ exists iff $c > c^*$.

Theorem 1.1 *(a) For any $c < c^*$, w.h.p. (G_1, G_2) does not have a doubly connected set of size greater than 2.*

(b) For any $c > c^*$, w.h.p. (G_1, G_2) has a doubly connected set, and the largest such set has size $\alpha(c)n + o(n)$.

At $c = c^*$, $\alpha = .5116...$ and so w.h.p. if a non-trivial doubly connected set exists then it is on more than half the vertices. In comparison, the giant component of $G_{n,p=c^*/n}$ has size roughly $.8866n$ and squaring tells us that the intersection of the giant components of G_1, G_2 would have size roughly $.786n$.

Remark: A doubly connected set of size 2 is an edge that is selected for both graphs. The expected number of such edges is $c^2/2$, and a straightforward Method of Moments argument (see eg. [6]) shows that the number is asymptotically distributed like a Poisson variable. Thus, for $c < c^*$, the probability that there is a doubly connected set of size 2 is $1 - e^{-c^2/2} + o(1)$.

Our proof extends to several copies of $G_{n,p}$ with varying edge-probabilities. Consider random graphs $G_1, ..., G_t$, on the same vertex set, where G_i is from $G_{n,p=c_i/n}$. If there is a solution to:

$$\alpha = \prod_{i=1}^{t}(1 - e^{-c_i\alpha}),$$

then w.h.p. there is a set S of size $\alpha n + o(n)$ such that S induces a connected subgraph in each of $G_1, ..., G_t$. If there is no such solution, then w.h.p. there is no such S on more than two vertices (in fact, on more than one vertex if $t \geq 3$). We omit the details of the straightforward proof adaptation.

The proof of our theorem will be reminiscent of studies of the k-core, the pure literal rule, and other similar problems [10, 8, 4, 5, 7]. There, one repeatedly removes vertices (literals, etc.) that have small degree, until a core remains. Here, we will repeatedly remove vertices that lie in small components of at least one of the two graphs. Analysis of the k-core process is enabled by the fact that, at each iteration, what remains is a random graph conditioned on certain degree sequence properties, which we know how to analyze. In our process, at each iteration, we find that what remains is (more-or-less) the giant component of a random graph on a smaller vertex set. This enables us to continue to analyze it.

2 Some intuition

We begin with a short intuitive explanation for α and c^*.

We can find whether (G_1, G_2) has a doubly connected set using an iterative stripping process. At each step, every vertex that is in a small component of either graph is removed from both graphs. W.h.p., at each stage each graph contains at most one component that is not small (i.e. the giant component). We can show that w.h.p. if this process removes all vertices then there is no doubly connected set; otherwise, the vertices that remain form the largest doubly connected set.

Let Θ_1, Θ_2 be the vertices that are removed because they lie in small components of G_1, G_2, respectively. Thus, the doubly connected component that we find is $S = V\backslash(\Theta_1 \cup \Theta_2)$. Note that

S is the giant component of $G_1 \backslash \Theta_2$ (the graph remaining after removing Θ_2 from G_1) and is also the giant component of $G_2 \backslash \Theta_1$.

Now make a leap of faith and suppose that, somehow, G_1 and G_2 come up with Θ_1, Θ_2 *independently.*

Suppose $|\Theta_1| = (1-\rho)n + o(n)$ for some constant $\rho > 0$; by symmetry, it is reasonable to assume that we also have $|\Theta_2| = (1-\rho)n + o(n)$. S is the vertex-set of the giant component of $G_1 \backslash \Theta_2$, and also of the giant component of $G_2 \backslash \Theta_1$. Since Θ_2 is independent of G_1, we can treat $G_1 \backslash \Theta_2$ as $G_{n',p=c/n}$ where $n' = n - |\Theta_2| = \rho n + o(n)$. So by Lemma 3.4 (below), $|S| = \alpha n + o(n)$ where

$$\alpha = \rho(1 - e^{-\alpha c}). \tag{1}$$

By the independence of Θ_1, Θ_2, our leap of faith can also lead us to assume that Θ_1 intersects $V \backslash \Theta_2$ in the same proportion that it intersects V. $\Theta_1 \backslash \Theta_2$ is simply the set of vertices not appearing on the giant component of $G_1 \backslash \Theta_2$, and so

$$\frac{\rho - \alpha}{\rho} = 1 - \rho; \qquad \text{i.e. } \alpha = \rho^2. \tag{2}$$

(1) and (2) yield $\alpha = (1 - e^{-\alpha c})^2$ and hence the definition of c^*.

Of course, in our process,Θ_1, Θ_2 are *not* formed independently. However, our proof can be viewed as building a different pair of sets Θ_1', Θ_2' with $\Theta_i \subseteq \Theta_i'$, which yields the same doubly connected set; i.e where $\Theta_1 \cup \Theta_2 = \Theta_1' \cup \Theta_2'$ and $S = V \backslash (\Theta_1' \cup \Theta_2')$ is the giant component of both $G_1 \backslash \Theta_2'$ and $G_2 \backslash \Theta_1'$. Furthermore, Θ_1', Θ_2' will be formed (essentially) independently, enabling an analysis similar to that above.

3 A stripping procedure

We start by proving that w.h.p. every doubly connected set with more than 2 vertices must have linear size.

Lemma 3.1 *For every $c > 0$, there exists a constant $\phi = \phi(c) > 0$ such that w.h.p. (G_1, G_2) does not have a doubly connected set of size greater than 2 and less than ϕn.*

Proof This is a very standard argument, using the principle that w.h.p. the subgraph induced by any small set of vertices must have very low edge-density.

Let X_a denote the number of doubly connected sets of size a. Each such set must contain a spanning tree in G_1 and in G_2. There are a^{a-2} spanning trees on a vertices, and each has exactly $a-1$ edges. So:

$$E(X_a) \leq \binom{n}{a} (a^{a-2})^2 \left(\frac{c}{n}\right)^{2a-2} < \frac{n^2}{a^4 c^2} \left(\frac{en}{a} a^2 \left(\frac{c}{n}\right)^2\right)^a < \frac{n^2}{a^4 c^2} \left(\frac{ec^2 a}{n}\right)^a.$$

From this, it is straightforward to obtain $\sum_{a=3}^{\phi n} E(X_a) = o(1)$ for $\phi < \frac{1}{ec^2}$. □

Lemma 3.1 implies that any doubly connected set of size at least 3 must be contained in the intersection of the giant components of G_1 and G_2. Furthermore, the following procedure will w.h.p. find it.

At each iteration i, S_i, resp. T_i is defined to be the vertex set of the largest component of the subgraph of G_1, resp. G_2 induced by V_i.

STRIP:

```
Initialize $V_1 = \{1, ..., n\}$.
For $i = 1$ to $\infty$
    Expose $S_i$.
    Expose $T_i$.
    if $S_i = T_i$ then HALT SUCCEED.
    else
        $V_{i+1} := V(S_i) \cap V(T_i)$.
        if $V_{i+1} = \emptyset$ then HALT FAIL.
```

If we Halt Fail, then w.h.p. there is no doubly connected set of linear size, as w.h.p. neither G_1 nor G_2 has more than one linear-sized component. If we Halt Succeed then $S_i = T_i$ is doubly connected, and is the maximum doubly connected set in (G_1, G_2).

This is the iterative procedure described in Section 2. If it halts succeed, then Θ_1, Θ_2 are $\cup_{i \geq 1} V_i \backslash S_i, \cup_{i \geq 1} V_i \backslash T_i$.

To analyse STRIP, we define:

$$\rho_1 = 1,$$

for $i \geq 1$, γ_i is the unique positive solution to

$$\gamma_i = \rho_i(1 - e^{-\gamma_i c}),$$

and

$$\rho_{i+1} = \gamma_i / \rho_i.$$

Lemma 3.2 *The sequences $\gamma_1, ...,$ and $\rho_1, ...$ are both strictly decreasing and*

(a) For any $c < c^*$, $\lim_{i \to \infty} \gamma_i = \lim_{i \to \infty} \rho_i = 0$.

(b) For any $c > c^*$, $\lim_{i \to \infty} \gamma_i = \alpha(c)$ and $\lim_{i \to \infty} \rho_i = \sqrt{\alpha(c)}$.

Proof It is easy to confirm that $\rho_2 < 1 = \rho_1$. Suppose $\rho_i < \rho_{i-1}$. Then it is straightforward to check that $\gamma_i < \gamma_{i-1}$ (in fact, Lemma 3.4 below implies this). Note that $\rho_{i+1} = 1 - e^{-\gamma_i c}$ and so $\gamma_i < \gamma_{i-1}$ implies $\rho_{i+1} < \rho_i$. So both sequences are decreasing. Since they are both positive, they have a limit. That limit must be a fixed point of the recursive equations. For $c < c^*, (0, 0)$ is the only fixed point. For $c \geq c^*$, $(\sqrt{\alpha}, \alpha)$ is the only other fixed point. A straightforward induction proves that $\rho_i > \sqrt{\alpha}, \gamma_i > \alpha$ for every i, and so $(\sqrt{\alpha}, \alpha)$ must be the limit. $\square$

In the next section, we will prove:

Lemma 3.3 *For any constant i, w.h.p.*

(a) $|V_i| = \rho_i^2 n + o(n)$;

(b) $|S_i|, |T_i| = \gamma_i n + o(n)$.

Lemmas 3.2, 3.3 show us that, by taking a sufficiently high constant number of iterations, S_i, T_i get quite close to what Theorem 1.1 would predict. Lemma 3.1 completes the proof of Theorem 1.1(b), by considering i high enough so that $\gamma_i < \phi$. The proof of Theorem 1.1(a) requires more work.

We close this section with some useful facts about the size of a giant component. We say $X \in Y \pm Z$ to mean that X is in the range $[Y - Z, Y + Z]$.

Lemma 3.4 *For $c\tau > 1$, w.h.p. the size of the largest component of $G_{n',p=c/n}$ where $n' = \tau n$ is in $\beta n \pm n^{3/5}$, where β is the solution to $\beta = \tau(1 - e^{-\beta c})$.*

Proof Note that $G_{n',p=c/n}$ is $G_{n',p=c\tau/n'}$. So the classical result by Erdos and Renyi[3] on the size of the giant component in $G_{n,p}$, along with the fact that the distribution of the size is asymptotically normal[12, 11, 2] implies that w.h.p. the size of the largest component is in $bn' \pm n^{3/5}$ where $b = 1 - e^{-c\tau b}$. The lemma follows by noting that $\beta = b\tau$. □

We are interested in how β changes with τ near $\tau = \sqrt{\alpha}$.

Lemma 3.5 *For every $c > c^*$, at $\tau = \sqrt{\alpha}$ we have*

$$\frac{\partial}{\partial \tau}\beta < 2\sqrt{\alpha}.$$

Proof We use β' to denote $\frac{\partial}{\partial\tau}\beta$. So

$$\beta' = (1 - e^{-\beta c}) + \tau\frac{\partial}{\partial \tau}(1 - e^{-\beta c}) = (1 - e^{-\beta c}) + \tau c e^{-\beta c}\beta'.$$

Therefore, applying $\beta = \tau(1 - e^{-\beta c})$, we have:

$$\beta' = \frac{1 - e^{-\beta c}}{1 - \tau c e^{-\beta c}}.$$

Next, recalling that c^* is a minimum of $\frac{\xi}{(1-e^{-\xi})^2}$, and differentiating with respect to ξ, we obtain that at the value of ξ corresponding to c^*:

$$1 - e^{-\xi} = 2\xi e^{-\xi}.$$

Recall that $\xi = \alpha c$, and that $\sqrt{\alpha} = 1 - e^{-\alpha c}$, this yields $\sqrt{\alpha} = 2\alpha c e^{-\alpha c}$ so $c\sqrt{\alpha}e^{-\alpha c} = \frac{1}{2}$. Noting that at $\tau = \sqrt{\alpha}$ we have $\beta = \alpha$, yields that at $\tau = \sqrt{\alpha}$:

$$\frac{\partial}{\partial \tau}\beta = \frac{1 - e^{-\alpha c}}{1 - \sqrt{\alpha} c e^{-\alpha c}} = \frac{\sqrt{\alpha}}{1 - \frac{1}{2}} = 2\sqrt{\alpha}.$$

Now the result follows from a straightforward check that $\frac{\partial}{\partial \tau}\beta$ decreases as τ increases from $\sqrt{\alpha}$. □

We close this section with the Chernoff Bound[**?**]. We use the version presented in [9]. Here, $BIN(n,p)$ is the sum of n independent variables, each equal to 1 with probability p and 0 otherwise.
The Chernoff Bound *For any* $0 \le t \le np$*:*

$$\Pr\left(|BIN(n,p) - np| > t\right) < 2e^{-\frac{t^2}{3np}}$$

4 Proof of Lemma 3.3

It is useful to consider what STRIP looks like from the perspective of G_1. At each iteration, G_1 removes all its small components, then is given a list T_i of vertices and removes all remaining vertices which are not in T_i. The following observation about T_i is crucial to our analysis:

Observation 4.1 *Given* V_i *and* $t_i = |T_i|$*, the set* T_i *is a uniform set of* t_i *vertices from* V_i*, and the choice of these vertices is independent of* G_1*.*

Proof At each step of STRIP, we can expose the vertex-sets S_i, T_i without exposing the edges. This implies that for any potential graph $H = G_2 \cap V_i$, every permutation of the vertices of H is equally likely to be $G_2 \cap V_i$, and this holds even after conditioning on G_1. □

This implies that if we are only viewing things from the perspective of G_1, then at each iteration, instead of exposing T_i and then deleting $V_i \backslash T_i$ from V_i, we can instead just expose $t_i = |T_i|$ and remove $\ell_i = |V_i| - t_i$ uniformly random vertices from V_i.

It will be convenient to keep track of an additional set $U_i \subseteq V_i$. Vertices are removed from U_i at the same proportional rate that G_2 causes them to be removed from V_i. G_1 does not cause any vertices to be removed from U_i. This set is useful because it extracts the effect that G_2 has on V_i and hence on S_i. Moreover, U_i is close to being independent of G_1; close enough to be useful. Very roughly speaking, in the limit as $i \to \infty$, $V - U_i$ can be thought of as Θ_2' from Section 2.

At each iteration i, S_i, resp. T_i is defined to be the vertex set of the largest component of the subgraph of G_1, resp. G_2 induced by V_i.

STRIP1:

```
Initialize $V_1 = U_1 = \{1, ..., n\}$.
For $i = 1$ to $\infty$
    Expose $S_i$.
    Expose $\ell_i = |V_i \backslash T_i|$.
    If $S_i = V_i$ and $\ell_i = 0$ then HALT SUCCEED
    Else
        Initialize $L_i = \emptyset$.
        Repeat $\ell_i$ times:
```

```
    Repeat until we choose a $u \in V_i$
        Pick a uniform vertex $u \in U_i$ (without replacement).
        Place $u$ into $L_i$.
$V_{i+1} := S_i \backslash L_i$.
$U_{i+1} := U_i \backslash L_i$.
If $V_{i+1} = \emptyset$ then HALT FAIL.
```

Note that, for each i, the only vertex of L_i that is in V_i is a uniform member of V_i. So by Observation 4.1, we can couple the choice of T_i in STRIP with the choice of L_i in STRIP1 so that $T_i = L_i \cap V_i$. Under this coupling, STRIP and STRIP1 produce the same sets S_i, T_i, V_i.

Note that, if the largest component of $G_1 \cap V_i$ has linear size, then w.h.p. it is equal to the largest component of $G_1 \cap U_i$. It will be convenient to focus on the latter graph.

Lemma 4.2 *For any constant i, w.h.p. $|U_i| = \rho_i n + o(n)$.*

Proof of Lemmas 3.3, 4.2: We will prove the lemmas by induction. More specifically, we prove there are two sequences of constants $\eta_1, \eta_2 ...$ and $\eta'_1, \eta'_2 ...$ such that w.h.p. $|S_i|, |T_i| \in \gamma_i n \pm \eta_i n^{2/3}$, $|V_i| \in \rho_i^2 n \pm \eta'_i n^{2/3}$, and $|U_i| \in \rho_i n \pm \eta'_i n^{2/3}$.

We start with $\eta'_1 = 0$; η_1 will be implicitly defined below. Now we proceed by induction.

V_{i+1} is determined by removing $\ell_i = |V_i| - |T_i|$ uniform vertices from V_i. Since V_{i+1} is what remains in S_i, and since $S_i \subseteq V_i$, we have

$$\mathbf{Exp}(|V_{i+1}| = |S_i|(1 - \frac{\ell_i}{V_i}) = \frac{|T_i||S_i|}{|V_i|},$$

which is in the range $\frac{\gamma_i^2}{\rho_i^2} n \pm \frac{1}{2}\eta'_{i+1} n^{2/3}$, for sufficiently large η'_{i+1} so long as the induction hypothesis holds for $|T_i|, |S_i|, |V_i|$. For any $0 \leq a \leq |S_i|$, the probability that $|V_{i+1}| = |S_i| - a$ is

$$\binom{|S_i|}{a}\binom{|V_i| - |S_i|}{\ell_i - a} / \binom{|V_i|}{\ell_i}.$$

From this, it is straightforward to show that w.h.p. $|U_i|$ is within $\frac{1}{2}\eta'_i n^{2/3}$ of its mean. Recalling $\rho_{i+1} = \frac{\gamma_i}{\rho_i}$, this establishes that w.h.p. $|U_{i+1}| \in \rho_{i+1} n \pm \eta'_{i+1} n^{2/3}$.

The induction step for U_{i+1} is similar. This time, we have $\mathbf{Exp}(|U_{i+1}|) = \frac{|T_i||U_i|}{|V_i|}$ which is in the range $\frac{\gamma_i}{\rho_i} n \pm \frac{1}{2}\eta'_{i+1} n^{2/3}$ for sufficiently large η'_{i+1}, so long as $|T_i| \in \gamma_i n \pm \eta_i n^{2/3}$, $|V_i| \in \rho_i^2 n \pm \eta'_i n^{2/3}$, and $|U_i| \in \rho_i n \pm \eta'_i n^{2/3}$. $|U_{i+1}|$ is determined by the number of uniformly random vertices that we remove from U_i before removing ℓ_i from V_i. The probability that this number is a is

$$\frac{\ell_i}{a}\binom{|V_i|}{\ell_i}\binom{|U_i| - |V_i|}{a - \ell_i} / \binom{|U_i|}{a}.$$

Explanation: consider the first a vertices removed from U_i. The event occurs iff (i) exactly ℓ_i of them are from V_i and (ii) one of the vertices from V_i is the ath vertex removed. Again, from this it is straightforward to show that w.h.p. $|U_i|$ is within $\frac{1}{2}\eta'_i n^{2/3}$ of its mean. Recalling $\rho_{i+1} = \frac{\gamma_i}{\rho_i}$, this establishes that w.h.p. $|U_{i+1}| \in \rho_{i+1}n \pm \eta'_{i+1}n^{2/3}$.

S_i is the largest component of $G_1 \cap U_i$. It would be convenient if $G_1 \cap U_i$ were distributed like $G_{n',p=c/n}$ where $n' = |U_i|$. But it is not, since there is some dependency between G_1 and $|U_i|$. So instead, we sandwich $G_1 \cap U_i$ between two graphs which really are from the $G_{n,p}$ model.

We consider two sets of vertices U^-, U^+, which are defined to be uniformly random subsets of V of sizes $\rho_i n - 2\eta'_i n^{2/3}, \rho_i n + 2\eta'_i n^{2/3}$. We couple these two sets with our process as follows: Choose a sequence of $n - |U^-|$ uniform members of $\{1, ..., n\}$, without replacement; this is $V \backslash U^-$. The first $n - |U^+|$ members of the sequence form $V \backslash U^+$. During the running of STRIP1, each time we choose a uniform member of U_i to place into L_i, we simply take the next member of the sequence.

If $|U^-| \leq |U_i| \leq |U^+|$, then our coupling succeeds and $U^- \subseteq U_i \subseteq U^+$; by our induction hypothesis, this is indeed the case. Let S^-, S^+ be the giant components of $G_1 \cap U^-, G_1 \cap U^+$ respectively. Then $S^- \subseteq S_i \subseteq S^+$.

Because the vertices of U^- are selected uniformly at random, we can select them before exposing the edges of $G_1 \cap U^-$. Therefore, setting $n' = |U^-| = \rho_i n - 2\eta'_i n^{2/3}$, we see that $G_1 \cap U^-$ is distributed as $G_{n',p=c/n}$. So Lemma 3.4 yields w.h.p. $|S^-| \in \beta n \pm n^{3/5}$ where β is the positive solution to

$$\beta = \tau(1 - e^{-\beta c}), \qquad \text{with } \tau = \rho_i - 2\eta'_i n^{-1/3}.$$

Recalling that γ_i is the solution to $\gamma_i = \rho_i(1 - e^{-\gamma_i c})$, we see that for η_i sufficiently large (in terms of η'_i and the value of $\frac{\partial}{\partial \tau} a$ at $\tau = \rho_i$), we have $\beta \geq \gamma_i - \frac{1}{2}\eta_i n^{-1/3}$ and so w.h.p. $|S^-| \geq \gamma_i n - \eta_i n^{2/3}$. Similarly, we obtain that w.h.p. $|S^+| \leq \gamma_i n + \eta_i n^{2/3}$, thus obtaining our required bound on $|S_i|$. The same bound holds for $|T_i|$ by symmetry.

To recap, we have shown that the bounds hold:

- w.h.p. for U_{i+1} if they hold for U_i, T_i, V_i;
- w.h.p. for V_{i+1} if they hold for S_i, T_i, V_i;
- w.h.p. for S_i if they hold for U_i;

and, by symmetry, if they w.h.p. hold for S_i then they w.h.p. hold for T_i. For any constant i, this induction requires that $4i$ w.h.p. events hold; since i is a constant, the union of those events holds w.h.p. This completes the induction and so establishes the lemma. □

5 Proof of Theorem 1.1

Proof of Theorem 1.1(a): Suppose $c < c^*$. By Lemma 3.2 we can take I large enough that $\gamma_I < \phi$ where ϕ is the constant from the statement of Lemma 3.1. By Lemma 3.3, w.h.p. $|S_I| < \phi n$ and so any doubly connected component must have size less than ϕn. Lemma 3.1 then yields Theorem 1.1(a). □

To prove part (b), we will run STRIP1 until some large constant I so that $|V_I|, |S_I|$ are very close to their limits α, and $|U_I|$ is very close to its limit $\sqrt{\alpha}$. From this point on, Lemma 3.5 implies that each time we delete a vertex u from U_i, we expect to remove fewer than $2\sqrt{\alpha}$ vertices from the giant component, S_i. Note that the probability that u itself is in S_i is roughly $\sqrt{\alpha}$. So each time we delete a vertex from S_i we expect to reduce the size of S_i by less than 2; i.e. the expected number of additional vertices to be removed is less than one. Because S_i, V_i are almost the same size, each time we delete a vertex from V_i, we expect to cause $|V_i \backslash S_i|$ to increase by less than one.

It follows that w.h.p. STRIP will halt very soon. We can view the vertices of $V_i \backslash T_i$ as a queue of vertices that must be removed from G_1 and the vertices of $V_i \backslash S_i$ as a queue of vertices that must be removed from G_2. Each time we remove a vertex from one queue, it results in an expected increase in the other queue of less than one. So the total size of these queues has a negative drift and with high probability the queues empty quickly. Forthwith the details.

Inspired in part by the approach in [4], we consider a process that continues to remove vertices from U_i even after we find a doubly connected set.

At each iteration i, S_i, resp. T_i is defined to be the vertex set of the largest component of the subgraph of G_1, resp. G_2 induced by V_i.

```
KEEP-STRIPPING:
Initialize $V_1 = U_1 = \{1, ..., n\}$.
For $i = 1$ to $\infty$
    Expose $S_i$.
    Expose $\ell_i = |V_i \backslash T_i|$.
    If $S_i = V_i$ and $\ell_i = 0$ then
      (*) Repeat until $U_i = \emptyset$ :
            Pick a uniform vertex $u \in U_i$
            $U_{i+1} = U_i \backslash \{u\}$;     $i := i + 1$.
    Else
      Initialize $L_i = \emptyset$.
      Repeat $\ell_i$ times:
            Repeat until we choose a $u \in V_i$
                Pick a uniform vertex $u \in U_i$ (without replacement).
                Place $u$ into $L_i$.
      $V_{i+1} := S_i \backslash L_i$.
      $U_{i+1} := U_i \backslash L_i$.
      If $V_{i+1} = \emptyset$ then HALT FAIL.
```

So once the procedure enters (*), we simply remove vertices from U_i one-at-a-time. S_i, T_i, V_i remain unchanged, and because we entered (*), we have $S_i = T_i = V_i$.

Lemma 5.1 *For any $c > c^*$ and $\delta > 0$, there exist constants $I = I(c, \delta)$ such that w.h.p.*

(a) $\alpha n < |V_I| < (\alpha + \delta)n$;

(b) STRIP halts after removing fewer than δn vertices from V_I.

Proof Recall from Lemma 3.4 that w.h.p. the size of the largest component in $G_{n',p=c/n}$ where $n' = \tau n$ is in $\beta n \pm n^{3/5}$ where β is the solution to $\beta = \tau(1 - e^{-\beta c})$. By Lemma 3.5 and the continuity of $\frac{\partial}{\partial\tau}\beta$, there exists $\xi, \zeta > 0$ such that for any $\tau \in [\sqrt{\alpha} - \xi, \sqrt{\alpha} + \xi]$, we have $\frac{\partial}{\partial\tau}\beta < (2 - \zeta)\sqrt{\alpha}$.

We can assume that $\delta < \xi$ and is sufficiently small in terms of z. By Lemma 3.2, we can choose I so that $\sqrt{\alpha} < \rho_I < \sqrt{\alpha} + \frac{\delta}{2}$, $\alpha < \gamma_I < \alpha + \frac{\delta}{2}$, $\rho_{I-1} - \rho_I < \delta^2$ and $\gamma_{I-1} - \gamma_I < \delta^2$. By Lemmas 3.3 and 4.2, we have that w.h.p. $|U_I| = \rho_I n + o(n), |V_I| = \rho_I^2 n + o(n)$ and both $|S_I|, |T_I| = \gamma_I n + o(n)$.

We will run KEEP-STRIPPING until U_i has size exactly $(\rho_I - \delta)n$; thus we have removed $\delta n + o(n)$ vertices from U_I. We have to be careful what we mean by this, since all the vertices in L_i are removed from U_i at once. So:

- If, at the end of iteration i, $|U_{i+1}| = \lfloor(\rho_I - \delta)n\rfloor$ then we Halt and set $U^* = U_{i+1}$, $S^* = S_{i+1}$ and $V^* + V_{i+1}$.
- If, at any point during iteration i, $|U_i| - |L_i| = \lfloor(\rho_I - \delta)n\rfloor$ then we Halt and set $U^* = U_i \backslash L_i$, we let S^* be the largest component of the subgraph of G_1 induced by U^*, and we let $V^* = S^* \backslash L_i$.

We let J denote the iteration during which, or at the end of which, we halted. By Lemma 3.3, w.h.p. $J \geq I$.

Recall that STRIP1, KEEP-STRIPPING and U_i were based on viewing the STRIP from the perspective of G_1. Now consider defining analogous procedures that view it from the perspective of G_2; i.e. we define U_i', STRIP1' and KEEP-STRIPPING' by replacing G_1 by G_2 throughout the definitions of U_i, STRIP1 and KEEP-STRIPPING. We define J' to be the analogue of J, and without loss of generality, assume $J' \leq J$.

Throughout all iterations of KEEP-STRIPPING, each vertex removed from U_i was selected uniformly from U_i. And the stopping condition depended only on the number of vertices removed. Therefore, U^* is a uniformly random set of $\lfloor(\rho_I - \delta)n\rfloor$ vertices from $U_1 = \{1, ..., n\}$. Thus, we can expose the subgraph of G_1 induced by U^* by first choosing the vertices of U^* and then choosing the edges; i.e. we can treat it as $G_{n'=\lfloor(\rho_I - \delta)n\rfloor,p=c/n}$.

We will prove that w.h.p. we enter line (*) before halting; i.e. w.h.p. we find a doubly-connected set before removing δn vertices from U_I. So suppose that we do not enter line (*); we will reach a contradiction.

By Lemma 3.4, w.h.p. S^* has size $gn + o(n)$ where $g = (\rho_I - \delta)(1 - e^{-gc})$. Since $\delta < \xi$, we have $\sqrt{\alpha} \leq \rho_I \leq \sqrt{\alpha} + \xi$. We also have $\gamma_{I-1} - \gamma_I < \delta^2$. So we can apply our bound on $\frac{\partial}{\partial\tau}\beta$ to show $g \geq \gamma_{I-1} - (2 - \zeta)\sqrt{\alpha}(\delta + \delta^2) + o(1) > \gamma_{I-1} - (2 - \frac{\zeta}{2})\sqrt{\alpha}\delta$, for δ sufficiently small in terms of ζ. The largest component of U^* is contained in S_J. Therefore w.h.p.

$$|S^*| \geq |S_{I-1}| - (2 - \frac{\zeta}{2})\sqrt{\alpha}\delta n. \tag{3}$$

We will now reach a contradiction by showing that w.h.p. if we do not enter (*) then the number of vertices removed from S_I, i.e. $|S_{I-1} \backslash S^*|$, is greater than $(2 - \zeta)\sqrt{\alpha}\delta n + o(n)$. We first focus on $\cup_{i \geq I} L_i \cap V_i$ which is a subset of $|S_{I-1} \backslash S^*|$ since for each $i \geq I$, we have $V_i \subseteq V_I \subseteq S_{I-1}$.

By our halting condition, we know that if we do not enter (*) then $\cup_{i\geq I} L_i = \delta n + o(n)$. The vertices of L_i are selected uniformly from U_i, and each choice is a member of V_i with probability at least

$$\frac{V_I - \delta n + o(n)}{U_I} = \frac{(\rho_I^2 - \delta)n + o(n)}{\rho_I n + o(n)} > \rho_I - \frac{2\delta}{\rho_I} > \sqrt{\alpha} - \frac{2\delta}{\sqrt{\alpha}},$$

since $\rho_I > \sqrt{\alpha}$. So the number of these vertices that are in V_i is dominated from below by the binomial variable $BIN(\delta n + o(n), \sqrt{\alpha} - \frac{2\delta}{\sqrt{\alpha}})$. The Chernoff Bound implies that w.h.p.

$$|\cup_{i=I}^{J} L_i \cap V_i| \geq \delta(\sqrt{\alpha} - \frac{3\delta}{\sqrt{\alpha}})n. \tag{4}$$

Recall that, by our coupling, $L_i \cap V_i \subseteq V_i \backslash T_i$. (Equality holds for every i except perhaps $i = J$. So the RHS of (4) is also a lower bound on $|\cup_{i=1}^{J} V_i \backslash T_i|$. By symmetry, and using the fact that $J' \leq J$, the same argument applied to KEEP-STRIPPING' implies that w.h.p.

$$|\cup_{i=1}^{J} V_i \backslash S_i| \geq \delta(\sqrt{\alpha} - \frac{3\delta}{\sqrt{\alpha}})n. \tag{5}$$

Note that since $V_i \subset S_{i-1}$, we have $\cup_{i=1}^{J} V_i \backslash S_i \subseteq S_{I-1} \backslash S^*$.

Finally, we wish to bound the number of vertices that were double-counted in (4), (5). Suppose some vertex u is in both $\cup_{i=I}^{J} V_i \backslash S_i$ and $\cup_{i=I}^{J} L_i \cap V_i$. Then u must be in $V_i \backslash S_i$ and in L_i for some $I \leq i \leq J$. We say that such a vertex is *hit twice*.

During iteration i, each of the $|L_i|$ vertices chosen from U_i is hit twice with probability

$$\frac{|V_i \backslash S_i|}{|U_i|} < \frac{|V_I \backslash S_J|}{|U^*|} < \frac{(\sqrt{\alpha} + \delta)^2 n - (\alpha - (2-\zeta)\sqrt{\alpha}\delta)n + o(n)}{(\rho_I - \delta)n} < \frac{5\delta}{\sqrt{\alpha}}.$$

So the total number of vertices that are hit twice is dominated from above by the binomial variable $BIN(\delta n + o(n), \frac{5\delta}{\sqrt{\alpha}})$. So the Chernoff Bound yields that w.h.p.

$$|\Big(\cup_{i=I}^{J} V_i \backslash S_i\Big) \cap \Big(\cup_{i=I}^{J} L_i \cap V_i\Big) \leq \frac{10\delta^2}{\sqrt{\alpha}} n. \tag{6}$$

(4), (5) and (6) imply that w.h.p.

$$|S^*| \leq |S_{I-1}| - 2 \times \delta(\sqrt{\alpha} - \frac{3\delta}{\sqrt{\alpha}})n + \frac{10\delta^2}{\sqrt{\alpha}} n < |S_{I-1}| - (2 - \frac{\zeta}{4})\sqrt{\alpha}\delta n,$$

for δ sufficiently small in terms of ζ. This contradicts (3).

Thus, w.h.p. we must enter (*) before $\delta n + o(n)$ vertices are removed from U_I. W.h.p. this will occur before δn vertices are removed from V_I. This proves the lemma. □

It follows that w.h.p.

$$|\cup_{i=I}^{J} V_i \backslash S_i| \leq |S_{I_1} \backslash S_J| - |\cup_{i=I}^{J} L_i \cap V_i| + \frac{10\delta^2}{\sqrt{\alpha}} n < ((2-\zeta)\sqrt{\alpha}\delta + \delta^2)n - \delta(\sqrt{\alpha} - 4\delta)n + \frac{10\delta^2}{\sqrt{\alpha}} n + o(n) < (1 - \frac{1}{2}\zeta)\sqrt{\alpha}\delta n,$$

since $5\delta^2 + \frac{10\delta^2}{\sqrt{\alpha}}n < \frac{1}{2}\zeta\sqrt{\alpha}\delta$ for δ sufficiently small in terms of ζ.

By symmetry, the same proof yields that

$$\sum_{i=I}^{J} \ell_i = \sum_{i=I}^{J} |V_i \backslash T_i| < (1 - \frac{1}{2}\zeta)\sqrt{\alpha}\delta n < \delta(\sqrt{\alpha} - 4\delta)n \leq | \cup_{i=I}^{J} L_i \cap V_i|,$$

for δ sufficiently small in terms of ζ.

But $\sum_{i=I}^{J} \ell_i = |\cup_{i=I}^{J} L_i \cap V_i|$. Contradiction. Therefore we must have reached line (*). Therefore STRIP1, and hence STRIP must have halted before δn vertices were removed from U_I, and hence before δn vertices were removed from V_I. □

The remaining part of our main theorem follows immediately:

Proof of Theorem 1.1(b): For any $\delta > 0$, Lemma 5.1 yields that STRIP will w.h.p. Halt Succeed and produce a set $S = T = V$ of size at most $(\alpha + \delta)n$ and at least $(\alpha - \delta)n$. □

References

[1] B. Bollobás, *Random Graphs.* 2nd edition. Cambridge University Press, 2001.

[2] B. Bollobás and O. Riordan, *Asymptotic normality of the size of the giant component via a random walk.* arXiv:1010.4595

[3] P. Erdos and A. Renyi. *On the evolution of random graphs.* Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960), 17 - 61.

[4] D. Fernholz and V. Ramachandran. *Cores and Connectivity in Sparse Random Graphs.* The University of Texas at Austin, Department of Computer Sciences, technical report TR-04-13 (2004).

[5] S. Janson and M. Luczak. *A simple solution to the k-core problem.* Random Structures Algorithms **30** (2007) 50 - 62 (2007).

[6] S. Janson, T. Łuczak and A. Ruciński. Random Graphs. Wiley, New York (2000).

[7] J.H.Kim. *Poisson cloning model for random graphs.* arXiv:0805.4133v1

[8] M. Molloy *Cores in random hypergraphs and boolean formulas.* Random Structures and Algorithms **27**, 124 - 135 (2005).

[9] M. Molloy and B. Reed, *Graph colouring and the probabilistic method*, Springer-Verlag, Berlin, 2001.

[10] B. Pittel, J. Spencer and N. Wormald, *Sudden emergence of a giant k-core in a random graph.* J.Comb.Th.(B) **67** (1996), 111 - 151.

[11] B. Pittel and N. Wormald, *Counting connected graphs inside-out.* J. Comb.Th. B **93** (2005), 127 - 172.

[12] V.E. Stepanov, *Some features of the structure of a random graph near a critical point.* (Russian) Teor. Veroyatnost. i Primenen. **32** (1987), 633657.