# The Giant Component*

Abbas Mehrabian
amehrabian@uwaterloo.ca

July 29, 2010

## 1 Introduction

What is the number of vertices in the largest connected component of the Erdös-Rényi random graph $\mathcal{G}(n,p)$ when $p = \Theta(1/n)$? If we parametrize $p = c/n$, $c$ fixed, then it turns out that this number increases from $O(\ln n)$ for $c < 1$ to $\Theta(n)$ for $c > 1$. This phenomenon is called the *Erdös-Rényi phase transition*, and has been studied extensively in random graph literature (see e.g. [2]). Here, following the presentation of [1], we study the size of largest component when $c < 1$ (called the subcritical case) and when $c > 1$ (called the supercritical case). The case $c = 1$ is more delicate and is not considered here; the interested reader is referred to [2]. Since our approach is based on branching processes, we first review them and state the needed results.

## 2 Branching Processes Preliminaries

A *branching process* has a sequence of independent nonnegative random variables $Z_1, Z_2, \dots$ as its underlying space, and is a sequence of random variables given by the recurrence relation

$$Y_0 = 1, \qquad Y_i = Y_{i-1} + Z_i - 1 \ (i > 0),$$

which gives the explicit formula

$$Y_i = 1 + \sum_{j=1}^{i} Z_j - i \qquad \forall i \geq 0.$$

Here is the intuition: there is a particle who gives birth to $Z_1$ other particles and dies. Now, each of the particles, give birth to some number of new particles and die themselves (one particle per time unit). Suppose that the $i$-th particles gives birth to $Z_i$ particles (independent of what others do). Then the number of alive particles at time $i$ is precisely $Y_i$.

It is possible that at some time, say time $i$, there is no alive particle. Hence we would have $Y_i = 0$. We call this event *extinction*. Let $T$ be the first index for which $Y_T = 0$, and let $T = \infty$ if extinction does

*Notes of a talk by the author given for the Advanced Random Graph Theory course, Department of Combinatorics and Optimization, University of Waterloo, Spring 2010

not happen. Thus

$$\text{if } i < T \text{ then (since } Y_i > 0) \sum_{j=1}^{i} Z_j \geq i. \tag{1}$$

We say that the process *dies* at time $T$. Note that if the process dies at time $T$, exactly $T$ particles were active during the process, hence we can call $T$ the *total population*. Here the story breaks down because $Y_{T+1} = Y_T + Z_T$ may be greater than zero, while we cannot have a new particle born where there is no alive particle! However, usually we will be working with the process in the $i \leq T$ range, so the intuition is valid.

In most of the applications, the variables $Z_i$ have the same distribution (each particle has the same ability in giving birth) but there are examples where this is not the case. Branching process are classified according to the distribution of the $Z_i$.

## 2.1 Binomial branching process

Let $c > 0$ be fixed. If every $Z_i$ has binomial distribution with parameters $n, p$, where $p \sim c/n$[1], then we have a *binomial branching process*. The $Z_i$ can be approximated (for large $n$) by Poisson variables with mean $c$, so there is a strong similarity between the binomial branching process and the Poisson branching process with mean $c$ (which is one in which every $Z_i$ is Poisson with mean $c$). In particular we have the following:

**Lemma 1.** *Assume that $c > 1$ and $\rho$ be the unique solution of $\rho e^c = e^{\rho c}$ in $(0, 1)$. Then we have*

$$\lim_{n\to\infty} \mathbf{Pr}[\text{extinction}] = \rho. \tag{2}$$

We define the following two functions of $n$:

$$k_0 = k_0(n) = \left\lfloor \frac{34c}{(c-1)^2} \ln n \right\rfloor$$
$$k_1 = k_1(n) = \left\lceil n^{2/3} \right\rceil$$

Using Chernoff's bounds we get the following:

**Lemma 2.**

$$\mathbf{Pr}\left[\exists i \geq k_0 \text{ s.t. } Y_i \leq \frac{(c-1)i}{2}\right] = o(n^{-2}) \tag{3}$$

$$\mathbf{Pr}\left[\text{extinction}, k_0 \leq T\right] = o(n^{-2}) \tag{4}$$

## 2.2 Graph branching process

Let $v_1$ be a vertex of the random graph $\mathcal{G}(n, p)$. Then $v_1$ has $Z_1 \overset{d}{\sim} Bin(n-1, p)$ neighbors[2], number them as $v_2, v_3, \ldots, v_{Z_0}$. We can view $v_1$ as a particle that gives birth to these $Z_0$ new particles and dies. Now, $v_2$

[1] $a \sim b$ means that $a = (1 + o(1)b)$

[2] $X \overset{d}{\sim} Bin(n, p)$ means that $X$ is a random variable having binomial distribution with paramteres $n, p$

can have up to $n-1-Z_0$ "non-discovered" neighbors, each with probability $p$. Let $Z_2 \stackrel{d}{\sim} Bin(n-1-Z_0, p)$ be the number of non-discovered neighbors of $v_2$, and number them as $v_{Z_0+1}, v_{Z_0+2}, \ldots, v_{Z_0+Z_1-1}$. We can view $v_2$ as a particle that gives birth to these $Z_1$ new particles and dies. Continue in the same way. Let us denote the number of alive particles by $G_i$ (instead of the usual $Y_i$) in the graph branching process. In the $i$-th step, when we are going to determine the offspring of $i$-th particle, we have $i-1$ already-dead particles, $G_i$ alive particles, and so $n-(i-1)-G_i$ non-discovered vertices, each of which can be a child of the $i$-th particle with probability $p$. Therefore, for the graph branching process, we have

$$G_0 = 1, \qquad G_i = G_{i-1} + Z_i - 1 \ (i > 0)$$
$$Z_i \stackrel{d}{\sim} Bin(n-(i-1)-G_{i-1}, p)$$

A moment of thought shows that $T$ is the size of the connected component containing $v_1$, hence its study gives insight into the size of the components.

As this process is a bit complicated, we usually approximate it with two processes $Y^+$ and $Y^-$, defined as following:

$$Y_0^+ = 1, \quad Y_i^+ = Y_{i-1}^+ + Z_i^+ - 1 \ (i > 0), \qquad Z_i^+ \stackrel{d}{\sim} Bin(n, p),$$
$$Y_0^- = 1, \quad Y_i^- = Y_{i-1}^- + Z_i^- - 1 \ (i > 0), \qquad Z_i^- \stackrel{d}{\sim} Bin(n - \lceil ck_1 \rceil, p).$$

We can couple all three processes nicely: for each $i$, flip $n$ fresh coins with $\mathbf{Pr}[heads] = p$. Let $Z_i^-$ be the number of heads among the first $n - \lceil ck_1 \rceil$ of them, $Z_i$ be the number of heads among the first $n-(i-1)-G_{i-1}$ of them, and $Z_i^+$ be the total number of heads. If $i + G_i \le ck_1$ then $n - \lceil ck_1 \rceil \le n-(i-1)-G_{i-1}$; so we have the following inequalities:

$$Z_i \le Z_i^+ \tag{5}$$
$$G_i \le Y_i^+ \tag{6}$$
$$\text{If } i + G_i \le ck_1 \text{ then } Y_i^- \le G_i \tag{7}$$

Using this coupling we get the following result for the graph branching process:

$$\text{if } p \sim c/n \text{ then } \mathbf{Pr}\left[\exists k_0 < i < k_1 \text{ s.t. } G_i < \frac{(c-1)i}{2}\right] = o(n^{-2}) \tag{8}$$

Here is the argument: if there exists $k_0 < i < k_1$ with $G_i < \frac{(c-1)i}{2}$, then $i + G_i < \frac{(c+1)i}{2} < ci \le ck_1$, thus $Y_i^- \le G_i$. But by (3) we know that $\mathbf{Pr}\left[\exists i \ge k_0 \text{ s.t. } Y_i \le \frac{(c-1)i}{2}\right] = o(n^{-2})$.

# 3 Largest Connected Component of The Random Graph

Now we study the size (number of vertices) of the largest connected component of the random graph $\mathcal{G}(n,p)$, where $p = c/n$ for some fixed $c \neq 1$.

## 3.1 Subcritical case: $c < 1$

In this (easy) case the size of the largest component of $\mathcal{G}(n,p)$ is $O(\ln n)$ by the following theorem.

**Theorem 1.** *For fixed $c < 1$ and $\epsilon > 0$ the number of vertices in the largest connected component of $\mathcal{G}(n, c/n)$ is a.s.*[3] *at most $\frac{2+\epsilon}{(1-c)^2} \ln n$.*

*Proof.* Let $v_1$ be a fixed vertex of $\mathcal{G}(n, c/n)$. We prove that the probability that $v_1$ is in a component of size $> i := \frac{2+\epsilon}{(1-c)^2} \ln n$ is $O(n^{-1-\epsilon/2})$, and the theorem follows by using the union bound. We can build a graph branching process by using $v_1$ as the root and use the idea of Section 2.2. In this process, $T$ (total population in the branching process) is the size of the component containing $v_1$. Therefore the component containing $v_1$ has size $> i$ if and only if $i < T$, and

$$\begin{aligned}
\mathbf{Pr}[i < T] \leq & \mathbf{Pr}[\sum_{j=1}^{i} Z_j \geq i] \qquad \text{(by 1)} \\
\leq & \mathbf{Pr}[\sum_{j=1}^{i} Z_j^+ \geq i] \qquad \text{(by 5)} \\
\leq & \exp\left(\frac{-(1-c)^2 i}{2}\right) \qquad \text{(by Chernoff's bounds)} \\
= & O(n^{-1-\epsilon/2}).
\end{aligned}$$

□

## 3.2 Supercritical case: $c > 1$

In this case the largest component have linear size! We need three lemmas before proving the main theorem.

**Lemma 3.** *Almost surely, There is no component with size in $(k_0, k_1)$.*

*Proof.* Let $v_1$ be a fixed vertex of $\mathcal{G}(n, c/n)$. We prove that the probability that $v_1$ is in a component of size $\in (k_0, k_1)$ is $o(n^{-1})$, and the lemma follows by using the union bound. Build a graph branching process by using $v_1$ as the root. We know that $T$ is the size of the component containing $v_1$, and $G_T = 0$ by the definition of $T$. Thus by (8) we have $\mathbf{Pr}[k_0 < T < k_1] = o(n^{-2})$. □

By the previous lemma, a.s. all components either have size $\leq k_0$ or have size $\geq k_1$. Let us call a component of former type *small* and a component of latter type *large*.

**Lemma 4.** *Almost surely, there is at most one large component.*

*Proof.* Let $v, v'$ be two fixed vertices of $\mathcal{G}(n, c/n)$. We prove that the probability that $v, v'$ are in two different large components is $o(n^{-2})$, and the lemma follows by using union bound. First, start a graph branching process using $v$ as the root, and stop it at time $k_1$. If the process is died at that time, then $v$ is not in a large component. Suppose this does not happen. Let $S$ denote the set of "dead" vertices at this time, so $N(S) - S$[4] is the set of "alive" vertices (whose children have not yet been determined). By (8), with probability $1 - o(n^{-2})$, $|N(S) - S| = G_{k_1} \geq (c-1)k_1/2$.

[3] a.s. stands for *almost surely*, i.e. with probability $1 - o(1)$

[4] For a subset $S$ of vertices, $N(S)$ is the set of vertices that have a neighbor in $S$

If $v'$ has already been discovered then $v, v'$ are not in two different large components. Otherwise, start another graph branching process on the graph $G'$ induced by $V(G) - S$ using $v'$ as the root. This new graph is a random graph with $n' = n - k_1 \sim n$ vertices and two vertices are joined with probability $p = c/n \sim c/n'$. It can be shown (using (8) by a proof similar to that of Lemma 3) that this new process either dies in less than $k_0(n')$ steps, or survives for at least $k_1(n')$ steps. In the former case $v'$ is not in a large component. In the latter case, stop the process at time $k_1/2 < k_1(n')$. Let $S'$ be the set of dead vertices, and $N(S') - S'$ be the set of alive vertices. By (8), with probability $1 - o(n^{-2})$, $N(S') - S' = G'_{k_1/2} \geq (c-1)k_1/4$.

Now, If $(N(S) - S) \cap (N(S') - S') \neq \emptyset$ then $v, v'$ are not in two different large components. Otherwise, note that there are at least $(c-1)^2 k_1^2/8 = \Theta(n^{4/3})$ possible locations for edges between $N(S) - S$ and $N(S') - S'$, none of which has been tested in either of the processes. If $v, v'$ are in different large components, then none of these edges exist. But the probability of this event is at most

$$(1 - c/n)^{\Theta(n^{4/3})} < \exp\left(-c\Theta(n^{1/3})\right) = o(n^{-2})$$

as required.

□

**Lemma 5.** *Let $X$ be the number of vertices in small components. Then $\mathbf{E}X = \rho n + o(n)$ and $\mathbf{Var}X = o(n^2)$, where $\rho$ is the unique solution in $(0,1)$ of $\rho e^c = e^{\rho c}$.*

*Proof.* Let $X_v$ be the indicator variable for "$v$ is in a small component." We need to show that $\mathbf{E}X_v = \rho + o(1)$ (and then use linearity of expectation). Consider the graph branching process $G_i$ using $v$ as root, together with its companions $Y_i^-, Y_i^+$. Notice that by (2), the extinction probability of $Y^+, Y^-$ is $\rho + o(1)$. If $v$ is in a small component then $T + G_T = T \leq k_0 < ck_1$ so $Y_T^- \leq G_T = 0$, i.e. $Y^-$ becomes extinct. Therefore $\mathbf{Pr}[X_v = 1] \leq \rho + o(1)$. On the other hand, $Y^+$ dies with probability $\rho + o(1)$. Moreover, if $Y^+$ dies then by (4), with probability $1 - o(n^{-2})$ it happens at time $i < k_0$. If $Y^+$ dies at time $i < k_0$ then $G_i \leq Y_i^+ = 0$ so $v$ is in a small component. Hence we get $\rho + o(1) \leq \mathbf{Pr}[X_v = 1]$. Consequently, $\mathbf{Pr}[X_v = 1] = \rho + o(1)$ as required.

To compute variance, we shall compute $\mathbf{Pr}[X_v = X_{v'} = 1]$ for distinct vertices $v, v'$. First, consider the graph branching process using $v$ as root. With probability $\rho + o(1)$, $v$ is in a small component and we have $T \leq k_0$. With probability $1 - o(1)$, $v'$ has not been discovered yet. Consider the graph branching process using $v'$ as root, on the graph obtained by removing the component containing $v$. This graph has $\sim n$ vertices and so the probability that $v'$ is in a small component is again $\rho + o(1)$. Hence we find $\mathbf{Pr}[X_v = X_{v'} = 1] = \rho^2 + o(1)$ for $v \neq v'$.

Therefore, noting $X = \sum_{v \in V} X_v$,

$$\begin{aligned}
\mathbf{Var}X =& \mathbf{E}[X^2] - \mathbf{E}[X]^2 \\
=& \sum_v \mathbf{E}[X_v^2] + \sum_{v \neq v'} \mathbf{E}[X_v X_{v'}] - (\rho n + o(n))^2 \\
\leq& n(\rho + o(1)) + n^2(\rho^2 + o(1)) - (\rho^2 n^2 + o(n^2)) = o(n^2).
\end{aligned}$$

□

**Theorem 2.** *Let $c > 1$ be fixed. In $\mathcal{G}(n, c/n)$, there is a.s. a unique connected component with more than $\frac{34c}{(c-1)^2} \ln n$ vertices. This component a.s. has $n - \rho n + o(n)$ vertices, where $\rho$ is the unique solution in $(0, 1)$ of $\rho e^c = e^{\rho c}$.*

*Proof.* The first part follows from Lemmas 3,4. For the second part, for any $\epsilon > 0$, using Lemma 5 and Chebyshev's inequality,

$$\mathbf{Pr}[|X - \mathbf{E}X| \geq \epsilon \mathbf{E}X] \leq \frac{\mathbf{Var}X}{(\epsilon \mathbf{E}X)^2} = \frac{o(n^2)}{\epsilon^2 \rho^2 n^2 + o(n^2)} = o(1).$$

Therefore, as $\mathbf{E}X = \Theta(n)$, we have $X = \rho n + o(n)$ with probability $1 - o(1)$. Hence the number of vertices in the unique large component is a.s. $n - \rho n + o(n)$. □

**Remark.** This unique largest connected component that has linear size, is called *the giant component* of the random graph.

# References


[1] N. Wormald, A Guide to Random Graphs, Lecture Notes.

[2] N. Alon, J. Spencer, The Probablistic Method, 3rd ed., Wiley, 2008.