The Diameter of Sparse Random Graphs

Fan Chung [*†] Linyuan Lu [*]

This paper is dedicated to the memory of Paul Erdős.

Abstract

We consider the diameter of a random graph $G(n,p)$ for various ranges of p close to the phase transition point for connectivity. For a disconnected graph G, we use the convention that the diameter of G is the maximum diameter of its connected components. We show that almost surely the diameter of random graph $G(n,p)$ is close to $\frac{\log n}{\log(np)}$ if $np \to \infty$. Moreover if $\frac{np}{\log n} = c > 8$, then the diameter of $G(n,p)$ is concentrated on two values. In general, if $\frac{np}{\log n} = c > c_0$, the diameter is concentrated on at most $2\lfloor \frac{1}{c_0} \rfloor + 4$ values. We also proved that the diameter of $G(n,p)$ is almost surely equal to the diameter of its giant component if $np > 3.6$.

1 Introduction

As the master of the art of counting, Erdős has had a far-reaching impact in numerous areas of mathematics and computer science. A recent example, perhaps least expected by Erdős, is the area of Internet computing. In a natural way, massive graphs that arise in the studies of the Internet share a number of similar aspects with random graphs, although there are significant differences (e.g., there can be vertices with large degrees in a sparse massive graph). Nevertheless, many of the methods and ideas [1, 2, 3, 3, 4, 6] that are used in modeling and analyzing massive graphs have been frequently traced to the seminal papers of Erdős and Rényi [13] in 1959.

One topic of considerable interest is to determine the diameter of a sparse random graph. These techniques and methods can also be used to examine the connected components and the diameter of Internet graphs [2, 15].

Let $G(n,p)$ denote a random graph on n vertices in which a pair of vertices appears as an edge of $G(n,p)$ with probability p. (The reader is referred to [8] for definition and notation in random graphs.) In this paper, we examine the diameter of $G(n,p)$ for all ranges of p including the range that $G(n,p)$ is not connected. For a disconnected graph G, the diameter of G is defined to be the diameter of its largest connected component.

We will first briefly survey previous results on the diameter of the random graph $G(n,p)$.

[*]University of California, San Diego
[†]Research supported in part by NSF Grant No. DMS 98-01446

In 1981, Klee and Larman [14] proved that for a fixed integer d, $G(n,p)$ has diameter d with probability approaching 1 as n goes to infinity if $(pn)^{d-1}/n \to 0$ and $(pn)^d/n \to \infty$. This result was later strengthened by Bollobás [7] and was proved earlier by Burtin [10, 11].

Bollobás [9] showed that the diameter of a random graph $G(n,p)$ is almost surely concentrated on at most four values if $pn - \log n \to \infty$. Furthermore, it was pointed out that the diameter of a random graph is almost surely concentrated on at most two values if $\frac{np}{\log n} \to \infty$ (see [8] exercise 2, chapter 10).

In the other direction, Łuczak [16] examined the diameter of the random graph for the case of $np < 1$. Łuczak determined the limit distribution of the diameter of the random graph if $(1-np)n^{1/3} \to \infty$. The diameter of $G(n,p)$ almost surely either is equal to the diameter of its tree components or differs by 1.

In this paper, we focus on random graphs $G(n,p)$ for the range of $np > 1$ and $np \le c \log n$ for some constant c. This range includes the emergence of the unique giant component. Since there is a phase transition in connectivity at $p = \log n/n$, the problem of determining the diameter of $G(n,p)$ and its concentration seems to be difficult for certain ranges of p. Here we intend to clarify the situation by identifying the ranges that results can be obtained as well as the ranges that the problems remain open.

For $\frac{np}{\log n} = c > 8$, we slightly improve Bollobás' result [9] by showing that the diameter of $G(n,p)$ is almost surely concentrated on at most two values around $\log n/\log np$. For $\frac{np}{\log n} = c > 2$, the diameter of $G(n,p)$ is almost surely concentrated on at most three values. For the range $2 \ge \frac{np}{\log n} = c > 1$, the diameter of $G(n,p)$ is almost surely concentrated on at most four values.

For the range $np < \log n$, the random graph $G(n,p)$ is almost surely disconnected. We will prove that almost surely the diameter of $G(n,p)$ is $(1+o(1))\frac{\log n}{\log(np)}$ if $np \to \infty$. Moreover, if $\frac{np}{\log n} = c > c_0$ for any (small) constant c and c_0, then the diameter of $G(n,p)$ is almost surely concentrated on finitely many values, namely, no more than $2\lfloor \frac{1}{c_0} \rfloor + 4$ values.

In the range of $\frac{1}{n} < p < \frac{\log n}{n}$, the random graph $G(n,p)$ almost surely has a unique giant component. We obtain a tight upper bound of the sizes of its small components if p satisfies $np \ge c > 1$. We then prove that the diameter of $G(n,p)$ almost surely equals the diameter of its giant component for the range $np > 3.513$. This problem was previously considered by Łuczak [16].

Here we summarize various results in the following table. The values of concentration for the diameter of $G(n,p)$, when it occurs, is near $\frac{\log n}{\log np}$. From the table, we can see that numerous questions remain, several of which will be discussed in the last section.

RANGE	$diam(G(n,p))$	REFERENCE
$\frac{np}{\log n} \to \infty$	Concentrated on at most 2 values	[8]
$\frac{np}{\log n} = c > 8$	Concentrated on at most 2 values	here
$8 \geq \frac{np}{\log n} = c > 2$	Concentrated on at most 3 values	here
$2 \geq \frac{np}{\log n} = c > 1$	Concentrated on at most 4 values	[9]
$1 \geq \frac{np}{\log n} = c > c_0$	Concentrated on at most $2\lfloor \frac{1}{c_0} \rfloor + 4$ values	here
$\log n > np \to \infty$	$diam(G(n,p)) = (1+o(1))\frac{\log n}{\log(np)}$	here
$np \geq c > 1$	The ratio $\frac{diam(G(n,p))}{\frac{\log n}{\log(np)}}$ is finite (between 1 and $f(c)$)	here
$np < 1$	$diam(G,p)$ equals the diameter of a tree component if $(1-np)n^{1/3} \to \infty$	[16]

Table 1: The diameter of random graphs $G(n,p)$.

2 The neighborhoods in a random graph

In a graph G, we denote by $\Gamma_k(x)$ the set of vertices in G at distance k from a vertex x:

$$\Gamma_k(x) = \{y \in G : d(x,y) = k\}$$

We define $N_k(x)$ to be the set of vertices within distance k of x:

$$N_k(x) = \cup_{i=0}^{k} \Gamma_i(x).$$

A main method to estimate the diameter of a graph is to examine the sizes of neighborhoods $N_k(x)$ and $\Gamma_k(x)$. To bound $|N_k(x)|$ in a random graph $G(n,p)$, the difficulties varies for different ranges of p. Roughly speaking, the sparser the graph is, the harder the problem is. We will first establish several useful lemmas concerning the neighborhoods for different ranges of p.

Lemma 1 *Suppose $np > 1$. With probability at least $1 - o(n^{-1})$, we have*

$$|\Gamma_i(x)| \leq 2i^2 \log n (np)^i \qquad \text{for all } 1 \leq i \leq n.$$

$$|N_i(x)| \leq 2i^3 \log n (np)^i \qquad \text{for all } 1 \leq i \leq n.$$

Lemma 2 *Suppose $p > \frac{c \log n}{n}$ for a constant $c \leq 2$. Then with probability at least $1 - o(n^{-1})$, we have*

$$|\Gamma_i(x)| \leq \frac{9}{c}(np)^i \qquad \text{for all } 1 \leq i \leq n$$

$$|N_i(x)| \leq \frac{10}{c}(np)^i \qquad \text{for all } 1 \leq i \leq n$$

Lemma 3 *Suppose $p \geq \frac{\log n}{n}$. For any $\varepsilon > 0$, with probability at least $1 - \frac{1}{\log^2 n}$, we have*

$$|\Gamma_i(x)| \leq (1+\varepsilon)(np)^i \qquad \textit{for all } 1 \leq i \leq \log n$$

$$|N_i(x)| \leq (1+2\varepsilon)(np)^i \qquad \textit{for all } 1 \leq i \leq \log n$$

Let X_1, X_2 denote two random variables. If $Pr[X_1 > a] \leq Pr[X_2 > a]$ for all a, we say X_1 dominates X_2, or X_2 is dominated by X_1. We will need the following fact.

Lemma 4 *Let $B(n,p)$ denote the binomial distribution with probability p in a space of size n.*

1. Suppose X dominates $B(n,p)$ For $a > 0$, we have

$$Pr(X < np - a) \leq e^{-\frac{a^2}{2np}} \tag{1}$$

2. Suppose X is dominated by $B(n,p)$. For $a > 0$, we have

$$Pr(X > np + a) \leq e^{-\frac{a^2}{2np} + \frac{a^3}{(np)^3}} \tag{2}$$

We will repeatedly use Lemma 4 in the following way. For a vertex x of $G(n,p)$, we consider $\Gamma_i(x)$ for $i = 1, 2, \ldots$. At step i, let X be the random variable of $|\Gamma_i(x)|$ given $|\Gamma_{i-1}(x)|$. We note that X is not exactly a binomial distribution. However, it is close to one if $|\Gamma_{i-1}(x)|$ is small. To be precise, X is dominated by a random variable with the binomial distribution $B(t,p)$ where $t = n|\Gamma_{i-1}(x)|$. On the other hand, if $|N_i(x)| < m$, then X dominates a random variable $B(t',p)$ where $t' = (n-m)|\Gamma_{i-1}(x)|$. Thus an upper bound and lower bound of $|\Gamma_i(x)|$ can be obtained. For different ranges of p, we will derive different estimates in Lemmas 1 - 3.

Proof of Lemma 1: We consider p satisfying $np > 1$. We want to show that with probability at least $1 - o(n^{-1})$, we have

$$|\Gamma_i(x)| \leq 2i^2 \log n (np)^i \qquad \text{for all } 1 \leq i \leq n.$$

$$|N_i(x)| \leq 2i^3 \log n (np)^i \qquad \text{for all } 1 \leq i \leq n.$$

First we will establish the following:

Claim 1 :

With probability at least $1 - ie^{-\lambda^2/2 + \frac{\lambda^3}{(\log n)^{1.5}}}$, we have

$$|\Gamma_i(x)| \leq a_i \log n (np)^i \qquad \text{for all } 1 \leq i \leq n.$$

where a_i $(1 \leq i \leq k)$ satisfies recurrence formula,

$$a_i = a_{i-1} + \frac{\lambda}{\sqrt{\log n}} \frac{\sqrt{a_{i-1}}}{(np)^{i/2}} \qquad \text{for all } 1 \leq i \leq n$$

with initial condition $a_0 = 1$.

We prove this claim by induction on i. Clearly, for $i = 0$, $|\Gamma_0(x)| = 1 < \log n$, it is true. Suppose that it holds for i. For $i+1$, $|\Gamma_{i+1}(x)|$ is dominated by the binomial distribution $B(t, p)$ where $t = |\Gamma_i(x)|(n - |N_i(x)|)$. With probability at least $1 - ie^{-\lambda^2/2 + \frac{\lambda^3}{(\log n)^{1.5}}}$, we have

$$|\Gamma_i(x)|(n - |N_i(x)|) \le a_i \log n (np)^i n.$$

By Lemma 4 inequality (2), with probability at least $1-(i+1)e^{-\lambda^2/2+\frac{\lambda^3}{(\log n)^{1.5}}}$ (since $a_i \log n(np)^{i+1} > \log n$), we have

$$\begin{aligned} |\Gamma_{i+1}(x)| &\le a_i \log n (np)^i np + \lambda \sqrt{a_i \log n (np)^{i+1}} \\ &\le \log n (np)^{i+1} \left(a_i + \frac{\lambda}{\sqrt{\log n}} \frac{\sqrt{a_i}}{(np)^{(i+1)/2}}\right) \\ &= a_{i+1} \log n (np)^{i+1} \end{aligned}$$

By choosing $\lambda = \sqrt{5 \log n}$, we have

$$1 - ne^{-\lambda^2/2+\frac{\lambda^3}{(\log n)^{1.5}}} = 1 - ne^{-2.5 \log n + 5^{1.5}} = 1 - o(n^{-1}).$$

Now we show by induction that $a_i \le 2i^2$ for all $1 \le i \le n$. Suppose that $a_j \le 2j^2$, for all $1 \le j \le i$. Then

$$\begin{aligned} a_{i+1} &= 1 + \frac{\lambda}{\sqrt{\log n}} \sum_{j=0}^{i} \frac{\sqrt{a_j}}{(np)^{(j+1)/2}} \\ &< 1 + \sqrt{5}\left(1 + \sum_{j=1}^{i} \sqrt{a_j}\right) \\ &\le 1 + \sqrt{5}\left(1 + \sum_{j=1}^{i} \sqrt{2} j\right) \\ &\le 1 + \sqrt{5}\left(1 + \sqrt{2}(i^2 + i)/2\right) \\ &< 2(i+1)^2 \end{aligned}$$

We have completed the proof of Lemma 1. $\square$

For $p > \frac{c \log n}{n}$, where $c \le 2$ is a constant, the upper bound for $|\Gamma_i(x)|$ can be improved.

Proof of Lemma 2:

We here focus on the range $p > \frac{c \log n}{n}$ for a constant $c \le 2$. We want to show that with probability at least $1 - o(n^{-1})$, we have

$$|\Gamma_i(x)| \le \frac{9}{c}(np)^i \qquad \text{for all } 1 \le i \le n,$$
$$|N_i(x)| \le \frac{10}{c}(np)^i \qquad \text{for all } 1 \le i \le n.$$

We will first prove the following claim.

Claim 2: With probability at least $1 - ie^{-\lambda^2/2 + \frac{\lambda^3}{(np)^{1.5}}}$, we have

$$|\Gamma_i(x)| \le a_i(np)^i \qquad \text{for all } 1 \le i \le n.$$

where a_i $(1 \le i \le n)$ satisfies the following recurrence formula,

$$a_i = a_{i-1} + \lambda \frac{\sqrt{a_{i-1}}}{(np)^{i/2}} \qquad \text{for all } 1 \le i \le n$$

with initial condition $a_0 = 1$.

Obviously, for $i = 0$, $|\Gamma_0(x)| = 1 = a_0$, it holds. Suppose that it holds for i. For $i+1$, $|\Gamma_{i+1}(x)|$ is dominated by the binomial distribution $B(t,p)$ where $t = |\Gamma_i(x)|(n - |N_i(x)|)$. With probability at least $1 - ie^{-\lambda^2/2 + \frac{\lambda^3}{(np)^{1.5}}}$, we have

$$|\Gamma_i(x)|(n - |N_i(x)|) \le a_i(np)^i n.$$

By Lemma 4 inequality (2), with probability at least $1 \ge 1 - (i+1)e^{-\lambda^2/2 + \frac{\lambda^3}{(np)^{1.5}}}$ (since $a_i(np)^{i+1} > np$), we have

$$\begin{aligned}
|\Gamma_{i+1}(x)| &\le a_i(np)^i np + \lambda\sqrt{a_i(np)^{i+1}} \\
&\le (np)^{i+1}(a_i + \frac{\lambda\sqrt{a_i}}{(np)^{(i+1)/2}}) \\
&= a_{i+1}(np)^{i+1}
\end{aligned}$$

We choose $\lambda = \sqrt{5\log n}$ and we have

$$1 - ne^{-\lambda^2/2 + \frac{\lambda^3}{(pn)^{1.5}}} = 1 - ne^{-2.5\log n + (\frac{5}{c})^{1.5}} = 1 - o(n^{-1}).$$

Now we show by induction that $a_i \le \frac{9}{c}$. Suppose that $a_j \le \frac{9}{c}$, for all $1 \le j \le i$. Then

$$\begin{aligned}
a_{i+1} &= 1 + \lambda \sum_{j=0}^{i} \frac{\sqrt{a_j}}{(np)^{(j+1)/2}} \\
&\le 1 + \sqrt{5\log n} \sum_{j=0}^{\infty} \frac{\sqrt{9}}{\sqrt{c}(np)^{(j+1)/2}} \\
&= 1 + \sqrt{45} \frac{\sqrt{\log n}}{\sqrt{c}(\sqrt{np} - 1)} \\
&\le 1 + \sqrt{45} \frac{\sqrt{\log n}}{\sqrt{c}(\sqrt{c\log n} - 1)} \\
&\le 1 + \frac{7}{c} \\
&\le \frac{9}{c}
\end{aligned}$$

for $c \leq 2$. Thus,

$$|N_i(x)| = \sum_{j=0}^{i} |\Gamma_i(x)| \leq \sum_{j=0}^{i} \frac{9}{c}(np)^j \leq \frac{10}{c}(np)^i$$

by using the fact that $np \geq c \log n$. Lemma 2 is proved. □

If we only require having probability $1 - o(1)$ instead, the preceding upper bound can be strengthened as follows.

Proof of Lemma 3: Suppose $p \geq \frac{\log n}{n}$. We want to show that for any $k < \log n$ and any $\varepsilon > 0$, with probability at least $1 - \frac{1}{\log^2 n}$, we have

$$|\Gamma_i(x)| \leq (1+\varepsilon)(np)^i \qquad \text{for all } 1 \leq i \leq k$$

$$|N_i(x)| \leq (1+2\varepsilon)(np)^i \qquad \text{for all } 1 \leq i \leq k$$

provided n is large enough.

We will first show the following:

Claim 3: With probability at least $1 - ie^{-\lambda^2/2 + \frac{\lambda^3}{(np)^{1.5}}}$, we have

$$|\Gamma_i(x)| \leq a_i(np)^i \qquad \text{for all } 1 \leq i \leq n.$$

where a_i $(1 \leq i \leq n)$ satisfies recurrence formula,

$$a_i = a_{i-1} + \lambda \frac{\sqrt{a_{i-1}}}{(np)^{i/2}} \qquad \text{for all } 1 \leq i \leq n$$

with initial condition $a_0 = 1$.

By choosing $\lambda = 3\sqrt{\log \log n}$, we have

$$1 - ke^{-\frac{\lambda^2}{2} + \frac{\lambda^3}{2(pn)^{1.5}}} = 1 - ko(\frac{1}{\log^4 n}) = 1 - o(\frac{1}{\log^3 n}).$$

since $np \geq \log n$.

By induction, we will prove

$$a_i < (1+\varepsilon) \qquad \text{for all } 0 \leq i \leq k.$$

Certainly it holds for $i = 0$, since $a_0 = 1 < 1 + \varepsilon$.

Suppose that $a_j < 1+\varepsilon$, for all $1 \le j \le i$. Then

$$\begin{aligned} a_{i+1} &= 1+\lambda \sum_{j=0}^{i} \frac{\sqrt{a_j}}{(np)^{(j+1)/2}} \\ &< 1+\lambda \sum_{j=0}^{i} \frac{\sqrt{1+\varepsilon}}{(np)^{(j+1)/2}} \\ &\le 1+\lambda\sqrt{1+\varepsilon}\frac{1}{\sqrt{np}-1} \\ &\le 1+\varepsilon \end{aligned}$$

by using the assumption $\lambda = 3\sqrt{\log\log n} = o(\sqrt{np})$.

Therefore, with probability at least $1 - o(\frac{1}{\log^3 n})$, we have

$$|\Gamma_i(x)| \le (1+\varepsilon)(np)^i \qquad \text{for all } 1 \le i \le k$$

and

$$\begin{aligned} |N_i(x)| &= 1+\sum_{j=1}^{i} |\Gamma_j(x)| \\ &\le 1+(1+\varepsilon)\sum_{j=1}^{i}(np)^j \\ &= 1+(1+\varepsilon)\frac{(np)^{i+1}-np}{np-1} \\ &\le (1+2\varepsilon)(np)^i \end{aligned}$$

for n large enough. □

3 The diameter of the giant component

Łuczak asked the interesting question of determining if the diameter of the giant component is the diameter of a random graph $G(n,p)$. We will answer this question for certain ranges of p. This result is needed later in the proof of the main theorems.

Lemma 5 *Suppose $1 < c \le np < logn$, for some constant c. Then almost surely the sizes of all small components are at most*

$$(1+o(1))\frac{\log n}{np-1-\log(np)}$$

Proof: When $p > 1 + 2(2\log n)^{1/2} n^{-1/3}$, Bollobás [8] shows that a component of size at least $n^{2/3}$ in $G_{n,p}$ is almost always unique (so that it is the giant component) in the sense that all other components are at most of size $n^{2/3}/2$. Suppose that x is not in the giant component. We compute the probability that x lies in a component of size $k+1 < n^{\frac{2}{3}}$. Such a connected component must contain a spanning tree. There are $\binom{n-1}{k}$ ways to select other k vertices. For these $k+1$ vertices, there are exactly $(k+1)^{k-1}$ spanning trees rooted at x. Hence, the probability that a spanning tree exists is at most

$$\binom{n-1}{k}(k+1)^{k-1}p^k(1-p)^{k(n-n^{2/3})} < \sqrt{2\pi k}e^k(np)^k e^{-knp(1-n^{-1/3})}$$

The above probability is $o(n^{-2})$ if $k > \frac{3\log n}{np-1-\log(np)}$. It is $o(n^{-1}e^{-\sqrt{\log n}})$ if $k > \frac{\log n + 2\sqrt{\log n}}{np-1-\log(np)}$. Hence, the probability that x lies in a component of size $k+1 \geq \frac{\log n+2\sqrt{\log n}}{np-1-\log(np)}$ is at most

$$n \times o(n^{-2}) + \frac{3\log n}{np-1-\log(np)} \times o(n^{-1}e^{-\sqrt{\log n}}) = o(n^{-1}).$$

This implies that almost surely all small components is of size at most

$$\frac{\log n + 2\sqrt{\log n}}{np-1-\log(np)} = (1+o(1))\frac{\log n}{np-1-\log(np)}.$$

Theorem 1 *Suppose that $np > 3.513$, then almost surely the diameter of $G(n,p)$ equals the diameter of its giant component.*

Proof: From Lemma 5, the diameter of small components is at most $(1+o(1))\frac{\log n}{np-1-\log(np)}$. On the other hand, by Lemma 1, for any vertex x, with probability at least $1-o(n^{-2})$,

$$|N_i(x)| = \sum_{j=0}^{i} |\Gamma_j(x)| \leq 2i^3 \log n (np)^i.$$

This implies the diameter of $G(n,p)$ is at least

$$(1+o(1))\frac{\log n}{\log np}.$$

When $np > 3.513$, we have $np - 1 - \log(np) > \log(np)$. Hence, the diameter of $G(n,p)$ is strictly greater than the sizes of all small components. This completes the proof of the theorem. $\square$

We can now prove a lower bound for $|\Gamma_i(x)|$.

Lemma 6 *Suppose $np \geq c > 1$ with some constant c. For each vertex x in the giant component (if $G(n,p)$ is not connected), with probability at least $1-o(n^{-1})$, we have*

$$|\Gamma_i(x)| \geq \frac{1}{(\sqrt{np}-1)^2}(np)^{i-i_0}\log n$$

for i satisfying $i_0 \leq i \leq \frac{3}{5}\frac{\log n}{\log(np)}$ where $i_0 = \frac{(\frac{10np}{(\sqrt{np}-1)^2}+1)\log n}{np-\log(2np)}$.

Proof: First we prove that with probability at least $1 - o(n^{-1})$, there exists a i_0 satisfying

$$|\Gamma_{i_0}(x)| \geq \frac{9 \log n}{(\sqrt{np} - 1)^2} = d.$$

If $i \leq \frac{3}{5} \frac{\log n}{\log(np)}$, then by Lemma 1, with probability at least $1 - o(n^{-1})$, we have $|\Gamma_i(x)| \leq n^{2/3}$. Now we compute the probability that $|N_i(x)| = k + 1 < n^{2/3}$. We want to show for some k_0, the probability that $|\Gamma_i(x)| < d$ and $|N_i(x)| > k_0$ is $o(n^{-1})$.

We focus on the neighborhood tree formed by breadth-first-search starting at x. There are $\binom{n-1}{k}$ ways to select other k vertices. For these $k + 1$ vertices, there are exactly $(k + 1)^{k-1}$ trees rooted at x. Suppose $|\Gamma_i(x)| < d$. The probability that such a tree exists is at most

$$\binom{n-1}{k}(k+1)^{k-1}p^k(1-p)^{(k-d)(n-n^{2/3})} < e^k(np)^k e^{-(k-d)np(1-n^{-1/3})}$$

Let $k_0 = \frac{dnp + \log n + 2\sqrt{\log n}}{np - 1 - \log(np)}$. The above probability is $o(n^{-2})$ if $k > \frac{dnp + 3\log n}{np - 1 - \log(np)}$. It is $o(n^{-1}e^{-\sqrt{\log n}})$ if $k > k_0$. Hence, the probability that $|\Gamma_i|(x)| < d$ and $|N_i(x)| = k + 1 > k_0 + 1$ is at most

$$n \times o(n^{-2}) + \frac{dnp + 3\log n}{np - 1 - \log(np)} \times o(n^{-1}e^{-\sqrt{\log n}}) = o(n^{-1}).$$

Let i_0 denote the least integer i satisfying $|\Gamma_i(x)| \geq d$. The above arguments give a crude upper bound for i_0.

$$i_0 \leq k_0 = \frac{dnp + \log n + 2\sqrt{\log n}}{np - 1 - \log(np)} \leq \frac{(\frac{10np}{(\sqrt{np}-1)^2} + 1)\log n}{np - 1 - \log(np)}.$$

Now, we want to prove that $|\Gamma_i(x)|$ grows quickly after $i = i_0$. Namely, with probability at least $1-o(n^{-1})$, we have

$$|\Gamma_i(x)| \geq \frac{1}{(\sqrt{np} - 1)^2}(np)^{i-i_0} \log n$$

for all i satisfying $\frac{3}{5} \frac{\log n}{\log(np)} \geq i > i_0$.

Claim 4: With probability at least $1 - o(n^{-1}) - (i - i_0)e^{-\lambda^2/2}$, we have

$$|\Gamma_i(x)| \geq a_i(np(1 - n^{-1/3}))^{i-i_0} \log n$$

for all $i_0 \leq i \leq \frac{3}{5} \frac{\log n}{\log(np)}$. Here a_i satisfies the following recurrence formula,

$$a_i = a_{i-1} - \frac{\lambda}{\sqrt{\log n}} \frac{\sqrt{a_{i-1}}}{(np(1 - n^{-1/3}))^{(i-i_0)/2}}$$

for all $i_0 \leq i \leq \frac{3}{5} \frac{\log n}{\log(np)}$, with initial condition $a_{i_0} = \frac{\lambda^2}{\log n} \frac{1.7}{(\sqrt{np(1-n^{-1/3})}-1)^2}$.

We choose $\lambda = \sqrt{5 \log n}$. Clearly, for $i = i_0$, $|\Gamma_{i_0}(x)| \geq d \geq a_{i_0}$, the statement of the claim is true. Suppose that it holds for i. For $i + 1$, $|\Gamma_{i+1}(x)|$ dominates a random variable with the binomial distribution

$B(t,p)$ where $t = |\Gamma_i(x)|(n - n^{-2/3})$ with probability at least $1 - o(n^{-1}) - ie^{-\lambda^2/2}$. By Lemma 4 part 1, with probability at least $1 \geq 1 - (i+1)e^{-\lambda^2/2}$, we have

$$\begin{aligned}
|\Gamma_{i-i_0+1}(x)| &\geq a_i(np(1-n^{-2/3}))^{i-i_0}(\log n)np(1-n^{-2/3}) - \lambda\sqrt{a_i(np(1-n^{-2/3}))^{i-i_0+1}\log n} \\
&\geq (np(1-n^{-2/3}))^{i-i_0+1}\log n(a_i - \frac{\lambda\sqrt{a_i}}{\sqrt{\log n}(np(1-n^{-2/3}))^{(i-i_0+1)/2}}) \\
&= a_{i+1}(np(1-n^{-2/3}))^{i+1}
\end{aligned}$$

Here,

$$1 - o(n^{-1}) - ie^{-\lambda^2/2} = 1 - o(n^{-1}) - ne^{-2.5\log n} = 1 - o(n^{-1}).$$

Since $a_i < a_{i_0}$ for $i > i_0$, we have

$$\begin{aligned}
a_i &= a_{i_0} - \frac{\lambda}{\sqrt{\log n}}\sum_{j=i_0}^{i-1}\frac{\sqrt{a_j}}{(np(1-n^{-2/3}))^{(j-i_0+1)/2}} \\
&\geq a_{i_0} - \sqrt{5}\sum_{j=i_0}^{i-1}\sqrt{a_{i_0}}(np(1-n^{-2/3}))^{(j+1)/2} \\
&\geq a_{i_0} - \sqrt{5a_{i_0}}\frac{1}{\sqrt{np(1-n^{-2/3})}-1} \\
&\geq \frac{2}{(\sqrt{np(1-n^{-2/3})}-1)^2}
\end{aligned}$$

Hence, for $i \geq i_0$,

$$\begin{aligned}
|\Gamma_i(x)| &\geq a_i(np(1-n^{-1/3}))^{i-i_0}\log n \\
&\geq \frac{2}{(\sqrt{np(1-n^{-2/3})}-1)^2}(np(1-n^{-1/3}))^{i-i_0}\log n \\
&\geq \frac{1}{(\sqrt{np}-1)^2}(np)^{i-i_0}\log n
\end{aligned}$$

□

If $np > c\log n$, the statement in Lemma 6 can be further strengthened.

Lemma 7 *Suppose $p \geq \frac{c\log n}{n}$ for some constant $c \leq 2$. Then, for each vertex x in the giant component (if $G(n,p)$ is not connected), for each i satisfying $i_0 \leq i \leq \frac{2}{3}\frac{n}{\log(np)}$, with probability at least $1 - o(n^{-1})$, we have*

$$|\Gamma_i(x)| \geq \frac{5}{c}(np)^{i-i_0}$$

where i_0 satisfies $i_0 \leq \lfloor\frac{1}{c}\rfloor + 1$.

Proof: We first prove the following statement, which is similar to the claim in the proof of previous lemma. However, we use a different proof here to obtain an improvement.

With the probability at least $1-o(n^{-1})$, there exists a $i_0 \leq \lfloor \frac{1}{c} \rfloor + 1$ satisfying

$$|\Gamma_{i_0}(x)| \geq d,$$

where $d = \frac{20}{c}$.

Let $k = \lfloor \frac{1}{c} \rfloor$. Since x is in the giant component, $|\Gamma_k(x)| \geq 1$. There exists a path $xx_1 \ldots x_k$ satisfying $x_j \in \Gamma_j(x)$ for $1 \leq j \leq k$. We write $x_0 = x$. Let $f(x_j)$ denote the number of vertices y, which $x_j y$ forms a edge but y is not one of those vertices $x_0, x_1, \ldots, x_k$. We compute the probability that $f(x_j) \leq d$ as follows.

$$\begin{aligned}
Pr(f(x_j) \leq d) &= \sum_{l=0}^{d} \binom{n-k-1}{l} p^l (1-p)^{n-l} \\
&\leq \sum_{l=0}^{d} \frac{(np)^l}{l!} e^{-(n-l-k-1)p} \\
&\leq (np)^d e^{-(n-d-k-1)p} \sum_{l=0}^{d} \frac{1}{l!} \\
&\leq (c \log n)^d e^{-c(1-\frac{d+k+1}{n}) \log n} e \\
&= o(n^{-c+\epsilon})
\end{aligned}$$

for any small $\varepsilon > 0$.

Here, $f(x_j)$'s are independent random variables. The probability that $f(x_j) \leq d$ for all $0 \leq j \leq k$ is at most

$$o((n^{-c+\epsilon})^{k+1}) = o(n^{-1})$$

if ε is small enough.

With probability at least $1-o(n^{-1})$, there is an index $1 \leq i_0 \leq k+1$ satisfying $f(x_{i_0-1}) \geq d$. Hence, $\Gamma_{i_0}(x) \geq d$.

By Lemma 1, with probability at least $1-o(n^{-1})$, we have $|N_i(x)| \leq n^{\frac{3}{4}}$ for all $1 \leq i \leq \frac{2}{3} \frac{\log n}{\log(np)}$.

For $i = i_0 + 1$, we have

$$\begin{aligned}
Pr\left(\Gamma_{i_0+1}(x) \leq \frac{1}{2}|\Gamma_{i_0}(x)|(n-|N_{i_0}(x)|)p\right) &\leq e^{-|\Gamma_{i_0}(x)|(n-|N_{i_0}(x)|)p/8} \\
&\leq e^{-dc(1-n^{-1/4}) \log n/8} \\
&= o(n^{-dc/9}) \\
&= o(n^{-1})
\end{aligned}$$

since $d \geq \frac{10}{c}$.

Hence with probability at least $1 - o(n^{-1})$,

$$|\Gamma_{i_0+1}(x)| \geq \frac{1}{2}|\Gamma_{i_0}(x)|(n - |N_{i_0}(x)|)p \geq \frac{1}{3}dnp$$

For $i = i_0 + 2$, $|\Gamma_{i_0+2}(x)|$ dominates a random variable with the binomial distribution $B(t, p)$ where $t = |\Gamma_{i_0+1}(x)(n - |N_{i_0+1}(x)|)$. Hence

$$Pr\left(|\Gamma_{i_0+2}(x)| < \Gamma_{i_0+1}(x)(n - |N_{i_0+1}(x)|)p - \lambda\sqrt{\Gamma_{i_0+1}(x)(n - |N_{i_0+1}(x)|)p}\right) < e^{-\frac{\lambda^2}{2}}$$

Hence, with probability at least $1 - o(n^{-1}) - e^{-\frac{\lambda^2}{2}}$,

$$\begin{aligned}
|\Gamma_{i_0+2}(x)| &\geq \Gamma_{i_0+1}(x)(n - |N_{i_0+1}(x)|)p - \lambda\sqrt{\Gamma_{i_0+1}(x)(n - |N_{i_0+1}(x)|)p} \\
&\geq \Gamma_{i_0+1}(x)(n - n^{3/4})p - \lambda\sqrt{\Gamma_{i_0+1}(x)np} \\
&\geq \frac{1}{3}d(np)^2(1 - n^{-1/4} - \frac{3\lambda}{\sqrt{(np)^2}})
\end{aligned}$$

By induction on $i \geq i_0 + 2$, we can show that with probability at least $1 - o(n^{-1}) - ie^{-\frac{\lambda^2}{2}}$,

$$|\Gamma_i(x)| \geq \frac{d}{3}(np)^{i-i_0}\prod_{j=2}^{i-i_0}(1 - n^{-1/3} - \frac{3\lambda}{\sqrt{(np)^j}})$$

We choose $\lambda = \sqrt{3\log n}$. Since $i < \log n$, we have

$$1 - o(n^{-1}) - (i - i_0)e^{-\frac{\lambda^2}{2}} = 1 - o(n^{-1}) - in^{-1.5} = 1 - o(n^{-1})$$

Therefore, with probability at least $1 - o(n^{-1})$,

$$\begin{aligned}
|\Gamma_i(x)| &\geq \frac{d}{3}(np)^{i-i_0}(1 - in^{-1/4} - \sum_{j=2}^{i-i_0}\frac{3\lambda}{\sqrt{(np)^j}}) \\
&\geq \frac{d}{3}(np)^{i-i_0}(1 - in^{-1/4} - \frac{3\lambda}{(np)}\frac{1}{1 - (np)^{-1/2}}) \\
&\geq \frac{d}{3}(np)^{i-i_0}(1 - O(\frac{1}{\sqrt{\log(n)}})) \\
&\geq \frac{d}{4}(np)^{i-i_0} \\
&= \frac{5}{c}(np)^{i-i_0}
\end{aligned}$$

for n large enough. □

Lemma 8 *Suppose $p \geq \frac{c\log n}{n}$ for some constant $c > 2$. For each vertex x belonging to the giant component (if $G(n,p)$ is not connected), and each i satisfying $1 \leq i \leq \frac{2}{3}\frac{n}{\log(np)}$, with probability at least $1 - o(\frac{1}{n})$, we have*

$$|\Gamma_i(x)| \geq c_1(np)^i$$

where $c_1 = 1 - \sqrt{\frac{2}{c}} - \varepsilon$.

Proof: Let δ be a small positive number. For $i = 1$, we have

$$\begin{aligned} Pr\left(\Gamma_1(x) \leq (c_1+\delta)np\right) &= Pr\left(\Gamma_1(x) \leq np - (1-c_1-\delta)np\right) \\ &\leq e^{-(1-c_1-\delta)^2 np/2} \\ &\leq e^{-(1-c_1-\delta)^2 c\log n/2} \\ &= n^{-(1-c_1-\delta)^2 c/2} \\ &= o(n^{-1}) \end{aligned}$$

where δ is a small value satisfying $(1-c_1-\delta)^2 c/2 > 1$. (It is always possible to choose such a $\delta > 0$, by the assumption on c_1). Hence with probability at least $1 - o(n^{-2})$, we have

$$|\Gamma_1(x)| \geq (c_1+\delta)np.$$

To obtain a better concentration result in the range of $c > 8$, more work is needed here. However, the arguments are similar to those in Lemmas 6 and 7. For $i = 2$, $|\Gamma_2(x)|$ dominates a random variable with the binomial distribution $B(t,p)$ where $t = |\Gamma_1(x)|(n - n^{1/4})$. We have

$$Pr\left(|\Gamma_2(x)| < \Gamma_1(x)(n-n^{1/4})p - \lambda\sqrt{\Gamma_1(x)(n-n^{1/4})p}\right) < e^{-\frac{\lambda^2}{2}}$$

Hence, with probability at least $1 - o(n^{-1}) - e^{-\frac{\lambda^2}{2}}$, we have

$$|\Gamma_2(x)| \geq (c_1+\delta)(np)^2(1 - n^{-1/4} - \frac{\lambda}{\sqrt{c_1(np)^2}})$$

By induction on $i \geq 2$, it can be shown that with probability at least $1 - o(n^{-2}) - ie^{-\frac{\lambda^2}{2}}$,

$$|\Gamma_i(x)| \geq (c_1+\delta)(np)^i \prod_{j=2}^{i}(1 - n^{-1/4} - \frac{\lambda}{\sqrt{c_1(np)^j}})$$

By choosing $\lambda = \sqrt{5\log n}$, we have

$$1 - o(n^{-1}) - ie^{\frac{\lambda^2}{2}} = 1 - o(n^{-1}) - in^{-2.5} = 1 - o(n^{-1})$$

since $i < \log n$.

Therefore, with probability at least $1 - o(n^{-1})$, we have

$$
\begin{aligned}
|\Gamma_i(x)| &\geq (c_1+\delta)(np)^i(1 - in^{-1/4} - \sum_{j=2}^{i} \frac{\lambda}{\sqrt{c_1(np)^j}}) \\
&\geq (c_1+\delta)(np)^i(1 - in^{-1/4} - \frac{\lambda}{\sqrt{c_1}(np)} \frac{1}{1-(np)^{-1/2}}) \\
&\geq (c_1+\delta)(np)^i(1 - O(\frac{1}{\sqrt{\log(n)}})) \\
&\geq c_1(np)^i
\end{aligned}
$$

for n large enough.

4 The main theorems

We first state the main theorems that we will prove in this section:

Theorem 2 *If $p \geq \frac{c \log n}{n}$ for some constant $c > 8$, the diameter of random graph $G(n,p)$ is almost surely concentrated on at most two values at $\frac{\log n}{\log np}$.*

Theorem 3 *If $p \geq \frac{c \log n}{n}$ for some constant $c > 2$, then the diameter of random graphs $G(n,p)$ is almost surely concentrated on at most three values at $\frac{\log n}{\log np}$.*

Theorem 4 *If $p \geq \frac{c \log n}{n}$ for some constant c, then we have*

$$
\left\lceil \frac{\log(\frac{cn}{11})}{\log(np)} \right\rceil \leq diam(G(n,p)) \leq \left\lceil \frac{\log(\frac{33c^2}{400} n \log n)}{\log(np)} \right\rceil + 2\lfloor \frac{1}{c} \rfloor + 2.
$$

The diameter of random graph $G(n,p)$ is almost surely concentrated on at most $2\lfloor \frac{1}{c} \rfloor + 4$ values.

Theorem 5 *If $\log n > np \to \infty$, then almost surely we have*

$$
diam(G(n,p)) = (1+o(1))\frac{\log n}{\log np}
$$

Theorem 6 *Suppose $np \geq c > 1$ for some constant c. Almost surely we have*

$$
(1+o(1))\frac{\log n}{\log np} \leq diam(G(n,p)) \leq \frac{\log n}{\log np} + 2\frac{(10c/(\sqrt{c}-1)^2+1)}{c - \log(2c)} \frac{\log n}{np} + 1
$$

Before proving Theorems 2 - 6, we first state two easy observations that are useful for establishing upper and lower bounds for the diameter.

Observation 1: Suppose there is an integer k, satisfying one of the following two conditions

1. When $G(n,p)$ is connected, there exists a vertex x satisfying, almost surely

$$|N_k(x)| < (1-\varepsilon)n.$$

2. When $G(n,p)$ is not connected, almost surely for all vertices x

$$|N_k(x)| < n^{1-\varepsilon}$$

(Here $n^{1-\varepsilon}$ can be replaced by any lower bound of the giant component.) Then we have

$$diam(G(n,P)) > k.$$

Observation 2: Suppose there are integers k_1 and k_2, satisfying

$$|\Gamma_{k_1}(x)||\Gamma_{k_2}(x)|p > (2+\epsilon)\log n$$

for all pairs of vertices (x,y) in the giant component. If $\Gamma_{k_1}(x) \cap \Gamma_{k_2}(x) \neq \emptyset$, then $d(x,y) \leq k_1 + k_2$. If $\Gamma_{k_1}(x) \cap \Gamma_{k_2}(x) = \emptyset$, the probability that there is edge between them is at least

$$1-(1-p)^{|\Gamma_{k_1}(x)||\Gamma_{k_2}(x)|} \geq 1 - e^{-|\Gamma_{k_1}(x)||\Gamma_{k_2}(x)|p} = 1 - o(n^{-2})$$

Since there are at most n^2 pairs, almost surely

$$d(x,y) \leq k_1 + k_2 + 1.$$

Thus the diameter of the giant component is at most $k_1 + k_2 + 1$.

Proof of Theorem 2: $G(n,p)$ is almost surely connected at this range. By Lemma 3, almost surely there is a vertex x satisfying

$$|N_i(x)| \leq (1+2\varepsilon)(np)^i \qquad \text{for all } 1 \leq i \leq \log n$$

Here, we choose $k = \lfloor \frac{\log(\frac{1-\varepsilon}{1+2\varepsilon}n)}{\log(np)} \rfloor$. Hence almost surely, we have

$$diam(G(n,p)) \geq \lceil \frac{\log(\frac{1-\varepsilon}{1+2\varepsilon}n)}{\log(np)} \rceil \qquad \text{for any } \varepsilon$$

by using observation 1.

On the other hand, by Lemma 8, almost surely for all vertices x,

$$|\Gamma_i(x)| \geq c_1(np)^i$$

where $c_1 = 1 - \sqrt{\frac{2}{c}} - \varepsilon$.

Now we choose

$$k_1 = \lceil \frac{\log(\frac{\sqrt{2(1+\varepsilon)n\log n}}{c_1})}{\log(np)} \rceil \quad \text{and} \quad k_2 = \lceil \frac{\log(\frac{2(1+\varepsilon)n\log n}{c_1^2})}{\log(np)} - k_1 - 1 \rceil$$

as in observation 2. We note that $k_1 \approx k_2 \approx \frac{1}{2}\frac{\log(\frac{2n\log n}{c_1^2})}{\log(np)} < \frac{2}{3}\frac{\log n}{\log(np)}$ both satisfy the condition of Lemma 8. Almost surely we have

$$|\Gamma_{k_1}(x)||\Gamma_{k_2}(y)|p \geq c_1(np)^{k_1}c_1(np)^{k_2}p \geq 2(1+\varepsilon)\log n$$

Hence, we have

$$diam(G(n,p)) \leq k_1 + k_2 + 1 = \left\lceil \frac{\log(\frac{2(1+\varepsilon)n\log n}{c_1^2})}{\log(np)} \right\rceil.$$

Therefore, we have proved that almost surely

$$\left\lceil \frac{\log(\frac{1-\varepsilon}{1+2\varepsilon}n)}{\log(np)} \right\rceil \leq diam(G(n,p)) = \left\lceil \frac{\log(\frac{2(1+\varepsilon)}{c_1^2}n\log n)}{\log(np)} \right\rceil \qquad \text{for any } \varepsilon.$$

The difference of the upper bound and lower bound is at most

$$\begin{aligned}
\left\lceil \frac{\log(\frac{2(1+\varepsilon)}{c_1^2}n\log n)}{\log(np)} \right\rceil - \left\lceil \frac{\log(\frac{1-\varepsilon}{1+2\varepsilon}n)}{\log(np)} \right\rceil &\leq \left\lceil \frac{\log(\frac{2(1+\varepsilon)(1+2\varepsilon)\log n}{c_1^2(1-\varepsilon)})}{\log(np)} \right\rceil \\
&\leq \left\lceil \frac{\log(\frac{2(1+\varepsilon)(1+2\varepsilon)\log n}{c_1^2(1-\varepsilon)})}{\log(c\log n)} \right\rceil \\
&\leq 1
\end{aligned}$$

when $\varepsilon \to 0$.

Therefore, the diameter of $G(n,p)$ is concentrated on at most two values in this range. □

Proof of Theorem 3: The proof is quite similar to that of Theorem 2 and will be omitted. It can be shown that

$$\left\lceil \frac{\log(\frac{1-\varepsilon}{1+2\varepsilon}n)}{\log(np)} \right\rceil \leq diam(G(n,p)) = \left\lceil \frac{\log(\frac{2(1+\varepsilon)}{c_1^2}n\log n)}{\log(np)} \right\rceil \qquad \text{for any } \varepsilon.$$

It is not difficult to check that in this range the difference between upper bound and lower bound is 2 instead of 1, for $c > 2$. Therefore, the diameter of $G(n,p)$ is concentrated on at most three values at this range.

Proof of Theorem 4: In this range, $G(n,p)$ may be disconnected. However, the diameter of $G(n,p)$ is determined by the diameter of its giant component by using Theorem 1. By Lemma 2, almost surely for all vertices x, we have

$$|N_i(x)| \le \frac{10}{c}(np)^i$$

We choose $k = \left\lfloor \frac{\log \frac{cn}{11}}{\log(np)} \right\rfloor$. Note that in this range, the size of giant component is $(1-o(1))n$. $|N_k(x)| \le \frac{10}{11}n$ is less than the giant component. Hence, we have

$$diam(G(n,p)) \ge \left\lfloor \frac{\log(\frac{cn}{11})}{\log(np)} \right\rfloor + 1$$

On the other direction, by Lemma 7, almost surely for a vertices x in giant component, there exists an $i_0 \le \lfloor \frac{1}{c} \rfloor + 1$ satisfies

$$|\Gamma_i(x)| \ge \frac{5}{c}(np)^{i-i_0}$$

We choose

$$k_1 = \lceil \frac{\log(\sqrt{\frac{33c^2}{400} n \log n})}{\log(np)} + i_0 \rceil \quad \text{and} \quad k_2 = \lceil \frac{\log(\frac{33c^2}{400} n \log n)}{\log(np)} - k_1 - 1 + i_0 \rceil$$

$k_1 \approx k_2 \approx \frac{1}{2} \frac{\log(\frac{33c^2 n \log n}{400})}{\log(np)} + i_0 < \frac{2}{3} \frac{\log n}{\log(np)}$. The condition of Lemma 7 is satisfied. Almost surely

$$|\Gamma_{k_1}(x)||\Gamma_{k_2}(y)| \ge \frac{5}{c}(np)^{k_1 - i_0} \frac{5}{c}(np)^{k_2 - i_0} p \ge 2.0625 \log n$$

Hence, almost surely we have

$$diam(G(n,p)) \le k_1 + k_2 + 1 = \left\lceil \frac{\log(\frac{33c}{400} n \log n)}{\log(np)} + 2i_0 \right\rceil.$$

Therefore, almost surely

$$\left\lceil \frac{\log \frac{cn}{11}}{\log(np)} \right\rceil \le diam(G(n,p)) \le \left\lceil \frac{\log(\frac{33c^2}{400} n \log n)}{\log(np)} \right\rceil + 2\lfloor \frac{1}{c} \rfloor + 2.$$

The difference of the upper bound and lower bound is at most

$$\begin{aligned}
&\left\lceil \frac{\log(\frac{33c^2}{400} n \log n)}{\log(np)} \right\rceil \ + \ 2\lfloor \frac{1}{c} \rfloor + 2 - \left\lceil \frac{\log \frac{cn}{11}}{\log(np)} \right\rceil \\
&\le \ \left\lceil \frac{\log(\frac{363c}{400} \log n)}{\log(np)} \right\rceil + 2\lfloor \frac{1}{c} \rfloor + 2 \\
&\le \ \left\lceil \frac{\log(\frac{363c}{400} \log n)}{\log(c \log n)} \right\rceil + 2\lfloor \frac{1}{c} \rfloor + 2 \\
&\le \ 2\lfloor \frac{1}{c} \rfloor + 3
\end{aligned}$$

Therefore, if $n \ge \frac{c \log n}{n}$, the diameter of $G(n,p)$ is concentrated on at most $2\lfloor \frac{1}{c} \rfloor + 4$ values. $\square$

Proof of Theorem 5: By Lemma 1, for almost all x and i, we have

$$|N_i(x)| \leq 2i^3 \log n(np)^3$$

We now choose $k = \left\lfloor \frac{\log n - 4 \log\log n}{\log(np)} \right\rfloor$. Hence, we have

$$diam(G(n,p)) > k + 1 = (1 + o(1))\frac{\log n}{\log(np)}$$

On the other hand, by Lemma 6, there exists an i_0 satisfying $i_0 \leq \frac{(\frac{10np}{(\sqrt{np}-1)^2}+1)\log n}{np-1-\log(np)} = o(\frac{\log n}{\log(np)})$. For almost vertices x, we have

$$|\Gamma_i(x)| \geq \frac{1}{(\sqrt{np}-1)^2}(np)^{i-i_0}\log n.$$

We can then choose

$$k_1 \approx k_2 \approx \frac{1}{2}\frac{\log n}{\log(np)} + i_0$$

Therefore, $|\Gamma_{k_1}(x)| \approx |\Gamma_{k_1}(x)| < n^{2/3}$. The condition of Lemma 6 is satisfied. Hence we have

$$diam(G(n,p)) \leq k_1 + k_2 + 1 \approx \frac{\log n}{\log(np)} + 2i_0 + 1 = (1 + o(1))\frac{\log n}{\log(np)}$$

We obtain

$$diam(G(n,p)) = (1 + o(1))\frac{\log n}{\log(np)}.$$

□

Proof of Theorem 6: The proof is very similar to that of Theorem 5, so we will only sketch the proof here. It can be shown that

$$diam(G(n,p)) \geq (1 + o(1))\frac{\log n}{\log(np)}.$$

In the other direction, we choose

$$k_1 \approx k_2 \approx \frac{1}{2}\frac{\log n}{\log(np)} + i_0.$$

But now $i_0 \leq \frac{(\frac{10c}{(\sqrt{c}-1)^2}+1)\log c}{c-\log(2c)}\frac{\log n}{\log np}$. Hence

$$diam(G(n,p)) \leq \frac{\log n}{\log np} + 2\frac{(10c/(\sqrt{c}-1)^2+1)}{c-1-\log(c)}\frac{\log n}{np} + 1$$

5 Problems and remarks

We have proved that the diameter of $G(n,p)$ is almost surely equal to its giant component if $np > 3.5128$. Several questions here remain unanswered:

Problem 1: Is the diameter of $G(n,p)$ equal to the diameter of its giant component ?

Of course, this question only concerns the range $1 < p \leq 3.5128$. There are numerous questions concerning the diameter in the evolution of the random graph. The classical paper of Erdős and Rényi [13] stated that all connected components are trees or unicyclic in this range. What is the the distribution of the diameters of all connected components? Is there any "jump" or "double jumps" as the connectivity [13] in the evolution of the random graphs during this range for p?

In this paper we proved that almost surely the diameter of $G(n,p)$ is $(1+o(1))\frac{\log n}{\log np}$ if $np \to \infty$. When $np = c$ for some constant $c > 1$, we can only show that the diameter is within a constant factor of is $\frac{\log n}{\log np}$. Can this be further improved?

Problem 2: Prove or disprove

$$diam(G(n, \frac{c}{n})) = (1+o(1))\frac{\log n}{\log c}$$

for constant $c > 1$.

Our method for bounding the diameter through estimating $|N_i(x)|$ does not seem to work for this range. This difficulty can perhaps be explained by the following observation. The probability that $|\Gamma_1(x)| = 1$ is approximately $\frac{c}{e^c}$, a constant. Hence, the probability that

$$|\Gamma_1(x)| = |\Gamma_2(x)| = \ldots = |\Gamma_l(x)| = 1$$

is about $(\frac{c}{e^c})^l$. For some l up to $(1-\varepsilon)\frac{\log n}{c - \log c}$, this probability is at least $n^{1-\varepsilon}$. So it is quite likely that this may happen for vertex x. In other words, there is a nontrivial probability that the random graph around x is just a path starting at x of length $c \log n$. The i-th neighborhood $N_i(x)$ of x, for $i = c\log n$, does not grow at all!

In Theorems 2 and 3 we consider the case of $p > \frac{c \log n}{n}$. Does the statements still hold for $p = \frac{c\log n}{n}$?

Problem 3: Is it true that the diameter of $G(n,p)$ is concentrated on $2k+3$ values if $p = \frac{\log n}{kn}$?

It is worth mentioning that the case $k = 1$ is of special interest.

For the range $np = 1 + n^{-c}$, Lemma 1 implies $diam(G(n,p)) \geq (\frac{1}{1-3c} + o(1))\frac{\log n}{\log(np)}$. Can one establish a similar upper bound?

Problem 4: Is it true that

$$diam(G(n,p)) = \Theta(\frac{\log n}{\log(np)})$$

for $np = 1 + n^{-c}$ and $c < \frac{1}{3}$?

Łuczak [16] proved that the diameter of $G(n,p)$ is equal to the diameter of a tree component in the subcritical phase $(1-np)n^{1/3} \to \infty$. What can we say about the diameter of $G(n,p)$ when $(1-np)n^{1/3} \to c$, for some constant c? The diameter problem seems to be hard in this case.

A related problem is to examine the average distance of graphs instead of the diameter which is the maximum distance. The problem on average distance of a random graph with given degree sequence has applications in so-called *small world graphs* [12, 17]. Research in this direction can be found in [12].

References

[1] L. A. Adamic and B. A. Huberman, Growth dynamics of the World Wide Web, *Nature* , **401**, September 9, 1999, pp. 131.

[2] W. Aiello, F. Chung and L. Lu, A random graph model for massive graphs, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, (2000) 171-180.

[3] W. Aiello, F. Chung and L. Lu, Random evolution in massive graphs, *Handbook on Massive Data Sets*, (Eds. J. Abello et. al.), to appear.

[4] Réka Albert, Hawoong Jeong, and Albert-László Barabási, Diameter of the World Wide Web, *Nature* **401** (1999) 130-131.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley and Sons, New York, 1992.

[6] Albert-Laszlo Barabasi, Reka Albert, Emergence of scaling in random networks, Science **286** (1999) 509-512.

[7] B. Bollobás, The Diameter of Random Graphs. *IEEE Trans. Inform. Theory* **36** (1990), no. 2, 285-288.

[8] B. Bollobás, *Random Graphs* Academic Press, (1985).

[9] B. Bollobás, The evolution of sparse graphs, *Graph Theory and Combinatorics* (Cambridge 1983), 35-57, Academic Press, London-New York, 1984.

[10] J. D. Burtin, Extremal metric characteristics of a random graph I, *Theo. Verojatnost. i Primenen* **19** (1974), 740-754.

[11] J. D. Burtin, Extremal metric characteristics of a random graph II, *Theo. Verojatnost. i Primenen* **20** (1975), 82-99.

[12] F. Chung and L. Lu, Random graphs and small world graphs, preprint.

[13] P. Erdős and A. Rényi, On random graphs. I, *Publ. Math. Debrecen* **6** (1959), 290-291.

[14] V. Klee and D. Larman, Diameters of random graphs. *Canad. J. Math.* **33** (1981) 618-640.

[15] Linyuan Lu, The Diameter of Random Massive Graphs, *Proceedings of the Twelfth ACM-SIAM Symposium on Discrete Algorithms* (SODA 2001), 912-921.

[16] Tomasz Łuczak, Random Trees and Random Graphs, *Random structures & algorithms* **13** (1998), 485-500.

[17] D. J. Watts, *Small Worlds*, Princeton Unversity Press, 1999.