The Diameter of Random Massive Graphs *

Linyuan Lu †

October 20, 2000

Abstract

Many massive graphs (such as the WWW graph and Call graphs) share certain universal characteristics which can be described by so-called the "power law". Here we determine the diameter of random power law graphs up to a constant factor for almost all ranges of parameters. These results show a strong evidence that the diameters of most massive graphs are about logarithm of their sizes up to a constant factor.

1 Introduction

Searching contents on the Internet is directly related to the diameter of the WWW graph, which is, by definition, the minimum number of "clicks" needed to be performed to "jump" between any two documents. The diameter can not be computed by brute-force search because of the huge size (estimated to be at least 800 million documents [22]) and rapid growth (at least 100% a year) of the Web. Recently, however, some structure of the web has come to light, which may enable us to describe graph properties of the Web qualitatively. Several groups of researchers [1, 5, 7, 8, 20, 21, 19] have made the crucial observation that the WWW graph obeys the *power law* (i.e., the number of nodes, y, of a given degree x is proportional to $x^{-\beta}$ for some constant $\beta > 0$). Aiello et al. [3] proposed a random graph model for generating graphs with degrees satisfying given power law distribution, which is called power law graphs. By using techniques in random graph theory, the sizes of connected components are determined for many cases. There are several other models [4, 7, 8, 20, 21, 19], which generate power law graphs for some ranges.

In this paper, we consider a variant of random power law graph model (specified in section 2), which is an analogy of classic random graph model $G(n,p)$. $G(n,p)$ is a random graph on n vertices in which a pair of vertices appears as an edge with probability p. The studies of graph properties of $G(n,p)$ can be traced back to the seminal papers of Erdős and Rényi [15] in 1959. As many other graph properties, the diameter of $G(n,p)$ is well-studied. For a disconnected graph G, we use the convention that the diameter of G is the maximum diameter of its connected components. There is rich literature of computing the diameter of $G(n,p)$ [9, 10, 11, 13, 14, 16, 23]. In general, the sparser the graph is, the harder the problem is. Recently Chung and Lu [16] further extended these techniques to examine the diameter of sparse graphs.

Relatively little is known about the diameter of massive graphs. Albert et al. [5] estimated the average distance between any two documents of the Web is about 19. Broder et al. [12] reported on experiments on local and global properties of the web graph using two Altavista crawls each with over 200M pages and 1.5 billion links. They showed that the diameter of this part of the WWW graph is over 500 and the diameter of its giant strongly connected component is at least 28. These experimental results can be very effective in predicting the actual diameter of the WWW graph, if we could answer the following question: —How does the diameter of the WWW graph grow when its size increases?

In this paper, we try to answer these kinds of questions by examining the diameter of random power law graphs for all ranges of the powers. We determine the diameter of random power law graphs up to a constant factor (Theorem 3.1, 3.2) for various ranges of β. We also derive the complete distribution of their connected components (Theorem 3.3, 3.4).

There are some limitations when these theoretical results are applied. The model here is an undirected graph model, while most massive graphs are directed graphs. These results can be applied to massive graphs in a sense that people are often interested in the diameter and connected components of a massive graph as an undirected graph. The power parameter β's of most massive graphs are in the range $\beta > 2$. (For example, Kumar et al. [20] and Barabási et al. [7, 8] independently reported that the value of β of the WWW graph is approximately 2.1 for in-degree power law and 2.7 for the out-degree one.) Theorem 3.2 provides a

*Supported in part by NSF grant DMS-9803515 and DMS-9801446

†University of California, San Diego

strong evidence that diameters of most massive graphs are about logarithm of their sizes up to a constant factor for β in this range.

The rest of the paper is organized as follows. First, a random power law graph model is defined and some facts are stated in section 2. Results are stated in section 3. Methods are described in section 4 and several technical lemmas are also given there. We prove the main Theorems 3.3, 3.4 on connected components in section 5. From section 6, we compute the diameter. In section 6 and 7, we deal with the ranges $\beta < 2$ and $\beta > 4$, respectively. The last range $2 < \beta \le 4$ is covered in section 8. Some open problems are given in the last section.

2 Model and facts

Aiello at el. [3] introduced a new random power law graphs model, which can be described as follows. The model in [3] has two parameters α and β. Here α is the logarithm of the graph size and β is the log-log (negative) growth rate. So the model consists of random graphs satisfying the following: The number of vertices with degree x, y, satisfies

$$\log y = \alpha - \beta \log x.$$

(Here all logarithms are with base e.)

We consider a variant of the model in [3]. The model here is more complicated to state than the one in [3], but is easier to analyze due to its edge independence. In general, results from one model can be expected from the other model, and vice versa. However, the difference of the two models does exist. For example, in [3] it is stated that there is no giant component when $\beta > 3.4785$. But Theorem 3.3 states that there is always a unique giant component for ranges considered in this model. It is because that our model allows isolated vertices and has fewer vertices with a single edge. So the chance of forming a giant component increases.

The random power law graphs model is defined as follows.

Model: Given n weighted vertices with weights $w_1, \dots, w_n$, a pair of vertices (i, j) appears as an edge with probability $w_i w_j p$ independently. Here these parameters $w_1, \dots, w_n$ and p satisfy

1. $\#\{i | 1 \le w_i < 2\} = \lfloor e^\alpha \rfloor - r$. $\#\{i | k \le w_i < k+1\} = \lfloor \frac{e^\alpha}{k^\beta} \rfloor$ for $k = 2, 3, \dots, \lfloor e^{\frac{\alpha}{\beta}} \rfloor$. Here α is a value minimizing $\left| n - \sum_{k=1}^{\lfloor e^{\frac{\alpha}{\beta}} \rfloor} \lfloor \frac{e^\alpha}{k^\beta} \rfloor \right|$, and $r = n - \sum_{k=1}^{\lfloor e^{\frac{\alpha}{\beta}} \rfloor} \lfloor \frac{e^\alpha}{k^\beta} \rfloor$.
2. $p = \frac{1}{\sum_{i=1}^n w_i}$.

We remark that $|r| \le e^{\frac{\alpha}{\beta}} = o(e^\alpha)$ for $\beta > 1$. The above definition is well-defined. However, if $0 < \beta \le 1$, we may have $r = \Omega(e^\alpha)$. For simplicity and convenience, we only consider the model for those n such that $r = o(e^\alpha)$. (An alternative solution is to modify condition 1 to $\#\{i | k \le w_i < k+1\} = \lfloor \frac{e^\alpha}{k^\beta} \rfloor - 1$. for $1 \le k \le r$.)

For $1 \le k \le e^{\frac{\alpha}{\beta}}$, we have

$$\#\{i | k \le w_i < k+1\} \approx \frac{e^\alpha}{k^\beta}$$

Since for each vertex v_i, the expected degree of v_i is $\sum_j w_i w_j p \approx w_i$. Hence, the degree sequence of this model roughly follows the power law distribution with parameter (α, β).

Since α is $\Theta(\log n)$ where n is the graph size and n goes to infinity. Only β does matter. We simply denote this random power law graph model by G_β.

If we use real numbers instead of rounding down to integers, it may cause some error terms in further computation. However, the error terms can be easily bounded. For convenience, we will use real numbers with the understanding the actual numbers are their integer parts.

We can deduce the following facts for our graph:
(1) The maximum weight of the graph is about $e^{\frac{\alpha}{\beta}}$.
(2) The vertices number n is related to α as follows

$$n \approx \sum_{k=1}^{e^{\frac{\alpha}{\beta}}} \lfloor \frac{e^\alpha}{k^\beta} \rfloor \approx \begin{cases} \zeta(\beta) e^\alpha & \text{if } \beta > 1 \\ \alpha e^\alpha & \text{if } \beta = 1 \\ C_0' e^{\frac{\alpha}{\beta}} & \text{if } 0 < \beta < 1 \end{cases}$$

where $\zeta(t) = \sum_{n=1}^\infty \frac{1}{n^t}$ is the Riemann Zeta function and C_0' is a constant satisfying $\frac{\beta}{1-\beta} < C_0' \le \frac{1}{1-\beta}$.
(3) The total weights of G_β (or the volume of G_β) can be computed by

$$vol(G) = \sum_{i=1}^n w_i = 1/p \approx \begin{cases} C_1 e^\alpha & \text{if } \beta > 2 \\ \alpha e^\alpha & \text{if } \beta = 2 \\ C_1' e^{\frac{2\alpha}{\beta}} & \text{if } 0 < \beta < 2 \end{cases}$$

where C_1 and C_1' are two constants satisfying $\zeta(\beta - 1) \le C_1 \le \zeta(\beta - 1) + \zeta(\beta)$ and $\frac{\beta}{2(2-\beta)} < C_1' \le \frac{1}{2-\beta}$ respectively.
(4) The expected number of edges E can be computed as follows:

$$E = \sum_{i<j} w_i w_j p \approx \frac{1}{2} \sum_i w_i \approx \frac{1}{2} vol(G_\beta)$$

(5) The higher moments of the weights distribution are as follows.

$$I_k = \sum_i w_i^k \approx \begin{cases} C_k e^\alpha & \text{if } \beta > k+1 \\ \frac{\alpha}{\beta} e^\alpha & \text{if } \beta = k+1 \\ C_k' e^{\frac{(k+1)\alpha}{\beta}} & \text{if } 0 < \beta < k+1 \end{cases}$$

where C_k and C_k' are two constants satisfying $\zeta(\beta - k) \le C_k \le \zeta(\beta - k) + \zeta(\beta - k + 1)$ and $\frac{\beta}{(k+1)(k+1-\beta)} < C_k' \le \frac{1}{k+1-\beta}$ respectively.

3 Our results

We have

THEOREM 3.1. *When $0 < \beta < 2$, the diameter of the random power law graph G_β is almost surely at most $2\lfloor\frac{1}{2-\beta}\rfloor + 5$.*

THEOREM 3.2. *When $\beta > 2$, the diameter of the random power law graph G_β is almost surely $\Theta(\log n)$.*

We also get the following results on connected components.

THEOREM 3.3. *For all ranges of β, the random power law graph G_β has a unique giant component.*

THEOREM 3.4. *For all ranges of $\beta > 0$, almost surely all components of the random power law graph G_β other than the giant component have size at most $O(\log n)$. Moreover we have*

1. *When $\beta > 2$, there exist two constant $0 < c_1 < c_2 < 1$ satisfying that almost surely the expected number of connected components of size k is at least $c_1^k n$ and at most $c_2^k n$, provided $k = o(n)$. This implies that almost surely the second largest component has size $\Theta(\log n)$.*
2. *When $\beta = 2$, almost surely the second largest component has size $\Theta(\frac{\log n}{\log\log n})$.*
3. *When $\beta < 2$, almost surely the second largest component has size $\Theta(1)$.*

4 Methods and useful lemmas

We will prove that there is always a giant component in G_β (Theorem 3.3) and further determine the distribution of the sizes of other components (Theorem 3.4). The diameters of small components are bounded by their sizes. The diameter of the giant component is obtained by analyzing its structure, which is different for different ranges of β.

We use the following notation. The volume of a set of vertices S is defined as $\sum_{v_i \in S} w_i$. We denote it by $vol(S)$.

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:

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

For undefined terminology, the reader is referred to [10].

A main method to estimate the diameter of a giant component is to examine the volume of neighborhoods $N_k(x)$ and $\Gamma_k(x)$. We will use several useful lemmas concerning the volume of the neighborhoods (while some of the routine proofs will be omitted.)

LEMMA 4.1. *Let $X_1, \dots, X_n$ be independent random variables with*

$$Pr(X_i = 1) = p_i, \qquad Pr(X_i = 0) = 1 - p_i$$

Let $X = \sum_{i=1}^n a_i X_i$. Let $E(X) = \sum_{i=1}^n a_i p_i$, $\nu = \sum_{i=1}^n a_i^2 p_i$. Then we have

$$Pr(X < E(X) - \lambda) \le e^{-\lambda^2/2\nu}$$

Now we will use this lemma to bound the neighborhoods of a vertex in G. For a subset S of vertices, we denote $\Gamma(S) = \{v | vv' \text{ is an edge in G for some } v' \in S\}$.

LEMMA 4.2. *Let S be a subset of vertices with total weight s satisfying $s\sum w_i^2 p = o(\sum w_i)$ and $s > (\frac{2}{t^2} + o(1))\frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2}\log n$. Then almost surely we have*

$$vol(\Gamma(S)) > (1-t)s\sum w_i^2 p.$$

Proof: For every vertex v_i in S and $v_j \notin S$, let $X_{i,j}$ be the indicated random variable that $v_i v_j$ forms an edge in G. Since all pairs are independent to each other, we can use Lemma 4.1 to estimate $Y = \sum_{v_i \in S, v_j \notin S} w_j X_{i,j}$. However Y is not exactly equal to $vol(\Gamma(S))$ but it is close enough when $vol(S)$ is small. Since $E(y) = \sum_{v_i \in S, v_j \notin S} w_i w_j^2 p \approx s\sum_i w_i^2 p$ and $\nu = \sum_{v_i \in S, v_j \notin S} w_i w_j^3 p \approx s \sum_i w_i^3 p$.

We apply Lemma 4.1 by choosing $\lambda = ts\sum w_i^2 p$. We have

$$\begin{aligned} Pr(vol(\Gamma(S)) &< (1-t)s\sum w_i^2 p) \\ &\le e^{(ts\sum w_i^2 p)^2/(2(1-o(1))\sum s w_i^3 p)} \\ &\le e^{-(1+o(1))\log n} \\ &= o(n^{-1}). \end{aligned}$$

Hence, almost surely $vol(\Gamma(S)) > (1-t)s\sum w_i^2 p$. □

Let $s_0 = \Theta(\frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2}\log n)$. If there is a i_0 satisfying $vol(\Gamma_{i_0}(v)) > s_0$, then for $i \ge i_0$, the volume of its i-th neighborhood will continue to grow. In this case, v belongs to the giant component. We call s_0 *the growing threshold* of G_β. Any vertex with weight greater than s_0 is in the core of the giant component.

Now we compute the growing threshold according to different ranges.

The range:	Growing threshold
$0 < \beta < 2$	$\Theta(\alpha)$
$\beta = 2$	$\Theta(\alpha^2)$
$2 < \beta < 3$	$\Theta(\alpha e^{(1-2/\beta)\alpha})$
$\beta = 3$	$\Theta(\frac{1}{\alpha} e^{\alpha/3})$
$3 < \beta < 4$	$\Theta(\alpha e^{(4/\beta-1)\alpha})$
$\beta = 4$	$\Theta(\alpha^2)$
$\beta > 4$	$\Theta(\alpha)$

5 Connected components

In this section, we will deal with the connected components. We will prove Theorem 3.3 and 3.4 here.

Proof of Theorem 3.3: Let v be the vertex with the largest weight $e^{\frac{\alpha}{\beta}}$ in G. Note that $vol(v)$ is bigger than the growing threshold. Hence, its neighborhoods grow continuously until its volume reaches $\epsilon vol(G)$. Any two components with volume greater than $\epsilon vol(G)$ are connected. There is a unique giant component. □

Now we want to upper bound the sizes of components other than the giant component.

Proof of Theorem 3.4: We are going to compute the probability of a component of size k. We assume that the component has vertices set $S = \{v_{i_1}, v_{i_2} \dots, v_{i_k}\}$ with weights $w_{i_1}, w_{i_2}, \dots, w_{i_k}$. We also assume the component is not the giant component so that $vol(S) = w_{i_1} + w_{i_2} + \cdots + w_{i_k} = o(vol(G))$. The probability that there are no edges going out from S is

$$\begin{aligned} \prod_{v_i \in S, v_j \notin S} & (1 - w_i w_j p) \\ \approx & \; e^{-p \sum_{v_i \in S, v_j \notin S} w_i w_j} \\ = & \; e^{-pvol(S)(vol(G)-vol(S))} \\ \approx & \; e^{-vol(S)}. \end{aligned}$$

Now we compute the probability of edges inside S. A connected graph on S contains at least one spanning tree T. The probability of existence of T can be computed by

$$Pr(T) = \prod_{(v_{i_j} v_{i_l}) \in E(T)} w_{i_j} w_{i_l} p.$$

Hence the probability of existing a connected spanning graph on S is at most

$$\sum_T Pr(T) = \sum_T \prod_{(v_{i_j} v_{i_l}) \in E(T)} w_{i_j} w_{i_l} p,$$

where T is indexed over all spanning trees on S.

By a generalized version of well-known matrix-tree Theorem, the above summation equals to the determinant of any $k-1$ by $k-1$ principal sub-matrix of the matrix $D - A$, where A is the weight matrix

$$A = \begin{pmatrix} 0 & w_{i_1} w_{i_2} p & \cdots & w_{i_1} w_{i_k} p \\ w_{i_2} w_{i_1} p & 0 & \cdots & w_{i_2} w_{i_k} p \\ \vdots & \vdots & \ddots & \vdots \\ w_{i_k} w_{i_1} p & w_{i_k} w_{i_2} & \cdots & 0 \end{pmatrix}$$

and D is the diagonal matrix $diag((vol(S) - w_{i_1}^2), \dots, (vol(S) - w_{i_k}^2))$. By computing the determinant, we conclude that

$$\sum_T Pr(T) = w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1}.$$

Let X_k be the random variable of the number of the components with size k. Hence, the expected value $E(X_k)$ is at most

$$f(k) = \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)}$$

where the summation is over all k-vertices set S.

We will show that $f(k)$ is very small if k is big enough.

We upper bound $f(k)$ as follows. Note that the function $x^{2k-2} e^{-x}$ reaches its maximum value at $x = 2k - 2$.

$$\begin{aligned} f(k) &= \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)} \\ &\le \sum_S \frac{p^{k-1}}{k^k} vol(S)^{2k-2} e^{-vol(S)} \\ &\le \sum_S \frac{p^{k-1}}{k^k} (2k-2)^{2k-2} e^{-(2k-2)} \\ &\le \frac{n^k}{k!} \frac{p^{k-1}}{k^k} (2k-2)^{2k-2} e^{-(2k-2)} \\ &\approx \frac{(k-1)^2}{4p} (np)^k 2^{2k} e^{-(k-2)} (1 - \frac{1}{k})^{2k} \\ &\le \frac{(k-1)^2}{4p} (\frac{4np}{e})^k \\ &< \frac{n^3}{4np} (\frac{4np}{e})^k \end{aligned}$$

The above inequality is useful when $4np < e$. If $k > \frac{5 \log n - \log(4np)}{2 - \log(4np)}$, we have

$$Pr(X_k \ge 1) \le E(X_k) \le f(k) \le \frac{1}{n^2}.$$

It implies that almost surely all components other than the giant one have size at most

$$\frac{5 \log n - \log(4np)}{1 - \log(4np)}$$

We recall

$$np = \begin{cases} \leq \frac{\zeta(\beta)}{\zeta(\beta-1)} & \text{for } \beta > 2 \\ \Theta(\frac{1}{\log n}) & \text{for } \beta = 2 \\ \Theta(\frac{1}{n^{2-1/\beta}}) & \text{for } 1 < \beta < 2 \\ \Theta(\frac{\log^2 n}{n}) & \text{for } \beta = 1 \\ \Theta(\frac{1}{n}) & \text{for } 0 < \beta < 1 \end{cases}$$

We have

1. When $0 < \beta < 2$, we have $\log np = -\Theta(\log n)$. Almost surely the second largest component can have at most

$$\frac{5 \log n - \log(4np)}{1 - \log(4np)} = \Theta(1)$$

vertices.

2. When $\beta = 2$, we have $\log np = -\Theta(\log \log n)$. Almost surely the second largest component can have size at most

$$\frac{5 \log n - \log(4np)}{1 - \log(4np)} = \Theta(\frac{\log n}{\log \log n}).$$

3. When $2 < \beta < 2.8$, we have $np \leq \frac{\zeta(\beta)}{\zeta(\beta-1)} < e/4$. Almost surely the second largest component can have at most

$$\frac{5 \log n - \log(4np)}{1 - \log(4np)} = \Theta(\log n).$$

Above method is no longer useful when $\beta > 2.8$. A different estimation of $f(k)$ is needed here. We assume that $\beta > 2.5$. We have

$$np \leq \frac{\zeta(\beta)}{\zeta(\beta - 1)} < 1.$$

We choose

$$\delta = \frac{\zeta(\beta - 1)}{\zeta(\beta)} - 1.$$

We have $0 < \delta < 1$ since $\beta > 2.5$.

We split $f(k)$ into two parts as follows:

$$f_1(k) = \sum_{vol(S)<(1+\delta)k} w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)}$$

$$f_2(k) = \sum_{vol(S)\geq(1+\delta)k} w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)}$$

Now we upper bound the first part $f_1(k)$. Note that $x^{2k-2}e^{-x}$ is an increasing function when $x < 2k-2$. We get

$$vol(S)^{2k-2} e^{-vol(S)} \leq ((1 + \delta)k)^{2k-2} e^{-(1+\delta)k}$$

since $vol(S) < (1 + \delta)k < 2k - 2$. We have

$$\begin{aligned} & f_1(k) \\ = & \sum_{vol(S)<(1+\delta)k} w_{i_1} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)} \\ \leq & \sum_{vol(S)<(1+\delta)k} \frac{p^{k-1}}{k^k} vol(S)^{2k-2} e^{-vol(S)} \\ \leq & \sum_{vol(S)<(1+\delta)k} \frac{p^{k-1}}{k^k} ((1 + \delta)k)^{2k-2} e^{-(1+\delta)k} \\ \leq & \binom{n}{k} \frac{p^{k-1}}{k^k} ((1 + \delta)k)^{2k-2} e^{-(1+\delta)k} \\ \leq & \frac{n^k}{k!} \frac{p^{k-1}}{k^k} ((1 + \delta)k)^{2k-2} e^{-(1+\delta)k} \\ \approx & \frac{1}{(1 + \delta)^2 k^2 p} (np)^k (1 + \delta)^{2k} e^{-\delta k} \\ \leq & \frac{1}{p} (\frac{(1 + \delta)^2 np}{e^\delta})^k \end{aligned}$$

Now we upper bound the second part $f_2(k)$. Note that $x^{k-2}e^{-x}$ is a decreasing function when $x > k - 2$. We have

$$vol(S)^{k-2} e^{-vol(S)} \leq ((1 + \delta)k)^{k-2} e^{-(1+\delta)k}$$

since $vol(S) \geq (1 + \delta)k > k - 2$. We get

$$\begin{aligned} & f_2(k) \\ = & \sum_{vol(S)\geq(1+\delta)k} w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)} \\ \leq & \sum_{vol(S)\geq(1+\delta)k} w_{i_1} \cdots w_{i_k} p^{k-1} ((1 + \delta)k)^{k-2} e^{-(1+\delta)k} \\ \leq & \sum_{S} w_{i_1} w_{i_2} \cdots w_{i_k} p^{k-1} ((1 + \delta)k)^{k-2} e^{-(1+\delta)k} \\ < & \frac{vol(G)^k}{k!} p^{k-1} ((1 + \delta)k)^{k-2} e^{-(1+\delta)k} \\ \approx & \frac{1}{(1 + \delta)^2 k^2 p} (1 + \delta)^k e^{-\delta k} \\ \leq & \frac{1}{p} (\frac{(1 + \delta)}{e^\delta})^k \end{aligned}$$

Hence, we have

$$\begin{aligned} f(k) &= f_1(k) + f_2(k) \\ &\leq \frac{1}{p} (\frac{(1 + \delta)^2 np}{e^\delta})^k + \frac{1}{p} (\frac{(1 + \delta)}{e^\delta})^k \\ &\leq \frac{2}{p} (\frac{(1 + \delta)}{e^\delta})^k \end{aligned}$$

since $(1 + \delta)np \leq 1$.

If $k > \frac{3\log n}{\delta - \log(1+\delta)}$, we have

$$Pr(X_k \geq 1) \leq E(X_k) \leq f(k) \leq \frac{2}{n^3 p}.$$

The probability that the second largest component has size greater than $\frac{3\log n}{\delta-\log(1+\delta)}$ is at most $n\frac{2}{p}\frac{1}{n^3} = o(1)$. Hence, almost surely the second largest component has size at most

$$\frac{3\log n}{\delta - \log(1+\delta)} = \Theta(\log n).$$

Hence for all $\beta > 2$, the size of the second largest component is $O(\log n)$.

Now we are going to derive a lower bound of $E(X_k)$. We can assume that $k = O(\log n)$.

We first compute the probability that a component on S is exact a tree. Given a tree T on S, the conditional probability that a component on S is exactly T given it contains T is at least

$$\prod_{v_{i_j}, v_{i_l} \in S} (1 - w_{i_j} w_{i_l} p) \approx e^{-vol(S)^2 p} = e^{-o(1)} \geq \frac{1}{2}.$$

Hence $E(X_k)$ is at least $\frac{1}{2}f(k)$. We give the lower bound of $f(k)$.

$$\begin{aligned} f(k) &= \sum_S w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1} e^{-vol(S)} \\ &\geq \sum_{vol S \leq 2k} vol(S)^{k-2} p^{k-1} e^{-vol(S)} \\ &\geq \sum_{vol S \leq 2k} (2k)^{k-2} p^{k-1} e^{-2k} \\ &\geq \binom{e^\alpha}{k} (2k)^{k-2} p^{k-1} e^{-2k} \\ &\approx \frac{1}{4k^2 p} (\frac{2e^\alpha p}{e})^k \end{aligned}$$

When $\beta > 2$, both np and $e^\alpha p$ are constants. So there is a constant $c_1 < 1$, satisfying

$$\frac{1}{8k^2 p}(\frac{2e^\alpha p}{e})^k \geq n c_1^k$$

for all $k \geq 1$ provided $k = O(\log n)$.

Hence $E(X_k) \geq \frac{1}{2} f(k) \geq c_1^k$.

When $\beta = 2$, Let $k_0 = \lfloor \frac{\log n}{2(1-\log(2e^\alpha p))} \rfloor = \Theta(\frac{\log n}{\log\log n})$. We have

$$E(X_k) \geq \frac{1}{2} f_{k_0} \geq \frac{1}{8k_0^2 p} n^{-1/2} = \frac{\sqrt{n}}{8k_0^2 np} = \Omega(n^{1/3}).$$

Hence there is a component of size $k_0 = \Theta(\frac{\log n}{\log\log n})$. So the second largest component has size of $\Theta(\frac{\log n}{\log\log n})$.

The proof of Theorem 3.4 is finished. □

By Theorem 3.4, the sizes of small components are small enough so that the diameter of the giant component dominates (up to some constant factor). Thus we reduce the problem to computing the diameter of the giant component. We will consider 3 different ranges in next 3 sections.

6 The diameter for $0 < \beta < 2$.

There is a dense core inside the giant component, whose diameter is at most 3. Outside the core, there are some tree-like tails of finite length. Hence, the diameter is finite in this case.

Proof of Theorem 3.1: Let $t = \frac{1}{2}(1 - \frac{1}{2-\beta}\frac{1}{\lfloor \frac{1}{2-\beta} \rfloor + 1})$. Let S be the set of all vertices with weights greater than $\Theta(e^{t\alpha/\beta})$. First we will show that the diameter of S is at most 3. For any $v_1, v_2 \in S$, by Lemma 4.2, we have

$$vol(\Gamma(v_1)) \geq c e^{(t+1)\alpha/\beta},$$

$$vol(\Gamma(v_2)) \geq c e^{(t+1)\alpha/\beta}.$$

If $\Gamma(v_1) \cap \Gamma(v_2) \neq \emptyset$, then $d(v_1, v_2) \leq 2$. If $\Gamma(v_1) \cap \Gamma(v_2) = \emptyset$, the probability that no edge between $\Gamma(v_1)$ and $\Gamma(v_2)$ is at most

$$\prod_{i \in \Gamma(v_1), j \in \Gamma(v_2)} (1 - w_i w_j p) \approx e^{-vol(\Gamma(v_1)) vol(\Gamma(v_2)) p}$$

$$\leq e^{-c^2/C_1' e^{2t\alpha/\beta}} = o(n^{-2}).$$

Hence almost surely, $d(v_1, v_2) \leq 3$.

Next we will show that any vertex v in the giant component is connected to a vertex in the core S by a $\lfloor \frac{1}{2-\beta} \rfloor + 1$-path. Let $k = \lfloor \frac{1}{2-\beta} \rfloor + 1$. It is enough to show that $\Gamma_k(v) \cap S \neq \emptyset$. The worst case is that v has minimum weight 1. The probability that $\Gamma(v) \cap S = \emptyset$ is less than

$$\begin{aligned} P(\Gamma(v) \cap (V \setminus S) \neq \emptyset) &\leq \sum_{i=1}^{e^{t\alpha/\beta}} i \frac{e^\alpha}{i^\beta} p \\ &\approx C e^\alpha (e^{t\alpha/\beta})^{2-\beta} e^{-2\alpha/\beta} \\ &\leq C e^{-(2-\beta)(1-t)\alpha/\beta} \end{aligned}$$

for some constant C. The probability that $\Gamma_k(v) \cap S = \emptyset$ is less than

$$C^k e^{-(2-\beta)k(1-t)\alpha/\beta} = o(n^{-1}).$$

Hence, the diameter of the giant component is at most $2k + 3 = 2\lfloor \frac{1}{2-\beta} \rfloor + 5$. □

7 The diameter for $\beta > 4$

The range $\beta > 2$ is more difficult than $\beta < 2$. The giant component has a small dense core with diameter of $\Theta(\log n)$. There are also some tree-like tails. For $\beta > 4$, this is the picture of the giant component. But for $2 \le \beta \le 4$, there is a layer between the core and tails. We will deal with the middle layer in next section.

We need the following two lemmas to estimate the lengths of those tails.

LEMMA 7.1. *When $\beta > 2$, with probability at least $1 - o(n^{-1})$, for all vertices v in the giant component and any constant C, there is an index $i_0 = O(\log n)$ satisfying $vol(\Gamma_{i_0}(v)) \ge C \log n$.*

Similarly, we have following lemma for $\beta = 2$.

LEMMA 7.2. *When $\beta = 2$, with probability at least $1 - o(n^{-1})$, for all vertices v in the giant component and any constant C, there is an index $i_0 = O(\frac{\log n}{\log\log n})$ satisfying $vol(\Gamma_{i_0}(v)) \ge C \frac{\log n}{\log \log n}$.*

The proofs will be given at the end of this section.

When $\beta > 4$, the growth thresholds for neighborhoods are low enough so that the tails and the core are directly connected. We can apply Lemma 7.2 to prove the following partial version of Theorem 3.2.

THEOREM 7.1. *If $\beta > 4$, the diameter of G_β is $\Theta(\alpha) = \Theta(\log n)$.*

Proof: By Theorem 3.4, all other components except for the giant one have sizes $O(\log n)$. Therefore their diameters are at most $O(\log n)$. Hence it is enough to show that the diameter of the giant components is $\Theta(\log n)$.

For any vertices u and v in the giant component, the growing threshold in this range $\beta > 4$ is $\frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2} \log n = \Theta(\log n)$. By Lemma 7.1, there is a $i_0 = O(\log n)$, satisfying

$$vol(\Gamma_{i_0}(u)) \ge \frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2} \log n$$

Let $c = \sum w_i^2 p$. $c > 1$ is a constant in this range. Let $i_1 = \lfloor \frac{2 \log vol(G)}{3 \log c} \rfloor = \Theta(\log n)$. Then almost surely we have $vol(\Gamma_{i_0+i_1}(u)) \ge n^{2/3}$. Similarly, almost surely there is an $i_2 = O(\log n)$ and $i_3 = \Theta(\log n)$ satisfying $vol(\Gamma_{i_0+i_1}(v)) \ge n^{2/3}$. Hence u and v are connected by a path with length at most

$$i_0 + i_1 + i_2 + i_3 + 1 = \Theta(\log n).$$

Thus the diameter of the giant component is $O(\log n)$.

On the other hand, starting from a vertex v with weight less than 2. The probability that $\Gamma(v)$ is again a vertex with weight less than 2 is about $e^{\alpha} p e^{e^{\alpha} p}$, which is a constant. Hence, it is quite possible that the volume of $\Gamma_i(v)$ doesn't grow for $\Theta(\log n)$ steps. Thus the diameter of G is at least $\Theta(\log n)$. □

Proof of Lemma 7.1 and Lemma 7.2: We consider a breadth-first searching starting at v. We get $\Gamma_1(v)$ by exposing the neighbors of v, then $\Gamma_2(v)$ by exposing the neighbors of all $u \in \Gamma_1(v)$, and so on. Denote $d = C \log n$ if $\beta > 2$ and $d = C \frac{\log n}{\log\log n}$ if $\beta = 2$. Let $d' = C' \log n$ if $\beta > 2$ and $d' = C' \frac{\log n}{\log \log n}$ if $\beta = 2$. Here $C' > C$ is a constant depending on C and will be chosen later. Two possibilities could happen.

1. We get an $i_0 \le d'$, satisfying $vol(\Gamma_{i_0}(v)) \ge d$. Or
2. For all $1 \le i \le d'$, $vol(\Gamma_i(v)) \le d$.

It is enough to show that the probability of the second case is $o(n^{-1})$, provided we choose C' larger enough.

In the second case, $\Gamma_i(v) > 0$ for all $1 \le i \le d'$. Otherwise v is not in the giant component. We call $N_i(v)$ as a "partial component" and $\Gamma_i(v)$ a "open" set of the partial component. We choose i satisfying $d' \le \#(N_i(v)) \le 2d'$. Let $S_1 = \Gamma_i(v)$ be the open set. So a partial component with size k ($d' \le k \le 2d'$) exists and the volume of the open set is at most d.

Let us compute the probability of a partial component with size k and the volume of the open set d. As in the proof of Theorem 3.4, we assume that the component is on vertices set $S = \{v_{i_1}, v_{i_2} \dots, v_{i_k}\}$ with weights $w_{i_1}, w_{i_2}, \dots, w_{i_k}$. We have

$$vol(S) \le k e^{\alpha/\beta} \le 2d' e^{\alpha/\beta} = o(vol(G)).$$

$$\frac{vol(S_1)}{vol(S)} \le \frac{d}{k} \le \frac{d}{d'} = \frac{C}{C'}.$$

The probability that there are no edges going out from $S \setminus S_1$ to $V \setminus S$ is

$$\begin{aligned}
&\prod_{v_i \in S \setminus S_1, v_j \notin S} (1 - w_i w_j p) \\
\approx\ & e^{-p \sum_{v_i \in S\setminus S_1, v_j \notin S} w_i w_j} \\
=\ & e^{-p(vol(S) - vol(S_1))(vol(G) - vol(S))} \\
\approx\ & e^{-((vol(S) - vol(S_1))} \\
\le\ & e^{-((vol(S) - d)} \\
\le\ & e^{-(vol(S)(1 - \frac{C}{C'}))}
\end{aligned}$$

since $vol(S) = o(vol(G))$ and $p = 1/vol(G)$.

Now we compute the probability of edges inside S. As in the proof of Theorem 3.4, this probability is at most

$$w_{i_1} w_{i_2} \cdots w_{i_k} vol(S)^{k-2} p^{k-1}.$$

Since all weights are greater than or equal to 1, the size of S_1 is at most d. The choices of S_1 is at most

$$\sum_{1\le i\le d}\binom{k}{i} \le d(\frac{k}{d})^d \le (\frac{3C'}{C})^{kC/C'}$$

Hence the expected number of a partial component with size k and open volume $\le d$ is at most

$$\sum_S w_{i_1}\cdots w_{i_k} vol(S)^{k-2}p^{k-1}e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}.$$

We denote by $\bar{f}(k)$ the above expression, and will bound it in a similar way as in considering $f(k)$.

$$\begin{aligned}
&\bar{f}(k)\\
&= \sum_S w_{i_1}\cdots w_{i_k} vol(S)^{k-2}p^{k-1}\\
&\qquad \times e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}\\
&\le \sum_S \frac{p^{k-1}}{k^k} vol(S)^{2k-2}e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}\\
&\le \sum_S \frac{p^{k-1}}{k^k}(\frac{2k-2}{1-C/C'})^{2k-2}e^{-(2k-2)}(\frac{3C'}{C})^{kC/C'}\\
&= \binom{n}{k}\frac{p^{k-1}}{k^k}(\frac{2k-2}{1-C/C'})^{2k-2}e^{-(2k-2)}(\frac{3C'}{C})^{kC/C'}\\
&\le \frac{(k-1)^2}{4(1-C/C')^2p}(\frac{4np(3C'/C)^{C/C'}}{e(1-C/C')^2})^k
\end{aligned}$$

When $2 \le \beta < 2.8$, we have $np \le \frac{\zeta(\beta)}{\zeta(\beta-1)} < e/4$. We can choose constant C'' big enough so that $\frac{4np(3C'/C)^{C/C'}}{e(1-C/C'')^2} < 1$. Let $c = \frac{4(3C'/C)^{C/C'}}{e(1-C/C'')^2}$. Let $C' = max\{\frac{3\log n}{-\log cnp}, C''\}$. Then we have

$$\bar{f}(k) \le \frac{(k-1)^2}{4(1-C/C')^2p}n^{-3} = o(n^{-1}).$$

So the probability that the second case occurs is $o(n^{-1})$ as we claim at the beginning of the proof.

A different estimate of $\bar{f}(k)$ is needed for $\beta > 2.8$. We assume that $\beta > 2.5$. We have

$$np \le \frac{\zeta(\beta)}{\zeta(\beta-1)} < 1.$$

We choose

$$\delta = \frac{\zeta(\beta-1)}{\zeta(\beta)} - 1.$$

We have $0 < \delta < 1$ since $\beta > 2.5$.

We split $\bar{f}(k)$ into two parts as follows:

$$\begin{aligned}
\bar{f}_1(k) &= \sum_{vol(S)<(1+\delta)k} w_{i_1}w_{i_2}\cdots w_{i_k} vol(S)^{k-2}p^{k-1}\\
&\qquad e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}\\
\bar{f}_2(k) &= \sum_{vol(S)\ge(1+\delta)k} w_{i_1}w_{i_2}\cdots w_{i_k} vol(S)^{k-2}p^{k-1}\\
&\qquad e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}
\end{aligned}$$

Now we will bound the first part $\bar{f}_1(k)$. Note that $x^{2k-2}e^{-x(1-C/C')}$ is an increasing function when $x < (2k-2)/(1-C/C')$. Since $vol(S) < (1+\delta)k < (2k-2)(1-C/C')$, we have

$$vol(S)^{2k-2}e^{-vol(S)(1-C/C')}$$

$$\le ((1+\delta)k)^{2k-2}e^{-(1+\delta)(1-C/C')k}.$$

We have

$$\begin{aligned}
&\bar{f}_1(k)\\
&\le \sum_{vol(S)<(1+\delta)k}\frac{p^{k-1}}{k^k}vol(S)^{2k-2}\\
&\qquad e^{-(vol(S)(1-\frac{C}{C'}))}(\frac{3C'}{C})^{kC/C'}\\
&\le \sum_{vol(S)<(1+\delta)k}\frac{p^{k-1}}{k^k}((1+\delta)k)^{2k-2}\\
&\qquad e^{-(1+\delta)(1-\frac{C}{C'})k}(\frac{3C'}{C})^{kC/C'}\\
&\le \binom{n}{k}\frac{p^{k-1}}{k^k}\frac{((1+\delta)k)^{2k-2}}{e^{(1+\delta)(1-\frac{C}{C'})k}}(\frac{3C'}{C})^{kC/C'}\\
&\le \frac{n^k}{k!}\frac{p^{k-1}}{k^k}\frac{((1+\delta)k)^{2k-2}}{e^{(1+\delta)(1-\frac{C}{C'})k}}(\frac{3C'}{C})^{kC/C'}\\
&\approx \frac{(np)^k(1+\delta)^{2k}}{(1+\delta)^2k^2p}e^{(-\delta(1-\frac{C}{C'})+\frac{C}{C'})k}(\frac{3C'}{C})^{k/C'}\\
&\le \frac{1}{p}(\frac{(1+\delta)^2np(\frac{3C'}{C})^{C/C'}}{e^{\delta(1-\frac{C}{C'})-\frac{C}{C'}}})^k
\end{aligned}$$

We are going to bound the second part $\bar{f}_2(k)$. Note that $x^{k-2}e^{-x(1-C/C')}$ is a decreasing function when $x > \frac{k-2}{1-C/C'}$. We choose $C' > (1+\frac{1}{\delta})C$ so that $vol(S) \ge (1+\delta)k > \frac{k-2}{1-C/C'}$. We have

$$vol(S)^{k-2}e^{-vol(S)(1-C/C')}\le((1+\delta)k)^{k-2}e^{-(1+\delta)k(1-C/C')}$$

since $vol(S) \geq (1+\delta)k > k-2$. We have

$$\begin{aligned}
&\bar{f}_2(k) \\
&\leq \sum_{vol(S)\geq(1+\delta)k} w_{i_1}w_{i_2}\cdots w_{i_k}((1+\delta)k)^{k-2} \\
&\quad e^{-(1+\delta)k(1-C/C')}p^{k-1}(\frac{3C'}{C})^{kC/C'} \\
&\leq \sum_S w_{i_1}w_{i_2}\cdots w_{i_k}p^{k-1}((1+\delta)k)^{k-2} \\
&\quad e^{-(1+\delta)k(1-C/C')}(\frac{3C'}{C})^{kC/C'} \\
&< \frac{vol(G)^k}{k!}p^{k-1}((1+\delta)k)^{k-2} \\
&\quad e^{-(1+\delta)k(1-C/C')}(\frac{3C'}{C})^{kC/C'} \\
&\approx \frac{1}{(1+\delta)^2k^2p}(1+\delta)^k e^{(-\delta(1-\frac{C}{C'})+\frac{C}{C'})k}(\frac{3C'}{C})^{kC/C'} \\
&\leq \frac{1}{p}(\frac{(1+\delta)(\frac{3C'}{C})^{C/C'}}{e^{\delta(1-\frac{C}{C'})-\frac{C}{C'}}})^k
\end{aligned}$$

Hence, we have

$$\begin{aligned}
&\bar{f}(k) = \bar{f}_1(k) + \bar{f}_2(k) \\
&\leq \frac{1}{p}(\frac{(1+\delta)^2np(\frac{3C'}{C})^{C/C'}}{e^{\delta(1-\frac{C}{C'})-\frac{C}{C'}}})^k + \frac{1}{p}(\frac{(1+\delta)(\frac{3C'}{C})^{C/C'}}{e^{\delta(1-\frac{C}{C'})-\frac{C}{C'}}})^k \\
&\leq \frac{2}{p}(\frac{(1+\delta)(\frac{3C'}{C})^{C/C'}}{e^{\delta(1-\frac{C}{C'})-\frac{C}{C'}}})^k
\end{aligned}$$

since $(1+\delta)np \leq 1$. Let $g(x) = \frac{(1+\delta)(\frac{3}{x})^x}{e^{\delta(1-x)-x}}$. Since $\lim_{x\to 0} g(x) = \frac{1+\delta}{e^\delta} < 1$, we can choose small x' satisfying $g(x') < 1$. We choose $C' = max\{(1+\frac{1}{\delta})C, \frac{C}{x'}, \frac{3\log n}{-\log g(x')}\}$. Then

$$\bar{f}(k) \leq \frac{2}{p}n^{-3} = o(n^{-1}).$$

Hence for $\beta > 2.5$, the probability of the second case is at most $o(n^{-1})$ as we claim in the beginning of the proof. □

8 The diameter for $2 < \beta \leq 4$

In this range, the giant component has three layers — the core, middle layer and tree-like tails. In the previous section, we have shown the diameter of the core is $\Theta(\log n)$ and tree-like tails are of length $O(\log n)$. The remaining case is to measure the "thickness" of the middle layer. We will show that it is also of $O(\log n)$.

LEMMA 8.1. *Let S be a subset of vertices with total weight $s = o(vol(G))$. Then when $2 < \beta \leq 4$, there are two constants c_1 and $c_2 > 1$ satisfying that if $s > c_1 \log n$, then with probability at least $1 - o(n^{-1})$,*

$$vol(\Gamma(S)) > c_2 s$$

Now we will prove the rest part of Theorem 3.2.

THEOREM 8.1. *When $2 < \beta \leq 4$, almost surely the diameter of the power law random graph G_β is $\Theta(n)$.*

Proof: By Theorem 3.4, all other components except for the giant one have sizes $O(\log n)$. Therefore their diameters are at most $O(\log n)$. Hence it is enough to show that the diameter of the giant component is $\Theta(\log n)$.

For any vertices u and v in the giant component, the growing threshold in this range $4 \geq \beta > 2$ is $\frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2}\log n$. By Lemma 8.1, there are two constants c_1 and $c_2 > 1$ satisfying if $vol(S) > c_1 \log n$, then $vol(\Gamma(S)) > c_2 s$. By Lemma 7.1, there is an $i_0 = O(\log n)$, satisfying

$$vol(\Gamma_{i_0}(u)) \geq c_1 \log n$$

Apply Lemma 8.1 to $S = \Gamma_{i_0}(u)$ and let $i_1 = \lceil \frac{\log \frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2}\log n}{\log c_2} \rceil = O(\log n)$, then we have $vol(\Gamma_{i_0+i_1}(u)) \geq \frac{\sum w_i^3 \sum w_i}{(\sum w_i^2)^2}\log n$. Let $c = \sum w_i^2 p$ and $i_1 = \lfloor \frac{2\log vol(G)}{3\log c} \rfloor = O(\log n)$. Then almost surely we have $vol(\Gamma_{i_0+i_1+i_2}(u)) \geq n^{2/3}$. Similarly, almost surely there is an $i_3 = O(\log n)$, $i_4 = O(\log n)$ and $i_5 = O(\log n)$ satisfying $vol(\Gamma_{i_3+i_4+i_5}(v)) \geq n^{2/3}$. Hence u and v are connected by a path with length at most

$$i_0 + i_1 + i_2 + i_3 + i_4 + i_5 + 1 = O(\log n).$$

Hence the diameter of the giant component is $O(\log n)$.

On the other hand, starting from a vertex v with weight less than 2. The probability that $\Gamma(v)$ is again a vertex with weight less than 2 is about $e^\alpha p e^{e^\alpha p}$ which is a constant. Hence, it is quite possible that the volume of $\Gamma_i(v)$ doesn't grow for $\Theta(\log n)$ steps. Hence the diameter of G is at least $\Theta(\log n)$.

So the diameter of G is $\Theta(\log n)$. □

In the rest of this section, we will prove Lemma 8.1.

Proof of Lemma 8.1: Since $\sum w_i^2 p > 1$, we can choose a constant x_0 satisfying

$$\sum_{w_i \leq x_0} w_i^2 p > 1.$$

Denote $c_3 = \sum_{w_i \leq x_0} w_i^2 p$, which is a constant greater than 1 since $\beta > 2$. Denote by V_0 the set of all vertices with weights at most x_0.

For every vertex v_i in S and $v_j \in V_0 \setminus S$, let $X_{i,j}$ be the indicated random variable that $v_i v_j$ forms an edge in G. Since all pairs are independent to each other, we can use Lemma 4.1 to estimate $Y = \sum_{v_i \in S, v_j \in V_0 \setminus S} w_j X_{i,j}$.

Now we will use Y to estimate $vol(\Gamma(S))$.

$$E(Y) = \sum_{v_i \in S, v_j \in V_0 \setminus S} w_i w_j^2 p \approx s \sum_{w_i \le x_0} w_i^2 p \approx c_3 s.$$

$$\nu = \sum_{v_i \in S, v_j \in V_0 \setminus S} w_i w_j^3 p \approx s \sum_{w_i \le x_0} w_i^3 p.$$

Applying Lemma 4.1, we choose $\lambda = ts \sum w_i^2 p$. Then we have

$$\begin{aligned} Pr(vol(\Gamma(S)) \mid &< \frac{1+c_3}{2} s) \\ &\le e^{(\frac{1-c_3}{2} s)^2 / (2(1-o(1)) \sum_{w_i \le x_0} w_i^3 p)} \\ &\le e^{-(1+o(1)) \log n} \\ &= o(n^{-1}) \end{aligned}$$

provided $s > \frac{(1-c_3)^2}{8 \sum_{w_i \le x_0} w_i^3 p} \log n$.
Let $c_1 = \frac{(1-c_3)^2}{8 \sum_{w_i \le x_0} w_i^3 p}$ and $c_2 = \frac{1+c_3}{2}$. We are done. □

9 Open Problems

Most massive graphs fall into the range $\beta > 2$. Theorem 3.2 implies that there are two constants c_1 and c_2 satisfying

$$c_1 \le \frac{diameter}{\log n} \le c_2.$$

Currently there is a big gap between c_1 and c_2.

Open Problems:

1. Does the limit $\lim_{n \to \infty} \frac{diameter}{\log n}$ exist? If exists, what is the value?

2. As we know, there is a $\Theta(\log n)$-path connecting any pair of vertices (in the same component). Is there any deterministic algorithm constructing a $\Theta(\log n)$-path with running time $o(n^\epsilon)$ for any $\epsilon > 0$?

10 ACKNOWLEDGMENTS

I would like thank Fan Chung Graham for many invaluable discussions, William Aiello for many useful ideas of power law graphs, and Sam Buss for his encouragement.

References

[1] J. Abello, A. Buchsbaum, and J. Westbrook, *Proc. 6th European Symposium on Algorithms*, pp. 332–343, 1998.

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

[3] 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.

[4] W. Aiello, F. Chung and L. Lu, Random evolution in massive graphs, preprint.

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

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

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

[8] A. Barabási, R. Albert, and H. Jeong Scale-free characteristics of random networks: the topology of the world wide web, *Elsevier Preprint* August 6, 1999.

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

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

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

[12] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tompkins, and J. Wiener, "Graph Structure in the Web," *proceedings of the WWW9 Conference*, May, 2000, Amsterdam.

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

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

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

[16] Fan Chung and Linyuan Lu, The Diameter of Random Sparse Graphs, preprint. http://euclid.ucsd.edu/ llu/diameter.ps.

[17] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the internet topology, *Proceedings of the ACM SIGCOM Conference*, Cambridge, MA, 1999.

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

[19] J. Kleinberg, S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, The web as a graph: Measurements, models and methods, *Proceedings of the International Conference on Combinatorics and Computing*, 1999.

[20] S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, Trawling the web for emerging cyber communities, *Proceedings of the 8th World Wide Web Conference*, Edinburgh, Scotland, May 15–19, 1999.

[21] S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, Extracting large-scale knowledge bases from the web, *Proceedings of the 25th VLDB Conference*, Edinburgh, Scotland, 1999.

[22] Lawrence, S. and Giles, C. L. *Nature* **400**, (1999) 107-109.

[23] Tomasz Łuczak, Random Trees and Random Graph. *Random structures & algorithm* **13**, pp. 485-500.