ON THE DIAMETER OF A CLASS OF RANDOM GRAPHS

Thomas K. Philips
Don Towsley
Jack K. Wolf

COINS Technical Report 87-101
July 15, 1987

ON THE DIAMETER OF A CLASS OF RANDOM GRAPHS

by

Thomas K. Philips

IBM T.J. Watson Research Center

P.O. Box 704

Yorktown Heights, NY 10598

Don Towsley

The Department of Computer and Information Science

The University of Massachusetts

Amherst, MA 01003

and

Jack K. Wolf

The Center for Magnetic Recording Research

S-008

The University of California at San Diego

La Jolla, CA 92093

Abstract

The diameter of a class of directed random graphs in which the outdegree of each node is constrained to be exactly k is examined. The arcs in the graph are selected as follows. Each node connects itself to k other distinct nodes with outwardly directed arcs, all possible sets of of k nodes being chosen with equal probability. It is shown that the diameter of this random graph almost surely takes on only one of two values.

1. Introduction

The connectivity and Hamiltonicity of directed random graphs on N vertices in which the out-degree of each node is constrained to be k has been studied by Fenner and Frieze [2] and McDiarmid [5]. They have shown that

1. If $k \geq 2$, the graph is almost surely weakly connected.
2. If $k \geq (1+\varepsilon)\ln N,\ \varepsilon > 0$, the graph is almost surely strongly connected and almost surely possesses a Hamiltonian cycle.

In this paper, we show that if $k = c \ln N,\ c \geq 4.5$, with probability $\rightarrow 1$ as $N \rightarrow \infty$ the diameter of the graph takes on one of only two possible values- $\lceil \log_k\{(k-1)N+1\} - 1 \rceil$ and $\lceil \log_k\{(k-1)N+1\} \rceil$, where $\lceil x \rceil$, is the smallest integer greater than or equal to x.

2. Some Useful Inequalities

A number of inequalities that prove useful subsequently are listed here.

$$\binom{N}{k} < \frac{N^k}{k!}\exp\left[-\frac{\binom{k}{2}}{N} - \frac{\binom{k}{2}(2k-1)}{6N^2} - \frac{\left[\binom{k}{2}\right]^2}{3N^3} \cdots\right] < \frac{N^k}{k!} \tag{1}$$

$$\sqrt{2\pi N}\,[\frac{N}{e}]^N < N! < \sqrt{2\pi N}\,[\frac{N}{e}]^N \exp(\frac{1}{12N}). \tag{2}$$

For a binomial random variable with parameters n and p, define

$$b(n,i,p) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Then

$$\sum_{i=r}^{n} b(n,i,p) < \binom{n}{r} p^r \qquad (3)$$

Proofs of these inequalities may be found in [3].

2. Outline of the Proof

The diameter of these random graphs is determined in the following way. Some node is selected as a root. A breadth first search tree is constructed from this root and is used to compute a lower bound, $\hat{i}$, on the diameter in terms of N and k. It is shown that this lower bound holds for all values of k in some range (k_1, k_2). Upper bounds on the diameter are then derived for the smallest and largest possible value of k concommitant with the specified $\hat{i}$. In the first case $(k = k_1)$ it is shown that the diameter is at most $\hat{i} + 1$, while in the second $(k = k_2)$, it is shown to be at most $\hat{i}$. Lastly we conclude that if $k_1 < k < k_2$ the diameter can take only these two values.

The technique used to prove the upper bound is shown in figure 1. For convenience, the root node is given a label of 1. Nodes that lie at a distance i from the root node are referred to as the nodes at level i. Node 1 is of course the only node at level 0. At every level, the number of newly contacted nodes is estimated. At level $\hat{i} + 1$ every node in the graph is shown to be contacted.

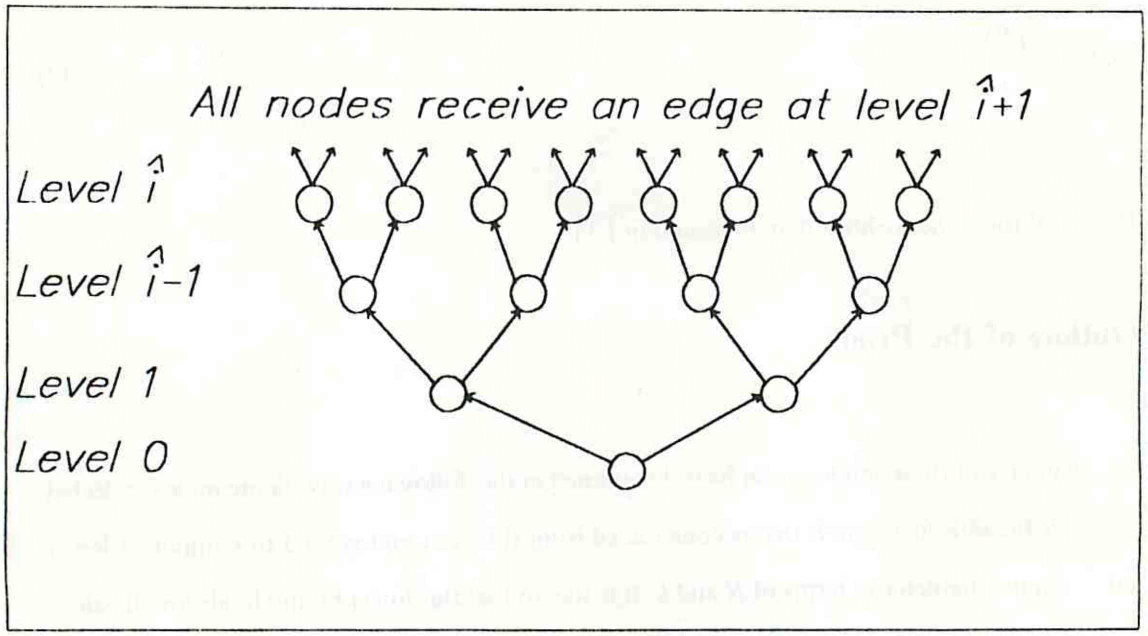

Figure 1. Finding the diameter of a Random Graph.

3. The Proof

Lemma 1:

$d(G) \geq \lceil \log_k\{(k-1)N+1\} - 1 \rceil, \triangleq \hat{i}$

Proof

Referring to figure 1, consider tracing links starting at node 1. As each node puts out k directed edges whose endpoints are necessarily distinct, at level i we may contact at most k^i new nodes. It follows that the number of nodes contacted at levels $0, 1, \ldots, i$ can then be at most

$$1 + k + k^2 + \cdots k^i = \frac{k^{i+1}-1}{k-1}.$$

To lower bound the diameter, note that if a path is to exist to all nodes by level $\hat{i}$ but not earlier, it must be that

$$\frac{k^{\hat{i}+1}-1}{k-1} \geq N \qquad (4)$$

and

$$\frac{k^{\hat{i}}-1}{k-1} < N \qquad (5)$$

These may now be rearranged to give

$$d(G) \geq \lceil \log_k\{(k-1)N+1\} - 1 \rceil \triangleq \hat{i}$$

as required.

Observation 1.

The lower bound holds for all values of k in some range. This range is next computed. Looking at (4) and (5), two extreme situations can be identified.

1. $$\frac{k_1^{\hat{i}+1}-1}{k_{1-1}} = N \text{ and } \frac{k_1^{\hat{i}}-1}{k_1-1} << N \qquad (6)$$

2. $$\frac{k_2^{\hat{i}+1}-1}{k_2-1} >> N \text{ and } \frac{k_2^{\hat{i}}-1}{k_2-1} = N-1. \qquad (7)$$

Clearly, k_1 and k_2 define the range of k concommitant with the specified value of $\hat{i}$. The diameter is more likely to exceed the lower bound in situation 1 than in situation 2, as $k_1 < k_2$. We shall show that in the first situation, the diameter is almost surely $\hat{i}+1$, while in the second case it is almost surely $\hat{i}$. As the diameter of this random graph is a non increasing function of k, it follows that for k such that $k_1 < k < k_2$, the diameter must be either $\hat{i}$ or $\hat{i}+1$.

Lemma 2:

If (6) is satisfied and $k = c\ln N$, then $\lim_{N\to\infty} Pr[d(G) = \hat{i}\,] = 0$.

Proof:

Construct the random graph as follows. Starting at node 1, choose k endpoints at random and join node 1 to them by a directed edge. From each of these newly contacted nodes choose k endpoints at random and once again join their "parent" node to them with directed edges. If $d(G)$ is not to exceed $\hat{i}$ every node other than node 1 must have an indegree of 1. It is next shown that by the time the first $N - 1$ edges are added, there is almost surely a node with indegree greater than 1.

$$Pr[\text{First N-1 edges contact new nodes}\,] = \frac{\binom{N-1}{k}\binom{N-k-1}{k}\cdots\binom{k+(N-1)\bmod k}{k}}{\binom{N-1}{k}\binom{N-1}{k}\cdots\binom{N-1}{k}}$$

$$< \frac{\binom{2k-1}{k}}{\binom{N-1}{k}}$$

$$< \left(\frac{2k}{N}\right)^k$$

If $k = c\ln N$, $\lim_{N\to\infty} Pr[\text{First N-1 edges contact new nodes}] = 0$, and the lemma follows.

Lemma 3:

If (6) is satisfied at least $k^2 - 1$ new nodes are contacted at level 2.

Proof:

Clearly, the probability of wasting an edge by contacting a previously contacted node is increased if we sample with replacement, or choose edges independently.

$$Pr[\text{ The } i^{th} \text{ edge contacts a previously contacted node }] < \frac{i-1}{N-1} < \frac{k^2-1}{N-1}\,,\ 1 < i < k^2.$$

Therefore,

$$Pr[\text{ 2 or more edges contact previously contacted nodes }] < \sum_{1 \le i < j \le k^2} \frac{(i-1)(j-1)}{(N-1)^2}$$

$$< k^4 \times \left[\frac{k^2-1}{N-1} \right]^2 \qquad (8)$$

$$= O(\frac{\log^8 N}{N^2})$$

if $k = c \ln N$.

It immediately follows that at most 1 edge is wasted at level 2. As no edges can be wasted at level 1, we have that the number of newly contacted nodes at level 2 is at least $k^2 - 1$ with probability $1 - O(\frac{\log^8 N}{N^2})$.

Lemma 4:

If (6) is satisfied, at level i, $3 \le i \le \lfloor \frac{\hat{i}}{2} \rfloor$ at most k edges are lost at every level.

Proof:

At level i there can be at most k^i edges available to contact new nodes. The probability of a specified edge contacting a previously contacted node is upper bounded by $\frac{k^{i+1}-1}{(k-1)N}$, as at most $\frac{k^{i+1}-1}{k-1}$ nodes can be contacted at levels up to and including i.

Let X be a binomial random variable with parameters k^i and $\frac{k^{i+1}-1}{(k-1)N}$. Then for $l > 0$

$Pr[\text{ \# of lost edges } > l] < Pr[X > l]$

From (3), it follows that

$$
\begin{aligned}
Pr[X \geq k] &< \binom{k^i}{k}\left[\frac{k^{i+1}-1}{(k-1)N}\right]^k \\
&< \binom{k^i}{k}\left[\frac{k^{i+1}-1}{k^{\hat{i}+1}-1}\right]^k \\
&< \frac{k^{ik}}{k!}\left[\frac{k^i}{k^{\hat{i}}}\right]^k \\
&< \frac{1}{k!}\left[\frac{k^{2i}}{k^{\hat{i}}}\right]^k \\
&< \frac{1}{k!}
\end{aligned}
\tag{9}
$$

From Stirlings approximation (2) and the fact that $(\ln N)^{lnN} = N^{lnlnN}$ we see that if $k = clnN$, $c > 1$ then $\frac{1}{k!} = O(N^{1-lnlnN})$.

Lemma 5:

Let $\hat{i} \geq 6$ and $k = clnN$, $c > 1$. Then if (6) is satisfied, at level i, $\lfloor \frac{\hat{i}}{2} \rfloor + 1 \leq i \leq \hat{i} - 3$, the number of edges lost is at most k^{i-2}.

Proof:

Following the proof of Lemma 4, we may define a set of binomial random variables $\{X_i\}$ with parameters k^i and $\frac{k^{i+1}-1}{(k-1)N}$.

These random variables have the property that

$Pr[\ \#$ of lost edges at level $i > l] < Pr[X_i > l]$

Then,

$$Pr[\ \#\text{ of lost edges at level } i > k^{i-2}\ ] < \binom{k^i}{k^{i-2}} \left[\frac{k^{i+1}-1}{(k-1)N} \right]^{k^{i-2}}$$

$$< \frac{1}{k^{i-2}!} \left[\frac{k^{2i}}{k^{\hat{i}}} \right]^{k^{i-2}}$$

$$< \left[\frac{ek^{2i}}{k^{i-2+\hat{i}}} \right]^{k^{i-2}} \qquad (11)$$

$$< \left[\frac{e}{k} \right]^{k}$$

$$< \frac{e^{c \ln N}}{\ln N^{\ln N}}$$

$$= O(N^{c - \ln \ln N})$$

where the last step follows from Stirlings approximation.

Observation 2.

There is an N_1 such that for $N > N_1$ the probability of the loss at any level between $\lfloor \frac{\hat{i}}{2} \rfloor$ and $\hat{i} - 3$ exceeding $k^{i-2} = o(\frac{1}{N})$.

Lemma 6

If (6) is satisfied, $\hat{i} \geq 6$, and $k = c \ln N$, $c > 1$, at level $\hat{i} - 2$ the number of edges lost is at most $k^{\hat{i} - 4 + \varepsilon}$ for any fixed $\varepsilon > 0$.

Proof:

Once again following the technique employed in the last two lemmas, we have

$$Pr[\ \#\text{ of lost edges at level } \hat{i} - 2 > k^{\hat{i} - 4 + \varepsilon}\ ] < \left[\frac{e}{k^{\varepsilon}} \right]^{k^{\hat{i} - 4 + \varepsilon}}$$

$$< \left[\frac{e}{k^{\varepsilon}} \right]^{k} \qquad (12)$$

As $k = c \ln N$, $c > 1$, (12) can be further bounded to give

$$Pr[\ \#\text{ of lost edges at level } \hat{i} - 2 > k^{\hat{i}-4+\varepsilon}\ ] < \frac{N^{c}}{lnN^{\varepsilon lnN}} = N^{c-\varepsilon lnlnN} \tag{13}$$

Observation 3.

For any fixed $\varepsilon > 0$, there is an $N_2(\varepsilon)$ such that for $N > N_2(\varepsilon)$, $Pr[\ \#\text{ of lost edges at level } \hat{i} - 2 > k^{\hat{i}-4+\varepsilon}\ ] = o(\frac{1}{N})$.

Lemma 7

If $\hat{i} > 6$, at level $\hat{i} - 1$ the probability that the number of edges lost exceeds $k^{\hat{i}-2+\varepsilon}$ for any fixed $\varepsilon > 0$ is $O(N^{c-\varepsilon lnlnN})$ if $k = c \ln N$, $c > 1$.

Proof:

Identical to that of Lemma 6.

Observation 4.

For any fixed $\varepsilon > 0$, there is an $N_3(\varepsilon)$ such that for $N > N_3(\varepsilon)$, $Pr[\ \#\text{ of lost edges at level } \hat{i} - 1 > k^{\hat{i}-2+\varepsilon}\ ] = o(\frac{1}{N})$.

Lemma 8:

For sufficiently large N, if $\hat{i} \geq 7$, at least $.99k^{\hat{i}-1}$ new nodes are contacted with probability $1 - o(\frac{1}{N})$ at level $\hat{i} - 1$.

Proof:

This follows from the last three lemmas. The number of new nodes contacted at every level is computed as follows.

At level 0 we contact 1 new node.

At level 1 we contact k new nodes.

At level 2 we contact at least $k^2 - 1$ new nodes.

At level 3 we contact at least $(k^2 - 1)k - k = k^3 - 2k$ new nodes.

If continued upto level $\hat{i} - 3$ we find that the number of nodes contacted by level $\hat{i} - 3$ is lower bounded by

$$k^{\hat{i}-3} - (\lceil \frac{\hat{i}}{2} \rceil - 1)k^{\hat{i}-5} - k^{\hat{i}-6} - \cdots - k^{\lceil \frac{\hat{i}}{2} \rceil - 2} \quad (13)$$

We can weaken the bound by adding terms to get

$$\text{Number of nodes at level } \hat{i} - 3 \geq k^{\hat{i}-3} - (\lceil \frac{\hat{i}}{2} \rceil - 1)k^{\hat{i}-5} - k^{\hat{i}-6} - \cdots - 1$$
$$> k^{\hat{i}-3} - (\lceil \frac{\hat{i}}{2} \rceil - 1)k^{\hat{i}-5} - \frac{k^{\hat{i}-5} - 1}{k - 1} \quad (14)$$

Lemmas 6 and 7 may now be used to give

$$\text{Number of nodes at level } \hat{i} - 1 > k^{\hat{i}-1} - (\lceil \frac{\hat{i}}{2} \rceil - 1)k^{\hat{i}-3} - \frac{k^{\hat{i}-3} - k^2}{k - 1} - k^{\hat{i}-3+\varepsilon} - k^{\hat{i}-2+\varepsilon}$$
$$> k^{\hat{i}-1}\left[1 - \frac{\lceil \frac{\hat{i}}{2} \rceil}{k^2} - \frac{1}{(k-1) \times k^2} - \frac{1}{k^{2-\varepsilon}} - \frac{1}{k^{1-\varepsilon}}\right] \quad (15)$$

Notice that $\hat{i}$ grows as log log N while k grows logarithmically with N. If $k = c \ln N$, for any fixed c the bracketed term in (15) is an increasing function of N. It follows that for any fixed $\varepsilon > 0$ there is an $N_4(c, \varepsilon)$ such that for all $N > N_4$, the bracketed term is greater than .99.

If $N > \max\{N_1, N_2(\varepsilon), N_3(\varepsilon), N_4(c, \varepsilon)\}$, this inequality holds with probability $1 - o(\frac{1}{N})$.

Lemma 9:

If $k = c \ln N$, $c \geq 1$, at least $.49k^{\hat{i}}$ new nodes are contacted at level $\hat{i}$ with probability $1 - o(\frac{1}{N})$ for sufficiently large N.

Proof:

From the last lemma, under the conditions stated above, at least $.99k^{\hat{i}-1}$ nodes are contacted at level $\hat{i}-1$ with probability $1-o(\frac{1}{N})$. Each of these nodes puts out k edges and therefore there are at least $.99k^{\hat{i}}$ edges available to contact nodes at level $\hat{i}$.

Of course, not all of these edges will contact new nodes. A given edge may be wasted in one of two ways:

1. It may contact a node that was contacted at level $\hat{i}-1$ or lower, or
2. It may contact a node at level $\hat{i}$ that has already been contacted by some other edge.

We shall call the the first kind of loss a *backtracking loss* and the second kind an *overlap loss*.

For any fixed $\delta > 0$

$$
\begin{aligned}
Pr[\text{ Backtracking loss exceeds } \delta k^{\hat{i}}\,] &< \binom{.99k^{\hat{i}}}{\delta k^{\hat{i}}}\left[\frac{k^{\hat{i}}-1}{(k-1)N}\right]^{\delta k^{\hat{i}}} \\
&< \left[\frac{.99e}{\delta k}\right]^{\delta k} \\
&= O\left(N^{\delta\left(\ln\left[\frac{.99e}{\delta}\right]-\ln\ln N\right)}\right)
\end{aligned}
\tag{16}
$$

For any fixed $\delta > 0$, there is an $N_5(\delta)$ such that for $N > N_5(\delta)$, $Pr[\text{ Backtracking loss exceeds } \delta k^{\hat{i}}\,] = o(\frac{1}{N})$.

The overlap loss is bounded indirectly. The number of nodes that are <u>not</u> contacted at level i is upper bounded, resulting in a lower bound on the number of nodes that are contacted. By doing so we avoid the need to estimate the number of nodes that receive more than one edge. Now,

$$
Pr[\ \#\text{ of uncontacted nodes } \geq \gamma k^{\hat{i}}\,] < \binom{k^{\hat{i}}}{\gamma k^{\hat{i}}}\left[1-\frac{\gamma k^{\hat{i}}}{N}\right]^{.99k^{\hat{i}}}. \tag{17}
$$

The binomial coefficient is the number of different sets containing $yk^{\hat{i}}$ nodes, while the term that follows it is the probability that none of the nodes in a specified set receives an edge. (17) can be simplified using (1), (3) and the fact that $k^{\hat{i}} < N$ to give

$$
\begin{aligned}
Pr[\ \#\text{ of uncontacted nodes } \geq yk^{\hat{i}}\ ] &< \binom{k^{\hat{i}}}{yk^{\hat{i}}}(1-y)^{.99k^{\hat{i}}} \\
&= O\left(\frac{k^{\hat{i}\,yk^{\hat{i}}} \exp[\,-k^{\hat{i}}(\frac{y^2}{1.98} + \frac{y^3}{5.8806} + \frac{y^4}{11.6436})]}{yk^{\hat{i}}!}(1-y)^{.99k^{\hat{i}}} \right) \\
&= O\left(\left[\left(\frac{1}{y}\right)^{y} \exp(y - \frac{y^2}{1.98} - \frac{y^3}{5.8806} - \frac{y^4}{11.6436})(1-y)^{.99} \right]^{k^{\hat{i}}} \right).
\end{aligned}
$$

If $y = .504$

$$
\begin{aligned}
Pr[\ \#\text{ of uncontacted nodes } \geq .504k^{\hat{i}}\ ] &= O(.99956^{k^{\hat{i}}}) \\
&\sim O(.99956^{N}).
\end{aligned}
\qquad (18)
$$

It follows that we must almost surely contact $(1 - .504)k^{\hat{i}} = .496k^{\hat{i}}$ nodes at level $\hat{i}$. These $.496k^{\hat{i}}$ nodes include those to which edges are lost due to backtracking, and this loss must be accounted for. Clearly, we can find an N_6 such that for $N \geq N_6$, $Pr[\ \#\text{ of nodes } \textit{contacted} < .496k^{\hat{i}}\ ] = o(\frac{1}{N})$.

Now let $\varepsilon > 0$ be fixed, $\delta = .006$ and $N > \max\{N_1, N_2(\varepsilon), N_3(\varepsilon), N_4, N_5(.006), N_6\}$. Then we contact at least $(.496 - .006)k^{\hat{i}} = .49k^{\hat{i}}$ new nodes at level $\hat{i}$ with probability $1 - o(\frac{1}{N})$.

Theorem:

If (6) is satisfied , $\hat{i} \geq 7$, and $k = c \ln N$, $c \geq 4.5$, then for sufficiently large N, $d(G) = \hat{i} + 1$ with probability $1 - o(1)$.

Proof:

At level $\hat{i}$ at least $.49k^{\hat{i}}$ nodes were contacted with probability $1 - o(\frac{1}{N})$. These nodes in turn put out k edges each, giving us at least $.49k^{\hat{i}}$ edges with which we can contact nodes at level $\hat{i}+1$. As $N = \frac{k^{\hat{i}+1}-1}{k-1}$, there must be an N_7 such that for $N > N_7$, $k^{\hat{i}} > .99N$, implying that for $N > N_0 \triangleq \max\{N_1, N_2(\varepsilon), N_3(\varepsilon), N_4, N_5(.01), N_6, N_7\}$ at least $.49 \times .99N = .4851N$ edges are available to contact nodes at level $\hat{i}+1$ with probability $1 - o(\frac{1}{N})$.

We now contend that at level $\hat{i}+1$ <u>every</u> node in the graph receives an incoming edge. To show this, we need the following result due to Von Mises [4] which can also be found in Feller [1].

If r balls are placed into N boxes so that all possible configurations are equiprobable (i.e. have probability $\frac{1}{N^r}$), and if $N\exp(-\frac{r}{N})$ is bounded, then as $N \to \infty$ the probability that all the boxes are filled is asymptotically equal to $\exp(-N\exp(-\frac{r}{N}))$.

In our problem $r = .4851N \times 4.5 lnN = 2.1829 N lnN$, so that if $N > N_0$

$$\begin{aligned} Pr[\text{Every node receives an edge at level } \hat{i}+1] &> \exp(-N\exp(-2.1829 lnN)) - o(\tfrac{1}{N}) \\ &= \exp(-N^{-1.1829}) - o(\tfrac{1}{N}) \qquad (19) \\ &> 1 - N^{-1.1829} - o(\tfrac{1}{N}). \end{aligned}$$

It immediately follows that the longest path from any node (say 1) is at most $\hat{i}+1$ with probability at least $1 - N^{-1.1829} - o(\frac{1}{N})$. As any node could have been chosen as the root node, we have that

$$\begin{aligned} Pr[d(G) \le \hat{i}+1] &> 1 - N \times [N^{-1.1829} + o(\tfrac{1}{N})] \\ &= 1 - o(1). \end{aligned} \qquad (20)$$

In lemma 2, it was shown that the diameter was almost surely at least as large as $\hat{i}+1$. It follows that the diameter is exactly $\hat{i}+1$ with probability $1 - o(1)$.

Observation 5:

In the case when (7) is satisfied, a virtually identical argument shows that $d(G) = \hat{i}$ with probability $1 - o(1)$. As the diameter is a non increasing function of k it must be that for intermediate values of k, $d(G) = \hat{i}$ or $\hat{i} + 1$.

4. Summary:

The diameter of a class of random graphs has been investigated. It has been shown that if $k = c\ln N$, $c \geq 4.5$, for sufficiently large N the diameter takes on one of only two values. Explicit calculation shows the various constants $N_1, \ldots, N_9$ to be quite large. Simulation studies, on the other hand, show that the asymptotic behaviour is exhibited for graphs that contain as few as a hundred nodes or so. Interestingly, even for very small N the diameter hardly ever exceeds $\hat{i} + 2$. We believe this result to be true for all fixed $c > 1$, though we have not been able to prove it. The difficulty arises at level $\hat{i}$, where the bounds on the losses are very weak, and consequently it becomes impossible to prove that every node receives a link at level $\hat{i} + 1$.

5. *References*

1 . Feller, W., An Introduction to Probability Theory and its Applications , John Wiley and Sons, New York, 1968.

2 . Fenner, T.I., and Frieze, A.M., On the Connectivity of Random m-orientable Graphs and Digraphs, Combinatorica, 2 (1982), 347-359.

3 . Palmer, E.M. Graphical Evolution , John Wiley and Sons, New York, 1985, pp 127-136.

4 . Von Mises, R., Ü ber Aufteilungs- und Besetzungs-wahrscheinlichkeiten, Revue de la Faculte des Sciences de l'Universite d'Istanbul, N.S., (1939), 145-163.

5 . McDiarmid, C.J.H., General Percolation and Random Graphs, Adv. App. Probab., 13, (1981), 40-60