Technical Report CS 7608

The Largest Clique Size in a Random Graph*

David W. Matula

Department of Computer Science
Southern Methodist University
Dallas, Texas 75275

Abstract

The size of the largest complete subgraph in a random graph is shown to be a random variable with a very peaked density. The extent to which the probability mass for this random variable is in the neighborhood of a specified threshold function is investigated by closed form inequality bounds affording numeric examples and by sharp asymptotic formulas for the limiting behavior.

Key words: Random Graphs, Clique Number, Strong Second Moment Method, Bounds and Asymptotic Forms.

April 1976

*Partially supported by NSF Grant DCR75-10930.

I. Introduction and Summary

By a random graph $G_{n,p}$ we shall mean a graph on n vertices where each of the n(n-1)/2 edges occurs independently with probability p. The clique number of a graph is the largest number of vertices in any complete subgraph of the graph, and the random variable $Z_{n,p}$ denotes the clique number of the random graph $G_{n,p}$.

In 1970 this author [4] utilized the mean and standard deviation of the number of k-membered complete subgraphs of a random graph to derive the following bounds on the density function of $Z_{n,p}$,

$$(1)\qquad \left\{ \sum_{j=\max\{0,2k-n\}}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} p^{-j(j-1)/2} \right\}^{-1} \leq \text{Prob}\{Z_{n,p} \geq k\} \leq \binom{n}{k} p^{k(k-1)/2}$$

Computational investigation of (1) suggested a very peaked behavior for the density of $Z_{n,p}$, and in [5] we presented the asymptotic result that for the threshold function

$$z(n,1/p) = 2 \log_{1/p} n - 2 \log_{1/p} \log_{1/p} n + 2 \log_{1/p} \frac{e}{2} + 1$$

and any $\varepsilon > 0$, with $\lfloor x \rfloor$ denoting the greatest integer less than or equal to **x**,

$$(2)\qquad \lim_{n\to\infty} \text{Prob}\{\lfloor z(n,1/p)-\varepsilon \rfloor \leq Z_{n,p} \leq \lfloor z(n,1/p)+\varepsilon \rfloor\} = 1.$$

Thus $Z_{n,p}$ takes on one of at most two values depending on n,p with **probability** approaching unity as $n\to\infty$. A computational indication of the sharpness of the approach to the limit was given in [5] by noting that the random graph on 1000 vertices with edge probability .5 will have a largest clique of size exactly 15 over 80% of the time.

The distribution of $Z_{n,p}$ has recently been the subject of more extensive investigations. Grimmett and McDiarmid [3] provided a result on the tails of the distribution of $Z_{n,p}$ by showing that the sequence of random variable $\{Z_{n,p}\}$ satisfies

$$\frac{Z_{n,p}}{\log n} \to \frac{2}{\log 1/p} \qquad \text{as } n \to \infty$$

almost surely and in any mean. Bollobás and Erdös [1] describe a measure on infinite graphs $G_{N,p}$ $\{N=1,2,3,\ldots\}$ such that the induced subgraph $G_{n,p}$ on the initial segment of vertices $\{1,2,\ldots, n\}$ has the same probability as in our model of random graphs, and they then show that almost every infinite graph $G_{N,p}$ has at most a finite number of initial segment graphs $G_{n,p}$ for which the clique number of $G_{n,p}$ is not within a small neighborhood of the threshold function $z(n,1/p)$. They also obtain a tighter asymptotic upper bound on $\text{Prob}\{Z_{n,p} \leq (1-\varepsilon)z(n,1/p)\}$ than that directly implied by the left hand side of (1).

Our purposes in this paper are first to derive closed form applicable inequality bounds on $\text{Prob}\{\lfloor z(n,1/p)-\varepsilon\rfloor \leq Z_{n,p} \leq \lfloor z(n,1/p)+\varepsilon\rfloor\}$ which then yield an asymptotic result somewhat sharper than (2), and secondly to provide a non-trivial lower bound on the tail $\text{Prob}\{Z_{n,p} \geq \lfloor z(n,1/p)+\varepsilon\rfloor+1\}$.

In section II for N a non-negative integer valued random variable with mean E(N) and standard deviation $\sigma < \infty$, we derive what we term the <u>strong second moment inequality</u>.

$$(3) \qquad \frac{E^2(N)}{\sigma^2+E^2(N)} = \frac{E(N)}{E(N \mid \text{sampling by weight})} \leq \text{Prob}\{N\neq 0\} \leq E(N).$$

The lower bound formulation utilizing the concept of sampling by weight avoids explicit determination of the second moment and appears better suited to investigation of structures in random combinatorial configurations, and in section III is immediately applied to obtain the fundamental inequality (1). Efficient computational utilization of (1) is exhibited in example 1 where the spiked density function for $Z_{1000,.5}$ is shown to satisfy

$$\mathrm{Prob}\{Z_{1000,.5} \le 13\} \le .0224,$$

$$\mathrm{Prob}\{Z_{1000,.5} \le 14\} \le .1510,$$

$$.8171 \le \mathrm{Prob}\{Z_{1000,.5} = 15\} \le .9807,$$

$$.0193 \le \mathrm{Prob}\{Z_{1000,.5} = 16\} \le .0318,$$

$$\mathrm{Prob}\{Z_{1000,.5} \ge 17\} \le .00003.$$

Closed form approximations to the right and left hand sides of inequality (1) are then pursued and specific functions c_1, c_2, and c_3 bounded by constants independent of n are determined in theorems 3 and 4 which yield for specified ranges of n, p and δ,

$$\mathrm{Prob}\{Z_{n,p} \ge z(n,1/p)+\delta\} \le c_1\left(\frac{1}{n}\right)^{\delta}\left(\frac{2\log n}{e\log 1/p}\right)^{\delta+c_2},$$

$$\mathrm{Prob}\{Z_{n,p} \le z(n,1/p)-1-\delta\} \le c_3\left(\frac{1}{n}\right)^{\delta}\left(\frac{2\log n}{e\log 1/p}\right)^{\delta+1/2}.$$

Theorems 3 and 4 are then utilized to compute $\mathrm{Prob}\{Z_{10^{10},.25} = 30\} > .9997$, providing numerically sharper evidence of the spiked behavior of $Z_{n,p}$. Finally for $z(n,1/p)+\varepsilon < k \le 2z(n,1/p)$, inequalities are derived in theorem 5 which show that $\mathrm{Prob}\{Z_{n,p} \ge k\} \to \binom{n}{k}p^{k(k-1)/2}$ as $n\to\infty$, thus affording a non-trivial lower bound on $\mathrm{Prob}\{Z_{n,p} \ne \lfloor z(n,1/p)\rfloor\}$. Results on the asymptotic behavior of the density of $Z_{n,p}$ obtained from the inequalities of section III are summarized in section IV.

II. The Strong Second Moment Inequality

Let N be a non-negative integer valued random variable with finite mean E(N). The <u>associated weighted random variable</u> N_w is the random variable with density function $\{s_j\}$ given by

$$(4)\qquad s_j = \frac{jr_j}{\sum_{i=1}^{\infty} ir_i} \quad \text{for } j \geq 1, \text{ where } r_i = \text{Prob}\{N=i\},\ i \geq 0.$$

The choice of an integer prescribed by the density $\{s_j\}$ is termed <u>sampling by weight</u>, and the resulting expectation is given by

$$(5)\qquad E(N \mid \text{sampling by weight}) = E(N_w) = \frac{\sum_{j=1}^{\infty} j^2 r_j}{\sum_{i=1}^{\infty} ir_i} \quad \text{for } \sum_{j=1}^{\infty} j^2 r_j < \infty .$$

<u>Theorem 1 (Strong Second Moment Inequality)</u>:

Let N be a non-negative integer valued random variable with mean E(N), standard deviation $\sigma < \infty$, and let N_w be the associated weighted random variable with mean $E(N_w)$. Then Prob{N≠0} satisfies

$$(6)\qquad \frac{E(N)}{E(N_w)} = \frac{E^2(N)}{\sigma^2 + E^2(N)} \leq \text{Prob}\{N \neq 0\} \leq E(N).$$

Furthermore, the right hand inequality is tight if and only if Prob{N≥2} = 0, and the left hand inequality is tight if and only if Prob{N=k} > 0 for at most one non-zero value of k.

<u>Proof</u>: For the right hand inequality

$$\text{Prob}\{N \neq 0\} = \sum_{i=1}^{\infty} r_i \leq \sum_{i=1}^{\infty} ir_i = E(N),$$

where equality is obtained if and only if $r_i = 0$ for $i \geq 2$. To obtain the left hand inequality note, in general, that $2ij \leq i^2 + j^2$ with equality if and only if $i=j$. Thus

$$E^2(N) = \left(\sum_{i=1}^{\infty} i r_i\right)^2$$

$$= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} ij\ r_i r_j$$

$$\leq \frac{1}{2}\left(\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} (i^2+j^2) r_i r_j\right)$$

$$= \left(\sum_{j=1}^{\infty} j^2 r_j\right)\left(\sum_{i=1}^{\infty} r_i\right)$$

$$= \left(\sum_{j=1}^{\infty} j^2 r_j\right)\text{Prob}\{N\neq 0\},$$

so that

$$\frac{E(N)}{\sum_{j=1}^{\infty} \frac{j^2 r_j}{E(N)}} \leq \text{Prob}\{N\neq 0\}$$

with equality if and only if $r_k \neq 0$ for at most one value of $k \geq 1$. Utilizing (5), the theorem is complete.

Note that in contrast to the lower bound in (6), a standard application of the Chebyshev inequality yields only the weaker result Prob$\{N\neq 0\} \geq 1 - \sigma^2/E^2(N)$. Erdös and Spencer successfully utilize the Chebyshev inequality in several counting problems in [2] terming the technique the "second moment method". Thus we refer to the use of the sharper inequality (6) as the "strong second moment method" and will now contrast the two methods.

For our investigation of cliques in random graphs, we shall be particularly interested in bounding the Prob$\{N(n)\neq 0\}$ for a particular sequence $N(n)$, $n=1,2,\ldots$ of non-negative integer valued random variables. Letting $N(n)$ have mean μ_n and standard deviation σ_n for $n \geq 1$, the general case of asymptotically increasing means where $\mu_n \to \infty$, $\sigma_n/\mu_n \to 0$ as $n \to \infty$ is seen to provide the same asymptotic lower bound on Prob{N(n)≠0} by either (6) or the Chebyshev inequality. For asymptoically decreasing means where $\mu_n \to c \leq 1/2$ as $n \to \infty$ the Chebyshev lower bound is trivial, whereas if $\sum_{j=1}^{\infty} j^2 \text{Prob}\{N(n)=j\}/\text{Prob}\{N(n)=1\} \to 1$ as $n \to \infty$, the strong second moment method yields the asymptotic equality

$$\text{Prob}\{N(n)\neq 0\} = \mu_n + o(\mu_n).$$

The latter situation is obtained in our application to cliques in random graphs, as will be demonstrated in theorem 5.

The formulation of the lower bound in (6) in terms of $E(N_w)$ can be useful for many random combinatorial problems. For the randomly chosen combinatorial structure Y, let A_i, i=1,2,..., m denote the occurrence of the ith substructure in Y, and let N denote the number of A_i occurring in Y. Suppose the symmetries of Y are such that the occurrence of any particular A_i is not favored by knowledge that exactly k of the m substructures occur in a chosen structure Y. That is, assume

$$\text{Prob}\{A_1|N=k\} = \text{Prob}\{A_i|N=k\} = \frac{k}{m} \text{ for } 1 \leq i \leq m,\ 0 \leq k \leq m.$$

It is then readily shown that $E(N|A_1) = E(N_w)$, which proves the following corollary.

<u>Corollary 1.1</u>: Let N be the number of events $\{A_i\}$, i=1,2,..., m that occur when a sample point of the sample space Σ is chosen. If

$$\text{Prob}\{A_1|N=k\} = \text{Prob}\{A_i|N=k\} = \frac{k}{m} \text{ for } 1 \leq i \leq m,\ 0 \leq k \leq m,$$

then

$$(7) \qquad \frac{E(N)}{E(N|A_1)} \leq \text{Prob}\{N\neq 0\} \leq E(N).$$

This latter formulation of the strong second moment inequality is particularly appropriate for the investigation of cliques in random graphs.

III. Bounds on The Clique Number of a Random Graph

For $n \geq 1$, $0 \leq p \leq 1$, let the sample space Σ be composed of all graphs on n vertices. The probability of each particular graph $G_{n,p} \varepsilon \Sigma$ where $G_{n,p}$ has m edges is then $p^m(1-p)^{m(m-1)/2-m}$ in our formulation of a random graph. The clique number of a graph is the number of vertices in a largest complete subgraph of the graph. Let the random variable $Z_{n,p}$ denote the clique number of the random graph $G_{n,p}$. If a sample graph $G_{n,p} \varepsilon \Sigma$ has a complete subgraph on k vertices, then this fact assures that $Z_{n,p} \geq k$ for that sample graph $G_{n,p}$. Application of the strong second moment method yields the following bounds noted in [4].

Theorem 2: For $n \geq 1$, $0<p<1$, let $Z_{n,p}$ be the clique number of the random graph $G_{n,p}$. Then for $1 \leq k \leq n$,

$$\left\{ \sum_{j=\max\{0,2k-n\}}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} p^{-j(j-1)/2} \right\}^{-1} \leq \text{Prob}\{Z_{n,p} \geq k\} \leq \binom{n}{k} p^{k(k-1)/2} . \tag{8}$$

Proof: Let the index $i=1,2,\ldots,\binom{n}{k}$ correspond to the $\binom{n}{k}$ k-membered subsets of vertices of the random graph $G_{n,p}$. Let A_i denote the event that the random graph restricted to the ith k-membered subset of vertices is a complete graph. Thus $\text{Prob}\{A_i\} = p^{k(k-1)/2}$ for $1 \leq i \leq \binom{n}{k}$, and with N denoting the number of A_i that occur in a random graph,

$$E(N) = \binom{n}{k} p^{k(k-1)/2}. \tag{9}$$

It is immediate from the definition of a random graph that

$$\text{Prob}\{A_1 | N=\ell\} = \text{Prob}\{A_i | N=\ell\} = \frac{\ell}{\binom{n}{k}} \text{ for } 1 \leq i \leq \binom{n}{k},\ 0 \leq \ell \leq \binom{n}{k}.$$

The event A_1 determines that all edges on a particular set of k of the n vertices must occur, so the event A_i given A_1 has probability $p^{k(k-1)/2-j(j-1)/2}$ if the 1st and ith k-membered sets have j vertices is common.

Hence

$$(10) \qquad E(N|A_1) = \sum_{j=\max\{0,2k-n\}}^{k} \binom{n-k}{k-j}\binom{k}{j} p^{k(k-1)/2-j(j-1)/2} \quad .$$

Noting that $\text{Prob}\{N \neq 0\} = \text{Prob}\{Z_{n,p} \geq k\}$, equations (9), (10) and corollary 1.1 yield the theorem.

Theorem 2 is now utilized to show that the discrete valued random variable $Z_{n,p}$ can have a surprisingly peaked density.

Example 1:

Random Graph: $G_{1000,.5}$

Size: $n = 1000$

Edge probability: $p = 1/2$

Claim: The maximum clique size in a random graph on 1000 vertices with edge probability 1/2 is usually the single value 15, more specifically

$$\text{Prob}\{Z_{1000,.5} \leq 13\} \leq .0224,$$

$$\text{Prob}\{Z_{1000,.5} \leq 14\} \leq .1510,$$

$$.8171 \leq \text{Prob}\{Z_{1000,.5} = 15\} \leq .9807,$$

$$.0193 \leq \text{Prob}\{Z_{1000,.5} = 16\} \leq .0318,$$

$$\text{Prob}\{Z_{1000,.5} \geq 17\} \leq .00003.$$

To verify the claim the upper bound of (8) is utilized for k=16,17 to derive

$$\text{Prob}\{Z_{1000,.5} \geq 16\} \leq \binom{1000}{16}\left(\frac{1}{2}\right)^{16\times 15/2} = .0318,$$

$$\text{Prob}\{Z_{1000,.5} \geq 17\} \leq \binom{1000}{17}\left(\frac{1}{2}\right)^{17\times 16/2} = .000028.$$

Since $\binom{1000}{15}\left(\frac{1}{2}\right)^{15\times 14/2} = 16.96$, the upper bound of (8) is trivial for k=15.

To facilitate the computation of the lower bound in (8) assuming $n \geq 2k$, let $b = 1/p$ and

$$\alpha_j(n,k) = \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} b^{j(j-1)/2} \quad \text{for } 0 \leq j \leq k,$$

so then

$$\text{Prob}\{Z_{n,p} \geq k\} \geq \{\sum_{j=0}^{k} \alpha_j(n,k)\}^{-1}$$

Now

$$\alpha_0(n,k) = \frac{(n-k)(n-k-1)\ldots(n-2k+1)}{n(n-1)\ldots(n-k+1)},$$

$$\alpha_{j+1}(n,k) = \frac{(k-j)^2 b^j}{(j+1)(n-2k+j+1)}\alpha_j(n,k) \quad \text{for } 0 \leq j \leq k-1,$$

so the α_j may be readily computed by recursion as illustrated in table 1.

j	k=14: $\frac{(14-j)^2\ 2^j}{(j+1)(973+j}$	k=14: $\alpha_j(1000,14)$	k=15: $\frac{(15-j)2^j}{(j+1)(971+j)}$	k=15: $\alpha_j(1000,15)$	k=16: $\frac{(16-j)2^j}{(j+1)(969+j)}$	k=16: $\alpha_j(1000,16)$
0	.20143	.81980	.23171	.79587	.26418	.7710
1	.17351	.16514	.20164	.18441	.23159	.2036
2	.19692	.02865	.23158	.03718	.26913	.0472
3	.24795	.00564	.29568	.00861	.34773	.0127
4	.32753	.00139	.39712	.00254	.47358	.0044
5	.44171	.00045	.54644	.00101	.66255	.0020
6	.59764	.00020	.75800	.00055	.93772	.0013
7	.80000	.00012	1.04703	.00041	1.32786	.0013
8	1.04383	.00009	1.42367	.00043	1.86330	.0017
9	1.30346	.00010	1.88081	.00062	2.56523	.0032
10	1.51521	.00013	2.37234	.00117	3.42315	.0082
11	1.56097	.00019	2.78071	.00278	4.35374	.0282
12	1.27950	.00031	2.88473	.00774	5.13886	.1230
13	.59345	.00039	2.37862	.02234	5.36281	.6324
14		.00023	1.10890	.05315	4.44462	3.3918
15				.05894	2.08130	15.0754
16						31.3765
$\sum_{j=0}^{k} \alpha_j(1000,k)$		1.02283		1.17775		51.6839

Table 1: Computation of $\alpha_j(n,k)$ for n=1000, b=1/p=2, k=14,15,16.

Thus

$$\text{Prob}\{Z_{1000,.5} \geq 14\} \geq \frac{1}{1.02283} = .9776,$$

$$\text{Prob}\{Z_{1000,.5} \geq 15\} \geq \frac{1}{1.17775} = .8490,$$

$$\text{Prob}\{Z_{1000,.6} \geq 16\} \geq \frac{1}{51.6839} = .01934,$$

which along with the previous computed upper bounds verifies the claim.

This example demonstrates that the value 15 is essentially a threshold level for the occurrence/non-occurrence of complete subgraphs of the specified size in the random graph $G_{1000,.5}$. In general for $n \geq 1$, $0 < p = \frac{1}{b} < 1$, let the threshold function $z(n,b)$ be given by

$$z(n,b) = 2 \log_b n - 2 \log_b \log_b n + 2 \log_b \frac{e}{2} + 1. \tag{11}$$

The next two theorems demonstrate that the random graph $G_{n,p}$ will most likely have a complete subgraph of size k for any $k<z(n,1/p)-\varepsilon$ and most likely not have a complete subgraph of size k for any $k>z(n,1/p)+\varepsilon$ for any $\varepsilon>0$.

Theorem 3: Let $0 < p = \frac{1}{b} < 1$, and assume $n \geq 2$. Let $z(n,1/p)$ be the threshold function given by (11). For the integer $k \geq 1$ with $z(n,1/p) < k \leq n$, let $\delta = k - z(n,1/p) > 0$. Then

$$\text{Prob}\{Z_{n,p} \geq z(n,1/p) + \delta\} \leq \frac{(\frac{2}{e}\log_b n)^{\delta+2\log_b e}}{n^{\delta}} \; \frac{p^{(\delta+\delta^2)/2}}{e^{1+\delta}\sqrt{2\pi(z(n,1/p) + \delta)}} . \tag{12}$$

Proof: Proceeding from theorem 2, a reasonably tight upper bound on $\binom{n}{k} p^{k(k-1)/2}$ exhibiting the dependence on n,δ, and p is desired. Now $k! > (\frac{k}{e})^k\sqrt{2\pi k}$ for $k \geq 1$ by Stirling's formula, so

$$\binom{n}{k} < \frac{n^k}{k!} < (\frac{en}{k})^k \frac{1}{\sqrt{2\pi k}} . \tag{13}$$

Let $z = z(n,1/p)$, so with $k = z+\delta$,

$$k(k-1)/2 = (k+\delta)(z-1)/2 + (\delta+\delta^2)/2,$$

and from the defining equation (11) for $z(n,1/p)$,

$$p^{(z-1)/2} = \frac{2\log_b n}{en},$$

so

$$p^{k(k-1)/2} = \left(\frac{2\log_b n}{en}\right)^{k+\delta} p^{(\delta+\delta^2)/2}. \tag{14}$$

Substituting (13) and (14) into the right hand side of (7),

$$\text{Prob}\{Z_{n,p} \geq k\} \leq \left(\frac{2\log_b n}{k}\right)^k \left(\frac{2\log_b n}{en}\right)^\delta \frac{p^{(\delta+\delta^2)/2}}{\sqrt{2\pi k}} \tag{15}$$

From (11) with $k = z+\delta$,

$$2\log_b n = k-\delta + 2\log_b \log_b n - 2\log_b \frac{e}{2} - 1.$$

Noting that $1+x \leq e^x$ for all real x and $u^{\log_v w} = w^{\log_v u}$ for $u,v,w > 0$,

$$\left(\frac{2\log_b n}{k}\right)^k = \left(1 + \frac{2\log_b(\frac{2}{e}\log_b n)-1-\delta}{k}\right)^k$$

$$\leq e^{2\log_b(\frac{2}{e}\log_b n)-1-\delta}$$

$$= \frac{(\frac{2}{e}\log_b n)^{2\log_b e}}{e^{1+\delta}},$$

and substitution into (15) yields

$$\text{Prob}\{Z_{n,p} \geq z+\delta\} \leq \frac{(\frac{2}{e}\log_b n)^{\delta+2\log_b e}}{n^\delta} \frac{p^{(\delta+\delta^2)/2}}{e^{1+\delta}\sqrt{2\pi(z+\delta)}},$$

which is equation (12).

To investigate the lower bound in (8) note in example 1 (Table 1) for k=15,16 that the terms $a_j(n,k)$ are first decreasing and then increasing in j. The next lemma shows that for sufficiently large n with k in the neighborhood of $z(n,1/p)$, the sequence $\alpha_2(n,k)$, $\alpha_3(n,k),\ldots,$ $\alpha_k(n,k)$ is at first sharply decreasing through some term $\alpha_m(n,k)$, where $\frac{1}{4}(z-1) \leq m \leq \frac{3}{4}(z-1)+1$, and the sequence is then sharply increasing from $\alpha_m(n,k)$ (or $\alpha_{m+1}(n,k)$ if $\alpha_m(n,k) = \alpha_{m+1}(n,k)$) through $\alpha_k(n,k)$. This behavior is pronounced so that the sum $\sum_{j=2}^{k}\alpha_j(n,k)$ is dominated by the initial and terminal terms.

<u>Lemma</u>: For $0 < p \leq \frac{1}{b} < 1$, let n and k be chosen so that with the threshold function $z=z(n,b)$ given by (11),

(i) $n \geq (2 \log_b n)^6$,

(ii) $2 \log_b n \geq e\sqrt{b}$,

(iii) $z \geq 21 \left(\frac{b}{b-1}\right)$,

(iv) $\frac{7}{8}z \leq k \leq 2z$,

and let

(16) $$\alpha_j = \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} b^{j(j-1)/2} \quad \text{for } 2 \leq j \leq k.$$

Then

(17) $$\sum_{j=2}^{k} \alpha_j \leq (\alpha_2 + \alpha_k)\left(1 + \frac{1}{n^{1/12}}\right).$$

Proof: For n sufficiently large conditions (i)-(iii) must hold. From (ii) $\log_b \log_b n \geq \log_b \frac{e}{2} + \frac{1}{2}$, so from (11) $2 \log_b n \geq z$. Thus conditions (i)-(iv) tacitly require

(18) $$n^{1/6} \geq 2 \log_b n \geq z > 21,$$

(19) $$4 \log_b n \geq k \geq 19.$$

The defining equation (11) for z yields $b^{(z-1)/2} = en/(2 \log_b n)$, so from (16)

(20) $$\frac{\alpha_{j+1}}{\alpha_j} = \frac{(k-j)^2}{(j+1)(n-2k+j+1)} \left(\frac{en}{2 \log_b n}\right)^{\frac{2j}{z-1}} \quad \text{for } 2 \leq j \leq k-1 .$$

The range $j = 2,3,\ldots,k$ will be treated in three cases.

Case 1: Let $2 \leq j \leq (z-1)/4$. From (20), (19), (18), noting $n > 10^5 k$,

(21) $$\frac{\alpha_{j+1}}{\alpha_j} \leq \frac{k^2}{3(n-2k)} \left(\frac{en}{2 \log_b n}\right)^{1/2}$$

$$\leq \frac{4e^{1/2}}{3(1-2k/n)} \left[\frac{(2 \log_b n)^6}{n}\right]^{1/4} \frac{1}{n^{1/4}}$$

$$\leq \frac{3}{n^{1/4}} \quad \text{for } 2 \leq j \leq (z-1)/4 .$$

Let $m \leq k$ be the largest index such that $\alpha_2 > \alpha_3 > \ldots > \alpha_m$. Now $z > 21$ implies $m \geq 6$, so from (18), (19)

(22) $$\sum_{j=2}^{m} \alpha_j \leq \alpha_2 + \alpha_3 + k\alpha_4$$

$$\leq \alpha_2\left(1 + \frac{3}{n^{1/4}} + \frac{9k}{n^{1/2}}\right)$$

$$\leq \alpha_2\left(1 + \frac{7}{n^{1/4}}\right) .$$

<u>Case 2</u>: Let $(\frac{3}{4})(z-1) \le j \le k-1$. Utilizing (20), (19), (18),

$$(23) \qquad \frac{\alpha_{j+1}}{\alpha_j} \ge \frac{1}{(j+1)n}\left(\frac{en}{2\log_b n}\right)^{3/2}$$

$$\ge \frac{e^{3/2}\, n^{1/2}}{2\,(2\log_b n)^{5/2}}$$

$$\ge 2n^{1/12} \quad \text{for } \frac{3}{4}(z-1) \le j \le k-1 .$$

Let $m' \ge 2$ be the smallest index such that $\alpha_{m'} < \alpha_{m'+1} < \dots < \alpha_k$.
Since $z > 21$ and $k \ge \frac{7}{8}z$,

$$k-3 \ge \frac{7}{8}z - \left(\frac{z+3}{8}\right) \ge \frac{3}{4}(z-1) ,$$

so (23) holds for $k-3 \le j \le k-1$. Thus using (18), (19),

$$(24) \qquad \sum_{j=m'}^{k} \alpha_j \le \alpha_k + \alpha_{k-1} + \alpha_{k-2} + k\alpha_{k-3}$$

$$\le \alpha_k\left(1 + \frac{1}{2n^{1/12}} + \frac{1}{4n^{1/6}} + \frac{k}{8n^{1/4}}\right)$$

$$\le \alpha_k\left(1 + \frac{1}{n^{1/12}}\right) .$$

<u>Case 3</u>: $\frac{1}{4}(z-1) - 1 \le j \le \frac{3}{4}(z-1)$. Using (i)-(iv) and (20),

$$(25) \qquad \frac{\alpha_{j+2}/\alpha_{j+1}}{\alpha_{j+1}/\alpha_j} = \frac{(k-j-1)^2(j+1)(n-2k+j+1)}{(k-j)^2(j+2)(n-2k+j+2)}\, b$$

$$\ge \left(1 - \frac{8}{z}\right)^2 \left(1 - \frac{4}{z}\right)\left(1 - \frac{1}{z}\right) b$$

$$> \left(1 - \frac{21}{z}\right) b$$

$$\ge 1 \quad \text{for } \frac{1}{4}(z-1) - 1 \le j \le \frac{3}{4}(z-1) .$$

Along with the results of cases 1 and 2, this confirms that $\alpha_2, \alpha_3, \ldots, \alpha_k$ is strictly decreasing through some term α_m, $\frac{1}{4}(z-1) \le m \le \frac{3}{4}(z-1)+1$, and then strictly increasing from term α_m (or α_{m+1} if $\alpha_m = \alpha_{m+1}$) through α_k. Thus from (22), (24) and (18),

$$(26) \qquad \sum_{j=2}^{k} \alpha_j \le \left(1 + \frac{1}{n^{1/12}}\right)\left(\alpha_2 + \alpha_k\right) ,$$

proving the lemma.

<u>Theorem 4</u>: For $0 < p \le \frac{1}{b} < 1$, let n and k be chosen so that with the threshold function $z=z(n,b)$ given by (11),

(i) $n \ge (2 \log_b n)^6$.

(ii) $2 \log_b n \ge e\sqrt{b}$,

(iii) $z \ge 21 \left(\frac{b}{b-1}\right)$,

(iv) $\frac{7}{8}z \le k < z$.

Then for $\delta = z-k > 0$,

$$(27) \qquad \text{Prob}\{Z_{n,p} \le z-1-\delta\} \le \frac{(\frac{2}{e} \log_b n)^{\delta+1/2}}{n^{\delta}} 6b^{(\delta^2-\delta)/2} + \frac{(\log_b n)^4}{n^2} 10b$$

<u>Proof</u>: From the left hand side of (8),

$$\text{Prob}\{Z_{n,p} \ge k\} \ge 1 \Big/ \sum_{j=0}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} b^{j(j-1)/2}.$$

Since $\sum_{j=0}^{k} \binom{n-k}{k-j}\binom{k}{j} = \binom{n}{k}$ and $b > 1$,

$$\text{Prob}\{Z_{n,p} \le k-1\} \le \sum_{j=0}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} (b^{j(j-1)/2} - 1)$$

$$\le \sum_{j=2}^{k} \frac{\binom{n-k}{k-j}\binom{k}{j}}{\binom{n}{k}} b^{j(j-1)/2},$$

and by the preceeding lemma

(28) $$\text{Prob}\{Z_{n,p} \leq k-1\} \leq \sum_{j=2}^{k} \alpha_j \leq (\alpha_2 + \alpha_k)(1 + \frac{1}{n^{1/12}}) \ .$$

Conditions (i)-(iv) of theorem 4 yield $k < z \leq 2 \log_b n$, so from (16)

(29) $$\alpha_2 \leq \frac{k^4}{n^2}\frac{b}{2} \leq \frac{(2 \log_b n)^4}{n^2}\frac{b}{2} \ .$$

Note that $\alpha_k = 1/\binom{n}{k}p^{k(k-1)/2}$ is the inverse of the expected number of k-membered complete subgraphs.

Utilizing Stirling's formula noting $n^{1/6} > k \geq 19$,

$$\frac{1}{\binom{n}{k}} \leq \frac{k!}{(n-k)^k} \leq \left(\frac{k}{en}\right)^k \frac{\sqrt{2\pi k}\, e^{1/12k}}{\left(1 - \frac{k}{n}\right)^k} \leq \left(\frac{k}{en}\right)^k (2.52)k^{1/2} \ .$$

With $\delta = z-k > 0$, $k(k-1)/2 = (k-\delta)(z-1)/2 + (\delta^2-\delta)/2$ and from (11) $b^{(z-1)/2} = en/(2 \log_b n)$, so

(30) $$\alpha_k \leq b^{k(k-1)/2}\left(\frac{k}{en}\right)^k (2.52)k^{1/2}$$

$$\leq \left(\frac{2 \log_b n}{en}\right)^{\delta} (2.52)(2 \log_b n)^{1/2} b^{(\delta^2-\delta)/2}$$

$$\leq \frac{\left(\frac{2}{e} \log_b n\right)^{\delta + 1/2}}{n^{\delta}} \quad (4.16)\ b^{(\delta^2-\delta)/2}$$

Thus from (28), (29), (30) with $n > 21^6$ implied by (i)-(iv),

$$\text{Prob}\ \{Z_{n,p} \leq k-1\} \leq \frac{\left(\frac{2}{e} \log_b n\right)^{\delta + 1/2}}{n^{\delta}}\ 6b^{(\delta^2-\delta)/2} + \frac{(\log_b n)^4}{n^2}\ 10b$$

proving the theorem.

The following numeric example employs the results of theorems 3 and 4 to show how peaked the density of $Z_{n,p}$ can be for large n.

Example 2: Random Graph: $G_{10^{10}, .25}$

Size: $n = 10^{10}$

Edge Probability: $p = .25$

Threshold Function: $z = 30.608$

We wish to bound the probability that $Z_{10^{10}, .25}$ has the value 30. From inequality (12) we determine

$$\text{Prob}\{Z_{10^{10}, .25} \geq 31\} \leq .000145.$$

Now $2 \log_4 10^{10} = 33.219$ and conditions (i)-(iii) of theorem 4 are verified.

Thus from inequality (27),

$$\text{Prob}\{Z_{10^{10}, .25} \leq 29\} \leq .000068$$

and we obtain

$$\text{Prob}\{Z_{10^{10}, .25} = 30\} \geq .9997 .$$

Having established the fact that the density of $Z_{n,p}$ is peaked in the neighborhood of the threshold function $z(n,1/p)$, it is of interest to ask if inequality (8) yields any immediate insight into the tails of the density of $Z_{n,p}$. For values of $k > z(n,1/p)$ where $\binom{n}{k} p^{k(k-1)/2} << 1$, the discussion of the strong second moment method in section II suggests that both bounds of inequality (8) may be quite good if the cliques of size k do not come in bursts. In example 1 the bounds of (8) yielded the result $.0193 \leq \text{Prob}\{Z_{1000, .5} \geq 16\} \leq .0318$ for $k=16 > z(1000,2) = 15.18$. Utilizing the preceeding lemma, the next theorem shows that for $z(n,1/p) < k \leq 2z(n,1/p)$ with $\binom{n}{k} p^{k(k-1)/2} << 1$ and sufficiently large n, the two bounds of inequality (8) are indeed quite close.

Theorem 5: For $0 < p = \frac{1}{b} < 1$, let n and k be chosen so that with the threshold function $z = z(n,1/p)$ given by (11),

(i) $n \geq 2\,(\log_b n)^6$,

(ii) $2 \log_b n \geq e\sqrt{b}$,

(iii) $z \geq 21(\frac{b}{b-1})$,

(iv) $z < k \leq 2z$.

Then

$$\text{(31)} \qquad E(N)\ (1 - \frac{1}{n^{1/12}} - 2E(N)) \leq \text{Prob}\{Z_{n,p} \geq k\} \leq E(N)$$

where $E(N) = \binom{n}{k} p^{k(k-1)/2}$ is the expected number of k membered complete subgraphs of $G_{n,p}$.

Proof: From (8), (16) and (17)

$$\text{Prob}\{Z_{n,p} \geq k\} \geq \frac{1}{1 + \Sigma_{j=2}^{k} \alpha_j} \geq \frac{1}{1 + (\alpha_2 + \alpha_k)(1 + \frac{1}{n^{1/12}})} \ .$$

From (16), conditions (i)-(iv), and (18),

$$\alpha_2 \leq \frac{k^4 b}{2n^2} < 1 \ ,$$

and from (16)

$$\alpha_k^{-1} = \binom{n}{k} p^{k(k-1)/2} = E(N) \ .$$

Thus

$$\text{Prob}\{Z_{n,p} \geq k\} \geq \frac{1}{(2 + \alpha_k)(1 + \frac{1}{n^{1/12}})}$$

$$\geq \frac{\alpha_k^{-1}}{(1 + 2\alpha_k^{-1})(1 + \frac{1}{n^{1/12}})}$$

$$\geq E(N)\ (1 - \frac{1}{n^{1/12}} - 2E(N)) \ ,$$

and the theorem follows.

IV. Asymptotic Value of the Clique Number of a Random Graph

The bounds of theorems 3 and 4 may be utilized to derive the following asymptotic behavior for the clique number of a random graph.

Theorem 6: For $0 < p = \frac{1}{b} < 1$ and $n \geq 1$ let $z = 2 \log_b n - 2 \log_b \log_b n + 2 \log_b \frac{e}{2} + 1$. Then for any $\varepsilon > 0$, the maximum clique size $Z_{n,p}$ of the random graph $G_{n,p}$ satisfies

$$(35) \qquad \text{Prob}\{\lfloor z-\varepsilon \rfloor \leq Z_{n,p} \leq \lfloor z+\varepsilon \rfloor\} = 1-0\left(\frac{(\ln n)^{\varepsilon+c}}{n^{\varepsilon}} + \frac{(\ln n)^4}{n^2}\right)$$

where $c = \max \{\frac{1}{2}, \frac{2}{\ln b} - \frac{1}{2}\}$ and $\lfloor x \rfloor$ denotes the greatest integer in x.

Proof: For $n \geq 2$, let $\delta = \lfloor z+\varepsilon \rfloor + 1-z$, so from (12)

$$\begin{aligned}\text{Prob}\{Z_{n,p} \leq \lfloor z+\varepsilon \rfloor\} &= 1 - \text{Prob}\{Z_{n,p} \geq \lfloor z+\varepsilon \rfloor + 1\} \\ &= 1 - \text{Prob}\{Z_{n,p} \geq z+\delta\} \\ &\geq 1 - \frac{(\frac{2}{e}\log_b n)^{\delta+2 \log_b e}}{n^{\delta} z^{1/2}}\end{aligned}$$

Now ε is constant and $\varepsilon < \delta \leq 1 + \varepsilon$ for all n, so

$$\text{Prob}\{Z_{n,p} \leq \lfloor z+\varepsilon \rfloor\} \geq 1 - 0\left(\frac{(\ln n)^{\varepsilon + \frac{2}{\ln b} - \frac{1}{2}}}{n^{\varepsilon}}\right).$$

For n sufficiently large so that conditions (i)-(iii) of theorem 4 and $\delta^* = z - \lfloor z-\varepsilon \rfloor \leq z/8$ are satisfied, utilizing (27)

$$\text{Prob}\{\lfloor z-\varepsilon \rfloor \le Z_{n,p}\} = 1 - \text{Prob}\{Z_{n,p} \le \lfloor z-\varepsilon \rfloor - 1\}$$

$$= 1 - \text{Prob}\{Z_{n,p} \le z-1-\delta^*\}$$

$$\ge 1 - \frac{(\frac{2}{e}\log_b n)^{\delta^*+1/2}}{n^{\delta^*}} 6b^{(\delta^{*2}-\delta^*)/2} - \frac{(\log_b n)^4}{n^2} \qquad 10b.$$

Again ε is constant and $\varepsilon \le \delta^* < 1 + \varepsilon$ for all n above, so

$$\text{Prob}\{\lfloor z-\varepsilon \rfloor \le Z_{n,p}\} = 1 - O\left(\frac{(\ell n\ n)^{\varepsilon + \frac{1}{2}}}{n^{\varepsilon}} + \frac{(\ell n\ n)^4}{n^2}\right)$$

and the theorem follows.

In theorem 5 of the previous section the strong second moment method was utilized to obtain a lower bound on $\text{Prob}\{Z_{n,p} \ge k\}$ for $k > z$. This result is now utilized to provide an asymptotic lower bound on the tail $\text{Prob}\{Z_{n,p} \ge \lfloor z+\varepsilon \rfloor + 1\}$ of order comparable to the asymptotic upper bound on the tails implicit in theorem 6.

Theorem 7: For $0 < p = \frac{1}{b} < 1$, $n \ge 1$, let $z = z(n,1/p)$ be given by (11). For any $\varepsilon > 0$, there exist constants $c = c(\varepsilon,p)$, $M = M(\varepsilon,p)$ such that for all $n > M$,

$$\text{(36)} \qquad \text{Prob}\{Z_{n,p} \ge \lfloor z+\varepsilon \rfloor + 1\} > c\,\frac{(\ell n\ n)^{1/2+\varepsilon}}{n^{1+\varepsilon}} .$$

Proof: For $0 < p < 1$ and $n = 1,2,\ldots,$ let $k = k(n)$ be uniquely determined by $z+\varepsilon < k \le z+\varepsilon+1$. By Stirling's formula there exists $c_1 > 0$, $M_1 > 1$ such that

$$\binom{n}{k} > c_1 \left(\frac{en}{k}\right)^k \frac{1}{\sqrt{k}} \quad \text{for } n > M_1.$$

Let $\delta = \delta(n) = k-z \leq 1+\varepsilon$, so utilizing (14) there exists $c_2 > 0$ such that

$$p^{k(k-1)/2} \geq c_2 \left(\frac{2 \log_b n}{en}\right)^{k+\delta} \quad \text{for } n \geq 2.$$

Now $k < 2 \log_b n < 2k$ for sufficiently large n, so there exists $c_3 > 0$ and M_2 such that

$$\begin{aligned} E(N) &= \binom{n}{k} p^{k(k-1)/2} \\ &> c_1 c_2 \left(\frac{2 \log_b n}{en}\right)^{\delta} \frac{1}{\sqrt{k}} \quad \text{for } n > M_1 \\ &> c_3 \frac{(\ell n\ n)^{1/2+\varepsilon}}{n^{1+\varepsilon}} \quad \text{for } n > M_2. \end{aligned}$$

The choice of $k = k(n) > z+\varepsilon$ assures by theorem 3 that $E(N) \to 0$ as $n \to \infty$, so from theorem 5 $\text{Prob}\{Z_{n,p} \geq k\} \to E(N)$ as $n \to \infty$, and the theorem follows.

REFERENCES

[1] Bollobás, B. and Erdös, P., "Cliques in Random Graphs," to appear.

[2] Erdös, P. and Spencer, J., Probabilistic Methods in Combinatorics, Academic Press, New York, 1974.

[3] Grimmett, G. R. and McDiarmid, C. J. H., "On Coloring Random Graphs," Math Proc. Camb. Phil. Soc. 77 (1975), 313-324.

[4] Matula, D. W., "On the Complete Subgraphs of a Random Graph," Comb. Math and Its Appl. (Chapel Hill, N. C., 1970), 356-369.

[5] Matula, D. W., "The Employee Party Problem," Notices A.M.S. 19 (Feb. 1972), p. A-382.