COMBINATORICA
Akadémiai Kiadó — Springer-Verlag

COMBINATORICA 8 (1) (1988) 49—55

THE CHROMATIC NUMBER OF RANDOM GRAPHS

B. BOLLOBÁS

Received March 20, 1987

For a fixed probability p, $0<p<1$, almost every random graph $G_{n,p}$ has chromatic number

$$\left(\frac{1}{2}+o(1)\right)\log(1/(1-p))\frac{n}{\log n},$$

0. Introduction

One of the best known and most studied unsolved problems in the theory of random graphs is the determination of the chromatic number of almost every random graph. As usual, we write G_p for a random graph with vertex set $V=[n]=\{1, 2, ..., n\}$ in which the edges are chosen independently and with probability $p=p(n)$, $0<p<1$. In most of what follows, we shall take p to be fixed and write $d=1/(1-p)$. Putting it rather crudely, Grimmett and McDiarmid [8] proved that *a. e.* G_p is such that its chromatic number $\chi(G_p)$ satisfies

$$\frac{n}{2\log_d n}(1+o(1)) \leqq \chi(G_p) \leqq \frac{n}{\log_d n}(1+o(1)) \tag{1}$$

and a little sharper result was proved by Bollobás and Erdős [4].

The chromatic number of sparse graphs was studied in [2, 5, 11, 17] and the performance of sequential algorithms was analysed by McDiarmid [10] and Shamir and Upfal [18]. Recently, Shamir and Spencer [16] proved that for every $\omega(n)\to\infty$ there is an interval of length $\omega(n)n^{1/2}$ such that the chromatic number of *a. e.* G_p belongs to this interval. However, none of these results improved on (1); in particular, they did not rule out that both bounds in (1) are essentially best possible.

Very recently, the ratio 2 of the two bounds in (1) was improved substantially by Matula [12] who proved that *a. e.* G_p satisfies

$$\frac{n}{2\log_d n}(1+o(1)) \leqq \chi(G_p) \leqq \frac{2n}{3\log_d n}(1+o(1)). \tag{2}$$

AMS subject classification (1980): 05 C 15.

The main aim of this note is to improve (2) to an essentially best possible result by showing that $\chi(G_p)$ is, almost surely, equal to the lower bound in (1) (and so in (2)).

The notation and terminology we use are standard; the reader is referred to [3] for an introduction to random graphs.

1. Cliques in random graphs

Let $0<p<1$ be fixed and set $b=1/p$. It was proved in [4] (see also [3, p. 253]) that there is an integer-valued function $r=r_1(n)$ such that

$$2\log_b n-2\log_b\log n+2\log_b(e/2)+o(1)\leqq$$
$$\leqq r_1(n)\leqq 2\log_b n-2\log_b\log n+2\log_b(e/2)+1+o(1)$$

and there is a set $M\subset\mathbf{N}$ of density 1 such that if $n\in M$ then $cl(G_p)$, the *clique number* of G_p, is precisely r_1. Putting it more formally,

$$\lim_{\substack{n\to\infty\\ n\in M}} P(cl(G_{n,p})=r_1)=1.$$

We shall show that if r is substantially smaller than r_1 (although $r\sim r_1$) then the probability that $cl(G_p)\leqq r$ is exponentially small.

The proof of this assertion is based on a martingale inequality frequently used in combinatorial functional analysis (see, e.g., [9, 13, 14, 15, 17]) and applied to random graphs by Shamir and Spencer [16].

Let (Ω, F, P) be a probability triple and let $F_0\subset F_1\subset F_2\subset\ldots$ be an increasing sequence of sub $-\sigma-$fields of F. Let $X_0, X_1, \ldots$ be random variables on (Ω, F, P) such that X_k is F_k-measurable and $E(X_{k+1}|F_k)=X_k$. The following inequality is due to Freedman [7]; a similiar inequality was proved by Azuma [1].

Lemma 1. *If* $|X_{k+1}-X_k|\leqq c$ *for some* $c>0$, $b=mc^2$ *and* $a>0$ *then*

$$P(X_m\leqq E(X_m)-a)\leqq\left(\frac{b}{a+b}\right)^{a+b}e^a\leqq\exp\left\{-\frac{a^2}{2(a+b)}\right\}. \tag{3}$$

We shall apply Freedman's inequality to martingales on $\mathcal{G}(n,p)$, the probability space of all graphs on $V=[n]$, with the probability giving the random graphs G_p.

Let $E_0\subset E_1\subset\ldots\subset E_m=V^{(2)}$ be an increasing sequence of sets in $V^{(2)}$, the set of all *unordered pairs of vertices*, and let F_k be defined by the function $G\to E(G)\cap E_k$. Given a random variable X on $\mathcal{G}(n,p)$, define $X_k=E(X|F_k)$. Then $(X_k)_0^m$ is a martingale; we shall apply Lemma 1 to this martingale to prove our result.

Let us turn now to the study of cliques in G_p. Let $r=r(n)\in\mathbf{N}$ and let $Y_r=Y(n,r)=Y_r(G_p)=k_r(G_p)$ be the number of complete r-graphs in G_p. Then, trivially, the expectation of Y_r is

$$E_n=E(n,r)=E(Y_r)=\binom{n}{r}p^{\binom{r}{2}}.$$

Theorem 2. *Let* $r=r(n)\geqq 3$ *and* $\alpha=\alpha(n)\geqq 1/2$ *be such that* $E(r, n)=n^\alpha=o\left(n^2/(\log n)^4\right)$. *Then for* $0<c\leqq 1$ *we have*

$$P\left(Y_r \leqq (1-c)n^\alpha\right) \leqq \exp\left\{-\left(c^2+o(1)n^{2\alpha-2}\right)\right\}.$$

Proof. The choice of r_1 implies (see [3, pp. 252—253]) that if $r'\geqq r_1+1$ then $E(n, r')=o(1)$ and if $3\leqq r''\leqq r_1-3$ then $E(n, r'')/n^2\to\infty$. Hence $r_1-3<r<r_1+1$ so, in particular, $r\sim r_1$ and

$$c_1 n/(\log n)^2 \leqq p^{-r} \leqq c_2 n^2/(\log n)^2 \tag{4}$$

for some positive constants c_1 and c_2. Let $X=X(G_p)$ be the *maximal number of edge-disjoint* K^r *subgraphs in* G_p. Then, by definiton, $X\leqq Y_r$.

Let $E_0\subset E_1\subset\ldots\subset E_N=V^{(2)}$ be such that $|E_k|=k$, where, as usual, $N=\binom{n}{2}$. Let $(X_k)_0^N$ be the martingale defined by X and this sequence $(E_k)_0^N$, namely let $X_k(G)=E\left(X(G_p)|E(G_p)\cap E_k=E(G)\cap E_k\right)$. In particular, $X_N\equiv X$. Clearly $|X_{k+1}-X_k|\leqq 1$ for every k. Indeed, $E_{k+1}-E_k=\{uv\}$ for some pair uv, and the addition of uv to the edge set of a graph increases X by at most one, because at most one of the K^r s containing uv may be counted by X. Hence, by (3), for $a>0$ we have

$$P\left(X \leqq E(X)-a\right) \leqq \exp\left\{-\frac{a^2}{2(a+N)}\right\}. \tag{5}$$

All that remains is to estimate $E(X)$ from below. Let $X'(G)$ be the *number of* K^r *subgraphs in a graph G sharing no edge with another* K^r. Then $X'\leqq X$ so $E(X')\leqq E(X)$.

Let K_0 be the complete graph with vertex set $W=[r]=\{1, \ldots, r\}$ and let $A\subset G(n, p)$ be the event $K_0\subset G_p$. Furthermore, let Z_l be the number of K^r subgraphs of G_p having l vertices in W and let $Z=\sum_{l=2}^{r-1} Z_l$. By definition, A holds and $Z=0$ iff G_p contains K_0 and K_0 shares no edge with another K^r subgraph, so

$$E(X') = \binom{n}{r}P(Z=0|A) \geqq \binom{n}{r} P(A)\{1-E(Z|A)\}. \tag{6}$$

Also,

$$E(Z|A) = \sum_{l=2}^{r-1} E(Z_l|A) \tag{7}$$

and

$$E(Z_l|A) = \binom{r}{l}\binom{n-r}{r-l}p^{\binom{r}{2}-\binom{l}{2}} = F_l. \tag{8}$$

Then for $2\leqq l\leqq r-2$ we have

$$F_{l+1}/F_l = \frac{(r-l)^2}{(l+1)(n-2r+l+1)}p^{-l}.$$

This shows that $F_{l+1}/F_l=o(n^{-1/3})$ for $l\leqq(1/2)\log_b n$ and for $l\geqq(1/2)\log_b n$ it is an increasing function of l. Also $F_{r-1}/F_{r-2}\geqq c_3 n/(\log n)^2$ for some $c_3>0$.

Therefore

$$\sum_{l=2}^{r-1} F_l = (F_2 + F_{r-1})(1+o(1)).$$

Now

$$F_2 \leqq E(n,r)\frac{r^4}{n^2}p^{-1} = O\big((\log n)^4 n^{\alpha-2}\big) = o(1)$$

and, by (4),

$$F_{r-1} \leqq rnp^{r-1} = O\big((\log n)^3/n\big) = o(1).$$

Hence

$$\sum_{l=2}^{r-1} F_l = o(1)$$

and so, by (6), (7) and (8),

$$E(X') \geqq \big(1+o(1)\big)\binom{n}{r}P(A) = \big(1+o(1)\big)E(n,r) = \big(1+o(1)\big)n^{\alpha}. \tag{9}$$

The proof of the theorem is essentially complete. Set $a = E(X) - (1-c)n^{\alpha}$. Then by (9), $a \geqq E(X') - (1-c)n^{\alpha} = \big(c+o(1)\big)n^{\alpha}$. Hence, by (5),

$$P\big(Y_r \leqq (1-c)n^{\alpha}\big) \leqq P\big(X \leqq (1-c)n^{\alpha}\big) \leqq P\big(X \leqq E(X) - a\big) \leqq$$

$$\leqq \exp\left\{-\frac{a^2}{2(a+N)}\right\} = \exp\left\{-\big(c^2+o(1)\big)n^{2\alpha-2}\right\}. \quad \blacksquare$$

Corollary 3. *Let $n' = n'(n) = n^{\beta} \leqq n$ and $r' = r'(n) \geqq 3$ be such that $E(r', n') = n^{\alpha}$ and $2n^{3\beta}\log n \leqq n^{2\alpha} = o\big(n^{4\beta}/(\log n)^8\big)$. Then a. e. G_p is such that every n'-subset of vertices contains at least one complete graph of order r'.*

Proof. The number of n'-subsets of vertices is $\binom{n}{n'} < n^{n'}$ and the probability that a given n'-subset of vertices fails to contain a $K^{r'}$ is at most $\exp\{-2n^{2\alpha-2\beta}/3\} = o(n^{-n'})$. $\blacksquare$

2. The chromatic number of random graphs

A colouring of a graph is a partitioning of the vertex set into independent sets, i.e. sets spanning complete subgraphs of the complement. As the complement of a random graph G_p is a random graph G_q where $q = 1-p$, the following theorem is an easy consequence of Corollary 3.

Theorem 4. *Let $0 < p < 1$ be fixed and set $q = 1-p$, $d = 1/q$ and*

$$s_0 = [2\log_d n - \log_d \log_d n + 2\log_d(e/2) + 1].$$

Then a. e. G_p is such that

$$\frac{n}{s_0} \leqq \chi(G_p) \leqq \frac{n}{s_0}\left(1 + \frac{3\log\log n}{\log n}\right).$$

In particular, $\chi(G_p) = \frac{n}{s_0}\,(1+o(1))$.

Proof. The lower bound is well known (see [3, p. 265]) and is immediate from the fact that almost no G_q contains a K^{s_0+1}. Thus the content of the theorem is the upper bound. Set $s_1=s_0-\lfloor 5\log_d \log n\rfloor$ and let n_1 be the maximal natural number such that

$$E_q(n_1, s_1) = \binom{n_1}{s_1} q^{\binom{s_1}{2}} \geqq 2n_1^{5/3}.$$

Note that $E_q(n_1, s_1)$ is the expected number of K^{r_1} subgraphs in $G_{n_1,q}$, i.e. the expected number of independent s_1-sets of vertices in $G_{n_1,p}$. Then, trivially (see also [3, pp. 253—255]), $2n_1^{5/3}\leqq E_q(n_1, s_1)\leqq 3n_1^{5/3}$ and $n/(\log n)^3\leqq n_1\leqq n/(\log n)^2$.

Applying Corollary 3 to G_q with $n'=n_1$ and $r'=s_1$, we find that *a. e.* G_p is such that every set of n_1 vertices contains s_1 independent vertices. The chromatic number of such a G_p is at most n/s_1+n_1. Indeed, colour such a G_p by successively picking s_1-sets of independent vertices and giving them the same colour; when there is no independent s_1-set, give the remaining vertices distinct colours. By assumption we can always find an independent s_1-set, unless we are left with fewer than n_1 vertices. Thus altogether we do use at most n/s_1+n_1 colours. Since

$$\frac{n}{s_1}+n_1 < \frac{n}{s_1}\left(1+\frac{3\log\log n}{\log n}\right)$$

the theorem is proved. ∎

3, Varying probabilities

Though the main aim of this note is the study of the clique number and chromatic number of G_p for p constant, we sketch some results implied by the methods for other values of p. It is clear that in Theorem 2 we need not assume that p is constant; what we need is that the expected number of K^r subgraphs be about the expected number of K^r subgraphs sharing no edges with other K^r subgraphs. As p decreases and so r decreases this is easily satisfied but when $q=1-p$ is rather small, this is not true. In view of this we state the result corresponding to Theorem 2 for the complementary graph, i.e. for independent sets in G_p. The proof is just that of Theorem 2, mutatis mutandis.

Theorem 5. *Let* $0\leqq p=p(n)<1/2$, $q=1-p$, $s=s(n)\geqq 3$ *and* $E_q(n, s)=\binom{n}{s}q^{\binom{s}{2}}=n^{\alpha}$. *Suppose that* $n^{\alpha}=o\big(p^4n^2/(\log n)^4\big)$ *and* $snq^s=o(1)$. *Denote by* Z_s *the number of independent s-sets in* G_p. *Then for* $0<c\leqq 1$ *we have*

$$P(Z_s\leqq (1-c)n^{\alpha}) \leqq \exp\{-(c^2+o(1))n^{2\alpha-2}\}. \quad ∎$$

If p and r satisfy the conditions of Theorem 5 then p is not too small; for $\alpha\geqq 1$ the smallest it can be is about $n^{-1/4}$. Theorem 5 implies not only that, as it is well known, the independence number of *a. e.* G_p is almost determined but also that the probability that the independence number is a little smaller than it should be is exponentially small. As the result can easily be sharpened, we only sketch the proof.

Corollary 6. *Let $0<\theta<1/8$ be fixed, $p=n^{-\theta}$, $k\geqq 1$ a constant, $n_0=\lfloor n/(\log n)^k\rfloor$, $s=\lfloor n^\theta\{2(1-\theta)\log n-2\log\log n\}\rfloor$ and $s_0=\lceil s-(2k+1)\log\log n\rceil$. Then a. e. G_p is such that it does not contain s independent vertices but every set of n_0 vertices contains s_0 independent vertices.*

Proof. It is easily checked that $E_q(n,s)=o(1)$ so almost no G_p contains s independent vertices. Fix α such that $3/2<\alpha<2-4\theta$. The integers n_0 and s_0 have been chosen to guarantee $E_q(n_0,s_0)\geqq n^2$. Let n_1 be the minimal natural number such that $E_q(m,s_0)\geqq 2m^\alpha$ for all m, $n_1\leqq m\leqq n_0$. Straightforward calculations show that $n_1\geqq n/(\log n)^{k+1}$ and $E_q(n_1,s_0)\leqq 3n_1^\alpha$. The conditions of Theorem 5 are satisfied for $s=s_0$ and $n=n_1$ so the probability that a given set of n_1 vertices of G_p fails to contain s_0 independent vertices is at most $\exp\{-n_1^{2\alpha-2}\}$. Hence the probability that among some $n_1\leqq n_0$ vertices we do not have s_0 independent vertices is at most $\binom{n}{n_1}\exp\{-n_1^{2\alpha-2}\}=o(1)$. ∎

Theorem 7. *Let $0<\theta<1/3$ be fixed, let $p=n^{-\theta}$ and $s=\lfloor n^\theta\{2(1-\theta)\log n-2\log\log n\}\rfloor$. Then a. e. G_p is such that*

$$\frac{n}{s}\leqq\chi(G_p)\leqq\frac{n}{s}\left(1+\frac{4\log\log n}{(1-\theta)\log n}\right).$$

Proof. The lower bound is trivial since almost no G_p contains s independent vertices.

Turning to the upper bound, let $n_0=\lfloor n/(\log n)^2\rfloor$ and

$$s_0=\lceil s-5\log\log n\rceil\geqq s\Big/\left(1+\frac{3}{1-\theta}\frac{\log\log n}{\log n}\right).$$

By Corollary 6, *a. e.* G_p is such that every n_0-subset of vertices contains s_0 independent vertices.

Furthermore, the greedy algorithm (see [10] and [3, p. 266]) shows that *a. e.* G_p is such that every set of n_0 vertices can be coloured with at most $pn_0=n^{1-\theta}/(\log n)^2$ colours. Let us take such a G_p and colour it by successively selecting s_0 independent vertices and giving them a new colour. In this way after at most

$$n/s_0\leqq\frac{n}{s}\left(1+\frac{3\log\log n}{(1-\theta\log n}\right)$$

colours we are left with m vertices for some $n_0/2\leqq m\leqq n_0$. These m vertices can be coloured with at most $n^{1-\theta}/(\log n)^2$ colours. Since

$$\frac{n}{s}\left(1+\frac{3\log\log n}{(1-\theta)\log n}\right)+\frac{n^{1-\theta}}{(\log n)^2}\leqq\frac{n}{s}\left(1+\frac{4\log\log n}{(1-\theta)\log n}\right),$$

the proof is complete. ∎

Theorems 2 and 5 were proved by considering edge disjoint cliques, and independent sets sharing no pair of vertices. The range in Theorem 8 can be extended to $0<\theta<1/4$ if for suitable values of s_0 and t_0 we consider the maximal number of independent s_0-sets such that no pair of vertices is in more than t_0 of these sets.

References

[1] K. Azuma, Weighted sums of certain dependent random variables, *Tôhoku Math. J.*, 3 (1967), 357—367.

[2] B. Bollobás, The evolution of sparse graphs, *in Graph Theory and Combinatorics, Proc. Cambridge Combinatorial Conf. in honour of Paul Erdős (B. Bollobás, ed.)*, Academic Press, London, 1984, 35—37.

[3] B. Bollobás, *Random Graphs*, Academic Press, London, 1985, xvi+447 pp.

[4] B. Bollobás and P. Erdős, Cliques in random graphs, *Math. Proc. Cambridge Phil. Soc.*, **80** (1976), 419—427.

[5] B. Bollobás and A. G. Thomason, Random graphs of small order, *in Random Graphs, Annals of Discr. Math.*, 1985, 47—97.

[6] P. Erdős and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, New York and London, 1974,

[7] D. Freedman, On tail probabilities for martingales, *Ann. Probab.*, 3 (1975), 100—118.

[8] G. R. Grimmett and C. J. H. McDiarmid, On colouring random graphs, *Math. Proc. Cambridge Phil. Soc.*, 77 (1975), 313—324.

[9] W. B. Johnson and G. Schechtman, Embedding l_p^m into l_1^n, *Acta Math.*, **150** (1983), 71—85.

[10] C. J. H. McDiarmid, Colouring random graphs badly, *in Graph Theory and Combinatorics (R. J. Wilson, ed.), Pitman Research Notes in Mathematics*, 34 (1979), 76—86.

[11] C. J. H. McDiarmid, On the chromatic forcing number of a random graph, *Discrete Appl. Math.*, 5 (1983), 123—132.

[12] D. W. Matula, Expose-and-merge exploration and the chromatic number of a random graph, *Combinatorica*, 7 (1987), 275—284.

[13] B. Maurey, Construction de suites symétriques, *Compt. Rend. Acad. Sci. Paris* **288** (1979), 679—681.

[14] V. D. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, 1986, viii+156 pp.

[15] G. Pisier, On the dimension of l_p^n-subspaces of Banach spaces, for $1<p<2$, *Trans. Amer. Math. Soc.*, 276 (1983), 201—211.

[16] E. Shamir and J. Spencer, Sharp concentration of the chromatic number of random graphs $G_{n,p}$, *Combinatorica*, 7 (1987), 121—129.

[17] G. Schechtman, Random embeddings of euclidean spaces in sequence spaces, *Israel J. Math.*, 40 (1981), 187—192.

[18] E. Shamir and R. Upfal, Sequential and distributed graph colouring algorithms with performance analysis in random graph spaces, *J. of Algorithms* 5 (1984), 488—501.

[19] W. F. De La Vega, On the chromatic number of sparse random graphs, *in Graph Theory and Combinatorics, Proc. Cambridge Combinatorial Conf. in honour of Paul Erdős (B. Bollobás, ed.)*, Academic Press, London, 1984, 321—328.

Béla Bollobás

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, England