THE CHROMATIC NUMBER OF RANDOM GRAPHS

By

Tomasz Łuczak

IMA Preprint Series # 404

March 1988

THE CHROMATIC NUMBER OF RANDOM GRAPHS

TOMASZ ŁUCZAK†‡

Abstract. Let $\chi(G(n,p))$ denote the chromatic number of the random graph $G(n,p)$. We prove that for every constant $\epsilon > 0$ there exists a constant d_ϵ such that for $np(n) > d_\epsilon$, $p(n) \to 0$, the probability that

$$\frac{np}{2\log np} \leq \chi(G(n,p)) \leq (1+\epsilon)\frac{np}{2\log np}$$

tends to 1 as $n \to \infty$.

Let $G(n,p)$ be a random graph with vertex set $[n] = \{1, 2, \ldots, n\}$ in which each possible edge is present independently with the probability $p = p(n)$. We say that $G(n,p)$ has some property *a.s.* if the probability that it has this property tends to 1 as $n \to \infty$. In this paper we shall consider the asymptotical behaviour of the chromatic number of $G(n,p)$.

The case when p is a constant, $0 < p < 1$, was considered in [1] and [4] (in [1] $\chi(G(n,p))$ was determined also for $p(n)$ which tends to 0 slowly enough i.e. for $p(n) > n^{-\frac{1}{3}+\epsilon}$ where $\epsilon > 0$). We shall solve this problem for every function $p(n) \to 0$ such that $np(n)$ is greater than some large constant. Our main result is the following.

THEOREM. *For every positive constant ϵ there exists a constant d_ϵ such that if $d = d(n) = np(n) > d_\epsilon$ and $p(n) \to 0$ then a.s.*

$$\frac{d}{2\log d} < \chi(G(n,p)) < (1+\epsilon)\frac{d}{2\log d}\,.$$

The lower bound is an immediate consequence of the well known fact that a.s. the independence number of $G(n,p)$ is less than $\dfrac{2\log d}{d}n$. Thus it is enough to show the second inequality.

A subset S of $G(n,p)$ is *k-independent* if it can be split into some number of disjoint independent sets each of at least k elements. We shall start with the proof that for $k_0 = \lfloor (2-\epsilon)\,\dfrac{n\log\, d}{d} \rfloor$ $G(n,p)$ contains a.s. large k_0-independent set.

† Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, U.S.A. – post-doctoral fellowship.

‡ Institute of Mathematics, Adam Mickiewicz University, Poznań, Poland – permanent position.

LEMMA 1. *Let $\epsilon > 0, k_0 = \lfloor (2-\epsilon)\,\frac{n\log d}{d} \rfloor$. Then there exists a constant d_ϵ such that whenever $np(n) = d(n) > d_\epsilon$ then with the probability at least $1 - n^{-2}$ $G(n,p)$ contains k_0-independent set with at least $nd^{-\epsilon}$ vertices.*

Proof. Let $z = d^{0.9\epsilon}$ and $\mathcal{V} = \mathcal{V}(n,p,z)$ denote the partition of $[n]$ onto sets $V_1, \ldots, V_{\frac{n}{z}}$ such that $z - 1 \le |V_i| \le z + 1$ for $i = 1, \ldots, \frac{n}{z}$. We shall call a subset S of $G(n,p)$ *$\mathcal{V}$-disjoint* if no two elements of S are contained in the same set of partition $\mathcal{V}$.

Let X denote the size of the largest k_0-independent $\mathcal{V}$-disjoint set contained in $G(n,p)$. Now define X_i, $i = 1, 2, \ldots, z$, as the expectation of the largest k_0-independent $\mathcal{V}$-disjoint set in a graph $\mathcal{G}_i$ such that subgraphs induced by $\bigcup_{r=1}^{i} V_r$ in $\mathcal{G}_i$ and $G(n,p)$ are the same and all edges of $\mathcal{G}_i$ which are not contained in $\bigcup_{r=1}^{i} V_r$ are present in this graph independently with the probability p. Then $X = X_{\frac{n}{z}}$ and setting $X_0 = \mathrm{E}\,X$ we have

$$|X_i - X_{i-1}| \le 1 \qquad \text{for} \quad i = 1, \ldots, \frac{n}{z}\,.$$

Furthermore the sequence $X_0, \ldots, X_{\frac{n}{z}}$ is a martingale known as Doob's Martingale Process (see [5]). Hence, using a martingale inequality of Azuma (see [1], [5]), we obtain

$$\mathrm{Prob}\,\{|X - \mathrm{E}\,X| \ \ge t\} \le 2\exp\left(-\frac{zt^2}{2n}\right) = 2\exp\left(-\frac{d^{0.9\epsilon}t^2}{2n}\right)\,.$$

Thus, to prove Lemma 1, it is enough to show that the probability that $G(n,p)$ contains a k_0-independent $\mathcal{V}$-disjoint set with more than $2nd^{-\epsilon}$ elements is greater than $\exp(-nd^{-1.8\epsilon})$ (this observation was first made by Alan Frieze in [2]).

Let Y be the number of $\mathcal{V}$-disjoint k_0-independent sets of mk_0 elements, $m = \lceil \frac{2nd^{-\epsilon}}{k_0} \rceil$, which can be split into exactly m independent sets, each of k_0 elements. Then

$$P(Y > 0) \ge \frac{(\mathrm{E}\,Y)^2}{\mathrm{E}\,Y^2}$$

and

$$\frac{\mathrm{E}\,Y^2}{(\mathrm{E}\,Y)^2} \le \prod_{i=1}^{m} \sum_{\substack{k_1,k_2,\ldots,k_{m+1} \\ \sum_{j=1}^{m+1} k_j = k_0}} \frac{\binom{k_0}{k_1}\binom{k_0}{k_2}\cdots\binom{k_0}{k_m}\binom{\frac{n}{z}-(i-1)k_0}{k_{m+1}} z^{k_{m+1}}}{\binom{\frac{n}{z}-(i-1)k_0}{k_0} z^{k_0} (1-p)^{\sum_{j=1}^{m}\binom{k_j}{2}}}$$

$$\le \left[\sum_{l=0}^{k_0} \frac{a_l \binom{\frac{n}{z}-mk_0}{k_0-l} z^{k_0-l}}{\binom{\frac{n}{z}-mk_0}{k_0} z^{k_0}}\right]^m$$

$$\le \left[\sum_{l=0}^{k_0} \frac{a_l}{(n-k_0mz)^l}\cdot\frac{k_0!}{(k_0-l)!}\right]^m \ \le \ \left[\sum_{l=0}^{k_0} a_l \left(\frac{2k_0}{n}\right)^l\right]^m$$

where

$$(1) \qquad a_l = \sum_{\substack{k_1,k_2,\ldots,k_m \\ \sum_{j=1}^m k_j = l}} \binom{k_0}{k_1}\binom{k_0}{k_2}\cdots\binom{k_0}{k_m}(1-p)^{-\sum_{j=1}^{m}\binom{k_j}{2}} .$$

To estimate a_l divide all terms of the sum into two groups. We shall call a term *large* if some of indices $k_1, k_2, \ldots, k_m$ are larger than $0.1\epsilon k_0$ and *small* otherwise.

Denote by $k'_1, k'_2, \ldots, k'_r$ those from $k_1, k_2, \ldots, k_m$ which are greater than $0.1\epsilon k_0$. Since $\sum_{j=1}^m k_j = l$ so $r \le 10\epsilon^{-1}$ and we can choose $k'_1, k'_2, \ldots, k'_r$ in at most $(ml)^{10\epsilon^{-1}}$ ways. Moreover, for every k', k'' such that $k' > k'' > 0.1\epsilon k_0$ and $k' + k'' \le l \le k_0$, we have (for d large enough)

$$\frac{\binom{k_0}{k'}\binom{k_0}{k''}(1-p)^{-\binom{k'}{2}-\binom{k''}{2}}}{\binom{k_0}{k'+k''}(1-p)^{-\binom{k'+k''}{2}}} \le 2^{k_0}2^{k_0}\exp(-k'k'') < 1 ,$$

so

$$\binom{k_0}{k'_1}\binom{k_0}{k'_2}\cdots\binom{k_0}{k'_r}(1-p)^{-\sum_{j=1}^{r}\binom{k'_j}{2}} \le \binom{k_0}{l}\exp\left(\frac{l^2 d}{2n}\right) .$$

Furthermore, for every choice of $\bar{k}_1, \bar{k}_2, \ldots, \bar{k}_s$, one can easily get the following inequality

$$(2) \qquad \sum_{\substack{\bar{k}_1,\bar{k}_2,\ldots,\bar{k}_s \\ \sum_{j=1}^s \bar{k}_j = l \\ \max_{1\le j\le s} \bar{k}_j = b}} \binom{k_0}{\bar{k}_1}\binom{k_0}{\bar{k}_2}\cdots\binom{k_0}{\bar{k}_s}(1-p)^{-\sum_{j=1}^{s}\binom{\bar{k}_j}{2}} \le \binom{sk_0}{l}\exp\left(\frac{bld}{2n}\right) .$$

Hence, (1) and (2) imply that the sum of all large terms is estimated from above by

$$(ml)^{10\epsilon^{-1}}\binom{k_0}{l}\exp\left(\frac{l^2 d}{2n}\right)\binom{mk_0}{0.1\epsilon l}\exp\left(\frac{0.1\epsilon l^2 d}{2n}\right) \le \left(\frac{10k_0}{l}\exp\left(\frac{(1+0.2\epsilon)ld}{2n}\right)\right)^l$$

whereas the upper bound for the sum of all small terms, from (2), is given by

$$\binom{mk_0}{l}\exp\left(\frac{0.1\epsilon l^2 d}{2n}\right) \le \left(\frac{emk_0}{l}\exp\left(\frac{0.1\epsilon ld}{2n}\right)\right)^l .$$

So, finally, we obtain

$$\frac{\mathrm{E}\,Y^2}{(\mathrm{E}\,Y)^2} \le \left[\sum_{l=0}^{k_0}\left(\frac{10\ k_0^2\ m}{ln}\ \exp\left(\frac{0.1\epsilon l d}{2n}\right)\right)^l + \left(\frac{20\ k_0^2}{ln}\ \exp\left(\frac{(1+0.2\epsilon)ld}{2n}\right)\right)^l\right]^m .$$

Since the maximal term in the above sum is smaller than $\exp\left(\frac{100k_0^2 d^{0.15\epsilon} m}{n}\right)$ we have

$$\mathrm{Prob}(Y>0) > \exp\left(-\frac{200k_0^2 d^{0.15\epsilon} m^2}{n}\right) > \exp(-nd^{-1.8\epsilon}) .$$

This completes the proof of Lemma 1. □

LEMMA 2. *Let $0<\epsilon<0.1$. Then there is a constant d_ϵ such that for $np(n)=d(n)>d_\epsilon$ with the probability greater than $1-n^{-1}-d^{-0.1\epsilon}$ more than $n-nd^{-0.05\epsilon}$ vertices of $G(n,p)$ can be properly coloured with less than $(1+5\epsilon)\frac{d}{2\log d}$ colours.*

Proof. In the proof we shall use an "expose-and-merge" technique introduced by David Matula (for details see [3] and [4]).

For $A\subset[n]$ define $[A]^2=\{\{v,w\}: v,w\in A\}$ and let $\bar{k}_0=(2-0.05\epsilon)(1-\epsilon)\frac{n\log d}{d}$, $l_0=\frac{nd^{-1.1\epsilon}}{\bar{k}_0}$. Consider the following algorithm

Algorithm

$E := \emptyset$;
$F_0 := \emptyset$;
$W_0 := \emptyset$;
for $i=1$ **to** $d^{1.1\epsilon}-d^\epsilon$ **do**

begin

choose randomly $A_i\subset[n]\setminus W_{i-1}$ with $|A_i|=nd^{-\epsilon}$;

define $\mathcal{G}_i$ as the graph with the set of vertices A_i and the set of edges E_i, where the probability that $\{v,w\}\in E_i$ is independent and equal p for each $\{v,w\}\in[A_i]^2$;

choose a family $\{\tilde{\mathfrak{R}}_1^i,\tilde{\mathfrak{R}}_2^i,\ldots,\tilde{\mathfrak{R}}_{l_0}^i\}$ of disjoint independent sets from A_i, each of the size $\bar{k}_0$ – if it is not possible FAIL ;

$E_i' := E_i\setminus(E_i\cap F_{i-1})$;

$E := E\cup E_i'$;

$F_i := F_{i-1}\cup[A_i]^2$;

$W_i := W_{i-1}\cup\bigcup_{l=1}^{l_0}\tilde{\mathfrak{R}}_l^i$;

end

$\overline{F} := [n]^2 \setminus \bigcup_{i=1}^{d^{1.1\epsilon}-d^\epsilon} F_i$;

choose $\overline{E} \subset \overline{F}$ in such a way that each $e \in \overline{F}$ belong to $\overline{E}$ with the probability p independently of each other ;

$E := E \cup \overline{E}$;

output E ; $\{\tilde{\mathfrak{R}}_1^1, \tilde{\mathfrak{R}}_2^1, \ldots, \tilde{\mathfrak{R}}_{l_0}^{d^{1.1\epsilon}-d^\epsilon}\}$;

end

Let us observe first that the probability that $\{v, w\} \in E$ is equal p independently for each $\{v, w\} \in [n]^2$, so the graph $\tilde{G}$ with the set of vertices $[n]$ and the set of edges E can be treated as $G(n, p)$.

Obviously, we can consider each $\mathcal{G}_i$ as $G(\bar{n}, p)$ where $\bar{n} = nd^{-\epsilon}$. The average degree of such a random graph is given by $\bar{d} = nd^{-\epsilon}p = d^{1-\epsilon}$ and $\bar{k}_0 = (2 - 0.05\epsilon)\frac{\bar{n}\log\bar{d}}{\bar{d}}$. Thus, from Lemma 1, the probability that $\mathcal{G}_i$ does not contain $\bar{k}_0$-independent sets with $nd^{-1.1\epsilon} < \bar{n}\bar{d}^{-0.05\epsilon}$ elements is less than n^{-2} so the probability of FAIL in the Algorithm is less than n^{-1}.

Thus, with the probability $1 - o(1)$, the Algorithm finds

$$(d^{1.1\epsilon} - d^\epsilon)l_0 = \frac{n - nd^{-0.1\epsilon}}{\bar{k}_0} < (1 + 5\epsilon)\frac{d}{2\log d}$$

disjoint sets $\tilde{\mathfrak{R}}_1^1, \tilde{\mathfrak{R}}_2^1, \ldots, \tilde{\mathfrak{R}}_{l_0}^{d^{1.1\epsilon}-d^\epsilon}$ each of the size $\bar{k}_0$, so $\sum_{i,l}|\tilde{\mathfrak{R}}_l^i| = n - nd^{-0.1\epsilon}$. Note however, that although $\tilde{\mathfrak{R}}_l^i$ is an independent subset of $\mathcal{G}_i$ it is *not* necessarily independent as a subset of $\tilde{G}$. Let X denote the number of edges of $\tilde{G}$ contained in $\tilde{\mathfrak{R}}_l^i$ for some $1 \le i \le d^{1.1\epsilon} - d^\epsilon$, $1 \le l \le l_0$. We shall estimate from above the size of X.

Let $\{v, w\} \in [n]^2$ be such that $\{v, w\} \in E$ and $\{v, w\} \in \tilde{\mathfrak{R}}_l^i$ for some i, l. Denote by $i(v, w)$ the smallest number i for which $\{v, w\} \subset A_i$ and let $j(v, w)$, $l(v, w)$ be such that $\{v, w\} \subset \tilde{\mathfrak{R}}_{l(v,w)}^{j(v,w)}$. Notice that $i(v, w) < j(v, w)$ since $\{v, w\} \in E$ implies that $\{v, w\}$ is an edge of $\mathcal{G}_{i(v,w)}$ whereas the set $\tilde{\mathfrak{R}}_{l(v,w)}^{j(v,w)}$ contains no edges of $\mathcal{G}_{j(v,w)}$. Since for all i we have $|W_i| \le n - nd^{0.1\epsilon}$, so the probability that for chosen $i(v, w)$, $j(v, w)$, a pair $\{v, w\}$ is contained in both $A_{i(v,w)}$ and $A_{j(v,w)}$ is less than $(2d^{-0.9\epsilon})^4$. Now observe that each subset of $A_{j(v,w)}$ of k_0 elements is equally likely to be chosen as $\tilde{\mathfrak{R}}_l^{j(v,w)}$ with $1 \le l \le l_0$ (i.e. this event depends only on the structure of $\mathcal{G}_{j(v,w)}$ which is symmetric with respect to the labelling of vertices). Thus, since $|A_{j(v,w)}| = nd^{-\epsilon}$, the probability that both v, w are in the same set $\tilde{\mathfrak{R}}_{l(v,w)}^{j(v,w)}$ for some $l(v, w)$ is less than $d^{1.1\epsilon-1}$.

So finally we arrive at the following upper bound for the expectation of X

$$\mathrm{E}X \le \binom{n}{2}\binom{d^{1.1\epsilon} - d^\epsilon}{2} 16d^{-3.6\epsilon}p\,d^{1.1\epsilon-1} < nd^{-0.2\epsilon}\,.$$

Thus, from Markov inequality,

$$\text{Prob}(X > nd^{-0.1\epsilon}) \leq d^{-0.1\epsilon} .$$

Now, for all i, l, delete from $\tilde{\mathcal{R}}_l^i$ all these vertices which belongs to edges of $\tilde{G}$ which are contained in $\tilde{\mathcal{R}}_l^i$ and denote the set obtained in this way by $\mathcal{R}_l^i$. Then

$$\sum_{i,l} |\mathcal{R}_l^i| \geq \sum_{i,l} |\tilde{\mathcal{R}}_l^i| - 2X = n - nd^{-0.1\epsilon} - 2X$$

so

$$\text{Prob}\left(\sum_{i,l} |\mathcal{R}_l^i| \geq n - nd^{-0.05\epsilon}\right) > 1 - o(1) - d^{0.1\epsilon} .$$

This completes the proof of Lemma 2. □

Proof of Theorem. Lemma 2 implies that with the probability at least $1-o(1)-d^{-0.01\epsilon}$ we can colour $n - nd^{-0.005\epsilon}$ vertices of $G(n,p)$ using only $(1+0.5\epsilon)\frac{d}{2\log d}$ colours. One can easily check that for d large enough a.s. every subgraph of $G(n,p)$ on s vertices, $s \leq nd^{-0.005\epsilon}$, has less than $d^{1-0.001\epsilon}s$ edges. Hence, for large d, we have

$$\text{Prob}\left(\chi(G(n,p)) \leq (1+0.9\epsilon)\frac{d}{2\log d}\right) > 1 - o(1) - d^{-0.005\epsilon} , \tag{3}$$

and for $d(n) \to \infty$, the assertion follows.

Moreover, Shamir and Spencer proved in [5] that for every $d(n) < \log n$ there exists a function $u_d(n)$ such that a.s.

$$u_d(n) \leq \chi(G(n,p)) \leq u_d(n) + 5 .$$

Thus, since from (3) for $d(n)$ greater than some constant d_ϵ we have

$$\liminf_{n\to\infty} \text{Prob}\left(\chi(G(n,p)) \leq (1+0.9\epsilon)\frac{d}{2\log d}\right) > 0.5$$

so, for such $d(n)$, a.s.

$$\chi(G(n,p)) \leq (1+0.9\epsilon)\frac{d}{2\log d} + 5 \leq (1+\epsilon)\frac{d}{2\log d} .$$

This completes the proof of Theorem. □

Acknowledgements. I would like to thank Joel Spencer and Alan Frieze for discussions and many valuable suggestions.

REFERENCES

[1] B. Bollobás, *The chromatic number of random graphs*, Combinatorica, to appear.

[2] A. M. Frieze, *On the independence number of random graphs*, to appear.

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

[4] D. Matula and L. Kučera, *On chromatic number of random graphs*, in preparation.

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