Equitable coloring of random graphs

Michael Krivelevich* Balázs Patkós †

July 2, 2008

Abstract

An equitable coloring of a graph is a proper vertex coloring such that the sizes of any two color classes differ by at most one. The least positive integer k for which there exists an equitable coloring of a graph G with k colors is said to be the *equitable chromatic number* of G and is denoted by $\chi_=(G)$. The least positive integer k such that for any $k' \geq k$ there exists an equitable coloring of a graph G with k' colors is said to be the *equitable chromatic threshold* of G and is denoted by $\chi^*_=(G)$. In this paper we investigate the asymptotic behavior of these coloring parameters in the probability space $G(n,p)$ of random graphs. We prove that if $n^{-1/5+\epsilon} < p < 0.99$ for some $0 < \epsilon$, then almost surely $\chi(G(n,p)) \leq \chi_=(G(n,p)) = (1+o(1))\chi(G(n,p))$ holds (where $\chi(G(n,p))$ is the ordinary chromatic number of $G(n,p)$). We also show that there exists a constant C such that if $C/n < p < 0.99$, then almost surely $\chi(G(n,p)) \leq \chi_=(G(n,p)) \leq (2+o(1))\chi(G(n,p))$. Concerning the equitable chromatic threshold, we prove that if $n^{-(1-\epsilon)} < p < 0.99$ for some $0 < \epsilon$, then almost surely $\chi(G(n,p)) \leq \chi^*_=(G(n,p)) \leq (2+o(1))\chi(G(n,p))$ holds, and if $\frac{\log^{1+\epsilon} n}{n} < p < 0.99$ for some $0 < \epsilon$, then almost surely we have $\chi(G(n,p)) \leq \chi^*_=(G(n,p)) = O_\epsilon(\chi(G(n,p)))$.

Keywords: graph coloring, equitable coloring, random graphs
AMS subject classification: 05C15, 05C80

1 Introduction and main results

An equitable coloring of a graph is a proper vertex coloring such that the sizes of any two color classes differ by at most one. One of the first results about equitable colorings

*School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: krivelev@post.tau.ac.il. Research supported in part by a USA-Israeli BSF grant and a grant from the Israeli Science Foundation.

†Department of Mathematics and its Applications, Central European University, Budapest, 1051 Nádor u. 9., Hungary. Email: patkos@renyi.hu. Research supported in part by the European Network PHD, MCRN-511953.

is the celebrated Hajnal-Szemerédi theorem [5] (recently reproved in a much simpler way by Kierstead and Kostocka [10]) stating that every graph with maximum degree Δ has an equitable coloring with k colors for any $k \geq \Delta + 1$. Lih's paper [14] surveys some basic results on equitable colorings and how the bound of $\Delta + 1$ can be replaced by Δ for certain classes of graphs. Applications of the Hajnal-Szemerédi theorem and recent results on equitable colorings of graphs can be found in (among others) [1], [2], [9], [11], [12], [19]. Equitable coloring turned out to be useful in establishing bounds on tails of sums of dependent variables [6], [8], [18].

The property of being equitably colorable by k colors is not monotone in k, i.e. it is possible that a graph admits an equitable k-coloring but is not equitably $(k+1)$-colorable. Therefore there are two parameters of a graph related to equitable colorings. The least positive integer k for which there exists an equitable coloring of a graph G with k colors is said to be the *equitable chromatic number* of G and is denoted by $\chi_=(G)$, while the least positive integer k such that for any $k' \geq k$ there exists an equitable coloring of a graph G with k' colors is said to be the *equitable chromatic threshold* of G and is denoted by $\chi^*_=(G)$. (We follow the notation of [14], though equitable chromatic threshold is sometimes denoted by $eq(G)$.)

In this paper, we prove results on the asymptotic behavior of the above parameters in the random graph $G(n,p)$. By $G(n,p)$ we mean the probability space of all labeled graphs on n vertices, where every edge appears randomly and independently with probability $p = p(n)$. We say that $G(n,p)$ possesses a property $\mathcal{P}$ *almost surely*, or a.s. for brevity, if the probability that $G(n,p)$ satisfies $\mathcal{P}$ tends to 1 as n tends to infinity. Our main aim is to address the following conjecture:

Conjecture: *There exists a constant C such that if $C/n < p < 0.99$, then almost surely*

$$\chi(G(n,p)) \leq \chi_=(G(n,p)) \leq \chi^*_=(G(n,p)) = (1+o(1))\chi(G(n,p))$$

holds (the first two inequalities are true by definition).

Before proceeding to our results let us summarize the asymptotic behavior of $\chi(G(n,p))$ in one theorem which for some values of p was discovered by Bollobás [4] and independently by Matula and Kučera [17], and for other values of p by Łuczak [15] (see also Chapter 7 of [7]). Here and throughout the paper log stands for the logarithm in the natural base e.

Theorem 1.1 *The following statements are true for the chromatic number $\chi(G(n,p))$.*

(a) *If $\log^{-8} n < p < 0.99$, then almost surely*

$$\frac{n}{2\log_b n - \log_b \log_b(np)} \leq \chi(G(n,p)) \leq \frac{n}{2\log_b n - 8\log_b \log_b(np)},$$

where $b = \frac{1}{1-p}$.

(b) *There exists a constant $C_0 > 0$ such that for every $p = p(n)$ satisfying $C_0/n \le p \le \log^{-8} n$ almost surely*

$$\frac{np}{2\log(np) - 2\log\log(np)} \le \chi(G(n,p)) \le \frac{np}{2\log(np) - 40\log\log(np)}.$$

Note that if p tends to 0, then $2(\log_b n - \log_b \log_b(np))$ is asymptotically $\frac{2(\log(np) - \log\log(np))}{p}$. Also note that if p is as in case (a), then $\log_b n \gg \log_b \log_b(np)$, while if p is as in case (b), then $\log(np) \gg \log\log(np)$, so changing the coefficient of $\log_b \log_b(np)$ (or $\log\log(np)$ in the latter case) in the denominator has no effect on the asymptotics of the expressions in Theorem 1.1. In our theorems, all lower bounds on $\chi_=(G(n,p))$ or $\chi^*_=(G(n,p))$ follow from the lower bound of Theorem 1.1.

By analyzing a greedy algorithm we obtain the following theorem.

Theorem 1.2 *Almost surely the equitable chromatic threshold $\chi^*_=(G(n,p))$ satisfies the following inequalities.*

(a) *If $p < 0.99$ and $p = n^{-o(1)}$, then*

$$\frac{n}{2\log_b n - \log_b \log_b(np)} \le \chi^*_=(G(n,p)) \le \frac{n}{\log_b n - 2.1\log_b \log_b(np)},$$

where $b = \frac{1}{1-p}$.

(b) *If there exists a positive ϵ such that $n^{-1+\epsilon} \le p \le \log^{-8} n$, then*

$$\frac{np}{2\log(np) - \log\log(np)} \le \chi^*_=(G(n,p)) \le \frac{np}{\log(np) - 2.1\log\log(np)}.$$

(c) *If $p \to 0$ and there is a positive ϵ such that $p \ge \frac{\log^{1+\epsilon} n}{n}$, then for any ϵ' with $0 < \epsilon' < \frac{\epsilon}{1+\epsilon}$ we have*

$$\frac{np}{2\log(np) - \log\log(np)} \le \chi^*_=(G(n,p)) \le \frac{np}{(\frac{\epsilon}{1+\epsilon} - \epsilon')\log(np)}.$$

Applying case (c) of the above theorem with ϵ tending to infinity, we get the following result.

Corollary 1.3 *If $p < 0.99$ and $\log\log n \ll \log(np)$, then almost surely we have*

$$\chi(G(n,p)) \le \chi_=(G(n,p)) \le \chi^*_=(G(n,p)) \le (2 + o(1))\chi(G(n,p)).$$

Although the above result is not asymptotically tight, the algorithm used in the proof gives us almost surely an equitable coloring in polynomial time, while the other results just prove the existence of a such coloring.

The following theorem is a purely deterministic one, which we will use in our probabilistic proofs; we state it among the main results for it can be of independent interest.

Theorem 1.4 *Let G be a graph on n vertices in which every induced subgraph $G[U]$ with $|U| \geq m$ contains an independent set of size s. Suppose further that $\frac{n-\Delta(G)-m-ms^2}{s} \geq m$ holds, where $\Delta(G)$ denotes the maximum degree of the graph G. Then G can be properly colored using color classes only of size s and $s-1$.*

Our next result gives the asymptotic value of $\chi_=(G(n,p))$ for dense random graphs.

Theorem 1.5 *If $n^{-1/5+\epsilon} < p < 0.99$ for some $\epsilon > 0$, then the following holds almost surely:*

$$\chi(G(n,p)) \leq \chi_=(G(n,p)) \leq (1+o(1))\chi(G(n,p)).$$

Our last theorem gives the same upper bound as Theorem 1.2; its importance is that its proof works also when p tends to 0 very quickly (i.e., when $G(n,p)$ is very sparse).

Theorem 1.6 *There exists a constant C such that if $\frac{C}{n} \leq p \leq \log^{-8} n$, then a.s. $\chi_=(G(n,p)) \leq \frac{np}{(1-o(1))\log(np)}$.*

The rest of this paper is organized as follows: in the next section we introduce some notation and gather some basic facts that we will use in our proofs. In Section 3, we prove Theorem 1.2 analyzing a greedy algorithm. Theorem 1.5 will be proved in Section 4 which consists of two subsections. In the first subsection, we give the proof of Theorem 1.4 and deduce the "very dense case" of Theorem 1.5, in the second subsection we prove the "dense, but not very dense" case of Theorem 1.5. Section 4 contains the proof of Theorem 1.6.

2 Preliminaries

In this section, we introduce some (standard) notation and gather some basic results concerning binomial distributions, random graphs and equitable coloring that we will use in our proofs.

Let $\alpha(G)$ denote the independence number (the size of a largest independent set) of G. We will say that a graph G is *d-degenerate* if every subgraph of G contains a vertex of degree at most d and the *degeneracy number* of G is the smallest integer d such that G is d-degenerate. We will also say that the sets $S_1, S_2, ..., S_s$ are almost equal if

$||S_i| - |S_j|| \leq 1$ for any $1 \leq i, j \leq s$. The neighborhood of a set of vertices $U \subseteq V(G)$ is $\{v \in V(G) \setminus U : \exists u \in U \text{ such that } (u,v) \in E(G)\}$ and is denoted by $N(U)$.

The following well-known bound (see e.g. Theorem 2.1. in [7]) on the tails of binomial distributions will be used several times for proving some properties of random graphs.

Chernoff bound: *If X is a binomially distributed random variable with parameters n and p and $\lambda = np$, then for any $t \geq 0$ we have*

$$\mathbb{P}(X \geq \mathbb{E}X + t) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right)$$

and

$$\mathbb{P}(X \leq \mathbb{E}X - t) \leq \exp\left(-\frac{t^2}{2\lambda}\right) .$$

Also, the following result of Kostochka, Nakprasit and Pemmaraju will be quoted.

Theorem 2.1 (Kostochka, Nakprasit, Pemmaraju [11]) *For every $d, n \geq 1$, if a graph G is d-degenerate, has n vertices and satisfies $\Delta(G) \leq n/15$, then $\chi^*_=(G) \leq 16d$.*

Let us collect a couple of facts, easy inequalities that will be used during the proofs.

Fact 2.2 *If $p > \frac{\log^2 n}{n}$, then almost surely $\Delta(G) < 1.01np$.*

Corollary 2.3 *If $p > \frac{\log^2 n}{n}$, then almost surely for every set of vertices U of size $|U| = t$, we have $|N(U)| \leq 1.01npt$.*

For binomial coefficients we will have the following upper bound.

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k = \exp\left(O\left(k \log(n/k)\right)\right).$$

Let us finish this section with introducing the standard notations used for comparing the order of magnitude of two non-negative sequences. We will write $g(n) = o(f(n))$ ($g(n) = \omega(f(n))$) to denote the fact that $\lim_n \frac{g(n)}{f(n)} = 0$ ($\lim_n \frac{g(n)}{f(n)} = \infty$), while $g(n) = O(f(n))$ ($g(n) = \Omega(f(n))$) will mean that there exists a positive number K such that $\frac{g(n)}{f(n)} < K$ ($\frac{g(n)}{f(n)} > K$) for all integers n. Finally, we will write $g(n) \sim f(n)$ for $\lim_n \frac{g(n)}{f(n)} = 1$.

3 Coloring greedily

Proof of Theorem 1.2 We will use the following greedy algorithm: let us fix an integer k, the future number of color classes and a partition of the vertex set $V = V_1 \cup V_2 \cup ... \cup V_{\lceil \frac{n}{k} \rceil}$ such that $|V_1| = |V_2| = ... = |V_{\lfloor \frac{n}{k} \rfloor}| = k$ and if k does not divide n $V_{\lceil \frac{n}{k} \rceil} = V \setminus \bigcup_{i=1}^{\lfloor \frac{n}{k} \rfloor} V_i$. Our algorithm consists of $\lceil \frac{n}{k} \rceil$ rounds. In the ith round we expose all the edges having one endpoint in V_i and the other in $\bigcup_{j=1}^{i} V_j$. In such a way, our graph will be truly random after the $\lceil \frac{n}{k} \rceil$th round. Suppose that after round $(i-1)$ we have a proper coloring of the subgraph spanned by $\bigcup_{j=1}^{i-1} V_j$ in k colors so that each color class has exactly one vertex in each V_j and has thus $i-1$ vertices. We say that the ith round succeeds if it is possible to extend this coloring to a proper coloring of $G[\bigcup_{j=1}^{i} V_j]$ by adding one vertex of V_i to each of the k color classes; if this is impossible then the round fails. Observe that the edges inside V_i have no bearing on the outcome of the ith round and can thus be ignored. If all $\lceil \frac{n}{k} \rceil$ rounds are successful, then clearly we have produced an equitable coloring of G in k colors.

Formally we can define an auxiliary random bipartite graph G^i on $2k$ vertices: k vertices stand for the vertices in V_i and the k other vertices represent the color classes that we have already built. There is an edge between a vertex representing a color class C and a vertex representing a vertex $v \in V_i$ if and only if all the pairs exposed in this round between v and C are non-edges. So our auxiliary graph is a random bipartite graph with edge-probability $\rho_i = (1-p)^{i-1}$.

By Remark 4.3 in Chapter 4 in [7], we know that the probability that there is no perfect matching in our auxiliary graph (that is, we cannot extend our coloring of the original graph $G(n,p)$ properly) is

$$O(ke^{-k\rho_i}).$$

There are $\lceil \frac{n}{k} \rceil$ rounds and the probability of a failure (i.e. there is no matching in the auxiliary graph) is the biggest in the last round. So by the union bound, we get that the probability that our algorithm fails to give us an equitable coloring is

$$O(ne^{-k(1-p)^{\frac{n}{k}}}) = O(e^{\log n - k(1-p)^{\frac{n}{k}}}).$$

It is clear that the probability of a failure decreases as k increases, so again using the union bound, we obtain that the probability that our algorithm does not produce an equitable k'-coloring for at least one $k' \geq k$ is

$$O((n-k)ne^{-k(1-p)^{\frac{n}{k}}}) = O(e^{2\log n - k(1-p)^{\frac{n}{k}}}).$$

So if for some value of k we have $k(1-p)^{\frac{n}{k}} = \omega(\log n)$, then almost surely $\chi^*_=(G(n,p)) \leq k$.

Case (a)

Let $k = \frac{n}{\log_b n - 2.1 \log_b \log_b(np)}$, where $b = \frac{1}{1-p}$. Then we have

$$\begin{aligned} k(1-p)^{\frac{n}{k}} &= \frac{n}{\log_b n - 2.1 \log_b \log_b(np)} (1-p)^{\log_b n - 2.1 \log_b \log_b(np)} \\ &= \frac{n}{\log_b n - 2.1 \log_b \log_b(np)} \frac{\log_b^{2.1}(np)}{n} = \frac{\log_b^{2.1}(np)}{\log_b n - 2.1 \log_b \log_b(np)} . \end{aligned}$$

By the assumption that $p > n^{-\delta}$ for every $\delta > 0$, the above expression is asymptotically equal to $\log_b^{1.1} n = \Omega(\log^{1.1} n) \gg \log n$.

Case (b)
Let $k = \frac{np}{\log(np) - 2.1 \log\log(np)}$. Then we have

$$k(1-p)^{\frac{n}{k}} = \frac{np}{\log(np) - 2.1\log\log(np)}(1-p)^{\frac{1}{p}(\log(np)-2.1\log\log(np))} \sim$$

$$\frac{np}{\log(np) - 2.1\log\log(np)} e^{2.1\log\log(np) - \log(np)} \geq \log^{1.1}(np).$$

By our assumption on p, this last expression is asymptotically bigger than $\log n$.

Case (c)

Fix an ϵ satisfying the assumption of the theorem and let $k = \frac{np}{(\epsilon/(1+\epsilon) - \epsilon')\log(np)}$. Then we have

$$k(1-p)^{\frac{n}{k}} = \frac{np}{(\epsilon/(1+\epsilon) - \epsilon')\log(np)}(1-p)^{\frac{(\epsilon/(1+\epsilon)-\epsilon')\log(np)}{p}} \sim$$

$$\frac{np}{(\epsilon/(1+\epsilon) - \epsilon')\log(np)} e^{-(\epsilon/(1+\epsilon)-\epsilon')\log(np)} = \frac{np}{(\epsilon/(1+\epsilon) - \epsilon')\log(np)}(np)^{-(\epsilon/(1+\epsilon)-\epsilon')} \geq$$

$$\geq \log^{1+\epsilon'/2} n,$$

where for the last inequality we used the assumption $\frac{\log^{1+\epsilon} n}{n} < p$. □

4 The equitable chromatic number of dense random graphs

In this section we prove Theorem 1.4. The proof will be divided into two parts. In the following subsection we prove the theorem when $p > \log^{-8} n$. For this purpose, we first prove Theorem 1.3 and deduce this case of Theorem 1.4 as a corollary along with an application to (n, d, λ)-graphs (see the definition there). In the second subsection we prove the "dense, but not very dense" case.

4.1 The very dense case

Proof of Theorem 1.4: Using the property of G assured by the assumption of the theorem, we pick pairwise disjoint independent sets $I_1, I_2, ..., I_t$ of size s as far as the number of remaining vertices is less than m, i.e. $|V(G) \setminus \cup_{i=1}^{t} I_j| < m$. So we have $t \geq \frac{n-m}{s}$. Since we are allowed to have color classes of size $s-1$, we may remove one vertex from each I_j. We will use this to create independent sets of size s for each $v \in V(G) \setminus \bigcup_{j=1}^{t} I_j$. For the first such vertex v_1, let us pick vertices $u_1 \in I_{j_1}, u_2 \in$

$I_{j_2}, ..., u_m \in I_{j_m}$ such that $i_1 \neq i_2$ implies $j_{i_1} \neq j_{i_2}$ and v_1 is not adjacent to any of the u_i's. Then by the property of G, there exists an independent set of size $s-1$ among the u_is, which together with v_1 is an independent set of size s.

We would like to repeat this procedure for all vertices in $V(G) \setminus \bigcup_{j=1}^{t} I_j$. Since we are allowed to have independent sets only of size s and $s-1$ we cannot use the vertices in the color classes from which we have already removed a vertex. We have to make sure that we can pick the u_i's in the above mentioned way even for the last vertex v_l in $V(G) \setminus \bigcup_{j=1}^{t} I_j$. After the first phase (picking independent sets greedily due to the property of G), we had at least $n-m$ vertices in the I_j's. For each previous vertex in $V(G) \setminus \bigcup_{j=1}^{t} I_j$ we have used $s-1$ independent sets, each containing s vertices. We cannot use these vertices, just as we cannot use at most $\Delta(G)$ vertices connected to v_l. Since we have to pick vertices from different I_j's, the number of possible I_j's we are still allowed to pick from is at least $\frac{n-\Delta(G)-m-ms^2}{s}$. By the assumption of the theorem, this is at least m, so our procedure never fails. □

Now, we are ready to prove the very dense case of Theorem 1.5. In order to apply Theorem 1.4 to the random graph $G(n,p)$, we need to find the corresponding values of m and s. This was the crucial step in Bollobás' proof [4] for the asymptotic value of $\chi(G(n,p))$. Here we cite a result of Krivelevich, Sudakov, Vu and Wormald [13] which we will use setting $\epsilon = 0.01$. Let $k_0 = \max\{k : \binom{n}{k}(1-p)^{\binom{k}{2}} \geq n^4\}$. It is well known that $2\log_b n \geq k_0 \geq 2\log_b n - C' \log_b \log_b(np)$ for some constant C'.

Theorem 4.1 [13] *Let $p(n)$ satisfy $n^{-2/5}\log^{6/5} n \ll p(n) \leq 1-\epsilon$ for an absolute constant $0<\epsilon<1$. Then*

$$\mathbb{P}[\alpha(G(n,p)) < k_0] = e^{-\Omega\left(\frac{n^2}{k_0^4 p}\right)}.$$

Now changing n to $\frac{n}{\log_b^3 n}$ and applying the union bound we get that the probability that some subgraph of $G(n,p)$ on $\frac{n}{\log_b^3 n}$ vertices does not contain an independent set of size

$$2\log_b\left(\frac{n}{\log_b^3 n}\right) - C' \log_b \log_b\left(\frac{np}{\log_b^3 n}\right) \geq 2\log_b n - C'' \log_b \log_b(np)$$

is at most

$$\binom{n}{\frac{n}{\log_b^3 n}} e^{-\Omega\left(\frac{n^2}{k_0^4 p}\right)} \leq \exp\left(O(n) - \Omega\left(\frac{n^2}{k_0^4 p}\right)\right),$$

which tends to zero, since $\log_b n \ll n^\gamma$ and therefore $k_0 \ll n^\gamma$ for all $\gamma > 0$ if $p > \log^{-8} n$. Therefore, using Fact 2.2, we can apply Theorem 1.4 with $s = 2\log_b n - C'' \log_b \log_b(np)$ and $m = \frac{n}{\log_b^3 n}$. This finishes the proof of the very dense case of Theorem 1.5. □

For another application of Theorem 1.4 we need the following definition.

Definition. An (n,d,λ)-*graph* is a d-regular graph on n vertices with eigenvalues $d=\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ such that $\lambda \geq \max\{|\lambda_i| : 2 \leq i \leq n\}$.

Theorem 4.2 *Let G_n be a sequence of (n,d,λ)-graphs where $d(n) \leq 0.9n$ and $\frac{d^3}{n^2\lambda} = \Omega(n^\alpha)$ holds for some $\alpha > 0$. Then $\chi_=(G_n) = O(\frac{d}{\log d})$.*
Proof. We will need the following result of Alon, Krivelevich and Sudakov:

Theorem 4.3 (Alon, Krivelevich, Sudakov [3]) *Let G be an (n,d,λ)-graph such that $\lambda < d < 0.9n$. Then the induced subgraph $G[U]$ of G on any subset U, $|U| = m$, contains an independent set of size at least*

$$\alpha(G[U]) \geq \frac{n}{2(d-\lambda)} \log\left(\frac{m(d-\lambda)}{n(\lambda+1)} + 1\right).$$

We would like to apply Theorem 1.4 with $m = \frac{d^2}{n\log^3 n}$ and $s = c\frac{n\log n}{d}$ for some $c > 0$, so we have to verify that the conditions of the Theorem are met. By the assumption $\frac{d^3}{n^2\lambda} = \Omega(n^\alpha)$, we have $\lambda = o(d)$, so $\frac{n}{2(d-\lambda)} = \Theta\left(\frac{n}{d}\right)$. Again by $\frac{d^3}{n^2\lambda} = \Omega(n^\alpha)$, we have $\log\left(\frac{m(d-\lambda)}{n(\lambda+1)} + 1\right) = \Omega(\log n)$, therefore it is indeed true that every subgraph of G_n with m vertices contains an independent set of size s.

It remains to verify the inequality $\frac{n-\Delta(G)-m-ms^2}{s} \geq m$. $\Delta(G) = d \leq 0.9n$, $ms^2 = c\frac{n}{\log n} = o(n)$, so we have $\frac{n-\Delta(G)-m-ms^2}{s} = \Theta(\frac{n}{s}) = \Theta(\frac{d}{\log n}) \geq \frac{d^2}{n\log^3 n} = m$. Theorem 1.3 gives that $\chi_=(G_n) \leq \frac{n}{s-1} = \Theta\left(\frac{d}{\log n}\right) = \Theta\left(\frac{d}{\log d}\right)$, where the last equality follows from $\frac{d^3}{n^2\lambda} = \Omega(n^\alpha)$ (since this trivially implies $d > n^{2/3}$). $\square$

4.2 The dense, but not very dense case

Bollobás' argument [4] for determining the asymptotic behavior of $\chi(G(n,p))$ for dense random graphs was to find many independent sets of size close to $\alpha(G(n,p))$ and then to color the remaining small set of vertices with few additional colors. The coloring obtained this way is very much *not* equitable. To overcome this difficulty we will introduce the notion of an independent (t,k)-comb which informally consists of t pairwise disjoint independent sets each of size k and an additional transversal independent set of size k that contains k/t vertices from each of the pairwise disjoint independent sets (see the next subsection for the formal definition). After proving that every large enough subgraph of $G(n,p)$ contains an independent (t,k)-comb with t and k appropriately chosen, we will proceed as follows: we will pick independent (t,k)-combs $C_1, C_2, ..., C_s$ until the number of remaining vertices will be small. Then using Theorem 2.1, we will partition these remaining vertices into exactly as many independent sets $I_1, I_2, ..., I_s$ as the number of independent (t,k)-combs. Finally, we will match the I_i's to the independent (t,k)-combs in such a way that if C_i is matched to I_j, then there are hardly any edges between I_j and the transversal independent set of C_i. Thus we will be able to obtain an equitable coloring of $G(n,p)$ by partitioning all matched pairs into $t+1$ independent sets where the $(t+1)$-st set will be formed from the vertices of I_j and most vertices of the transversal independent set of C_i.

In the next subsection we prove several properties of $G(n,p)$ including the existence of the independent combs and in Subsection 4.2.2 we prove the remaining case of Theorem 1.5.

4.2.1 Properties of random graphs

In this subsection we collect all auxiliary lemmas that we will use in the proof of Theorem 1.5.

Lemma 4.4 *Let $p \geq 30/n$ and let $c > 0$ a constant. Then in $G(n,p)$ almost surely every $s \leq \frac{n}{\log^c(np)}$ vertices span at most $\frac{2np}{\log^c(np)}s$ edges. Therefore almost surely every subgraph of $G(n,p)$ of size s is $\frac{4np}{\log^c(np)}$-degenerate.*

Proof: Let $r = \frac{2np}{\log^c(np)}$. The probability of the existence of a subset $V_0 \subseteq V$ violating the assertion of the lemma is at most

$$\sum_{i=r}^{\frac{n}{\log^c(np)}} \binom{n}{i}\binom{\binom{i}{2}}{ri} p^{ri} \leq \sum_{i=r}^{\frac{n}{\log^c(np)}} \left[\frac{en}{i}\left(\frac{ei}{2r}\right)^r p^r\right]^i = \sum_{i=r}^{\frac{n}{\log^c(np)}} \left[\frac{e^2np}{2r}\left(\frac{eip}{2r}\right)^{r-1}\right]^i$$

$$\leq \sum_{i=r}^{\frac{n}{\log^c(np)}} \left[3\log^c(np)\left(\frac{ei\log^c(np)}{4n}\right)^{\frac{2np}{\log^c(np)}-1}\right]^i = \sum_{i=r}^{\frac{n}{\log^c(np)}} a_i,$$

where $a_i = \left[3\log^c(np)\left(\frac{ei\log^c(np)}{4n}\right)^{\frac{2np}{\log^c(np)}-1}\right]^i$.

Noting that

$$\frac{a_{i+1}}{a_i} = 3\log^c(np)\left(\frac{e\log^c(np)}{4n}\right)^{\frac{2np}{\log^c(np)}-1}\left(\frac{(i+1)^{i+1}}{i^i}\right)^{\frac{2np}{\log^c(np)}-1}$$

is monotone increasing with respect to i, we obtain that the terms of the last sum form a convex sequence, so either the first one or the last one is the largest. Therefore the sum is at most

$$\max\left\{\frac{n}{\log^c(np)}\left[3\log^c(np)\left(\frac{ep}{2}\right)^{\frac{2np}{\log^c(np)}-1}\right]^{\frac{2np}{\log^c(np)}}, \frac{n}{\log^c(np)}\left[3\log^c(np)\left(\frac{e}{4}\right)^{\frac{2np}{\log^c(np)}-1}\right]^{\frac{n}{\log^c(np)}}\right\},$$

which tends to zero as n tends to infinity. $\square$

Lemma 4.5 *For every pair of positive constants $0 < \gamma' < \gamma$ there exists another constant $C = C(\gamma', \gamma)$ such that if $p \geq C/n$ and $x = x(n), y = y(n)$ satisfy $pxy = \frac{n}{\log^{\gamma}(np)}, x \leq y \leq \frac{n}{\log^{\gamma}(np)}$, then the following holds almost surely in $G(n,p)$:*
for any two disjoint sets of vertices U_1 and U_2 with $|U_1| = x$ and $|U_2| = y$, the number of edges between U_1 and U_2 is at most $\frac{n}{\log^{\gamma'}(np)}$.

Proof: The expected number of edges between two fixed disjoint sets of the prescribed size is $pxy = \frac{n}{\log^{\gamma}(np)}$. Applying the Chernoff bound, we get that if C is chosen such that $2\frac{n}{\log^{\gamma}(np)} \leq \frac{n}{\log^{\gamma'}(np)}$, then the probability that for a fixed pair of sets there are too many edges between them is at most

$$\exp\left(-\frac{3n}{16\log^{\gamma'}(np)}\right).$$

Therefore, using the union bound for all possible pairs of sets, we get that the probability that there is a pair of sets contradicting the lemma is at most

$$\binom{n}{x}\binom{n}{y}\exp\left(-\frac{3n}{\log^{16\gamma'}(np)}\right) =$$

$$\exp\left(x\log(n/x) + x + y\log(n/y) + y - \frac{3n}{16\log^{\gamma'}(np)}\right) \leq \exp\left(4y\log(n/y) - \frac{3n}{16\log^{\gamma'}(np)}\right).$$

$y\log(n/y) \leq \frac{n}{\log^{\gamma''}(np)}$ for some $\gamma' < \gamma'' < \gamma$ if C is large enough, therefore the expression above tends to zero provided $\frac{4}{\log^{\gamma''}(np)} \leq \frac{3}{16\log^{\gamma'}(np)}$ which is true if C is chosen large enough. $\square$

Definition: A set $E_0 \subset \binom{V(G)}{2} \setminus E(G)$ ($E_0 \subset E(G)$) of non-edges (edges) forms an *independent (clique)* (t,k)*-comb* in a graph G if there exist $2t$ pairwise disjoint sets of vertices $I_1, ..., I_t, J_1, ..., J_t$ with $|I_i| = \frac{t-1}{t}k$, and $|J_i| = \frac{1}{t}k$ for every i ($1 \leq i \leq t$) such that $E_0 = \bigcup_{i=1}^{t}\binom{I_i \cup J_i}{2} \cup \binom{\cup J_i}{2}$.

An edge of a clique (t,k)-comb is called *horizontal* if it lies inside some $I_i \cup J_i$ and is called *crossing* otherwise (i.e. if it is an edge between some $J_{i_1} \neq J_{i_2}$). The number of (all) edges in a clique (t,k)-comb will be denoted by $E(t,k)$ (t and k will be omitted if their value is clear from context).

Let us write m in the form $m = k + s\frac{t-1}{t}k + l$ for some $0 \leq s \leq t-1$ and $0 \leq l < \frac{t-1}{t}k$. An *optimal* subcomb of a clique (t,k)-comb is an induced subgraph of a clique (t,k)-comb spanned by all vertices in $\bigcup_{j=1}^{t} J_j \cup \bigcup_{i=1}^{s} I_i$ and l additional vertices from I_{s+1}. If $m < k$, then an optimal subcomb is an induced subgraph spanned by m vertices each of which are in $I_i \cup J_i$ for some $1 \leq i \leq t$. The number of edges spanned by an optimal subcomb of m vertices will be denoted by $L(t,k,m)$.

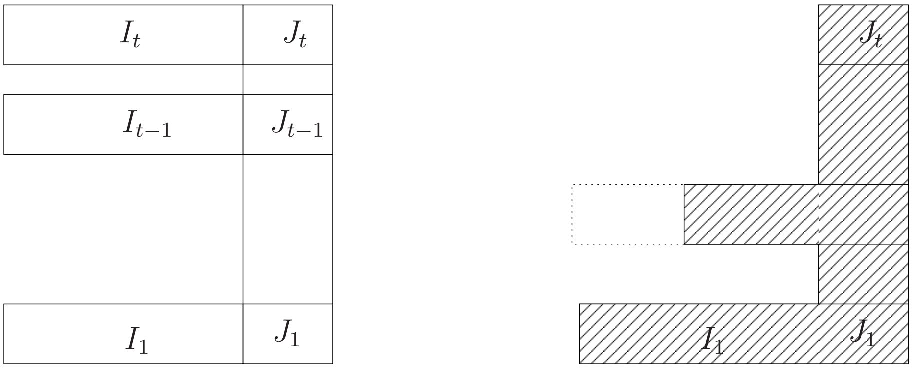

Figure 1: The vertex set of a (t,k)-comb and an optimal subcomb.

Lemma 4.6 *The number of edges spanned by any set of vertices U in a clique (t,k)-comb with $|U| = m$ is at most $L(t,k,m)$.*

Proof: If $m \le k$, then an optimal subcomb is a clique, thus the lemma is true trivially, therefore we can assume that $m > k$. Let U be a set of m vertices in a clique (t,k)-comb spanning the most number of edges among all such sets. We may assume that for any i $(1 \le i \le t)$, if $U \cap I_i \ne \emptyset$, then $J_i \subset U$. Indeed, if $u \in U \cap I_i$ and $v \in J_i, v \notin U$, than $U' = U \setminus \{u\} \cup \{v\}$ contains the same number of horizontal edges and contain at least as many crossing edges as U. We can assume that there is at most one I_i such that $U \cap I_i \ne \emptyset$ and $I_i \not\subset U$. Indeed, if there were two such I_i's, then we could pick a vertex v from the one of which the intersection with U is not larger than that of the other, and remove v from U and add a vertex to U among the vertices of the other I_i which are not yet in U. (It is clear that the number of spanned edges will increase.)

If U satisfies the above assumptions, and if U contains all vertices from $\bigcup_{j=1}^{t} J_j$, then U is an optimal substructure. If U does not contain all vertices from $\bigcup_{j=1}^{t} J_j$, then we can remove $\min\{|I_i \cap U|, |\bigcup_{j=1}^{t} J_j \setminus U|\}$ vertices from the only I_i which is not completely contained in U and add the same number of vertices to U from $\bigcup_{j=1}^{t} J_j \setminus U$. Again it is easy to see that the number of spanned edges cannot decrease.

As a last step, if still $\bigcup_{j=1}^{t} J_j \not\subset U$, we can remove $|\bigcup_{j=1}^{t} J_j \setminus U|$ vertices from any of the I_i's contained in U and add the vertices of $\bigcup_{j=1}^{t} J_j \setminus U$. $\square$

Lemma 4.7 *If there exists an $0 < \epsilon$ such that $n^{-1/5+\epsilon} \le p \le \log^{-8} n$ holds, then almost surely every subgraph of $G(n,p)$ with at least $\frac{n}{\log^7(np)}$ vertices contains an independent (t',k')-comb with $t' = \log(np) - 7\log\log(np)$ and $k' = \frac{1}{p}(2\log(np) - 24\log\log(np))$.*

Proof: Let X denote the number of independent (t,k)-combs in $G(n,p)$, where $t = \log(np)$ and $k = \frac{1}{p}(2\log(np) - 10\log\log(np))$. We establish an upper bound for the probability of the event that $X = 0$ by using the generalized Janson inequality (see e.g. Theorem 2.18 in [7]):

$$\mathbb{P}(X=0) \le \exp\left(-\frac{(\mathbb{E}X)^2}{\sum\sum_{A\cap B\neq\emptyset}\mathbb{E}(I_A I_B)}\right),$$

where I_A stands for the indicator variable that A is a set of non-edges forming an independent (t,k)-comb. (Formally we should apply the inequality for the indicator variable, that the set A of edges form a clique (t,k)-comb in $G(n,1-p)$, the "complement" of $G(n,p)$.)
Using Lemma 4.6, we have

$$\frac{\sum\sum_{A\cap B\neq\emptyset}\mathbb{E}(I_A I_B)}{(\mathbb{E}X)^2} \le \frac{\sum_{m=2}^{tk}\binom{tk}{m}\binom{n-tk}{tk-m}\binom{tk}{k}\binom{(t-1)k}{k}\cdots\binom{k}{k}\binom{k}{\frac{k}{t}}^t(1-p)^{E-L(t,k,m)}}{\binom{n}{tk}\binom{tk}{k}\binom{(t-1)k}{k}\cdots\binom{k}{k}\binom{k}{\frac{k}{t}}^t(1-p)^E}$$
$$= \frac{\sum_{m=2}^{tk}\binom{tk}{m}\binom{n-tk}{tk-m}(1-p)^{E-L(t,k,m)}}{\binom{n}{tk}(1-p)^E} = \frac{\sum_{m=2}^{tk}a_m}{\binom{n}{tk}(1-p)^E},$$

where $a_m=\binom{tk}{m}\binom{n-tk}{tk-m}(1-p)^{E-L(t,k,m)}$.

We want to prove that among the a_i's a_2 is the largest.

Using Lemma 4.6 and the definition of an optimal subcomb, we get that if $2\le m<k$, then

$$b_m=\frac{a_{m+1}}{a_m}=\frac{(tk-m)^2}{(m+1)(n-2tk+m+1)}(1-p)^{-m},$$

and if $m=k+s\frac{t-1}{t}k+l$ for some $0\le s\le t-1$ and $0\le l<\frac{t-1}{t}k$, then

$$b_m=\frac{a_{m+1}}{a_m}=\frac{(tk-m)^2}{(m+1)(n-2tk+m+1)}(1-p)^{-(l+\frac{k}{t})}.$$

Elementary calculations show that in each interval ($2\le m<k$ or $k+s\frac{t-1}{t}k\le m<k+(s+1)\frac{t-1}{t}k$) b_m decreases (with starting value less than 1) and then increases (with ending value larger than 1). This implies that the maximum of the a_m's is attained at a_2 or at some a_m, where $m=k+s\frac{t-1}{t}k$ for some $0\le s\le t$.

Before continuing the proof, let us remark that by the choice of t and k we have

$$\left(\frac{n}{t^3k}\right)^{tk}(1-p)^{(t+1)\frac{k^2}{2}}\ge 1. \tag{1}$$

Indeed,

$$\left(\frac{n}{t^3k}\right)^{tk}=\left(\frac{np}{2\log^4(np)-10\log^3(np)\log\log(np)}\right)^{\frac{1}{p}(2\log(np)-10\log\log(np))\log(np)}=$$

$$\exp\left(\frac{1}{p}\left[2\log^3(np) - 18\log^2(np)\log\log(np) - o(\log^2(np)\log\log(np))\right]\right),$$

while

$$(1-p)^{(t+1)\frac{k^2}{2}} \sim \exp\left(-\frac{1}{p}\left[2\log^3(np) - 20\log^2(np)\log\log(np) + o(\log^2(np)\log\log(np))\right]\right).$$

Let $c_s = a_{k+s\frac{t-1}{t}k}$. Then for $0 \le s \le t-1$

$$d_s = \frac{c_{s+1}}{c_s} = \frac{\binom{tk}{k+(s+1)\frac{t-1}{t}k}\binom{n-tk}{(t-1)k-(s+1)\frac{t-1}{t}k}}{\binom{tk}{k+s\frac{t-1}{t}k}\binom{n-tk}{(t-1)k-s\frac{t-1}{t}k}}(1-p)^{-\binom{k}{2}+\binom{k/t}{2}} \le$$

$$\le (t^2)^{\frac{t-1}{t}k}\frac{[(t-1)k - s\frac{t-1}{t}k]}{[n-(2t-1)k+(s+1)\frac{t-1}{t}k]}\cdot\ldots\cdot\frac{[(t-1)k-(s+1)\frac{t-1}{t}k+1]}{[n-(2t-1)k+s\frac{t-1}{t}k+1]}(1-p)^{-\binom{k}{2}+\binom{k/t}{2}},$$

which, by taking the $-\frac{t^2}{t-1}$th power of (1), is less than 1. This implies that the largest a_m is either a_2 or a_k.
To compare a_2 and a_k, observe that

$$\frac{a_k}{a_2} = \frac{\binom{tk}{k}\binom{n-tk}{(t-1)k}}{\binom{tk}{2}\binom{n-tk}{tk-2}}(1-p)^{-\binom{k}{2}+1} \le (t^2)^{tk}\frac{(tk-2)\ldots((t-1)k+1)}{(n-(t-1)k)\ldots(n-tk+3)}(1-p)^{-\binom{k}{2}+1},$$

which is (again using (1)) less than 1.

So we finally get that a_2 is the largest summand, therefore we have

$$\frac{\sum\sum_{A\cap B\neq\emptyset}\mathbb{E}(I_AI_B)}{(\mathbb{E}X)^2} \le \frac{(tk-1)a_2}{\binom{n}{tk}(1-p)^E} = \frac{(tk-1)\binom{tk}{2}\binom{n-tk}{tk-2}}{\binom{n}{tk}(1-p)} \le$$

$$\le \frac{(tk)^5}{(1-p)(n-2tk)^2} \le \frac{33\log^{10}(np)}{p^5n^2}.$$

By changing n to $\frac{n}{\log^7(np)}$, we get that the probability that there is a subgraph of $G(n,p)$ of size $\frac{n}{\log^7(np)}$ not containing an independent (t',k')-comb with

$$t' = \log\left(\frac{np}{\log^7(np)}\right) = \log(np) - 7\log\log(np),$$

$$k' = \frac{1}{p}\left(2\log(\frac{np}{\log^7(np)}) - 10\log\log(\frac{np}{\log^7(np)})\right) \ge \frac{1}{p}(2\log(np) - 24\log\log(np))$$

is at most

$$\binom{n}{\frac{n}{\log^7(np)}}\exp\left(-\frac{p^5n^2/\log^{14}(np)}{33\log^{10}(\frac{np}{\log^7(np)})}\right) \to 0.$$

$\square$

4.2.2 Proof of Theorem 1.5

It is enough to prove that we can color equitably any graph G having the properties assured by Lemma 4.4, 4.5 and 4.7 and the Corollary 2.3 with at most

$$\frac{np}{(1-\frac{1}{\log(np)-7\log\log(np)})(2\log(np)-24\log\log(np))}$$

colors, where in Lemma 4.5 we set $\gamma=5, \gamma'=4$, $x=\alpha\frac{n}{\log^7(np)}$ and $y=\beta\frac{\log^2(np)}{p}$.

Given such a graph G, we pick sequentially using Lemma 4.7 independent (t,k)-combs (which we will denote by S_i $1\le i\le s$) with $t=\log(np)-7\log\log(np)$, $k=\frac{1}{p}(2\log(np)-24\log\log(np))$ until we are left with at most $\frac{n}{\log^7(np)}$ vertices. Note that the number of combs s is $s=kt=\Theta(\frac{np}{\log^2(np)})$. With the help of Theorem 2.1 (the result of Kostochka et al.) we can partition the vertices left into almost equal independent sets $A_1,...,A_s$, i.e. the number of sets is equal to the number of independent combs. Indeed, the assumption $p\le\log^{-8}n$ assures that the maximum degree among the vertices left is at most $\frac{n}{\log^8(np)}$ which is much smaller than $\frac{n}{\log^7(np)}$, the number of remaining vertices, and Lemma 4.4 assures that the degeneracy number of the subgraph spanned by the remaining vertices is $O\left(\frac{np}{\log^7(np)}\right)$.

Note that the size of the independent sets $A_1,...,A_s$ is $\Theta\left(\frac{1}{p\log^5(np)}\right)$.

We are looking for a matching between the independent sets and the independent combs, such that if A_{i_1} is matched with S_{i_2}, then (with the notation of the definition of a (t,k)-comb) for any j $(1\le j\le t)$ at most $\frac{1}{2\log^2(np)}|I_j\cup J_j|=\frac{k}{2\log^2(np)}=\Theta\left(\frac{1}{p\log(np)}\right)$ vertices in $I_j\cup J_j$ have neighbors in A_{i_1}.

To ensure the existence of a such matching, we have to verify that Hall's condition holds. First note that any A_i can be matched to at least half of the independent combs. Indeed, if not, then in at least half of the independent combs there are at least $\Theta\left(\frac{1}{p\log(np)}\right)$ vertices having at least one neighbor in A_1. So altogether, there are $\frac{s}{2}\cdot\Theta\left(\frac{1}{p\log(np)}\right)=\Theta\left(\frac{n}{\log^3(np)}\right)$ vertices with this property, which contradicts the property assured by Corollary 2.3 which in this case states that there can be at most $O\left(\frac{n}{\log^5(np)}\right)$ such vertices. This gives that Hall's condition holds for every family consisting of at most half of the independent sets A_i $(1\le i\le s)$.

We claim that for any family $\mathcal{A}$ containing at least half of the independent sets $A_1,...,A_s$ and for any (t,k)-comb S, there is an $A\in\mathcal{A}$ such that A can be matched with S (this of course would imply that Hall's condition holds for $\mathcal{A}$, too). Suppose not. Then for any $A\in\mathcal{A}$ there are $\Omega\left(\frac{1}{p\log(np)}\right)$ edges between A and S. Therefore there are $s\cdot\Omega\left(\frac{1}{p\log(np)}\right)=\Omega\left(\frac{n}{\log^3(np)}\right)$ edges between S and $\bigcup_{A\in\mathcal{A}}A$ contradicting Lemma 4.6 according to which there should be at most $\frac{n}{\log^4(np)}$ such edges.

Having realized a matching with the property above, we would like to proceed as follows: for every pair of an independent set A and an independent (t,k)-comb S that are matched in our matching, we would like to partition $A \cup S$ into $t+1$ almost equal independent sets. Since all (t,k)-combs have the same size and the A_i's are almost equal, for any two matched pairs the size of $A \cup S$ may differ by at most 1, so the resulting independent sets will be almost equal.

Let us suppose that in our matching an independent set A is matched with an independent (t,k)-comb $S = \bigcup_{j=1}^{t} I_j \cup J_j$. By definition $|S| = tk$, therefore the independent sets we are looking for should be of size $\frac{tk+|A|}{t+1}$. We would like to have sets $J_i' \subset J_i$ $(1 \le i \le t)$ with $|J_1'| = \cdots = |J_t'|$ such that $\bigcup_{i=1}^{t} J_i' \cup A$ is independent and is of size $\frac{tk+|A|}{t+1}$. Therefore we need $|J_i'|$ to be

$$\frac{\frac{tk+|A|}{t+1} - |A|}{t} = \frac{k-|A|}{t+1}.$$

By the definition of our matching there are at most $\frac{k}{2\log^2(np)} \le \frac{k}{1.8t(t+1)}$ vertices in $I_i \cup J_i$ that have at least one neighbor in A. Even if all these vertices were in J_i, there would be $|J_i| - \frac{k}{1.8t(t+1)} = \frac{1.8(t+1)k-k}{1.8t(t+1)} \ge \frac{k}{t+1}$ vertices out of which we could form J_i'. $\square$

5 The order of the equitable chromatic number

In this section we will prove Theorem 1.6 using very similar ideas to those of the previous section. The main difference is that we are not able to prove the existence of large independent combs and therefore we have to use independent sets that we obtain by Łuczak's proof [15] of the chromatic number for sparse random graphs. In the next subsection we gather some technical lemmas corresponding to the ones of the dense case in Subsection 4.2.1 and then in Section 5.2 we prove Theorem 1.6.

5.1 Properties of random graphs

The Corollary from the Introduction is valid only if $p \ge \frac{\log^2 n}{n}$, therefore we need a similar statement that can be used in the sparse case as well.

Lemma 5.1 *For every constant $c > 0$ there exists a constant $C > 0$ so that the random graph $G(n,p)$ with $\frac{C}{n} \le p = p(n) = o(1)$ has almost surely the following property:*

For every set J of size $\frac{c}{p\log^{0.5}(np)}$ the number of vertices that are not adjacent to any vertex in J is at least $\frac{2}{3}n$, i.e. the neighborhood of J is of size at most $\frac{n}{3}$.

Proof: For a fixed set J of size $\frac{c}{p\log^{0.5}(np)}$ the expected number of such vertices is

$$(n-|J|)(1-p)^{\frac{c}{p\log^{0.5}(np)}} \sim (n-|J|)e^{-\frac{c}{\log^{0.5}(np)}} \ge (1-\frac{1}{100})n,$$

if n is large enough. So by the Chernoff bound we get that the probability that there are less than $\frac{2}{3}n$ such vertices for this fixed J is at most $\exp(-\frac{1}{32}n)$. Therefore, taking the union bound, the probability that there is a set contradicting the statement of the claim is at most

$$\binom{n}{\frac{c}{p\log^{0.5}(np)}}\exp\left(-\frac{1}{32}n\right) \leq (c^{-1}np\log^{0.5}(np))^{\frac{c}{p\log^{0.5}(np)}}\exp\left(-\frac{1}{32}n\right) = o(1)\,. \quad \square$$

Lemma 5.2 *There exists a constant C such that if $\frac{C}{n} \leq p \leq \log^{-8} n$ then almost surely the vertex set of $G(n,p)$ can be covered by pairwise disjoint independent sets larger than $\frac{1}{p}(2\log(np) - 38\log\log(np))$ and all of the same size with the exception of at most $4n\log^{-1.5}(np)$ vertices.*

Proof: Łuczak (using Matula's expose-and-merge method [16]) showed in [15] (see also Lemma 7.17 and Lemma 7.18 in Chapter 7 of [7]) that with probability at least $1 - o(\log^{-1}(np))$, one can pick pairwise disjoint sets of vertices $I_1, I_2, ..., I_t$ each of size $\frac{1}{p}(2\log(np) - 37\log\log(np))$ such that $|V(G) \setminus \bigcup_{i=1}^{t} I_i| \leq n\log^{-3}(np)$ and the total number of edges contained in some I_i $(1 \leq i \leq t)$ is at most $n\log^{-3}(np)$. It follows that $t \leq \frac{np}{\log(np)}$.

Let $\mathcal{A}$ be the set of those I_i's which contain more than $\frac{1}{p\log^{0.5}(np)}$ edges. Then $|\mathcal{A}| \leq np\log^{-2.5}(np)$ and so $|\bigcup_{I\in\mathcal{A}} I| \leq 2n\log^{-1.5}(np)$. The number of the remaining I_i's that do not contain more than $\frac{1}{p\log^{0.5}(np)}$ edges can be bounded by the total number of I_i's which is $t \leq np\log^{-1}(np)$. Therefore by deleting one vertex from each edge and possibly some additional vertices, we can get independent sets (all of the same size, which is larger than $\frac{1}{p}(2\log(np) - 38\log\log(np))$ if n is large enough) with removing at most $n\log^{-1.5}(np)$ vertices. So we covered the graph by independent sets of the same size with the exception of at most $2n\log^{-1.5}(np) + n\log^{-1.5}(np) + n\log^{-3}(np) \leq 4n\log^{-1.5}(np)$ vertices. $\square$

5.2 Proof of Theorem 1.6

The statements of Lemmas 4.4, 4.5, 5.1 and 5.2 hold almost surely, so it is enough to prove the theorem for graphs having properties assured by these lemmas, and this time in Lemma 4.5 we set $\gamma = 0.5, \gamma' = 0.3$, $x = \alpha\frac{n}{\log^{1.5}(np)}, y = \beta\frac{\log(np)}{p}$. By Lemma 5.2 we can cover all the vertices of such a graph by independent sets each of the same size, $\frac{1}{p}(2\log(np) - 38\log\log(np))$, with the exception of at most $4n\log^{-1.5}(np)$ vertices. Let us denote these independent sets by $\mathcal{I}$. By Lemma 4.4, the degeneracy number of the graph induced by the vertices not covered is $O(np\log^{-1.5}(np))$ (and the maximum degree is at most the maximum degree of the original graph, which is at most $\max\{1.01np, \log^2 n\} \ll n\log^{-1.5}(np)$ by the assumption on p), so by Theorem 2.1 (Kostochka et al. [11]), we can color them equitably with as many colors as the number of independent sets in $\mathcal{I}$. In such a way we get independent sets $J_1, J_2, ..., J_{|\mathcal{I}|}$, such that their size may differ by at most one and their size is at most $\frac{c}{p \log^{0.5}(np)}$ for some constant $c > 0$.

Our aim is to find a matching between the independent sets in $\mathcal{I}$ and the J_j's (let us denote the family containing them by $\mathcal{J}$) in such a way that whenever a J_j is matched with some I_i, then at least half of the vertices of I_i can be added to J_j preserving the stability of J_j. To assure the existence of such a matching, we have to verify that Hall's condition holds.

As a consequence of Lemma 5.1 we get that any $J \in \mathcal{J}$ can be matched with at least $\frac{1}{5}|\mathcal{I}|$ I's from $\mathcal{I}$. Indeed, if not then in at least $\frac{4}{5}|\mathcal{I}|$ I's at least half of the vertices cannot be added to J, so there are at least $\frac{2}{5}n$ vertices with this property - contradicting the lemma.

Now we are ready to check that Hall's condition holds for $\mathcal{I}$ and $\mathcal{J}$. If a subfamily $\mathcal{J}' \subset \mathcal{J}$ has size less than $\frac{1}{5}|\mathcal{J}|$ then by the previous paragraph (and considering only one set from $\mathcal{J}'$) the number of sets in $\mathcal{I}$ that can be matched with some set J in $\mathcal{J}'$ is at least $\frac{1}{5}|\mathcal{I}| \geq |\mathcal{J}'|$. Otherwise $|\mathcal{J}'| \geq \frac{1}{5}|\mathcal{J}|$, and we claim that all sets in $\mathcal{I}$ are can be matched with some set J in $\mathcal{J}'$. We only have to check this for $|\mathcal{J}'| = \frac{1}{5}|\mathcal{J}|$. In this case $|\bigcup_{J \in \mathcal{J}'} J| \leq \beta \frac{n}{\log^{1.5}(np)}$ and for any independent set $I \in \mathcal{I}$ we have $|I| = \gamma \frac{\log(np)}{p}$. Therefore by Lemma 4.5 the number of edges between I and $\bigcup_{J \in \mathcal{J}'} J$ is at most $\frac{n}{\log^{0.3}(np)}$. If no $J \in \mathcal{J}'$ can be matched with I, then there are $\Omega\left(\frac{\log(np)}{p}\right)$ edges between I and any $J \in \mathcal{J}'$, so there are $\Omega(n)$ edges between I and $\bigcup_{J \in \mathcal{J}'} J$ – a contradiction.

For any two matched pairs (I_1, J_1) and (I_2, J_2), the size of $I_1 \cup J_1$ and $I_2 \cup J_2$ differ by at most 1, so if we can split all matched pairs into two almost equal independent sets, then we partition the vertex set into almost equal independent sets each of size at least $\frac{1}{p}(\log(np) - 19 \log\log(np))$. By the assumption on the matching we may add at least half of the vertices of I to J, while to split $I \cup J$ into almost equal parts, we need only $\frac{|I \cup J|}{2} - |J|$ vertices. □

Acknowledgement. We would like to thank the anonymous referees for their careful reading and valuable comments.

References

[1] Alon, N.; Capalbo, M.; Kohayakawa, Y.; Rödl, V; Ruciński, A.; Szemerédi, E., *Universality and Tolerance*, Proc. 41^{st} IEEE FOCS, IEEE (2000), 14–21.

[2] Alon, N.; Füredi, Z. *Spanning subgraphs of random graphs*, Graphs and Combinatorics 8 (1992), 91–94.

[3] Alon, N.; Krivelevich, M.; Sudakov, B. *List coloring of random and pseudo-random graphs*, Combinatorica 19 (1999), no. 4, 453–472.

[4] Bollobás, B. *The chromatic number of random graphs*, Combinatorica 8 (1988), no. 1, 49–55.

[5] Hajnal, A.; Szemerédi, E. *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), pp. 601–623. North-Holland, Amsterdam, 1970.

[6] Janson, S. *Large deviations for sums of partly dependent random variables*, Random Structures and Algorithms 24 (2004), no. 3, 234–248.

[7] Janson, S.; Łuczak, T.; Ruciński, A. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[8] Janson, S.; Ruciński, A. *The infamous upper tail*, Random Structures and Algorithms 20 (2002), 317–342.

[9] Kostochka, A. V. *Equitable colorings of outerplanar graphs*, Discrete Math. 258 (2002), no. 1-3, 373–377.

[10] Kostochka, A.V.; Kierstead H. *A Short Proof of the Hajnal-Szemerédi Theorem on Equitable Coloring*, Combinatorics, Probability and Computing 17 (2008), no.2, 265–270.

[11] Kostochka, A. V.; Nakprasit, K.; Pemmaraju, S. V. *On equitable coloring of d-degenerate graphs*, SIAM J. Discrete Math. 19 (2005), no. 1, 83–95.

[12] Kostochka, A. V.; Pelsmajer, M. J.; West, D. B. *A list analogue of equitable coloring*, J. Graph Theory 44 (2003), no. 3, 166–177.

[13] Krivelevich, M.; Sudakov, B.; Vu, V.H.; Wormald, N.C. *On the probability of independent sets in random graphs*, Random Structures and Algorithms 22 (2003), 1–14.

[14] Lih, K.-W. *The equitable coloring of graphs*, *Handbook of combinatorial optimization*, Vol. 3, 543–566, Kluwer Acad. Publ., Boston, MA, 1998.

[15] Łuczak, T. *The chromatic number of random graphs*, Combinatorica 11 (1991), no. 1, 45–54.

[16] Matula, D. W. *Expose-and-merge exploration and the chromatic number of a random graph*, Combinatorica 7 (1987), no. 3, 275–284.

[17] Matula, D. W.; Kučera, L. *An expose-and-merge algorithm and the chromatic number of a random graph*, *Random graphs '87 (Poznań, 1987)*, 175–187, Wiley, Chichester, 1990.

[18] Pemmaraju, S. V. *Equitable coloring extends Chernoff-Hoeffding bounds*, Approximation, randomization, and combinatorial optimization (Berkeley, CA, 2001), 285–296, Lecture Notes in Comput. Sci., 2129, Springer, Berlin, 2001.

[19] Pemmaraju, S. V.; Nakprasit, K.; Kostochka, A. V. *Equitable colorings with constant number of colors*, Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003), 458–459, ACM, New York, 2003.