On the probability of independent sets in random graphs

Michael Krivelevich* Benny Sudakov † Van H. Vu ‡ Nicholas C. Wormald §

Abstract

Let k be the asymptotic value of the independence number of the random graph $G(n,p)$. We prove that if the edge probability $p(n)$ satisfies $p(n) \gg n^{-2/5} \ln^{6/5} n$ then the probability that $G(n,p)$ does not contain an independent set of size $k-c$, for some absolute constant $c>0$, is at most $\exp\{-cn^2/(k^4p)\}$. We also show that the obtained exponent is tight up to logarithmic factors, and apply our result to obtain new bounds on the choice number of random graphs. We also discuss a general setting where our approach can be applied to provide an exponential bound on the probability of certain events in product probability spaces.

1 Introduction

Let $G(n,p)$ denote as usual the probability space whose points are graphs on n labeled vertices $\{1,\ldots,n\}$, where each pair of vertices forms an edge randomly and independently with probability $p=p(n)$. We say that the random graph $G(n,p)$ possesses a graph property A *asymptotically almost surely*, or a.a.s. for short, if the probability that $G(n,p)$ satisfies A tends to 1 as the number of vertices n tends to infinity.

Define the following quantity:

$$k^* = \max\left\{k \in \mathbb{N} : \binom{n}{k}(1-p)^{\binom{k}{2}} \geq 1\right\} .$$

*Department of Mathematics, 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-Israel BSF Grant, by a grant from the Israel Science Foundation and by a Bergmann Memorial Grant.

†Department of Mathematics, Princeton University, Princeton, NJ 08540, USA and Institute for Advanced Study, Princeton, NJ 08540, USA. Email: bsudakov@math.princeton.edu. Research supported in part by NSF grants DMS-0106589, CCR-9987845 and by the State of New Jersey.

‡Department of Mathematics, UCSD, La Jolla, CA 92093. Email: vanvu@ucsd.edu, Web: http://www.math.ucsd.edu/˜vanvu. Research supported in part by Grant RB091G-Vu from UCSD, a Sloan fellowship and an NSF Grant

§Department of Mathematics and Statistics, University of Melbourne VIC 3010, Australia. Email: nick@ms.unimelb.edu.au. Web: http://www.ms.unimelb.edu.au/˜nick. Research supported by the Australian Research Council.

In words, k^* is the maximum integer k for which the expectation of the number of independent sets of size k in $G(n,p)$ is still at least 1.

It has been known for a long time ([6], [13]) that for large enough $p = p(n)$ (say, for $p(n) \geq n^{-\epsilon}$ for small enough constant $\epsilon > 0$) a.a.s. in $G(n,p)$ the independence number of G is asymptotically equal to k^*. In fact, using the so-called second moment method, one can prove that under the above assumptions the independence number of $G(n,p)$ is concentrated a.a.s. in two consecutive values, one of them being k^*. Now let us pick an integer k_0 slightly less than k^* (we will be more precise later) and ask the following: what is the probability that the random graph $G(n,p)$ does not contain an independent set of size k_0? This seemingly somewhat artificial question turns out to be of extreme importance for many deep problems in the theory of random graphs. An exponential estimate of the above probability provided a crucial ingredient in the seminal breakthrough of Bollobás [5], establishing the asymptotic value of the chromatic number of random graphs. Later, this problem became a fruitful playground for comparing the strength of various large deviation methods like martingales and the Janson and Talagrand inequalities. The reader may consult the survey paper of Spencer [15] for further details. More recent applications can be found in [10] and [11].

The main objective of the current paper is to provide a new, stronger estimate on the probability defined above. This estimate is obtained by combining hypergraph arguments, somewhat similar to those used by Bollobás in [5], and recent martingale results. We will prove that in a certain range of the edge probability $p(n)$, the probability that $G(n,p)$ does not contain an independent set of size k_0, with $k^* - k_0 \geq c$ for some absolute constant $c > 0$, is at most $\exp\{\Omega(-n^2/k_0^4 p)\}$. The exact formulation of this result and its proof are presented in Section 2. Somewhat surprisingly it turns out that the exponent in the estimate cited above is optimal up to a logarithmic factor. The proof of this is presented in Section 3. Then in Section 4 we demonstrate how our new bound can be used to extend the scope of the results of [10] and [11] about the asymptotic value of the choice number of random graphs to smaller values of $p(n)$. Our argument used to get an exponential bound for the probability defined above can in fact be viewed as an example of a general approach, for obtaining exponential bounds for probabilities of certain events in product probability spaces. This general approach, discussed in Section 5, can sometimes compete successfully with the well known Janson inequality. Section 6, the final section of the paper, is devoted to concluding remarks.

Throughout the paper we will use the standard asymptotic notation. In particular, $a(n) \ll b(n)$ means $a(n) = o(b(n))$, $\Omega(a(n))$ denotes a function $b(n)$ such that for some $C > 0$, for n sufficiently large $b(n) > Ca(n)$, and $\Theta(a(n))$ denotes a function which is both $O(a(n))$ and $\Omega(a(n))$. Also, $f(n) \sim g(n)$ means $\lim_{n\to\infty} f(n)/g(n) = 1$. For the sake of clarity of presentation we will systematically omit floor and ceiling signs at places where the choice of which is used does not affect the argument.

2 Independent sets in random graphs

Let $k_0 = k_0(n,p)$ be defined by

$$k_0 = \max\left\{k : \binom{n}{k}(1-p)^{\binom{k}{2}} \geq n^4\right\}. \tag{1}$$

One can show easily that k_0 satisfies $k_0 \sim 2\log_b(np)$ with $b = 1/(1-p)$. Also, it follows from known results on the asymptotic value of the independence number of $G(n,p)$ (see, e.g., [9]) that a.s. the difference between k_0 and the independence number of $G(n,p)$ is bounded by an absolute constant, as long as $p(n) \geq n^{-1/2+\epsilon}$ for a positive $\epsilon > 0$.

Theorem 2.1 *Let $p(n)$ satisfy $n^{-2/5}\ln^{6/5} n \ll p(n) \leq 1-\epsilon$ for an absolute constant $0 < \epsilon < 1$. Then*

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

Proof. In case p is a constant, the result of the theorem follows easily from Janson's inequality (see, e.g. [4], Chapter 10.3). Thus in the rest of the proof we will assume that $p = o(1)$.

Given a graph G on n vertices and an integer k_0, a collection $\mathcal{C}$ of pairs of vertices of G is called a *cover* if every independent set of size k_0 in G contains a pair from $\mathcal{C}$. We set $X = X(G)$ to be the minimum size of a cover in G. For the reader familiar with hypergraph terminology we can define $X(G)$ as follows. Given G, define a hypergraph $H = H(G)$ whose vertices are pairs of vertices of G and whose edges are formed by taking all pairs of vertices in every independent set of G of size k_0. Thus H is a $\binom{k_0}{2}$-uniform hypergraph on $\binom{n}{2}$ vertices, whose number of edges is equal to the number of independent sets of size k_0 in G. Then a cover in G corresponds to a vertex cover of the hypergraph H, and $X(G)$ is equal to the covering number of H.

When G is distributed according to $G(n,p)$, the quantity $X(G)$ becomes a random variable. Our aim will be first to estimate from below the expectation of X and then to show that X is concentrated. It may be noted that we use pairs of vertices in the definition of a cover, rather than single vertices, in order to achieve a better concentration in Lemma 2.6 below, whilst larger sets of vertices would not be suitable for Proposition 2.5.

Lemma 2.2 $E[X] = \Omega\left(\frac{n^2}{k_0^2}\right)$.

Proof. Let Y be a random variable counting the number of independent sets of size k_0 in $G(n,p)$. We denote by μ the expectation of Y. Then clearly

$$\mu = E[Y] = \binom{n}{k_0}(1-p)^{\binom{k_0}{2}} \geq n^4$$

by the definition of k_0.

For a pair $u, v \in V(G)$, let $Z_{u,v}$ be a random variable counting the number of k_0-subsets of V that contain u and v and span no edges except possibly the edge (u, v). (The edge (u, v) is permitted for ease of later analysis.) Thus, if $(u, v) \notin E(G)$, then $Z_{u,v}$ is equal to the number of independent sets of size k_0 that contain both u and v. If $\mu_0 = E[Z_{u,v}]$, then

$$\mu_0 = \binom{n-2}{k_0-2}(1-p)^{\binom{k_0}{2}-1} .$$

It is easy to see that, by definition, $\mu_0/\mu = \Theta(k_0^2/n^2)$. Next, we set

$$Z^+_{u,v} = \begin{cases} Z_{u,v}, & Z_{u,v} > 2\mu_0, \\ 0, & \text{otherwise} . \end{cases}$$

We also define $Z^+ = \sum_{u,v} Z^+_{u,v}$.

To finish the proof of the lemma, we use three propositions.

Proposition 2.3 *For every graph G, $X \geq \frac{Y-Z^+}{2\mu_0}$.*

Proof. Let $\mathcal{C}$ be an optimal cover in G, $|\mathcal{C}| = X$. Set $\mathcal{C}_0$ to be the set of pairs of vertices from $\mathcal{C}$ covering more than $2\mu_0$ independent sets of size k_0, and also set $\mathcal{C}_1 = \mathcal{C} \setminus \mathcal{C}_0$. Each pair u, v covers $Z_{u,v}$ independent sets of size k_0. Hence the set $\mathcal{C}_0$ covers at most $\sum_{\{u,v\}\in\mathcal{C}_0} Z_{u,v} \leq Z^+$ such independent sets. Then it follows that at least $Y - Z^+$ independent sets are covered by $\mathcal{C}_1$ only. As every pair in $\mathcal{C}_1$ participates in at most $2\mu_0$ independent sets of size k_0, we get $|\mathcal{C}_1| \geq (Y - Z^+)/(2\mu_0)$. Therefore $X \geq |\mathcal{C}_1| \geq (Y - Z^+)/(2\mu_0)$, as required. ∎

Proposition 2.4 *For each $u, v \in V(G)$ and all i with $2\mu_0 \leq i \leq \binom{n-2}{k_0-2}$, $Pr[Z_{u,v} \geq i] = O\left(\frac{k_0^4 p \mu_0^2}{n^2}\frac{1}{(i-\mu_0)^2}\right)$.*

Proof. Fix a pair $u, v \in V(G)$ and let U be the set of vertices in $V \setminus \{u, v\}$ not adjacent to either u or v. By definition the size of U is a binomially distributed random variable with parameters $n-2$ and $(1-p)^2$. Therefore by applying standard estimates for binomial distributions (see, e.g. [4, Theorems A.11 and A.13]) to the size of $V - U$ we obtain that

$$Pr\left[\left||U| - (1-p)^2 n\right| > \frac{np}{\ln^2 n}\right] < e^{-\Omega\left(\frac{np}{\ln^4 n}\right)}.$$

Denote by z_1 the value of the random variable $Z_{u,v}$ conditional on the particular set U of size $n_1, (1-p)^2 n - np/\ln^2 n \leq n_1 \leq (1-p)^2 n + np/\ln^2 n$. Let $\mathcal{S}$ be the family of all subsets of U of size $k_0 - 2$. For every $S \in \mathcal{S}$ let Z_S be the indicator random variable taking value 1 when S spans no edges of G, and value 0 otherwise. Clearly, $z_1 = \sum_{S\in\mathcal{S}} Z_S$. By definition, the expected value and the variance of z_1 are equal to

$$\mu_1 = \binom{n_1}{k_0-2}(1-p)^{\binom{k_0-2}{2}},$$

$$\sigma_1^2 = VAR[\sum_{S\in\mathcal{S}} Z_S] = \sum_{S\in\mathcal{S}} VAR[Z_S] + \sum_{S\neq S'\in\mathcal{S}} COV[Z_S, Z_{S'}] \ .$$

Clearly, if $S, S' \in \mathcal{S}$ have no common pairs of vertices, then the events $Z_S = 1$ and $Z_{S'} = 1$ are independent, implying $COV[Z_S, Z_{S'}] = 0$. Therefore we need to sum only over those pairs $S, S' \in \mathcal{S}$ for which $2 \le |S \cap S'| \le k_0 - 3$. This implies:

$$\begin{aligned}
\sigma_1^2 &\le E[z_1] + \binom{n_1}{k_0-2}\sum_{i=2}^{k_0-3}\binom{k_0-2}{i}\binom{n_1-k_0+2}{k_0-i-2}\left[(1-p)^{2\binom{k_0-2}{2}-\binom{i}{2}} - (1-p)^{2\binom{k_0-2}{2}}\right] \\
&= \mu_1 + \mu_1^2 \sum_{i=2}^{k_0-3} \frac{\binom{k_0-2}{i}\binom{n_1-k_0+2}{k_0-i-2}}{\binom{n_1}{k_0-2}}\left(\frac{1}{(1-p)^{\binom{i}{2}}} - 1\right) \ .
\end{aligned}$$

Denote the i-th summand of the last sum by $g(i)$, $2 \le i \le k-3$. One can check (see [4], Chapter 4.5 for a similar computation) that the dominating term is

$$g(2) = \frac{\binom{k_0-2}{2}\binom{n_1-k_0+2}{k_0-4}}{\binom{n_1}{k_0-2}}\left(\frac{1}{1-p} - 1\right) = O\left(\frac{k_0^4 p}{n_1^2}\right) .$$

Hence $\sigma_1^2 = O\left(\frac{k_0^4 p}{n_1^2}\mu_1^2\right)$. Next by applying Chebyshev's inequality we obtain that

$$Pr[z_1 \ge i] \le Pr[|z - \mu_1| \ge i - \mu_1] \le \frac{\sigma_1^2}{(i-\mu_1)^2}.$$

Using the fact that $n_1 = (1-p)^2 n + \Theta(np/\ln^2 n)$ and $k_0 = \Theta(\ln n/p)$ we obtain

$$\begin{aligned}
\frac{\mu_1}{\mu_0} &= \frac{\binom{n_1}{k_0-2}(1-p)^{\binom{k_0-2}{2}}}{\binom{n}{k_0-2}(1-p)^{\binom{k_0}{2}-1}} = \frac{n_1\cdots(n_1-k_0+3)}{n\cdots(n-k_0+3)}\,\frac{1}{(1-p)^{2k_0-4}} = \frac{(1+o(1))}{(1-p)^{2k_0-4}}\left(\frac{n_1}{n}\right)^{k_0-2} \\
&= (1+o(1))\left(\frac{n_1}{(1-p)^2 n}\right)^{k_0-2} = (1+o(1))\left(1+\Theta\left(\frac{p}{\ln^2 n}\right)\right)^{k_0-2} = 1+o(1).
\end{aligned}$$

Now to finish the proof note that

$$\begin{aligned}
Pr\Big[Z_{u,v} \ge i\Big] &\le Pr\left[Z_{u,v} \ge i \,\Big|\, |U| = (1-p)^2 n + \Theta\left(\frac{np}{\ln^2 n}\right)\right] + Pr\left[\Big||U| - (1-p)^2 n\Big| > \Theta\left(\frac{np}{\ln^2 n}\right)\right] \\
&\le \frac{\sigma_1^2}{(i-\mu_1)^2} + e^{-\Omega\left(\frac{np}{\ln^4 n}\right)} \\
&= O\left(\frac{k_0^4 p \mu_0^2}{n^2}\,\frac{1}{(i-\mu_0)^2}\right).
\end{aligned}$$

Here we used the estimate for σ_1^2 and the facts that $n_1 = (1+o(1))n$ and that the maximal possible value of i^2 is $\binom{n}{k_0-2}^2 = e^{o\left(np/\ln^4 n\right)}$. ∎

Proposition 2.5 $E[Z^+] = o(\mu)$.

Proof. We will use the following easily proven statement: If X is an integer random variable with finitely many values, then for every integer s,

$$\sum_{i>s} iPr[X=i] = sPr[X>s] + \sum_{i>s} Pr[X \geq i]\,. \tag{2}$$

For every pair (u,v) it now follows from the definition of $Z_{u,v}$ and Proposition 2.4 that

$$\begin{aligned} E[Z^+_{u,v}] &= \sum_{i>2\mu_0} iPr[Z_{u,v}=i] = 2\mu_0\, Pr[Z_{u,v} > 2\mu_0] + \sum_{i>2\mu_0} Pr[Z_{u,v} \geq i] \\ &= 2\mu_0\, O\left(\frac{k_0^4 p\mu_0^2}{n^2}\frac{1}{(2\mu_0-\mu_0)^2}\right) + \sum_{i>2\mu_0} O\left(\frac{k_0^4 p\mu_0^2}{n^2}\frac{1}{(i-\mu_0)^2}\right) \\ &= O\left(\frac{k_0^4 p\mu_0}{n^2}\right)\,. \end{aligned}$$

Then we derive from the definition of Z^+ and the linearity of expectation that $E[Z^+] = \sum_{u,v} E[Z^+_{u,v}] = O\left(n^2 \frac{k_0^4 p\mu_0}{n^2}\right) = O\left(\frac{k_0^6 p}{n^2}\mu\right)$. Now applying our assumption on the edge probability $p(n)$, we obtain the desired estimate. ∎

We can now complete the proof of Lemma 2.2. Recall that by Proposition 2.3, $X \geq (Y - Z^+)/(2\mu_0)$. Therefore, taking into account Proposition 2.5 and the definitions of μ and μ_0, we derive:

$$E[X] \geq \frac{E[Y]-E[Z^+]}{2\mu_0} = \frac{(1-o(1))\mu}{2\mu_0} = \Omega\left(\frac{n^2}{k_0^2}\right)\,,$$

as required. ∎

Lemma 2.6 *For every $n^2p > t > 0$, $Pr[X \leq E[X]-t] \leq e^{-\frac{t^2}{2n^2p}}$.*

Proof. Notice that X is an edge Lipschitz random variable, i.e. changing a graph G in one pair of vertices changes the value of X by at most one. This is due to the fact that if a pair (u,v) becomes a non-edge, then in the worst case it can be added to an optimal cover to produce a new cover. When applying the edge exposure martingale to X, the maximal variance in the martingale is $\binom{n}{2}p(1-p) \leq n^2p/2$. Therefore, the desired estimate of the lower tail of X follows from known results on graph martingales (see, e.g. [4], Th. 7.4.3). ∎

We are now in position to finish the proof of Theorem 2.1. Clearly, a graph G contains an independent set of size k_0 if and only if $X > 0$. From Lemmas 2.2 and 2.6 we obtain:

$$Pr[\alpha(G) < k_0] = Pr[X=0] = Pr\Big[X \leq E[X]-E[X]\Big] \leq e^{-\frac{(E[X])^2}{2n^2p}} = e^{-\Omega\left(\frac{n^2}{k_0^4 p}\right)}\,.$$

3 On the tightness of Theorem 2.1

In this section we show that the exponent in the bound of Theorem 2.1 is tight up to logarithmic factors.

Theorem 3.1 *Let $p(n) \le 1-\epsilon$ for an absolute constant $\epsilon > 0$. Define $k_0 = k_0(n,p)$ by (1). Then*

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

Proof. Set

$$\begin{aligned} T &= \frac{7n^2 \ln n}{k_0^2}, \\ M_0 &= \binom{n}{2} p + T . \end{aligned}$$

Our first goal will be to estimate from above the probability $Pr[\alpha(G) \ge k_0 || E(G)| = M]$, where $M \ge M_0$. Since the distribution of $G = G(n,p)$ conditional on the event $|E(G)| = M$ is identical to the distribution of graph with M random edges, we have

$$\begin{aligned} Pr[\alpha(G) \ge k_0 || E(G)| = M] &\le \frac{\binom{n}{k_0}\binom{\binom{n}{2}-\binom{k_0}{2}}{M}}{\binom{\binom{n}{2}}{M}} \le \binom{n}{k_0}\left(1-\frac{M}{\binom{n}{2}}\right)^{\binom{k_0}{2}} \\ &\le \binom{n}{k_0}\left(1-p-\frac{T}{\binom{n}{2}}\right)^{\binom{k_0}{2}} \\ &= \binom{n}{k_0}(1-p)^{\binom{k_0}{2}}\left(\frac{1-p-\frac{T}{\binom{n}{2}}}{1-p}\right)^{\binom{k_0}{2}} \\ &\le \binom{n}{k_0}(1-p)^{\binom{k_0}{2}}\left(1-\frac{T}{\binom{n}{2}}\right)^{\binom{k_0}{2}} , \end{aligned}$$

where we used the estimate $\binom{a-x}{b}\binom{a}{b}^{-1} \le \left(\frac{a-b}{a}\right)^x$ in the second inequality above. Returning to the definition (1) of k_0, we can notice that $\binom{n}{k_0}(1-p)^{\binom{k_0}{2}} \le n^6$. Therefore:

$$Pr[\alpha(G) \ge k_0 || E(G)| = M] \le n^6 e^{-\frac{T\binom{k_0}{2}}{\binom{n}{2}}} .$$

Substituting the definition of T, we get $Pr[\alpha(G) \ge k_0 || E(G)| = M] = n^6 n^{-7+o(1)} = o(1)$. As this estimate holds for every $M \ge M_0$, it follows that

$$Pr[\alpha(G) < k_0 || E(G)| \ge M_0] = 1 - o(1) .$$

Also, due to the standard estimates on the tails of a binomial random variable we have $Pr[|E(G)| \geq M_0] = e^{-\Theta\left(\frac{T^2}{n^2p}\right)}$. Combining the two estimates above and substituting the value of T, we thus obtain:

$$\begin{aligned} Pr[\alpha(G) < k_0] &\geq Pr[|E(G)| \geq M_0] Pr[\alpha(G) < k_0 || E(G)| \geq M_0] \geq (1 - o(1)) e^{-\Theta\left(\frac{T^2}{n^2p}\right)} \\ &= e^{-\Theta\left(\frac{n^2 \ln^2 n}{k_0^4 p}\right)} . \quad \blacksquare \end{aligned}$$

4 Applications to choosability of random graphs

The *choice number* $ch(G)$ of a graph G is the minimum integer k such that for every assignment of a set $S(v)$ of k colors to every vertex v of G, there is a proper coloring of G that assigns to each vertex v a color from $S(v)$. The choice number was introduced by Vizing [16] and independently by Erdős, Rubin and Taylor [8] and the study of this parameter received a considerable amount of attention in recent years.

In this section we consider the asymptotic behavior of the choice number of random graphs. In their original paper, Erdős, Rubin and Taylor [8] conjectured that almost surely $ch(G(n, 1/2)) = o(n)$. This was proved by Alon in [1]. Kahn proved (see [2]) that almost surely

$$ch(G(n, 1/2)) = (1 + o(1))\chi(G(n, 1/2)) = (1 + o(1))n/(2 \log_2 n).$$

His result was extended by Krivelevich [10], who determined the asymptotic value of $ch(G(n,p))$ when $p(n) \gg n^{-1/4}$. At the same time Alon, Krivelevich and Sudakov [3] and independently Vu [17] showed that for all values of the edge probability p almost surely the choice number of $G(n,p)$ has order of magnitude $\Theta(np/\ln(np))$ (see also [12] for better constants). Here we combine Theorem 2.1 and the ideas from [10] to prove the following result.

Theorem 4.1 *Let $0 < \epsilon < 1/3$ be a constant. If the edge probability $p(n)$ satisfies $n^{-1/3+\epsilon} \leq p(n) \leq 3/4$ then almost surely*

$$ch(G(n,p)) = (1 + o(1))\chi(G(n,p)) = (1 + o(1))\frac{n}{2 \log_b(np)},$$

where $b = 1/(1 - p)$.

Sketch of the proof. First note that a.a.s. every subset of vertices of $G(n,p)$ of size at least $m \sim n/\ln^4 n$ contains an independent set of size $k_0 = (1 - o(1))2 \log_b(np)$, where $b = 1/(1 - p)$. Indeed, from Theorem 2.1, the fact that $k_0 = O(\ln n/p)$ and the assumptions on the value of p, it follows that the probability that there exists a set of m vertices that does not span an independent set of size k_0 is at most

$$\binom{n}{m} e^{-\Omega\left(\frac{m^2}{k_0^4 p}\right)} \leq 2^n e^{-n^{1+3\epsilon - o(1)}} = o(1).$$

Next we sketch how, given a typical graph G in $G(n,p)$ and a family of lists $S_1, \dots, S_n$ each of size $n/k_0 + 3np/\ln^2 n$, we can color G from these lists. Our coloring procedure consists of two phases. As long as there exists a color c which appears in the lists of at least $n/\ln^4 n$ of yet uncolored vertices, we do the following. Denote by V_0 the set of those uncolored vertices whose color list contains c. Then $|V_0| \geq n/\ln^4 n$. Then V_0 spans an independent set I of size $|I| = k_0$. We color all vertices of I by c, discard I and delete c from all lists. The total number of deleted colors from each list $S(v)$ during the first phase cannot exceed n/k_0, as each time we remove a subset of size k_0.

Let U denote the set of all vertices that are still uncolored after the first phase has been completed. The lists of all vertices of U are still quite large, namely, $|S(u)| \geq 3np/\ln^2 n$ for each $u \in U$. For a color c denote by $W(c)$ the set of all vertices $u \in U$ for which c is included in the corresponding list of colors $S(u)$. We know that $|W(c)| \leq n/\ln^4 n$ for each color c. Thus we expect that the degree of a vertex u in the spanned subgraph $G[W(c)]$ is about $O(np/\ln^4 n) \ll |S(u)|$. If this indeed is the case for every color c and every vertex $u \in U$, then each color $c \in S(u)$ appears in the lists of only few neighbors of u. Then we can color the vertices of U simply by picking for each vertex a random color from its list. Unfortunately the graph $G[W(c)]$ can have a few vertices of degree much higher than $O(np/\ln^4 n)$. We color those vertices first and then treat the rest of U as indicated above. We omit technical details and some additional ideas required to complete the argument, and refer the reader to the paper of Krivelevich [10]. ∎

Next we consider a different model of random graphs — random regular graphs. For a positive integer-valued function $d = d(n)$ we define the model $G_{n,d}$ of random regular graphs consisting of all regular graphs on n vertices of degree d with the uniform probability distribution. Our aim here is to provide the asymptotic value of the choice number of $G_{n,d}$ for $d \gg n^{4/5}$. As in the case of $G(n,p)$ we need the following lemma.

Lemma 4.2 *For every constant $\epsilon > 0$, if $n^{4/5+\epsilon} \leq d \leq 3/4n$, then almost surely every subset of vertices of $G_{n,d}$ of size at least $m = n/\ln^4 n$ contains an independent set of size $k_0 = (1 - o(1))2\log_b d$, where $b = n/(n-d)$.*

Proof. Let $p = d/n$. We first need a lower bound on the probability that a random graph in $G(n,p)$ is regular. We use the result of Shamir and Upfal [14, equation (35)] with $\phi(n) = d$, $\theta = \frac{1}{2} + \delta$ for some $\delta > 0$, choosing $w(n) - \phi(n) = \lceil w(n)^{1-\delta} \rceil$, to deduce that the number of d-regular graphs on n vertices is at least

$$\binom{\binom{n}{2}}{nd/2} \exp\Big(- O(nd^{1/2+2\delta})\Big).$$

(Here there is a condition on $d(n)$; growing faster than $\log^2 n$ is sufficient.) It follows that for any fixed $\delta > 0$

$$Pr[G(n,d/n) \text{ is } d\text{-regular}] \geq \exp(-nd^{1/2+\delta}).$$

On the other hand as we have already mentioned in the proof of Theorem 4.1, the probability that the vertex set of $G(n, d/n)$ contains a subset of size m that does not span an independent set of size $k_0 = O(n \ln d/d)$ is at most

$$\binom{n}{m} \exp\left(-\Omega\Big(\frac{m^2}{k_0^4 p}\Big)\right) \leq 2^n \exp\left(-\Omega\Big(\frac{m^2}{k_0^4 p}\Big)\right) = \exp\left(-\Big(\frac{d^3}{n}\Big)^{1-o(1)}\right).$$

Comparing the last two exponents and using the assumption $d \geq n^{4/5+\epsilon}$, we observe that the probability that $G(n, d/n)$ is d-regular is much higher asymptotically than the probability that $G(n, d/n)$ contains a large subset without an independent set of size k_0. Therefore, almost surely if d lies in the range given in the assertion of the lemma, every subset of the vertices of $G_{n,d}$ of size at least $n/\ln^4 n$ spans an independent set of size k_0. ∎

Using this lemma, together with the ideas from [10] and the upper bound on the size of independent set in $G_{n,d}$ obtained in [11], one can deduce the following theorem:

Theorem 4.3 *For every constant $\epsilon > 0$, if $n^{4/5+\epsilon} \leq d \leq 3n/4$, then almost surely*

$$ch(G_{n,d}) = (1 + o(1))\chi(G_{n,d}) = (1 + o(1))\frac{n}{2\log_b d} ,$$

where $b = n/(n - d)$. ∎

Proof. The proof here is very similar to the proof of Theorem 4.1, and we therefore restrict ourselves to just a few words about it, leaving technical details to the reader.

To prove the lower bound for $ch(G_{n,d})$ observe that obviously $ch(G) \geq \chi(G) \geq |V(G)|/\alpha(G)$ for every graph G. Plugging in the estimate $\alpha(G) = (2 + o(1)) \log_b d$ for almost all graphs G in $G_{n,d}$, provided by Theorem 2.2 of [11], we get the required lower bound.

As for the upper bound, one can prove that almost surely the choice number of $G_{n,d}$ satisfies $ch(G) \leq n/k_0 + 3d/\ln^2 n$. The proof proceeds by essentially repeating the proof of Theorem 4.1 for the edge probability $p(n) = d/n$. Given a d-regular graph G on n vertices, satisfying the conclusion of Lemma 4.2 and having some additional properties, which hold almost surely in the probability space $G_{n,d}$, and also given color lists $\{S(v) : v \in V(G)\}$ of cardinality $|S(v)| = n/k_0 + 3d/\ln^2 n$, the coloring procedure starts by finding independent sets of size k_0 in frequent colors (i.e. colors appearing in at least $n/\ln^4 n$ lists). Once such a set is found in color c, we color all of its vertices by c, discard them and delete c from all lists.

After this part of the coloring procedure has finished, no color appears in more than $n/\ln^4 n$ vertices, and each uncolored vertex still has a list of at least $3d/\ln^2 n$ available colors. Moreover, for most uncolored vertices $v \in V$, most of the colors in the list $S(v)$ appear in the lists of $O(d/\ln^4 n)$ neighbors of v. We first treat few uncolored vertices which do not have the above stated property, and then color the rest by choosing colors at random from corresponding lists. For more details the reader is referred to [10]. ∎

5 A general setting

The aim of this section is to show that the approach exhibited in the proof of Theorem 2.1 can be applied in a much more general setting to obtain exponential bounds for probabilities of certain events. The bounds obtained can be better than those provided by the celebrated Janson inequality.

Let $H = (V, E)$ be a hypergraph with $|V| = m$ vertices and $|E| = k$ edges. We assume furthermore that H is r-uniform and D-regular. Form a random subset $R \subseteq V$ by

$$Pr[v \in R] = p_v \ ,$$

where these events are mutually independent over $v \in V$.

We want to estimate the probability $\mathbf{p}_0$ that the random set R does not contain any edge of H. Such an estimate is required frequently in applications of the probabilistic method. The following well-known theorem, proved first by Janson (see, e.g., [4], [9, Theorem 2.18]), usually gives an exponential bound for $\mathbf{p}_0$. To present this theorem, let Y be the number of edges of H spanned by R. We can represent Y as $I_1 + \cdots + I_k$ where I_j are the indicator functions of the edges of H. Let $\mu = E[Y]$, and write $I_i \sim I_j$ if the corresponding edges intersect. Set $\Delta = \sum_{i,j:\, I_i \sim I_j} E[I_i I_j]$.

Theorem 5.1 *We have*

$$\mathbf{p}_0 \leq \exp\left(-\frac{\mu^2}{\mu + \Delta}\right).$$

For any vertex v of H let Y_v denote the number of edges $f \in E(H)$ for which $v \in f$ and $f \setminus \{v\} \subseteq R$; set $\mu_v = E[Y_v]$. In many applications (especially those related to random graphs) the probabilities p_v all have the same value $\mathbf{p}$. In this case, $\mu = k\mathbf{p}^r$ and $\mu_v = \mu_0 = D\mathbf{p}^{r-1}$ for all v. Furthermore, it occurs frequently that the sum in Δ is dominated by the sum of those $E[I_i I_j]$ where the corresponding edges intersect in precisely one vertex. In such a case, $\Delta = \Theta(r\mu\mu_0)$. Assuming $\mu_0 \geq 1$, Janson's inequality gives

$$\mathbf{p}_0 \leq \exp\left(-\Theta\Big(\frac{\mu^2}{\mu + \Delta}\Big)\right) = \exp\left(-\Omega\Big(\frac{\mu}{r\mu_0}\Big)\right). \tag{3}$$

Our purpose here is to use the approach introduced in Section 2 to show that under a rather mild additional assumption (see Corollary 5.4), the following holds:

$$\mathbf{p}_0 \leq \exp\left(-\Omega\Big(\frac{\mu}{(1-\mathbf{p})r\mu_0}\Big)\right). \tag{4}$$

Inequality (4) is interesting for two reasons. First, in certain applications $\mathbf{p}$ is very close to 1 and therefore the term $1 - \mathbf{p}$ in the denominator yields a significant improvement. As we already saw in previous sections, this is exactly the case for the probability of independent sets in random graphs. For this problem an additional term $1 - \mathbf{p}$ is crucial, and the bound given by inequality (4) is almost sharp. Second, our proof is completely different from that of Janson (and also from the alternative proof by Boppana and Spencer [7]) and the method might therefore be of independent interest.

From now on we assume that $p_v = \mathbf{p}$ for all $v \in V(H)$. Let X denote the covering number of the spanned subhypergraph $H[R]$ (where the covering number of a hypergraph is the minimum number of vertices needed to cover all edges). Set $\tau = \min(E[X], m\mathbf{p}(1-\mathbf{p}))$.

Theorem 5.2 *The probability $\mathbf{p}_0$ that the hypergraph $H[R]$ has no edges satisfies:*

$$\mathbf{p}_0 \leq \exp\Big(-\Omega\Big(\frac{\tau^2}{m\mathbf{p}(1-\mathbf{p})}\Big)\Big).$$

Proof. Similarly to Lemma 2.6, using Theorem 7.4.3 [4], we have that for every $m\mathbf{p}(1-\mathbf{p}) > t > 0$ (the maximum variance in the martingale is $m\mathbf{p}(1-\mathbf{p})$),

$$Pr[X \leq E[X] - t] \leq e^{-\frac{t^2}{4m\mathbf{p}(1-\mathbf{p})}}.$$

Clearly, a hypergraph $H[R]$ contains no edges if and only if $X = 0$. Therefore

$$\mathbf{p}_0 = Pr[X = 0] \leq Pr[X \leq \mathbb{E}(X) - \tau] \leq e^{-\Omega\left(\frac{\tau^2}{m\mathbf{p}(1-\mathbf{p})}\right)}. \quad \blacksquare$$

It is well known that in a regular hypergraph, the covering number is at least the ratio between the number of edges and the degree. On the other hand, the expectation of the number of edges of $H[R]$ is μ and that of the degree of $H[R]$ is μ_0. Thus, it is reasonable to think that $E[X]$ is $\Omega(\mu/\mu_0)$. The following result shows that under an additional assumption, this is indeed the case.

Proposition 5.3 *Assume that $VAR[Y_v] = o\Big(\mu(\mu_0+1)/m\Big)$ for all v. Then $E[X] = \Omega\Big(\frac{\mu}{\mu_0+1}\Big)$.*

Remark. We need $\mu_0 + 1$ instead of μ_0 in order to deal with the case $\mu_0 \leq 1$. If $\mu_0 \geq 1$, we can replace $\mu_0 + 1$ by μ_0. Notice also that $\mu/\mu_0 = \frac{k\mathbf{p}}{D} = \frac{m\mathbf{p}}{r}$.

Now inequality (4) follows immediately from Theorem 5.2, Proposition 5.3 and the above remark. We note that the constant implicit in Ω is independent of $\mathbf{p}$, r and D.

Corollary 5.4 *Assume that $\mu_0 \geq 1$, $\mu/(\mu_0+1) \leq m\mathbf{p}(1-\mathbf{p})$ and $VAR[Y_v] = o(\mu\mu_0/m)$ for all v. Then*

$$\mathbf{p}_0 \leq \exp\Big(-\Omega\Big(\frac{\mu}{(1-\mathbf{p})r\mu_0}\Big)\Big).$$

We finish this section with the sketch of the proof of Proposition 5.3.

Sketch of the proof. Define $Z_v = Y_v$ if $Y_v \geq 2\mu_0 + 1$ and $Z_v = 0$ otherwise. Similarly to the proof of Proposition 2.3, by setting $Z^+ = \sum_v Z_v$, we obtain

$$X \geq \frac{Y - Z^+}{2\mu_0 + 1}. \tag{5}$$

Next, using the assumption that $VAR[Y_v] = o\Big(\mu(\mu_0+1)/m\Big)$, Chebyshev's inequality and applying the same techniques as in the proof of Proposition 2.5, we can show that

$$E[Z^+] = O(\sum_v VAR[Y_v]/\mu_0) = o(\mu).$$

Now it follows immediately that

$$\tau = E[X] \geq \frac{E[Y] - E[Z^+]}{2\mu_0 + 1} = (1 + o(1))\frac{\mu}{2\mu_0 + 1}.$$

This completes the proof. ∎

6 Concluding remarks

Consider the problem of estimating the probability that a random graph $G(n, p)$ has no cliques of cardinality t, with t fixed. In the setting of Section 5, define a hypergraph H whose vertices are the edges of K_n and whose edges are the t-cliques of $G(n, p)$. From Theorem 5.1, the probability that $G(n, p)$ has no t-cliques is at most $\exp(-\mu^2/(\mu + \Delta))$ where $\mu \sim p^{\binom{t}{2}} n^t/t!$ and $\Delta = \Theta(\mu^2/n^2 p)$. For fixed p with $\mu_0 > 1$ (where $\mu_0 = \Theta(\mu/n^2 p)$), the variance condition in Corollary 5.4 is easy to verify. Hence, for all such p, Corollary 5.4 gives virtually the same result as Janson's inequality, whilst its proof is entirely different. The argument also applies for graphs other than cliques but we do not elaborate in this direction.

An interesting and important open question is to estimate the probability that $G(n, p)$ does not contain an independent set of size $k = (1 - \epsilon)k_0$, where k_0 is defined in (1), and ϵ is a small constant, or even a function of n tending to 0 very slowly as n tends to infinity. We conjecture that

Conjecture.

$$Pr\Big[G(n, p) \text{ does not contain an independent set of size } k\Big] \leq \exp\Big(-\Omega(n^2/k^2 p)\Big).$$

This conjecture, if it holds, is best possible up to a logarithmic term in the exponent. It would immediately extend Theorem 4.1 to all $p = p(n) = \Omega(n^{-1+\epsilon})$ and also give a short proof for Łuczak's result on the chromatic number of $G(n, p)$ (by taking $\lim_{n\to\infty} \epsilon(n) = 0$).

Based on our method presented in this paper, to prove the above conjecture, it suffices to show that the expectation of the covering number of the corresponding hypergraph is $\Omega(n^2/k)$ (instead of $\Omega(n^2/k^2)$ as shown in the proof of Theorem 2.1). The following speculation might give the reader some intuition why this could be the case. Consider the complete hypergraph H_{com} consisting of all possible independent sets of size k. By Turán's theorem from extremal graph theory, the covering number of H_{com} is $\Omega(n^2/k)$. As k is much smaller than k_0, the expected number of independent sets of size k in $G(n,p)$ is huge (roughly $\binom{n}{k}^{\epsilon}$). So, the hypergraph corresponding to these independent sets looks typically like a fairly dense sub-hypergraph of H_{com} and one may hope that such a hypergraph should have covering number close to that of H_{com}, namely $\Omega(n^2/k)$.

References

[1] N. Alon, Choice numbers of graphs; a probabilistic approach, *Combinatorics, Probability and Computing* **1** (1992), 107-114.

[2] N. Alon, Restricted colorings of graphs, in *Surveys in Combinatorics 1993*, London Math. Soc. Lecture Notes Series 187 (K. Walker, ed.), Cambridge Univ. Press 1993, 1–33.

[3] N. Alon, M. Krivelevich and B. Sudakov, List coloring of random and pseudo-random graphs, *Combinatorica* **19** (1999), 453–472.

[4] N. Alon and J. Spencer, **The probabilistic method**, 2^{nd} ed., Wiley, New York, 2000.

[5] B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988), 49–55.

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

[7] R. Boppana and J. Spencer, A useful elementary correlation inequality, *J. Comb. Th. Ser. A* **50** (1989), 305–307.

[8] P. Erdős, A. L. Rubin and H. Taylor, Choosability in graphs, *Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium* XXVI, 1979, 125-157.

[9] S. Janson, T. Łuczak and A. Ruciński, **Random graphs**, Wiley, New York, 2000.

[10] M. Krivelevich, The choice number of dense random graphs, *Combinatorics, Probability and Computing* **9** (2000), 19–26.

[11] M. Krivelevich, B. Sudakov, V. H. Vu and N.C. Wormald, Random regular graphs of high degree, *Random Structures and Algorithms* 18 (2001), 346-363.

[12] M. Krivelevich and V. H. Vu, Choosability in random hypergraphs, *J. Comb. Theory Ser. B* 83 (2001), 241-257.

[13] D. Matula, The largest clique size in a random graph, Tech. Rep. Dept. Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.

[14] E. Shamir and E. Upfal, Large regular factors in random graphs, *Convexity and graph theory* (Jerusalem 1981), North Holland Math. Stud. 87, North Holland, Amsterdam-New York 1984, pp. 271–282.

[15] J. Spencer, Probabilistic methods in combinatorics, in *Proceedings Int. Congress of Math. Zürich, 1994*, Birkhäuser,Basel, 1375–1383.

[16] V. G. Vizing, Coloring the vertices of a graph in prescribed colors (in Russian), *Diskret. Analiz. No. 29, Metody Diskret. Anal. v. Teorii Kodov i Shem* 101 (1976), 3-10.

[17] V.H. Vu, On some degree conditions which guarantee the upper bound of chromatic (choice) number of random graphs, *Journal of Graph Theory* **31** 1999, 201–226.