8 Janson Inequalities

We present a collection of inequalities, known collectively as Janson inequalities (Janson 1990). These tools allow us to estimate **lower tail** large deviation probabilities.

A typical application of Janson's inequality allows us to upper bound the probability that a random graph $G(n, p)$ does not contain any copy of some subgraph. Compared to the second moment method from Chapter 4, Janson inequalities (which is applicable in more limited setups) gives much better bounds, usually with exponential decays.

8.1 Probability of non-existence

The following setup should be a reminiscent of both the second moment method as well as Lovász local lemma (the random variable model).

Setup 8.1.1 (for Janson's inequality: counting containments)
Let R be a random subset of $[N]$ with each element included independently (possibly with different probabilities).

Let $S_1, \dots, S_k \subseteq [N]$. Let A_i be the event that $S_i \subseteq R$. Let

$$X = \sum_i 1_{A_i}$$

be the number of sets S_i contained in the same set R. Let

$$\mu = \mathbb{E}[X] = \sum_i \mathbb{P}(A_i).$$

Write $i \sim j$ if $i \neq j$ and $S_i \cap S_j \neq \emptyset$. Let (as in the second moment method)

$$\Delta = \sum_{(i,j):i\sim j} \mathbb{P}(A_i A_j) = \sum_{(i,j):i\sim j} \mathbb{P}(S_i \cup S_j \subseteq R)$$

(note that (i, j) and (j, i) is each counted once).

The following inequality appeared in Janson, Łuczak, and Ruciński (1990).

Theorem 8.1.2 (Janson inequality I)
Assuming Setup 8.1.1,

$$\mathbb{P}(X = 0) \le e^{-\mu+\Delta/2}.$$

This inequality is most useful when $\Delta = o(\mu)$.

Remark 8.1.3. When $\mathbb{P}(A_i) = o(1)$ (which is the case in a typical application), Harris' inequality gives us

$$\begin{aligned}\mathbb{P}(X = 0) &= \mathbb{P}\left(\overline{A_1} \cdots \overline{A_k}\right) \ge \prod_{i=1}^{k} \mathbb{P}\left(\overline{A_i}\right) \\ &= \prod_{i=1}^{k}(1 - \mathbb{P}(A_i)) = \exp\left(-(1 + o(1)) \sum_{i=1}^{k} \mathbb{P}(A_i)\right) = e^{-(1+o(1))\mu}.\end{aligned}$$

In the setting where $\Delta = o(\mu)$, two bounds match to give $\mathbb{P}(X = 0) = e^{-(1+o(1)\mu}$.

Proof. Let

$$r_i = \mathbb{P}(A_i | \overline{A}_1 \cdots \overline{A}_{i-1}).$$

We have

$$\begin{aligned}\mathbb{P}(X = 0) &= \mathbb{P}(\overline{A}_1 \cdots \overline{A}_k) \\ &= \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2|\overline{A}_1) \cdots \mathbb{P}(\overline{A}_k|\overline{A}_1 \cdots \overline{A}_{k-1}) \\ &= (1 - r_1) \cdots (1 - r_k) \\ &\le e^{-r_1 - \cdots - r_k}\end{aligned}$$

It suffices now to prove that:

Claim. For each $i \in [k]$

$$r_i \ge \mathbb{P}(A_i) - \sum_{j<i:j\sim i} \mathbb{P}(A_i A_j).$$

Summing the claim over $i \in [k]$ would then yield

$$\sum_{i=1}^{k} r_i \ge \sum_i \mathbb{P}(A_i) - \frac{1}{2} \sum_i \sum_{j\sim i} \mathbb{P}(A_i A_j) = \mu - \frac{\Delta}{2}$$

and thus

$$\mathbb{P}(X = 0) \le \exp\left(-\sum_i r_i\right) \le \exp\left(-\mu + \frac{\Delta}{2}\right)$$

Proof of claim. Recall that i is given and fixed. Let

$$D_0 = \bigwedge_{j<i: j \nsim i} \overline{A}_j \qquad \text{and} \qquad D_1 = \bigwedge_{j<i: j \sim i} \overline{A}_j$$

Then

$$\begin{aligned} r_i &= \mathbb{P}(A_i | \overline{A}_1 \cdots \overline{A}_{i-1}) = \mathbb{P}(A_i | D_0 D_1) = \frac{\mathbb{P}(A_i D_0 D_1)}{\mathbb{P}(D_0 D_1)} \geq \frac{\mathbb{P}(A_i D_0 D_1)}{\mathbb{P}(D_0)} \\ &= \mathbb{P}(A_i D_1 | D_0) = \mathbb{P}(A_i | D_0) - \mathbb{P}(A_i \overline{D}_1 | D_0) \\ &= \mathbb{P}(A_i) - \mathbb{P}(A_i \overline{D}_1 | D_0) \quad \text{[by independence]} \end{aligned}$$

Since A_i and $\overline{D}_1$ are both increasing events, and D_0 is a decreasing event, by Harris' inequality (Corollary 7.1.6),

$$\mathbb{P}(A_i \overline{D}_1 | D_0) \leq \mathbb{P}(A_i \overline{D}_1) = \mathbb{P}\left(A_i \wedge \bigvee_{j<i: j \sim i} A_j\right) \leq \sum_{j<i: j \sim i} \mathbb{P}(A_i A_j)$$

This concludes the proof of the claim, and thus the proof of the theorem. □

Remark 8.1.4 (History). Janson's original proof was via analytic interpolation. The above proof is based on Boppana and Spencer (1989) with a modification by Warnke (personal communication). It has some similarities to the proof of Lovász local lemma from Section 6.1. The above proof incorporates ideas from Riordan and Warnke (2015), who extended Janson's inequality from principal up-set to general up-sets. Indeed, the above proof only requires that the events A_i are increasing, whereas earlier proofs of the result (e.g., the proof in Alon–Spencer) requires the full assumption of Setup 8.1.1, namely that each A_i is an event of the form $S_i \subseteq R_i$ (i.e., a ***principal up-set***).

Question 8.1.5

What is the probability that $G(n, p)$ is triangle-free?

In Setup 8.1.1, let $[N]$ with $N = \binom{n}{2}$ be the set of edges of K_n, and let $S_1, \ldots, S_{\binom{n}{3}}$ be 3-element sets where each S_i is the edge-set of a triangle. As in the second moment calculation in Section 4.2, we have

$$\mu = \binom{n}{3} p^3 \asymp n^3 p^3 \qquad \text{and} \qquad \Delta \asymp n^4 p^5.$$

(where Δ is obtained by considering all appearances of a pair of triangles glued along an edge).

If $p \ll n^{-1/2}$, then $\Delta = o(\mu)$, in which case Janson inequality I (Theorem 8.1.2 and Remark 8.1.3) gives the following.

Theorem 8.1.6
If $p = o(n^{-1/2})$, then

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) = e^{-(1+o(1))\mu} = e^{-(1+o(1))n^3p^3/6}.$$

Corollary 8.1.7
For a constant $c > 0$,

$$\lim_{n\to\infty} \mathbb{P}(G(n,c/n) \text{ is triangle-free}) = e^{-c^3/6}.$$

In fact, the number of triangles in $G(n,c/n)$ converges to a Poisson distribution with mean $c^3/6$. On the other hand, when $p \gg 1/n$, the number of triangles is asymptotically normal.

What about if $p \gg n^{-1/2}$, so that $\Delta \gg \mu$. Janson inequality I does not tell us anything nontrivial. Do we still expect the triangle-free probability to be $e^{-(1+o(1))\mu}$, or even $\le e^{-c\mu}$?

As noted earlier in Remark 7.2.3, another way to obtain a lower bound on the probability triangle-freeness is to consider the probability the $G(n,p)$ is empty (or contained in some fixed complete bipartite graph), in which case we obtain

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) \ge (1-p)^{\Theta(n^2)} = e^{-\Theta(n^2p)}$$

(the second step assumes that p is bounded away from 1. If $p \gg n^{-1/2}$, so the above lower bound better than the previous one: $e^{-\Theta(n^2p)} \gg e^{-(1+o(1))\mu}$.

Nevertheless, we'll still use Janson to bootstrap an upper bound on the triangle-free probability. More generally, the next theorem works in the complement region of the Janson inequality I, where now $\Delta \ge \mu$.

Theorem 8.1.8 (Janson inequality II)
Assuming Setup 8.1.1, if $\Delta \ge \mu$, then

$$\mathbb{P}(X = 0) \le e^{-\mu^2/(2\Delta)}.$$

The proof idea is to applying the first Janson inequality on a randomly sampled subset of events. This sampling technique might remind you of some earlier proofs, e.g., the proof of the crossing number inequality (Theorem 2.6.2), where we first proved a "cheap bound" that worked in a more limited range, and then used sampling to obtain a better bound.

Proof. For each $T \subseteq [k]$, let $X_T := \sum_{i \in T} 1_{A_i}$ denote the number of occurring events in T. We have

$$\mathbb{P}(X = 0) \le \mathbb{P}(X_T = 0) \le e^{-\mu_T + \Delta_T/2}$$

where

$$\mu_T = \sum_{i \in T} \mathbb{P}(A_i)$$

and

$$\Delta_T = \sum_{(i,j) \in T^2 : i \sim j} \mathbb{P}(A_i A_j)$$

Choose $T \subseteq [k]$ randomly by including every element with probability $q \in [0, 1]$ independently. We have

$$\mathbb{E}\mu_T = q\mu \qquad \text{and} \qquad \mathbb{E}\Delta_T = q^2 \Delta$$

and so

$$\mathbb{E}(-\mu_T + \Delta_T/2) = -q\mu + q^2\Delta/2.$$

By linearity of expectations, thus there is some choice of $T \subseteq [k]$ so that

$$-\mu_T + \Delta_T/2 \le -q\mu + q^2\Delta/2$$

so that

$$\mathbb{P}(X = 0) \le e^{-q\mu + q^2\Delta/2}$$

for every $q \in [0, 1]$. Since $\Delta \ge \mu$, we can set $q = \mu/\Delta \in [0, 1]$ to get the result. □

To summarize, the first two Janson inequalities tell us that

$$\mathbb{P}(X = 0) \le \begin{cases} e^{-\mu + \Delta/2} & \text{if } \Delta < \mu \\ e^{-\mu^2/(2\Delta)} & \text{if } \Delta \ge \mu. \end{cases}$$

Remark 8.1.9. If $\mu \to \infty$ and $\Delta \ll \mu^2$, then Janson inequality II implies $\mathbb{P}(X = 0) = o(1)$, which we knew from second moment method. However Janson's inequality gives an exponentially decaying tail bound, compared to only a polynomially decaying tail via the second moment method. The exponential tail will be important in an application below to determining the chromatic number of $G(n, 1/2)$.

Let us revisit the example of estimating the probability that $G(n, p)$ is triangle-free, now in the regime $p \gg n^{-1/2}$. We have

$$n^3 p^3 \asymp \mu \ll \Delta \asymp n^4 p^5.$$

So so for large enough n, Janson inequality II tells us

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) \le e^{-\mu^2/(2\Delta)} = e^{-\Theta(n^2 p)}$$

Since

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) \ge \mathbb{P}(G(n,p) \text{ is empty}) \ge (1-p)^{\binom{n}{2}} = e^{-\Theta(n^2 p)}$$

where the final step assumes that p is bounded away from 1, we conclude that

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) = e^{-\Theta(n^2 p)}$$

We summarize the results below (strictly speaking we have not yet checked the case $p \asymp n^{-1/2}$, which we can verify by applying Janson inequalities; note that the two regimes below match at the boundary).

Theorem 8.1.10
Suppose $p = p_n \le 0.99$. Then

$$\mathbb{P}(G(n,p) \text{ is triangle-free}) = \begin{cases} \exp\left(-\Theta(n^2 p)\right) & \text{if } p \gtrsim n^{-1/2} \\ \exp\left(-\Theta(n^3 p^3)\right) & \text{if } p \lesssim n^{-1/2} \end{cases}$$

***Remark* 8.1.11.** Sharper results are known. Here are some highlights.

1. The number of triangle-free graphs on n vertices is $2^{(1+o(1))n^2/4}$. In fact, an even stronger statement is true: almost all (i.e., $1-o(1)$ fraction) n-vertex triangle-free graphs are bipartite (Erdős, Kleitman, and Rothschild 1976).

2. If $m \ge C n^{3/2}\sqrt{\log n}$ for any constant $C > \sqrt{3}/4$ (and this is best possible), then almost all all n-vertex m-edge triangle-free graphs are bipartite (Osthus, Prömel, and Taraz 2003). This result has been extended to K_r-free graphs for every fixed r (Balogh, Morris, Samotij, and Warnke 2016).

3. For $n^{-1/2} \ll p \ll 1$, (Łuczak 2000)

$$-\log \mathbb{P}(G(n,p) \text{ is triangle-free}) \sim -\log \mathbb{P}(G(n,p) \text{ is bipartite}) \sim n^2 p/4.$$

   This result was generalized to general H-free graphs using the powerful recent method of hypergraph containers (Balogh, Morris, and Samotij 2015).

8.2 Lower tails

Previously we looked at the probability of non-existence. Now we would like to estimate lower tail probabilities. Here is a model problem.

Question 8.2.1

Fix a constant $0 < \delta \leq 1$. Let X be the number of triangles of $G(n, p)$. Estimate

$$\mathbb{P}(X \leq (1-\delta)\mathbb{E}X).$$

We will bootstrap Janson inequality I, $\mathbb{P}(X = 0) \leq \exp(-\mu + \Delta/2)$, to an upper bound on lower tail probabilities.

Theorem 8.2.2 (Janson inequality III)

Assume Setup 8.1.1. For any $0 \leq t \leq \mu$,

$$\mathbb{P}(X \leq \mu - t) \leq \exp\left(\frac{-t^2}{2(\mu + \Delta)}\right)$$

Note that setting $t = \mu$ we basically recover the first two Janson inequalities (up to an unimportant constant factor in the exponent):

$$\mathbb{P}(X = 0) \leq \exp\left(\frac{-\mu^2}{2(\mu + \Delta)}\right). \tag{8.1}$$

(Note that this form of the inequality conveniently captures Janson inequalities I & II.)

Proof. (by Lutz Warnke[1]) We start the proof similarly to the proof of the Chernoff bound, by applying Markov's inequality on the moment generating function. To that end, let $\lambda \geq 0$ to be optimized later. Let

$$q = 1 - e^{-\lambda}.$$

By Markov's inequality,

$$\begin{aligned}\mathbb{P}(X \leq \mu - t) &= \mathbb{P}\left(e^{-\lambda X} \geq e^{-\lambda(\mu - t)}\right) \\ &\leq e^{\lambda(\mu - t)}\, \mathbb{E}\, e^{-\lambda X} \\ &\leq e^{\lambda(\mu - t)}\, \mathbb{E}[(1-q)^X].\end{aligned}$$

[1] Personal communication

For each $i \in [k]$, let $W_i \sim \text{Bernoulli}(q)$ independently. Consider the random variable

$$Y = \sum_{i=1}^{k} 1_{A_i} W_i.$$

Conditioned on the value of X, the probability that $Y = 0$ is $(1-q)^X$ (i.e., the probability that $W_i = 0$ for each of the X events A_i that occurred). Taking expectation over X, we have

$$\mathbb{P}(Y = 0) = \mathbb{E}[\mathbb{P}(Y = 0|X)] = \mathbb{E}[(1-q)^X].$$

Note that Y fits within Setup 8.1.1 by introducing k new elements to the ground set $[N]$, where each new element is included according to W_i, and enlarging each S_i to include this new element. The relevant parameters of Y are

$$\mu_Y := \mathbb{E}Y = q\mu$$

and

$$\Delta_Y := \sum_{(i,j):i\sim j} \mathbb{E}[1_{A_i} W_i 1_{A_j} W_j] = q^2\Delta.$$

Then Janson inequality I applied to Y gives

$$\mathbb{P}(Y = 0) \le e^{-\mu_Y + \Delta_Y/2} = e^{-q\mu + q^2\Delta/2}.$$

Therefore,

$$\mathbb{E}[(1-q)^X] = \mathbb{P}(Y = 0) \le e^{-q\mu + q^2\Delta/2}.$$

Continuing the moment calculation at the beginning of the proof, and using that

$$\lambda - \frac{\lambda^2}{2} \le q \le \lambda,$$

we have

$$\begin{aligned} \mathbb{P}(X \le -\mu + t) &\le e^{\lambda(\mu - t)}\,\mathbb{E}[(1-q)^X] \\ &\le \exp\left(\lambda(\mu - t) - q\mu + q^2\Delta/2\right) \\ &\le \exp\left(\lambda(\mu - t) - \left(\lambda - \frac{\lambda^2}{2}\right)\mu + \lambda^2\frac{\Delta}{2}\right) \\ &= \exp\left(-\lambda t + \frac{\lambda^2}{2}(\mu + \Delta)\right) \end{aligned}$$

We optimize by setting $\lambda = t/(\mu + \Delta)$ to obtain $\le \exp\left(\frac{-t^2}{2(\mu+\Delta)}\right)$. □

***Example* 8.2.3 (Lower tails for triangle counts).** Let X be the number of triangles in

$G(n,p)$. We have $\mu \asymp n^3p^3$ and $\Delta \asymp n^4p^5$. Fix a constant $\delta \in (0,1]$. Let $t = \delta\mathbb{E}X$. We have

$$\mathbb{P}(X \le (1-\delta)\mathbb{E}X) \le \exp\left(-\Theta\left(\frac{-\delta^2 n^6 p^6}{n^3p^3 + n^4p^5}\right)\right) = \begin{cases}\exp\left(-\Theta_\delta(n^2p)\right) & \text{if } p \gtrsim n^{-1/2}, \\ \exp\left(-\Theta_\delta(n^3p^3)\right) & \text{if } p \lesssim n^{-1/2}.\end{cases}$$

The bounds are tight up to a constant in the exponent, since

$$\mathbb{P}(X \le (1-\delta)\mathbb{E}X) \ge \mathbb{P}(X=0) = \begin{cases}\exp\left(-\Theta(n^2p)\right) & \text{if } p \gtrsim n^{-1/2}, \\ \exp\left(-\Theta(n^3p^3)\right) & \text{if } p \lesssim n^{-1/2}.\end{cases}$$

***Example* 8.2.4** (No corresponding Janson inequality for upper tails)**.** Continuing with X being the number of triangles of $G(n,p)$, from on the above lower tail results, we might expect $\mathbb{P}(X \ge (1+\delta)\mathbb{E}X) \le \exp(-\Theta_\delta(n^2p))$, but actually this is false!

By planting a clique of size $\Theta(np)$, we can force $X \ge (1+\delta)\mathbb{E}X$. Thus

$$\mathbb{P}(X \ge (1+\delta)\mathbb{E}X) \ge p^{\Theta_\delta(n^2p^2)}$$

which is much bigger than $\exp\left(-\Theta(n^2p)\right)$. The above is actually the truth (Kahn–DeMarco 2012 and Chatterjee 2012):

$$\mathbb{P}(X \ge (1+\delta)\mathbb{E}X) = p^{\Theta_\delta(n^2p^2)} \quad \text{if } p \gtrsim \frac{\log n}{n},$$

but the proof is much more intricate. Recent results allow us to understand the exact constant in the exponent though new developments in large deviation theory. The current state of knowledge is summarized below.

Theorem 8.2.5 (Harel, Mousset, Samotij 2022)
Let X be the number of triangles in $G(n,p)$ with $p = p_n$ satisfying $n^{-1/2} \ll p \ll 1$,

$$-\log \mathbb{P}(X \ge (1+\delta)\mathbb{E}X) \sim \min\left\{\frac{\delta}{3}, \frac{\delta^{2/3}}{2}\right\} n^2p^2\log(1/p),$$

and for $n^{-1}\log n \ll p \ll n^{-1/2}$,

$$-\log \mathbb{P}(X \ge (1+\delta)\mathbb{E}X) \sim \frac{\delta^{2/3}}{2} n^2p^2\log(1/p).$$

***Remark* 8.2.6.** The leading constants were determined by Lubetzky and Zhao (2017) by solving an associated variational problem. Earlier results, starting with Chatterjee and Varadhan (2011) and Chatterjee and Dembo (2016) prove large deviation frameworks that gave the above theorem for sufficiently slowly decaying $p \geq n^{-c}$.

For the corresponding problem for lower tails, see Kozma and Samotij (2023) for an approach using relative entropy that reduces the rate problem to a variational problem. The exact leading constant is known only for sufficiently small $\delta > 0$, where the answer is given by "replica symmetry", meaning that the exponential rate is given by a uniform decrement in edge densities for the random graph. In contrast, for δ close to 1, we expect (though cannot prove) that the typical structure of a conditioned random graph is close to a two-block model (Zhao 2017).

8.3 Chromatic number of a random graph

Question 8.3.1

What is the chromatic number of $G(n, 1/2)$?

In Section 4.4, we used the second moment method to find the clique number ω of $G(n, 1/2)$. We saw that, with probability $1 - o(1)$, the clique number is concentrated on two values, and in particular,

$$\omega(G(n, 1/2)) \sim 2 \log_2 n \qquad \text{whp.}$$

The ***independence number $\alpha(G)$*** is the size of the largest independent set in G. The independence number $\alpha(G)$ is the equal to the clique number the complement of G. Since $G(n, 1/2)$ and its graph complement have the same distribution, we have $\alpha(G(n, 1/2)) \sim 2 \log_2 n$ whp as well.

Using the following lower bound on the chromatic number $\chi(G)$:

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)}$$

(since each color class is an independent set), we obtain that

$$\chi(G(n, 1/2)) \geq \frac{(1 + o(1))n}{\log_2 n} \qquad \text{whp.}$$

The following landmark theorem shows that the above lower bound on $\chi(G(n, 1/2))$ is asymptotically tight.

Theorem 8.3.2 (Chromatic number of a random graph — Bollobás 1988)
With probability $1 - o(1)$,

$$\chi(G(n, 1/2)) \sim \frac{n}{2 \log_2 n}.$$

Recall that $\omega(G(n, 1/2))$ is typically concentrated around the point k where the expected number of k-cliques $\binom{n}{k}2^{-\binom{k}{2}}$ is neither too large nor too close to zero. The next lemma show that this probability drops very quickly when we decrease k even by a constant.

Lemma 8.3.3
Let $k_0 = k_0(n)$ be the largest possible integer k so that $\binom{n}{k}2^{-\binom{k}{2}} \geq 1$. Then

$$\mathbb{P}(\alpha(G(n, 1/2)) < k_0 - 3) \leq e^{-n^{2-o(1)}}$$

Note that there is a trivial lower bound of $2^{-\binom{n}{2}}$ coming from an empty graph.

Proof. Let us prove the equivalent claim

$$\mathbb{P}(\omega(G(n, 1/2)) < k_0 - 3) \leq e^{-n^{2-o(1)}}.$$

Let $\mu_k := \binom{n}{k}2^{-\binom{k}{2}}$. For $k \sim k_0(n) \sim 2 \log_2 n$, we have

$$\frac{\mu_{k+1}}{\mu_k} = \frac{\binom{n}{k+1}}{\binom{n}{k}} 2^{-k} \sim \frac{n}{k} 2^{-(2+o(1)) \log_2 n} = \frac{1}{n^{1-o(1)}}.$$

Let $k = k_0 - 3$ and applying Setup 8.1.1 for Janson inequality with X being the number of k-cliques, we have

$$\mu = \mu_k > n^{3-o(1)}$$

and (details of the computation omitted)

$$\Delta \sim \mu^2 \frac{k^4}{n^2} = n^{4-o(1)}.$$

So $\Delta > \mu$ for sufficiently large n, and we can apply Janson inequality II:

$$\mathbb{P}(\omega(G(n, 1/2)) < k) = \mathbb{P}(X = 0) \leq e^{-n^{2-o(1)}}.$$ □

Proof of Theorem 8.3.2. The lower bound proof was discussed before the theorem statement. For the upper bound we will give a strategy to properly color the random graph with $(2 + o(1)) \log_2 n$ colors. We will proceed by taking out independent sets of size $\sim 2\log_2 n$ iteratively until $o(n/\log n)$ vertices remain, at which point we can use a different color for each remaining vertex.

Note that after taking out the first independent set of size $\sim 2\log_2 n$, we cannot claim that the remaining graph is still distributed as $G(n, 1/2)$. It is not. Our selection of the vertices was dependent on the random graph. We are not allowed to "resample" the edges after the first selection.

The strategy is to apply the previous lemma to see that every large enough subset of vertices has an independent set of size $\sim 2\log_2 n$.

Let $G \sim G(n, 1/2)$. Let $m = \lfloor n/(\log n)^2 \rfloor$, say. For any set S of m vertices, the induced subgraph $G[S]$ has the distribution of $G(m, 1/2)$. By Lemma 8.3.3, for

$$k = k_0(m) - 3 \sim 2\log_2 m \sim 2\log_2 n,$$

we have

$$\mathbb{P}(\alpha(G[S]) < k) = e^{-m^{2-o(1)}} = e^{-n^{2-o(1)}}.$$

Taking a union bound over all $\binom{n}{m} < 2^n$ such sets S,

$$\mathbb{P}(\text{there is an } m\text{-vertex subset } S \text{ with } \alpha(G[S]) < k) < 2^n e^{-n^{2-o(1)}} = o(1).$$

So the following statement is true in $G(n, 1/2)$ with probability $1 - o(1)$:

(*) Every m-vertex subset contains a k-vertex independent set.

Assume that G has property (*). Now we execute our strategy at the beginning of the proof:

1. While $\geq m$ vertices remain:
   i. Find an independent set of size k, and let it form its own color class
   ii. Remove these k vertices
2. Color the remaining $< m$ vertices each with a new color.

The result is a proper coloring. The number of colors used is

$$\frac{n}{k} + m \sim \frac{n}{2\log_2 n}. \qquad \square$$

Exercises

1. *3-AP-free probability.* Determine, for all $0 < p \leq 0.99$ (p is allowed to depend on n), the probability that $[n]_p$ does not contain a 3-term arithmetic progression, up to a constant factor in the exponent. (The form of the answer should be similar

to the conclusion in class about the probability that $G(n, p)$ is triangle-free. See 3 for notation.)

2. Prove that with probability $1 - o(1)$, the size of the largest subset of vertices of $G(n, 1/2)$ inducing a triangle-free subgraph is $\Theta(\log n)$.

3. *Nearly perfect triangle factor, again.* Using Janson inequalities this time, give another solution to Problem 11 in the following generality.

   a) Prove that for every $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that such that with probability $1 - o(1)$, $G(n, C_\varepsilon n^{-2/3})$ contains at least $(1/3 - \varepsilon)n$ vertex-disjoint triangles.

   b) (Optional) Compare the the dependence of the optimal C_ε on ε you obtain using the method in Problem 11 versus this problem (don't worry about leading constant factors).

4. $\star$*Threshold for extensions.* Show that for every constant $C > 16/5$, if $n^2 p^5 > C \log n$, then with probability $1 - o(1)$, every edge of $G(n, p)$ is contained in a K_4.

   (Be careful, this event is not increasing, and so it is insufficient to just prove the result for one specific p.)

5. *Lower tails of small subgraph counts.* Fix graph H and $\delta \in (0, 1]$. Let X_H denote the number of copies of H in $G(n, p)$. Prove that for all n and $0 < p < 0.99$,

$$\mathbb{P}(X_H \le (1 - \delta)\mathbb{E}X_H) = e^{-\Theta_{H,\delta}(\Phi_H)} \quad \text{where } \Phi_H := \min_{H' \subseteq H : e(H') > 0} n^{v(H')} p^{e(H')}.$$

   Here the hidden constants in $\Theta_{H,\delta}$ may depend on H and δ (but not on n and p).

6. $\star$ *List chromatic number of a random graph.* Show that the list chromatic number of $G(n, 1/2)$ is $(1 + o(1))\frac{n}{2 \log_2 n}$ with probability $1 - o(1)$.

   The *list-chromatic number* (also called *choosability*) of a graph G is defined to the minimum k such that if every vertex of G is assigned a list of k acceptable colors, then there exists a proper coloring of G where every vertex is colored by one of its acceptable colors.