Clique number, independence number

Recall that the clique number $\omega(G)$ is the size of the largest complete subgraph of G and the independence number (stability number) $\alpha(G)$ is the size of the largest independent set (i.e., the clique number of the complement of G).

Consider $G(n,p)$ with p fixed. In fact, to simplify notation, we take $p = 1/2$. The clique number ω (and thus α as well) may be estimated very precisely by the first and second moment methods:

Let N_k be the number of cliques of size k. Then

$$\mathbb{P}(\omega \geq k) = \mathbb{P}(N_k \geq 1) \leq \mathbb{E}N_k = \binom{n}{k} 2^{-\binom{k}{2}}$$

$$\leq \left(\frac{n e \sqrt{2}}{k} 2^{-k/2} \right)^k \to 0 \quad \text{if}$$

$$k = 2\log_2 n - 2\log_2\log_2 n + 2\log_2 e + 1$$

The lower bound follows from the second moment method. We need to show that $\frac{\mathrm{Var}(X_k)}{(\mathbb{E}X_k)^2} = \frac{\mathbb{E}(X_k^2)}{(\mathbb{E}X_k)^2} - 1 \to 0$ when the value of k is slightly decreased.

But

$$\frac{\mathbb{E}X_k^2}{(\mathbb{E}X_k)^2} - 1 = \frac{\binom{n}{k} 2^{-\binom{k}{2}} \sum_{i=0}^{k} \binom{k}{i}\binom{n-k}{k-i} 2^{-\binom{k}{2}+\binom{i}{2}}}{\left[\binom{n}{k} 2^{-\binom{k}{2}}\right]^2} - 1$$

$$\leq \frac{1}{\binom{n}{k}} \sum_{i=1}^{k} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}}$$

After careful bounding (Exercise!) we get the following:

Theorem W.h.p.,

$$2\log_2 n - 2\log_2\log_2 n + 2\log_2 e - \frac{5}{2} \leq \omega \leq 2\log_2 n - 2\log_2\log_2 n + 2\log_2 e + \frac{1}{2}$$

(In $G(n,p)$ the base of the logs is $\frac{1}{p}$ for the clique number and $\frac{1}{1-p}$ for the independence number.)

The chromatic number

Next we study the magnitude of the chromatic number χ of $\mathcal{G}(n,p)$ for fixed p. (Again, we may take $p=\frac{1}{2}$.)

Since in any proper coloring any set of vertices of the same color is an independent set, $\chi \geq \frac{n}{\alpha}$.
Since $\alpha \leq 2\log_2 n - 2\log_2\log_2 n$ w.h.p.,

$$\chi \geq \frac{n}{2\log_2 n\,(1-o(1))}.$$

It turns out that this bound is tight! (Bollobás '88)
The key ingredient is a sharp estimate for the probability of non-existence for large independent sets:

Lemma The probability that $\mathcal{G}(n,\frac{1}{2})$ has no independent set of size $2\log_2 n - 2.1\log_2\log_2 n$ is at most

$$\exp\left(-\frac{n^2}{66(\log_2 n)^5}\right)$$

Proof Let $k = 2\log_2 n - 2.1\log_2\log_2 n$ and use Janson's inequality to bound the probability of nonexistence of stable sets of size k. If X_k denotes the number of stable (independent) sets of size k, then

$$P(X_k=0) \leq \exp\left(-\frac{(\mathbb{E}X_k)^2}{\sum\limits_{\substack{A,B:|A|=|B|=k,\\ |A\cap B|\geq 2}}' \mathbb{E}X_A\cdot X_B}\right)$$

indicator that B is a stable set.

Just like before,

$$\frac{\sum'_{A,B} \mathbb{E} X_A X_B}{(\mathbb{E}X_k)^2} = \frac{\sum'^{\,k}_{i=2} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}}}{\binom{n}{k}} \underset{\text{Exercise.}}{\le} \frac{66(\log_2 n)^5}{n^2}$$

We may use the sharp estimate above in combination with the union bound:

$$\mathbb{P}\Big(\exists \text{ a subgraph of } G(n,\tfrac12) \text{ with } \ge \frac{n}{(\log n)^2} \text{ vertices and stable set of size } \underbrace{2\log_2 n - 7\log_2\log_2 n}_{\ge 2\log_2 \frac{n}{(\log n)^2} - 2.1\log_2\log_2 \frac{n}{(\log n)^2}}\Big)$$

$$\le 2^n \exp\left(-\frac{\left(n/(\log n)^2\right)^2}{66\left(\log \frac{n}{(\log n)^2}\right)^5}\right) \le 2^n \exp\left(-\frac{n^2}{(\log n)^{10}}\right) \to 0$$

upper bound on the # of subgraphs $\binom{n}{n/(\log n)^2}$

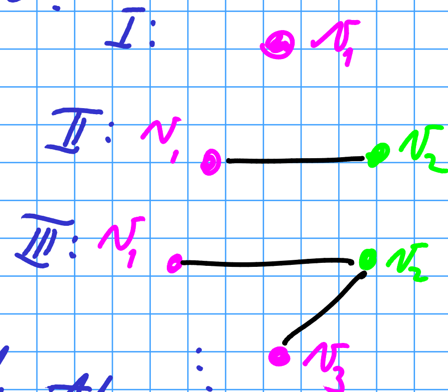
But then color the graph as follows: find a stable set of size $\ge 2\log_2 n - 7\log_2\log_2 n$ and give it color #1. Remove these vertices. In the remainder, again there is a stable set of the same size, so color them by color #2. We can continue until we have $\frac{n}{(\log n)^2}$ vertices remaining. The rest of the vertices will all get a different color. The number of colors used is at most

$$\frac{n}{2\log_2 n - 7\log_2\log_2 n} + \frac{n}{(\log_2 n)^2} \le \frac{n}{2\log_2 n - 8\log_2\log_2 n}$$

Thus, $\chi = \frac{n}{2\log_2 n}\left(1+o(1)\right)$.

Greedy coloring

Consider again $\mathcal{G}(n, \frac{1}{2})$ and color it greedily as follows: order the vertices $v_1, v_2, \dots, v_n$. Color v_1, then v_2, then v_3 etc. such that v_i gets the first available color not used by any of the vertices among $v_1, \dots, v_{i-1}$ that v_i is adjacent to.



Theorem W.h.p., the greedy algorithm colors $\mathcal{G}(n, \frac{1}{2})$ with at most $\dfrac{n}{\log_2 n - 3\log_2\log_2 n}$ colors.

This is within a factor of 2 of the optimal coloring!

Proof Let M denote the number of colors used by the algorithm and let $k = \left\lfloor \dfrac{n}{\log_2 n - 3\log_2\log_2 n} \right\rfloor$. Let A_i be the event that vertex i is the first to receive color $k+1$. Thus,

$$\mathbb{P}(M > k) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

To bound $\mathbb{P}(A_i)$, we may assume that exactly k colors have been used by vertices $1, \dots, i-1$ (otherwise the conditional probability of A_i is zero). Fix a coloring $C_1, \dots, C_k$ of $1, \dots, i-1$. Then

$$\mathbb{P}(A_i \mid C_1,\dots,C_k) = \prod_{j=1}^{k} \underbrace{\left(1-\left(\tfrac{1}{2}\right)^{|C_j|}\right)}_{\substack{=\text{prob that}\\ i \text{ is connected to}\\ \text{at least one vertex}\\ \text{from the color class } C_j}}$$

$$\le \left(1-\left(\tfrac{1}{2}\right)^{\sum_{j=1}^{k} C_j/k}\right)^{k}$$

(because $\log\left(1-\left(\tfrac{1}{2}\right)^{x}\right)$ is convex in $x \geq 0$)

$$= \left(1-2^{-i/k}\right)^{k} \le \left(1-2^{-n/k}\right)^{k} = o\left(\tfrac{1}{n}\right).$$

Thus, $\sum_{i=1}^{n} \mathbb{P}(A_i) \to 0$. $\blacksquare$

See the survey of

Krivelevich: Coloring random graphs: an algorithmic perspective. (2002)