Chromatic and Independence Numbers of $G_{n,\frac{1}{2}}$

We begin by determining the independence number of $G_{n,\frac{1}{2}}$. For any integer a, the expected number of independent sets of size a in $G_{n,\frac{1}{2}}$ is

$$\binom{n}{a}\left(\frac{1}{2}\right)^{\binom{a}{2}} \approx \left(\frac{en}{a}\right)^a \left(\frac{1}{2}\right)^{\binom{a}{2}} = \left(\frac{\sqrt{2} \times en}{a2^{a/2}}\right)^a.$$

This number will switch from big to small at roughly the point where $n = 2^{a/2}$; i.e. $a = 2\log_2 n$. (All logarithms here will have base 2.) More precisely, we let a^* be the (real) value that makes this expression evaluate to one:

$$a^* = 2\log n + 2\log\log n + 2\log e - 1.$$

It is easy to verify that there is some $z = o(1)$ such that when $a = a^* + z$, this expected number is $o(1)$. This implies:

Theorem 1 *Almost surely, the independence number of $G_{n,\frac{1}{2}}$ is at most $\lfloor a^* + o(1)\rfloor$.*

Conversely, we will prove:

Theorem 2 *Almost surely, the independence number of $G_{n,\frac{1}{2}}$ is at least $\lfloor a^* - o(1)\rfloor$.*

These two theorems imply that for every n, the independence number of $G_{n,\frac{1}{2}}$ is a.s. one of two specific integers. Furthermore, unless n is such that a^* evaluates to being within $o(1)$ of an integer, the independence number of $G_{n,\frac{1}{2}}$ is a.s. one specific integer, namely $\lfloor a^*\rfloor$. Thus, we know the (a.s.) independence number precisely for all but a vanishing proportion of values of n.

These theorems were proved independently by Bollobas and Erdos (1976) and Matula (1976).

To prove Theorem 2, we first choose some $w = o(1)$ such that setting $a = \lfloor a^* - w\rfloor$ causes the expected number to evaluate to n^ϵ for some $\epsilon > 0$. (It is easy to show that such a w must exist.) We let X denote the number of independent sets of size a in $G_{n,\frac{1}{2}}$ and we focus on the second moment of X.

We consider any ordering of the $\binom{n}{a}$ different sets of a vertices, and for each $1 \le i \le a$ we define X_i to be the indicator variable which is 1 iff the ith set is an independent set.

$$\begin{aligned}
\mathbf{Exp}(X^2) &= \mathbf{Exp}((\sum_i X_i)^2) \\
&= \mathbf{Exp}(\sum_i X_i^2 + \sum_{i\neq j} X_iX_j) \\
&= \mathbf{Exp}(\sum_i X_i) + \sum_i \mathbf{Pr}(X_i = 1)\sum_{j\neq i}\mathbf{Pr}(X_j = 1|X_i = 1) \\
&= \mathbf{Exp}(X) + \mathbf{Exp}(X)\sum_{j\neq 1}\mathbf{Pr}(X_j = 1|X_1 = 1)
\end{aligned}$$

The third line uses the fact that, since $X_i \in \{0,1\}$, we have $X_i^2 = X_i$. The fourth line uses the fact that, by symmetry, $\sum_{j\neq i}\mathbf{Pr}(X_j = 1|X_i = 1) = \sum_{j\neq 1}\mathbf{Pr}(X_j = 1|X_1 = 1)$.

Chebychev's inequality implies that if $\mathbf{Exp}(X^2) \le (1+\lambda)\mathbf{Exp}(X)^2$, then $\mathbf{Pr}(X = 0) \le \lambda$. Thus, our goal will be to show that this holds with $\lambda = o(1)$.

For each $0 \le k \le a$ we let I_k denote the total contribution made to the sum in the last line by all sets X_j that intersect X_1 in exactly k vertices. Thus, $I_k = \binom{a}{k}\binom{n-a}{a-k}\left(\frac{1}{2}\right)^{\binom{a}{2}-\binom{k}{2}}$. Using the fact that $a \approx 2\log n$, it is easy to evaluate that:

$$I_0 = \mathbf{Exp}(X)(1-o(1)); I_1 = \mathbf{Exp}(X) \times O(\log n/n); I_{a-1} = O(\log n/n); I_k = 1.$$

All other terms are negligible (details omitted), and so we get

$$\mathbf{Exp}(X^2) = \mathbf{Exp}(X)^2(1 + O(\log n/n)).$$

This proves Theorem 2.

Next, we turn to the chromatic number of $G_{n,\frac{1}{2}}$. In any proper colouring of G, every colour class must be an independent set. Hence, every colour class has size at most roughly $2\log_2 n$. Therefore, a.s.

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

Consider the following attempt to colour G with only $\frac{n}{2\log_2 n}(1+o(1))$ colours. G a.s. has an independent set A_1 of size $a^*(n) \approx 2\log_2 n$. Set $G_2 = G - A_1$ and $n_2 = |G_2| \approx n - 2\log_2 n$. Then, treat G_2 as $G_{n_2,\frac{1}{2}}$ and conclude that a.s. it has an independent set A_2 of size $a^*(n_2)$. Repeat this process until G has been partitioned into independent sets $A_1, ..., A_t$. This forms a t-colouring of G (the vertices of A_i all receive colour i.) Working out the appropriate recurrence relations yields that $t = \frac{n}{2\log_2 n}(1+o(1))$.

The problem is that we cannot treat G_2 as $G_{n_2,\frac{1}{2}}$. When we searched for a large independent set, A_1, in G, we had to *expose* all the edges of G and so we can no longer assume that those edges are random.

However, there is a way to find smaller independent sets $A_1, ...$ in a way that allows us to treat $G_2, ...$ as random graphs $G_{n_2,\frac{1}{2}}, ...$ We find A_1 using the following greedy algorithm. We start by setting $U = V(G)$, the entire set of vertices. U will denote the vertices still eligible to be selected for the independent set. We pick an arbitrary vertex $v_1 \in U$. Then we expose the neighbours of v_1 and remove them from U. Then we pick an arbitrary vertex $v_2 \in U$; we expose the neighbours of v_2 in U and remove them from U. We repeat until $U = \emptyset$. Clearly this produces an independent set $v_1, v_2, ...$ At each step, the size of U is roughly cut in half, so we end up choosing an independent set A_1 of size approximately $\log_2 n$. Furthermore, this exposes no information about any edges between two vertices not in A_1, so we can treat G_2 as $G_{n_2,\frac{1}{2}}$ (where $n_2 = |G_2| \approx n - \log_2 n$) and repeat the process. This proves that a.s.

$$\chi(G_{n,\frac{1}{2}}) \leq \frac{n}{\log_2 n}(1+o(1)).$$

This was first proved by Grimmett and McDiarmid in 1975. (We saw here just an outline of the proof - it takes a bit of work to formalize it.)

Finally in 1988, Bollobas tightened this gap by proving

Theorem 3 *A.s.* $\chi(G_{n,\frac{1}{2}}) \leq \frac{n}{2\log_2 n}(1+o(1))$.

Independently, and shortly thereafter, Matula and Kucera proved the same result using a different (and more complicated) argument. We will outline the Bollobas argument here. The key lemma is:

Lemma 4 *A.s.* $G_{n,\frac{1}{2}}$ *has the property that every set of at least* $n/\log^3 n$ *vertices contains an independent set of size* $2\log_2 n - 10\log\log n$.

Lemma 4 implies Theorem 3 because we can repeatedly remove our independent set/colour class A_i from G_i, until $|G_i| < n/\log^3 n$. (By Lemma 4, we don't need to be able to treat G_i as a random graph; all we need to use is the fact that it has at least $n/\log^3 n$ vertices.) This will use fewer than $n/(2\log_2 n - 10\log\log n) = \frac{n}{2\log_2 n}(1+o(1))$ colours. Then we use (fewer than) $n/\log^3 n = o(\frac{n}{2\log_2 n})$ colours to colour the remaining vertices such that each vertex gets its own colour.

Lemma 4 follows from:

Lemma 5 *The probability that $G_{n,\frac{1}{2}}$ has no independent set of size $a^* - 4$ is at most* $\exp(-n^{1.1})$.

This implies Lemma 4 as follows: It is enough to show that Lemma 4 holds for all sets of exactly $n/\log^3 n$ vertices. Set $n' = n/\log^3 n$ and observe that $a^*(n') - 4 > 2\log_2 n - 10\log\log n$. Therefore, the probability that a particular set S of size $n/\log^3 n$ has no independent set of size $2\log_2 n - 10\log\log n$ is, by Lemma 5, at most $\exp(-n'^{1.1}) < \exp(-n)$. Therefore, the expected number of such sets that violate Lemma 4 is less than $\binom{n}{n/\log^3 n}e^{-n} < 2^n e^{-n} = o(1)$. Therefore, a.s. there are no such violating sets.

If we attempt to prove Lemma 5 using the second moment method, we only obtain a probability bound of $1/poly(n)$ which is not nearly enough. Instead Bollobas used Azuma's Inequality. (This was one of the first few uses of Azuma's Inequality on random graphs.) Instead of stating Azuma's Inequality in its fullest power, we use the following watered down version:

Simple Concentration Bound *Suppose that X is a random variable determined by m independent trials $T_1, ..., T_m$. Suppose also that for every set of possible outcomes of the trials, if we alter the outcome of one trial, we change the value of X by at most c. Then for any $t > 0$,*

$$\mathbf{Pr}(|X - \mathbf{Exp}(X)| > t) < 2e^{-t^2/2c^2 m}.$$

In this setting, we will take $m = \binom{n}{2}$ and the m independent trials to be the choices for which of the m potential edges are present.

One might try to apply the SCB to X, the number of independent sets of size $a^* - 4$, but it is easy to see that changing the outcome of one trial, i.e. changing the presence of one edge, might affect X by up to $\binom{n}{a^*-6}$ which is a value of c that is way to big to be useful. So instead, Bollobas defined the variable Y to be the size of a largest collection of independent sets $S_1, ..., S_Y$ such that each S_i has size $a^* - 4$ and no two sets intersect in more than one vertex. Clearly, changing the presence of one edge can affect Y by at most 1. Thus, we can apply the SCB to Y with $c = 1$. One can compute that $\mathbf{Exp}(Y) = \mathbf{Exp}(X)(1 - o(1)) > n^3$. (We omit the details, but note that $\mathbf{Exp}(Y) > \mathbf{Exp}(X) - \sum_{k\geq 2} I_k$.) Thus the SCB, with $t = \mathbf{Exp}(Y)$ yields

$$\mathbf{Pr}(Y = 0) < 2e^{-(n^3)^2/2(1)^2 m} = e^{-O(n^2)},$$

which is sufficent.