Pick an n-vertex graph uniformly at random. Pick another one. Will it have the same chromatic number? Or if not, how different are their chromatic numbers likely to be if n is large?

The chromatic number of a random graph is a random variable, so what we are really asking is: Is this random variable essentially deterministic? That is, is the weight of the distribution concencentrated on one value, or on just a few values which are close together?

Until recently we did not know whether or not this is the case. In this blog post I'll describe a new result showing that, at least for infinitely many n, the chromatic number of a random n-vertex graph is not concentrated on fewer than about $\sqrt{n}$ consecutive values.

Random graphs

A uniform n-vertex graph is generated by the random graph $G_{n,1/2}$ where we include each possible edge independently with probability $1/2$, and more generally we can ask about the chromatic number $\chi(G_{n,p})$ of $G_{n,p}$.

Results about $\chi(G_{n,p})$ generally fall into one of two categories: Can we find (i) upper and lower bounds for its typical value, and (ii) bounds on how much it varies about this typical value?

The typical value

On (i), we have the well-known 1987 result of Bollobás who showed that, with high probability (whp), $\chi(G_{n,1/2}) \sim n/(2\log_2 n)$. This was improved and generalised several times, the current best bounds are from 2016:

$$\chi(G_{n,1/2}) = \frac{n}{2\log_2 n - 2\log_2\log_2 n - 2} + o\left(\frac{n}{\log^2 n}\right) \text{ whp.} \qquad (1)$$

(I am writing out this result because it will become important when talking about question (ii) later.)

Upper concentration bounds

So how about the width of the distribution of $\chi(G_{n,p})$ — what is the length of the shortest interval (or rather sequence of intervals) which is likely to contain $\chi(G_{n,p})$?

Of course (1) is already a weak concentration type result, giving an explicit such interval of length $o\left(\frac{n}{\log^2 n}\right)$. However, it turns out that if we don't have to specify the interval, we can do a lot better.

The starting point for question (ii) is the classic 1987 result of Shamir and Spencer, who showed that for any function $p(n)$, $\chi(G_{n,p})$ is whp contained in some sequence of intervals of length about $\sqrt{n}$. If $p \to 0$ quickly enough, however, much sharper concentration holds: in 1997, Alon and Krivelevich reduced

the interval length to only 2 in the case $p < n^{-1/2-\varepsilon}$. So here the chromatic number behaves almost like a deterministic function; almost all of the weight of the distribution is on two consecutive values.

The opposite question

In view of strong results showing that the chromatic number is sharply concentrated, in the late 1980s Bollobás raised the following question (later popularised by him and Erdős): Can we find any examples where $\chi(G_{n,p})$ is *not* very sharply concentrated? This is trivially true for $p = 1 - 1/(10n)$ (as noted by Alon and Krivelevich), but how about non-trivial examples, and what about the most natural special case, $p = 1/2$?

This question remained open for quite some time, for a simple reason: while we have a number of standard tools to prove upper bounds on the concentration of a random variable (for example the martingale concentration argument of Shamir and Spencer), we have few ways of giving *lower* concentration bounds. (Unless we can work out the entire approximate distribution, for example by proving asymptotic normality.)

A lower bound for concentration

I will sketch the main ideas needed for the following result:

Theorem 1 (H. 2019; H., Riordan 2021)**.** *Let $\varepsilon > 0$, and let $[s_n, t_n]$ be a sequence of intervals such that $\chi(G_{n,1/2}) \in [s_n, t_n]$ whp. Then there are infinitely many values n such that*

$$\ell_n := t_n - s_n \geqslant n^{1/2-\varepsilon}.$$

Of course we can also replace ε with some function $o(1)$ which tends to 0 slowly. Up to this $o(1)$-term, the lower bound matches Shamir and Spencer's upper bound.

A word of caution: The theorem only says that there are *some* n so that $\chi(G_{n,1/2})$ is not very sharply concentrated. It does not tell us what these values n are, and no bound is given for the other n. In fact, we do not believe the conclusion of Theorem 1 is true for every n — more on this later!

Basic proof strategy

The proof needs **two ingredients**:

(1) A (weak) concentration type result

A result that says that, whp, $|\chi(G_{n,1/2}) - f(n)| \leqslant \Delta(n)$ for some well-behaved function $f(n)$ and an error term Δ. Here, Δ is much larger than

the scale on which we are trying to prove non-concentration. We also need a lower bound on the slope of f,

$$\frac{\mathrm{d}}{\mathrm{d}n} f(n) > \frac{1}{\alpha} + \delta,$$

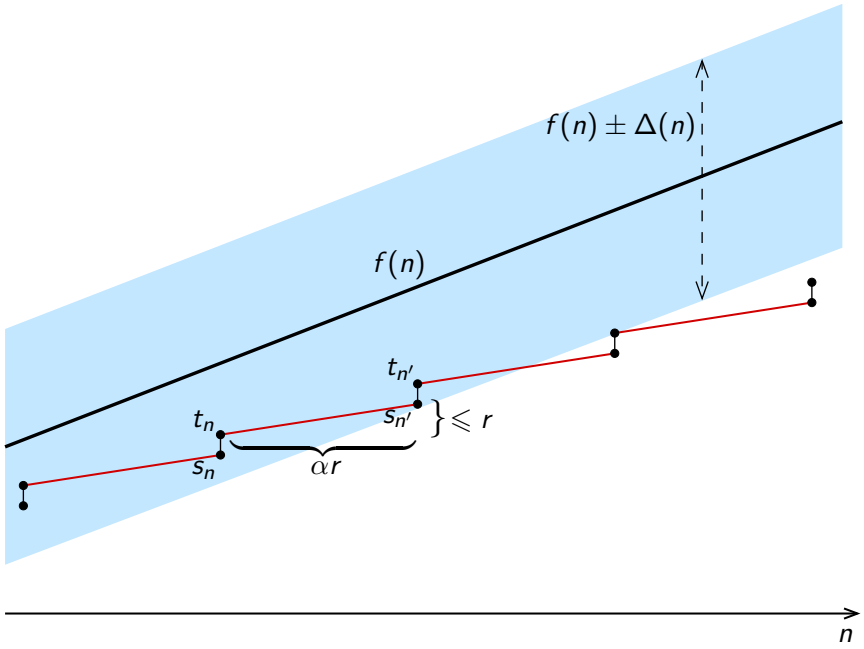
where we will specify $\alpha = \alpha(n)$ later.

(2) A coupling result

A coupling that shows that, for n and some slightly larger n', $\chi(G_{n,1/2})$ and $\chi(G_{n',1/2})$ are close together. More specifically, we need a coupling of $G_{n,1/2}$ and $G_{n',1/2}$ with $n' = n + \alpha r$ (with α as above) so that with probability at least $1/4$, say,

$$\chi(G_{n',1/2}) \leqslant \chi(G_{n,1/2}) + r.$$

Now suppose $[s_n, t_n]$ is a sequence of intervals as in Theorem 1. As shown in the picture, we now know that $\chi(G_{n,1/2})$ is concentrated around a function $f(n)$ with slope more than $1/\alpha$ (blue area), but it also follows from the coupling that the slope *between* the concentration intervals (red lines) is at most $r/(\alpha r) = 1/\alpha$.



If *all* intervals were short, then as we increase n, eventually this would lead to a contradiction: an interval $[s_n, t_n]$ lying outside the blue area. So there must be at least one long interval.

Working out the numbers, under certain reasonable conditions there is an interval of length about $\alpha\delta r$. We will have α of order $\log n$, δ of order $1/\log^2 n$ and r close to $\sqrt{n}$.

Independence number and large independent sets

So what's the function $\alpha(n)$? This turns out to be the independence number of $G_{n,1/2}$, that is, the size of the largest independent vertex set. Matula and independently Bollobás and Erdős proved that the independence number of $G_{n,1/2}$ behaves almost deterministically: for most n, whp it takes an explicitly known deterministic value $\alpha = \alpha(n) \approx 2\log_2 n$.

What does this have to do with $\chi(G_{n,1/2})$? Each colour class in a colouring (i.e. all the vertices of one particular colour) is independent, because neighbouring vertices must be coloured differently. So there are at most $\alpha(n)$ vertices in each colour class, and therefore $\chi(G_{n,1/2}) \geqslant n/\alpha(n)$. We saw in (1) that this easy lower bound is in fact asymptotically correct.

We call an independent set of size α an α-set. It is plausible that the optimal colouring of $G_{n,1/2}$ contains all or almost all α-sets as colour classes. Roughly speaking, this is because the expected number of k-colourings is dominated by colourings with as many α-sets as possible.

Let X_α count the number of α-sets, then it can be shown that X_α is approximately Poisson with mean $\mu_\alpha = n^\theta$, where $\theta = \theta(n) \in [o(1), 1 + o(1)]$. In particular, X_α typically varies by about $\sqrt{\mu_\alpha}$. If we must use all α-sets in the colouring, then it seems plausible that the overall number of colours we end up with also varies by roughly this amount. (Divided by $\log n$, as we will see in a moment.)

The weak concentration result

Luckily we already have a suitable estimate for $\chi(G_{n,1/2})$ in (1). Writing $f(n)$ for this estimate, then for most n,

$$\frac{\mathrm{d}}{\mathrm{d}n} f(n) = \frac{1}{\alpha} + \Theta\left(\frac{1}{\log^2 n}\right).$$

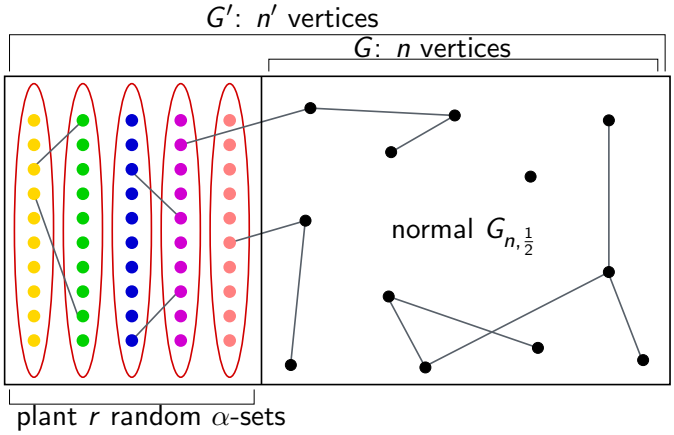
So what is the effect of having one extra α-set in $G_{n,1/2}$? Using one colour for this α-set, we need to colour the remaining $n - \alpha$ vertices. So one extra α-set should save us about

$$\alpha\left(\frac{\mathrm{d}}{\mathrm{d}n} f(n)\right) - 1 = \Theta\left(\frac{1}{\log n}\right)$$

colours. So $\sqrt{\mu_\alpha}$ extra α-sets should reduce $\chi(G_{n,1/2})$ by about $\sqrt{\mu_\alpha}/\log n$.

The coupling

We construct the coupling in the following way: take $n' = n + r\alpha$ vertices (we will pick r later). Choose r vertex sets of size α uniformly at random and make them independent sets, and then include every edge outside these r α-sets independently with probability $1/2$.



The inner graph G on n vertices is simply $G_{n,1/2}$. It is also clear that, if G' is the outer graph on n' vertices, then

$$\chi(G') \leqslant \chi(G) + r.$$

This is because any colouring of G can be extended to a colouring of G' by giving a new colour to each of the r α-sets.

The trouble is that G' is not distributed as $G_{n',1/2}$ — we have disturbed the distribution by planting the r α-sets. The key point is that, as long as r is not too large — less than the standard deviation $\sqrt{\mu_\alpha}$ of X_α — then G' and $G_{n',1/2}$ are very similar. This makes sense on an intuitive level: the random graph doesn't 'notice' the extra α-sets as long as we have planted fewer than the natural fluctuation $\sqrt{\mu_\alpha}$.

So we let $r = o(\sqrt{\mu_\alpha})$ (or in fact $r = \varepsilon\sqrt{\mu_\alpha}$ works). For the formal proof, we bound the total variation distance of the two random graph models.

Tying it all together

The two ingredients above can be combined to show that for every n, there is some nearby n^* with concentration interval length $\ell_{n^*} \geqslant C\sqrt{\mu_\alpha(n^*)}/\log n^*$, with $\mu_\alpha(n^*) \approx \mu_\alpha$. If we pick n so that μ_α is close to n, Theorem 1 follows.

How close to Shamir and Spencer's upper bound $\sqrt{n}$ can we actually get? At the moment, nothing better than $n^{1/2-o(1)}$ for some unspecified function $o(1)$ is possible. The main bottleneck is the size of the error term $\Delta = o(n/\log^2 n)$ in (1).

Konstantinos Panagiotou and I have been working on an improved estimate for $\chi(G_{n,1/2})$ in a paper we are currently writing up. Assuming this result, Oliver Riordan and I can prove that there are infinitely many n such that

$$\ell_n \geqslant c\frac{\sqrt{n}\log\log n}{\log^3 n}.$$

The best upper concentration bound for $p = 1/2$ is $\sqrt{n}/\log n$, due to Alon.

Which of these is closer to the truth? We conjecture that for the worst case n, the width of the distribution of $\chi(G_{n,1/2})$ matches our lower bound up to the constant. However, the concentration behaviour seems to be very different for different n, as described below.

The Zigzag conjecture

Let $g(n)$ be the standard deviation of $\chi(G_{n,1/2})$. We already gave a heuristic argument that $g(n) \geqslant C\sqrt{\mu_\alpha}/\log n$, coming from fluctuations in the number X_α of α-sets.

There is another conjectured lower bound coming from fluctuations in $X_{\alpha-1}$, the number of $(\alpha-1)$-sets, namely

$$g(n) \geqslant C\frac{\sqrt{n}}{\sqrt{\mu_\alpha}\log^{5/2} n}.$$

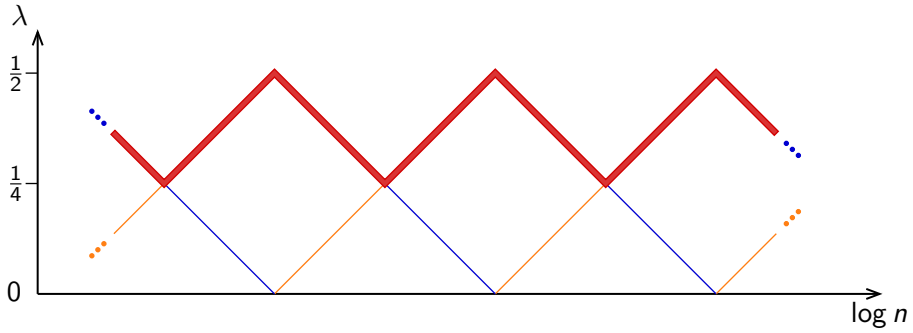
The Zigzag conjecture, made recently by Bollobás, Heckel, Morris, Panagiotou, Riordan and Smith, states that at least for most n, $g(n)$ is essentially the maximum of these two lower bounds. Ignoring $n^{o(1)}$-terms, we have the following simplified statement.

Conjecture 2 (Zigzag conjecture)**.** *Define θ by letting $\mu_\alpha = n^\theta$. Let*

$$\lambda = \max\left(\frac{\theta}{2}, \frac{1-\theta}{2}\right).$$

Then, if $g(n)$ denotes the standard deviation of $\chi(G_{n,1/2})$,

$$g(n) = n^{\lambda+o(1)}.$$



It is not hard to show that the function $\lambda(n)$ 'zigzags' between $1/4 + o(1)$ and $1/2 + o(1)$, roughly linearly in $\log n$, as shown in the picture (the orange and blue lines are the lower bounds coming from α- and $(\alpha-1)$-sets, respectively).

There's a detailed heuristic explanation of the conjecture in the paper. We also think that we have a pretty good idea of the behaviour of $g(n)$ at the extreme points. In particular, we believe that in the 'worst case', $g(n)$ is of order $n^{1/2}\log\log n/\log^3 n$, while in the 'best case' it's of order $n^{1/4}/\log^{7/4} n$.