How does the chromatic number of a random graph vary?

Oliver Riordan

University of Oxford

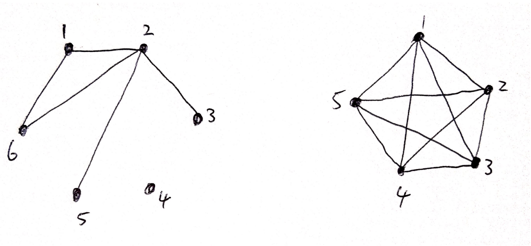
COLMEA, PUC-Rio, Gávea

22 November 2023

Joint work (and slides!) with Annika Heckel.

Graphs

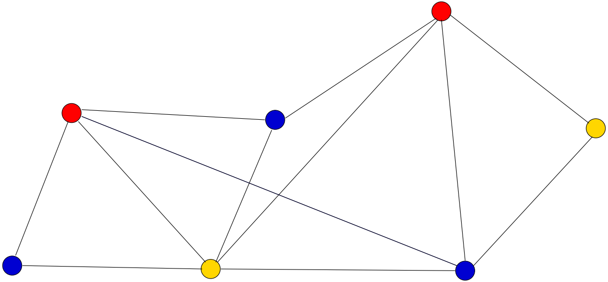
A **graph** consists of a set V of **vertices** (nodes) and a set E of **edges** (links), where an edge is just a pair of distinct vertices.



The data defining a graph G is: what are its vertices, and which pairs of vertices are adjacent, i.e., joined by an edge.
Given a set V of n vertices, there are thus $2^{\binom{n}{2}}$ possible graphs on V.

What is a colouring?

Colouring of G: Colour vertices so that neighbours get different colours.



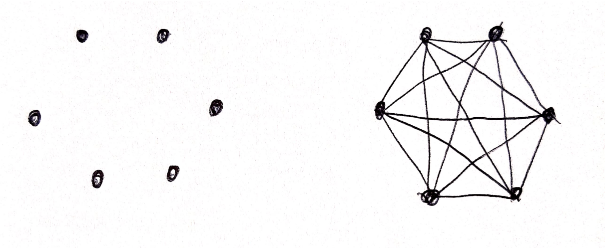
Chromatic number $\chi(G)$: Minimum number of colours we need.

Sounds like a game..., but important in applications.

Chromatic numbers

What can we say about the chromatic numbers of all graphs on a set V of size n?

Range of values is 1 to n:



What is the typical value? What is the typical spread?

Chromatic number of random graphs

Pick a graph on $V = \{1, 2, \ldots, n\}$ uniformly at random. Or,

Consider $G_{n,1/2}$: choose a graph on V by including each possible edge independently with probability $1/2$.

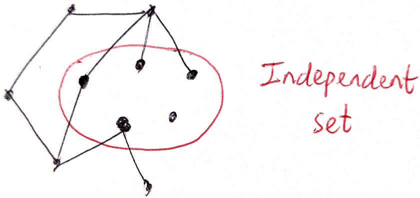
What can we say about $\chi(G_{n,p})$?

Value?

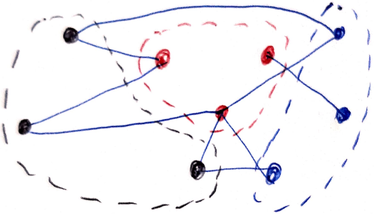
Upper and lower bounds?

Concentration?

How much does $\chi(G_{n,p})$ vary?
Upper and lower bounds?



An independent set in G is a set of vertices spanning no edges.



Colouring is the same as partitioning into independent sets.

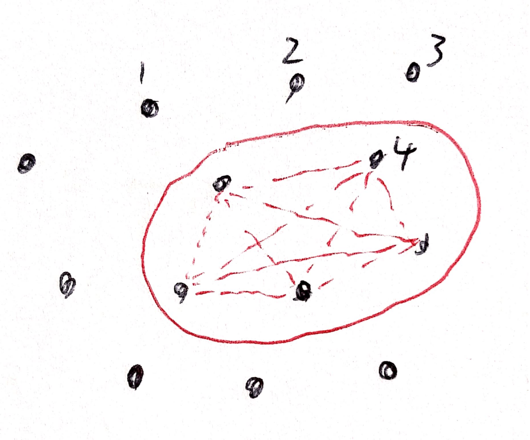
Independence number

The independence number $\alpha(G)$ is the size of a largest independent set in G.

What do we expect $\alpha(G_{n,1/2})$ to be?

For each k, consider the random variable

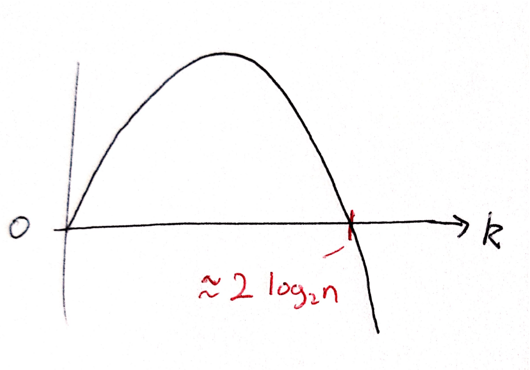
$$X_k = \text{number of } k\text{-vertex independent sets in } G_{n,1/2}.$$



$$\mathbb{E}X_k = \binom{n}{k}(1-p)^{\binom{k}{2}} = \binom{n}{k}2^{-\binom{k}{2}}.$$

$$\mathbb{E}X_k = \binom{n}{k} 2^{-\binom{k}{2}} \approx n^k 2^{-k^2/2}$$

$$\log_2(\mathbb{E}X_k) \approx k \log_2 n - k^2/2.$$

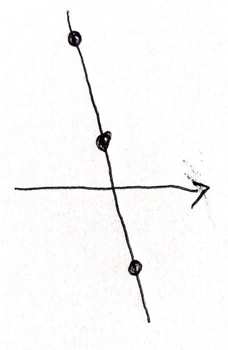


Sad parabola!

Up to some size $\alpha(n) \approx 2\log_2 n$ we have $\mathbb{E}[X_k] \geqslant 1$, and usually $\mathbb{E}[X_k]$ very large.

For $k > \alpha(n)$ we have $\mathbb{E}[X_k] < 1$, and usually very small.

For k near the crossing point we have $\mathbb{E}[X_{k+1}]/\mathbb{E}[X_k] \approx n^{-1}$.

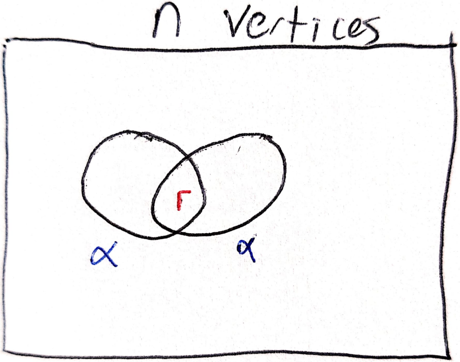


Usually $\mathbb{E}[X_{\alpha+1}]$ is small, so whp (with high probability) no independent sets of this size.
Usually $\mathbb{E}[X_\alpha]$ is large... does that mean X_α is?

Consider the *second moment* or *variance* of X_α.

Involves summing over pairs S, S' of sets of size α.

Only relevant parameter: $r = |S \cap S'|$.



Easy calculation: $\mathrm{var}[X_\alpha] \approx \mathbb{E}[X_\alpha]$, and in fact close to Poisson distribution.

Conclusion: (usually) whp $\alpha(G_{n,1/2}) = \alpha(n)$ for a known value $\alpha \approx 2\log_2 n$.

In a colouring with c colours, $n \leqslant \alpha c$. Thus whp

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

Further bounds

Grimmett + McDiarmid 1975:

$$(1 - o(1))\frac{n}{2\log_2 n} \leqslant \chi(G_{n,\frac{1}{2}}) \leqslant (1 + o(1))\frac{n}{\log_2 n} \text{ whp.}$$

Bollobás 1987: $$\chi(G_{n,\frac{1}{2}}) \sim \frac{n}{2\log_2 n} \text{ whp.}$$

Improvements: McDiarmid '90, Panagiotou & Steger '09, Fountoulakis, Kang & McDiarmid '10.

Heckel 2016:

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

Explicit interval of length $o\left(\frac{n}{\log^2 n}\right)$ which contains $\chi(G_{n,\frac{1}{2}})$ whp.

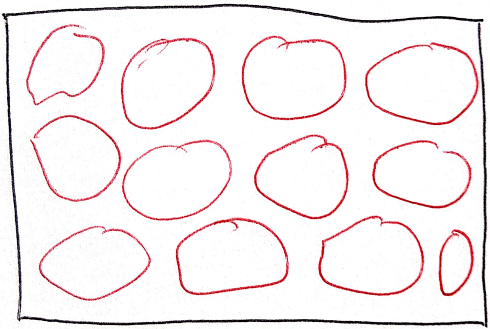
Idea

Directly study the random variable $X =$ number of colourings.

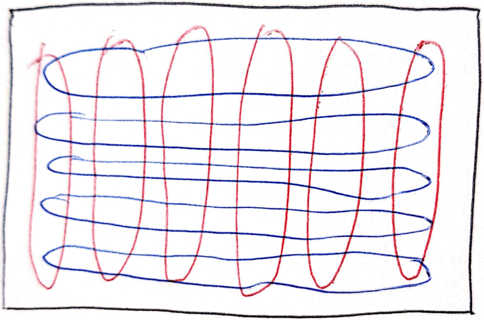
Calculate $\mathbb{E}X$ and $\mathrm{var}[X]$.

The first is (relatively) easy:

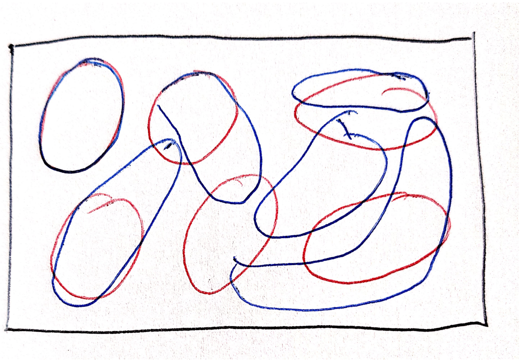
One partition



Two partitions



Two partitions



Complications?

What about other edge probabilities p?

In general, p can vary with n, and we expect very different behaviour if $p \to 1$ or $p \to 0$.

What about constant $p \in (0,1)$? Surely the same?

NO! Annika showed $p > 1 - 1/e^2 \approx 0.865$ is different!

Recent developements

Theorem (Heckel, Panagiotou 23)

If $\mu_\alpha \geqslant n^{0.1}$, whp

$$\chi_{\alpha-1}(G_{n,\frac{1}{2}}) = \mathbf{k}_{\alpha-1} + O(n^{0.99}).$$

- $\chi(G_{n,1/2})$ and $\chi_{\alpha-1}(G_{n,1/2})$ differ by at most about μ_α

$\rightarrow$ This gets us **sharper upper and lower bounds** for $\chi(G_{n,1/2})$.

Proof is clever and complicated!

What can we say about $\chi(G_{n,p})$?

Value?

Upper and lower bounds?

Concentration?

How much does $\chi(G_{n,p})$ vary?
Upper and lower bounds?

Concentration?

Shamir, Spencer 1987: For *any* function $p = p(n)$, $\chi(G_{n,p})$ is whp contained in a sequence of intervals of length about $\sqrt{n}$.

Proof uses martingales!

Martingales

Informally: a sequence of random variables (M_t) where the expected value of M_{t+1} 'at time t' is M_t.

E.g., total winnings after a sequence of fair bets: $M_{t+1} = M_t + S_t W_{t+1}$, where S_t (amount bet) depends on what's happened so far, but conditional on the past, W_{t+1} is ± 1 with equal probability.

Formally, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ a sequence of σ-algebras (information available at time t), and

- M_t is $\mathcal{F}_t$-measurable (known at time t),
- $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = M_t$.

(We are in a finite probabliity space here, so no problems with integrability.)

Martingales

Key property: a martingale 'behaves much like' a sum of independent random variables. In particular

Theorem (Hoeffding, Azuma)

Let $(M_t)_{t=0}^n$ be a martingale such that $|M_{t+1} - M_t| \leqslant C$ for every t. Then for any x

$$\mathbb{P}(|M_n - M_0| \geqslant x) \leqslant e^{-x^2/(2C^2 n)}.$$

In other words, similar Gaussian tails to sum of n independent RVs with variance C^2, whose sum has variance $C^2 n$.

Lipschitz functions

A function defined on graphs is 1-(vertex)-Lipschitz if, whenever $G - v = G' - v$, then $|f(G) - f(G')| \leqslant 1$.

E.g., chromatic number! We have

$$\chi(G - v) \leqslant \chi(G) \leqslant \chi(G - v) + 1.$$

$$\chi(G - v) \leqslant \chi(G') \leqslant \chi(G - v) + 1.$$

So $|\chi(G) - \chi(G')| \leqslant 1$.

Vertex exposure martingales

Let $f : \mathcal{G} \to \mathbb{R}$ be a function defined on graphs. Let $G = G_{n,p}$ be random.

Let $\mathcal{F}_t$ be the information describing which edges among vertices $1, \ldots, t$ are present. Then $(\mathcal{F}_t)$ is a filtration, so $M_t := \mathbb{E}[f(G) \mid \mathcal{F}_t]$ is a martingale!

Called the 'vertex exposure martingale'.

Note that $M_0 = \mathbb{E}[f(G)]$ (no information), while $M_n = f(G)$: complete information.

If f is 1-Lipschitz, $|M_{t+1} - M_t| \leqslant 1$. (Easy argument.)
Shamir–Spencer follows!

Concentration?

Shamir, Spencer 1987: For *any* function $p = p(n)$, $\chi(G_{n,p})$ is whp contained in a sequence of intervals of length about $\sqrt{n}$.

$p = 1 - \frac{1}{10n}$: not concentrated on fewer than $\Theta(\sqrt{n})$ values

$p \leqslant \frac{1}{2}$: slight improvement to $\frac{\sqrt{n}}{\log n}$ (Alon)

(Clever idea: use $f(G)$ minimum number of vertices to delete until can colour with k colours, for suitable k.)

$p < n^{-\frac{1}{2}-\varepsilon}$: 2 values ('two-point concentration')
(Alon, Krivelevich 97, Łuczak 91)

$\rightarrow \chi(G_{n,p})$ behaves almost deterministically

The opposite question

Question (Bollobás, Erdős)

Can we show that $\chi(G_{n,\frac{1}{2}})$ is not concentrated on 100 consecutive values?

Any non-trivial examples of non-concentration?

"even the weakest results claiming lack of concentration would be of interest"

The opposite question

Theorem (Heckel 2019)

$\chi(G_{n,\frac{1}{2}})$ is not contained whp in any sequence of intervals of length $n^{1/4-\varepsilon}$ for any fixed $\varepsilon > 0$.

More formally:

Theorem (Heckel 2019)

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

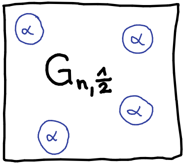
$$t_n - s_n \geqslant n^{1/4-\varepsilon}.$$

Intuition

Intuition: An optimal colouring of $G_{n,\frac{1}{2}}$ contains all or almost all independent α-sets as colour classes.

$\chi(G_{n,\frac{1}{2}})$ **should vary at least as much as** X_α **(roughly).**

$X_\alpha = \#$ independent α-sets

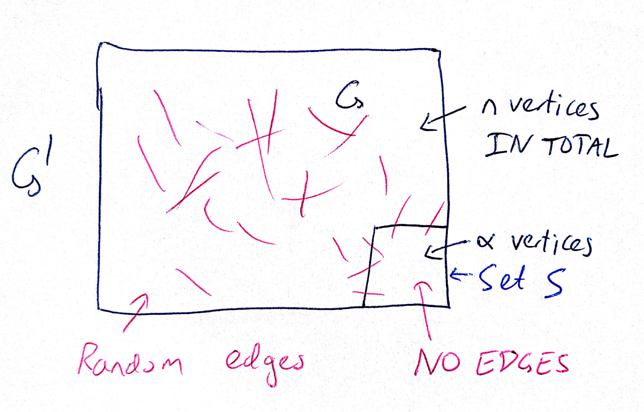


$$X_\alpha \underset{\text{roughly}}{\sim} \mathrm{Po}(\mu) \quad \to \text{varies by } \pm\sqrt{\mu}$$

where $\mu = n^\theta, \quad 0 \leqslant \theta(n) \leqslant 1.$

Proof idea

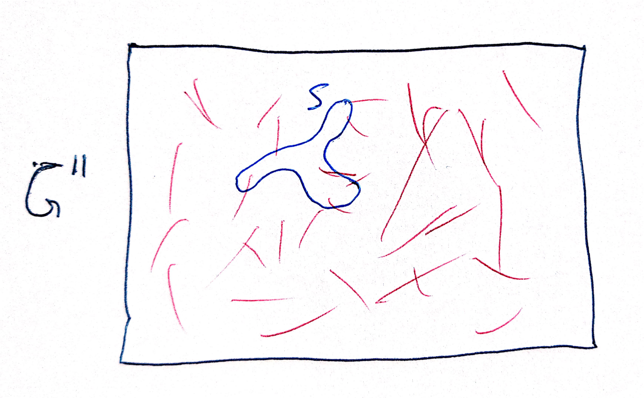
Plant an extra independent α-set:



Starting from $G = G_{n-\alpha,1/2}$, obtain a new random graph G', with $\chi(G') \leqslant \chi(G) + 1$.

Hide the hole

Does G' look like $G_{n,1/2}$? Not quite - it has a hole.
Hide the hole! Shuffle the vertices.



Distribution of G''

For a possible outcome H (graph on n vertices), what is $\mathbb{P}(G'' = H)$?

S must get mapped to an independent set in H of size α.

There are $X_\alpha(H)$ of these.

Given any one,

$$\mathbb{P}(G'' = H \text{ with } S \text{ mapped here}) = \frac{1}{\binom{n}{\alpha}}(1/2)^{\binom{n}{2}-\binom{\alpha}{2}}.$$

KEY: $\mathbb{P}(G'' = H)$ is proportional to $X_\alpha(H)$.

Size-biased distribution

Given any random variable Z, and size parameter $s(Z)$, the size-biased distribtuion Z^* has

$$\mathbb{P}(Z^* = z) = \frac{\mathbb{P}(Z = z)s(z)}{\mathbb{E}s(Z)}.$$

Common in various statistics contexts.

Our G'' has distribution of $G_{n,1/2}$ size-biased by X_α.

Easy check: since X_α is concentrated, size-biasing makes little difference:

$$d_{\mathrm{TV}}(G'', G_{n,1/2}) \leqslant \frac{1}{2}\mathbb{E}\left[\frac{|X_\alpha - \mu|}{\mu}\right] = O\left(\frac{1}{\sqrt{\mu}}\right)$$

where $\mu = \mathbb{E}[X_\alpha]$.

Key Lemma

$$d_{\mathrm{TV}}\left(G_{n,\frac{1}{2}}, G''\right) = O\left(\frac{1}{\sqrt{\mu}}\right),$$

where $\mu = \mathbb{E}[X_\alpha]$.

Proof:

$$\begin{aligned}
d_{\mathrm{TV}}\left(G_{n,\frac{1}{2}}, G''\right) &= \frac{1}{2}\sum_G \left|\mathbb{P}\left(G'' = G\right) - \mathbb{P}\left(G_{n,\frac{1}{2}} = G\right)\right| \\
&= \frac{1}{2}\sum_G \left|\frac{X_\alpha(G)}{\binom{n}{\alpha}}\left(\frac{1}{2}\right)^{\binom{n}{2}-\binom{\alpha}{2}} - \left(\frac{1}{2}\right)^{\binom{n}{2}}\right| \\
&= \frac{1}{2}\sum_G \left(\frac{1}{2}\right)^{\binom{n}{2}} \frac{\left|X_\alpha(G) - \binom{n}{\alpha}\left(\frac{1}{2}\right)^{\binom{\alpha}{2}}\right|}{\binom{n}{\alpha}\left(\frac{1}{2}\right)^{\binom{\alpha}{2}}} \\
&= \frac{1}{2}\mathbb{E}\left[\frac{|X_\alpha - \mu|}{\mu}\right] = O\left(\frac{1}{\sqrt{\mu}}\right)
\end{aligned}$$

□

Coupling

Outcome: can couple $G_{n-\alpha,1/2}$ and $G_{n,1/2}$ so that

$$\chi(G_{n,1/2}) \leqslant \chi(G_{n-\alpha,1/2}) + 1$$

with failure probability $O(1/\sqrt{\mu})$.

For r up to around $\sqrt{\mu}$ can chain: there is a coupling so that

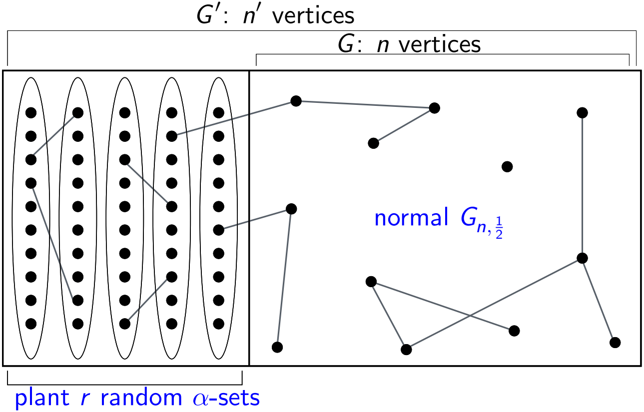
$$\chi(G_{n+r\alpha,1/2}) \leqslant \chi(G_{n,1/2}) + r$$

with failure probability $\leqslant 0.1$, say.

The coupling result

Coupling of $G_{n,1/2}$ and $G_{n',1/2}$ with $n' = n + \alpha r$ so that

$$\mathbb{P}\Big(\chi(G_{n',1/2}) \leqslant \chi(G_{n,1/2}) + r\Big) > 0.9.$$



Proof ingredients

Ingredient 1: A (weak) concentration type result

$$|\chi(G_{n,1/2}) - f(n)| \leqslant \Delta(n) \text{ whp}$$

where $f(n)$ is some function with slope

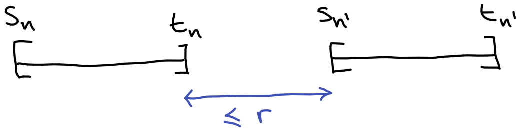
$$\frac{\mathrm{d}}{\mathrm{d}n} f(n) > \frac{1}{\alpha} + \delta.$$

Ingredient 2: A coupling result

Couple $G_{n,1/2}$ and $G_{n',1/2}$ with $n' = n + \alpha r$ (same α as above) so that

$$\mathbb{P}\Big(\chi(G_{n',1/2}) \leqslant \chi(G_{n,1/2}) + r\Big) > 0.9.$$

Trick: Suppose that $\chi(G_{n,\frac{1}{2}}) \in [s_n, t_n]$ whp.



Ingredient 2: A coupling result

Couple $G_{n,1/2}$ and $G_{n',1/2}$ with $n' = n + \alpha r$ (same α as above) so that

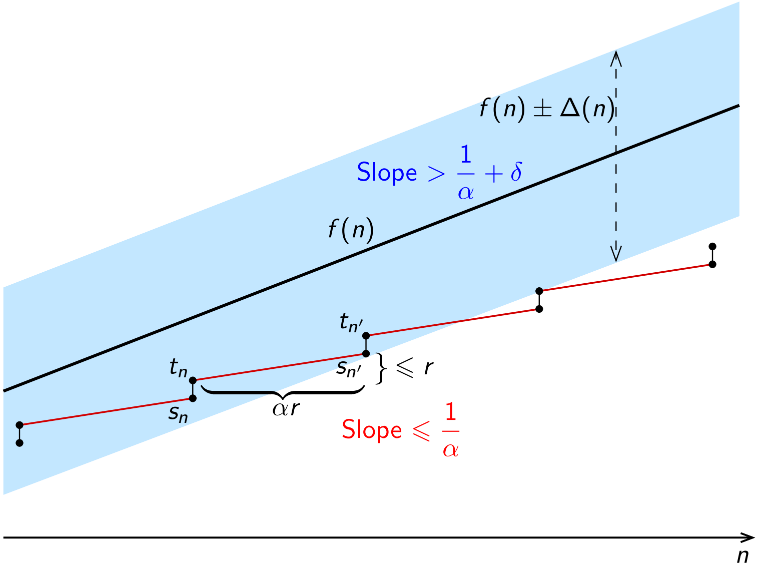
$$\mathbb{P}\Big(\chi(G_{n',1/2}) \leqslant \chi(G_{n,1/2}) + r\Big) > 0.9.$$

Trick: Suppose that $\chi(G_{n,\frac{1}{2}}) \in [s_n, t_n]$ whp.

Why? Because with probability at least 0.8,

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

But $s_{n'}$ and t_n are not random.



If all intervals **short**: Contradiction!

So there is at least one long interval. (Length $\approx \alpha\delta r$)

Ingredient 1: The (weak) concentration type result

Want:

$$\chi(G_{n,\frac{1}{2}}) = f(n) \pm \Delta(n)$$

$$\frac{\mathrm{d}f}{\mathrm{d}n} \geqslant \frac{1}{\alpha} + \delta$$

Heckel 2016:

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

then (unless μ_α is very close to n)

$$\frac{\mathrm{d}f}{\mathrm{d}n} \geqslant \frac{1}{\alpha} + \underbrace{\Theta\left(\frac{1}{\log^2 n}\right)}_{\delta(n)}$$

Remark: something very odd about this proof!

Theorem (Heckel, R. 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

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

Same idea, but (quite a bit) more calculation. With even more (+Heckel–Panagiotou):

Theorem (Heckel, R. 2021/3)

Concentration interval length of $\chi(G_{n,1/2})$ is at least

$$C\,\frac{n^{1/2}\log\log n}{\log^3 n}$$

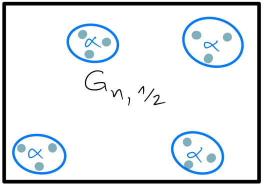
for infinitely many n.

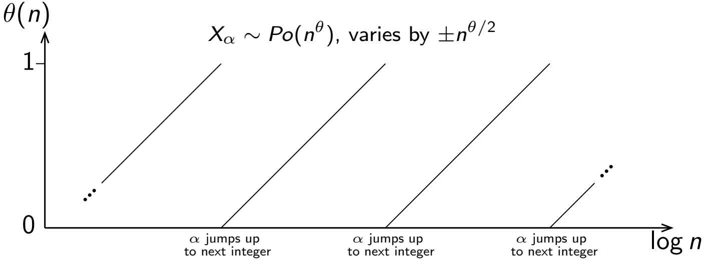
What going on? Number of α-sets

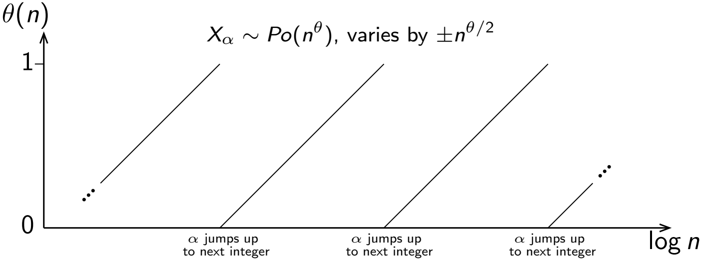
$X_\alpha = \#$ independent α-sets

$X_\alpha \underset{\text{roughly}}{\sim} \mathrm{Po}(\mu_\alpha)$

$\mu_\alpha = n^\theta, \qquad 0 \leqslant \theta(n) \leqslant 1.$







Benefit per α-set: $\approx 1/\log n$ colours.

Conjecture 1: $\chi(G_{n,\frac{1}{2}})$ is not concentrated on fewer than $n^{\theta/2}/\log n$ values.

Theorem(Heckel, R. 2021)

Suppose that $\chi(G_{n,\frac{1}{2}}) \in [s_n, t_n]$ whp for some sequence $[s_n, t_n]$ of intervals. Then for every n with $\theta(n) < 1 - \varepsilon$, there is some $n^* \sim n$ such that

$$t_{n^*} - s_{n^*} \geqslant C(\varepsilon) \cdot \frac{(n^*)^{\theta(n^*)/2}}{\log n^*}.$$

How about $(\alpha - 1)$-sets?

$$X_{\alpha-1} = \#\text{ independent } (\alpha-1)\text{-sets}$$

$$X_{\alpha-1} \underset{\text{roughly}}{\sim} \mathrm{Po}(\mu_{\alpha-1})$$

$$\mu_{\alpha-1} = n^{1+\theta+o(1)}, \qquad 0 \leqslant \theta(n) \leqslant 1.$$

New heuristic: The chromatic number is close to (or at least varies like) the **first moment threshold** (smallest k so that expected number of k-colourings $\geqslant 1$).

Benefit per $(\alpha-1)$-set: $$\frac{n}{\mu_{\alpha-1}\log^3 n}$$

Conjecture 2: $\chi(G_{n,\frac{1}{2}})$ is not concentrated on fewer than

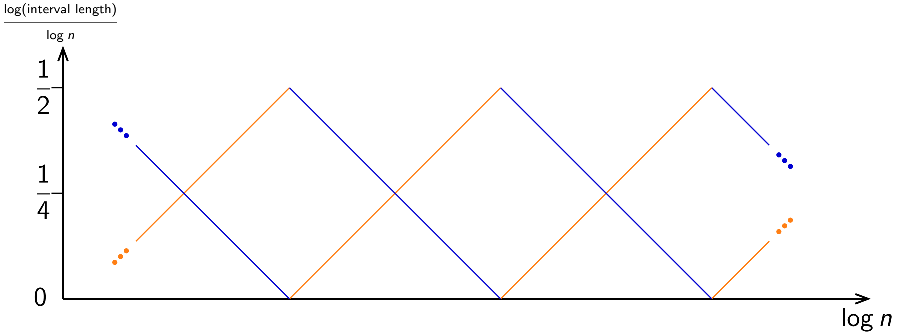
$$\sqrt{\mu_{\alpha-1}} \cdot \frac{n}{\mu_{\alpha-1}\log^3 n} \quad \approx \quad \frac{n^{(1-\theta)/2}}{\log^{5/2} n}$$

values.

Conjectured lower bounds on concentration

From α-sets: $n^{\theta/2+o(1)}$

From $(\alpha - 1)$-sets: $n^{(1-\theta)/2+o(1)}$



Is that all? $\chi(G_{n,1/2}) \approx \dfrac{n}{\alpha_0 - 3.89}$, so what about $(\alpha - 2)$-sets, $(\alpha - 3)$-sets, ...?

Bounded colourings: Two-point concentration

$\chi_t(G)$: t-**bounded chromatic number** — all colour classes $\leqslant t$ vertices

Theorem (Heckel, Panagiotou 23)

Let $m = \left\lfloor \frac{1}{2}\binom{n}{2} \right\rfloor$. There is some integer $k(n)$ so that, whp,

$$\chi_{\alpha-2}(G_{n,m}) \in \{k(n), k(n)+1\}$$

$\mathbf{k}_t$: t-**bounded first moment threshold** — smallest k so that expected number of t-bounded colourings is $\geqslant 1$

$$k(n) \in \{\mathbf{k}_{\alpha-2} - 1, \mathbf{k}_{\alpha-2}\}$$

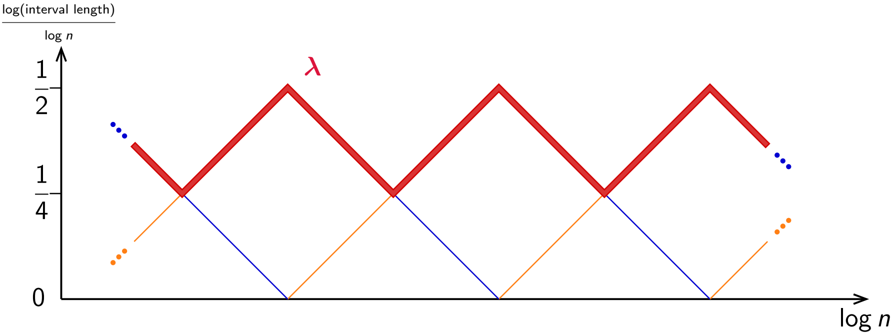
The Zigzag Conjecture

Recall $\mu_\alpha = n^\theta$, $0 \leqslant \theta \leqslant 1$.

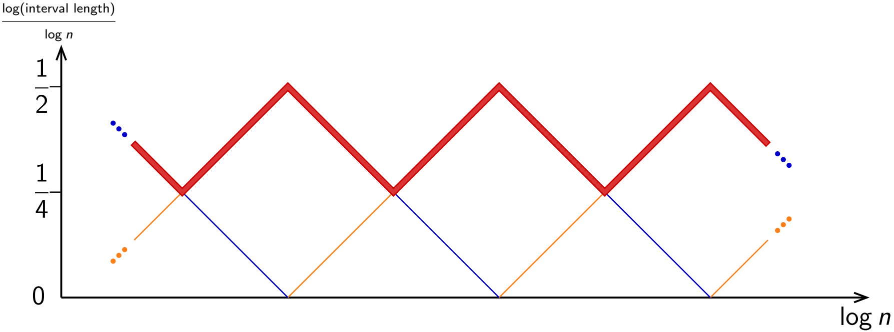
Conjecture 3 (Bollobás, Heckel, Morris, Panagiotou, R., Smith):
Let

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

then the correct concentration interval length for $\chi(G_{n,\frac{1}{2}})$ is $n^{\lambda+o(1)}$.



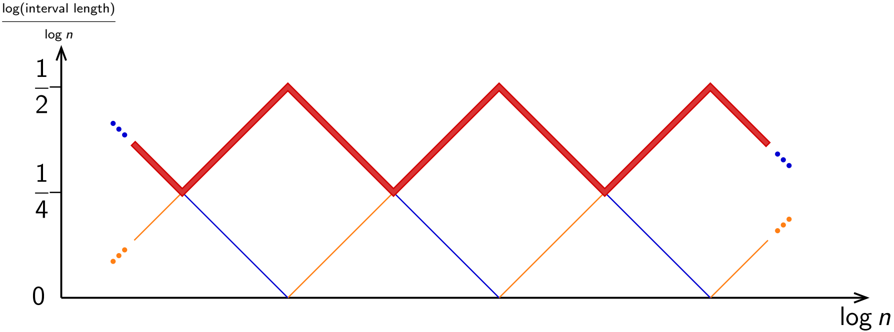
Peaks and valleys



Conjecture 4: The top of the zigzag is of order $\dfrac{n^{1/2} \log \log n}{\log^3 n}$.

Conjecture 5: The bottom of the zigzag is of order $\dfrac{n^{1/4}}{\log^{7/4} n}$.

Asymptotic distribution



Conjecture 6: Gaussian limiting distribution.

(And we can read out a formula for the conjectured standard deviation from our heuristics.)

Open questions

- The proof only finds some n^* near n where the chromatic number is not too concentrated. Can we prove something for every n?
- Does the correct concentration interval length zigzag between $n^{1/4+o(1)}$ and $n^{1/2+o(1)}$? What about the other conjectures?
- Alon's upper bound: $\frac{\sqrt{n}}{\log n}$. Our lower bound: $\frac{\sqrt{n}\log\log n}{\log^3 n}$. Show that this is optimal?
- Other ranges of p?

  $p < n^{-\frac{1}{2}-\varepsilon}$: two-point concentration. How "far down" does non-concentration go?
  $p \to 1$: Infinitely many 'jumps' in concentration behaviour? (Recent conjecture by Surya and Warnke 2022)

Thank you!