Topics in Random Graphs

Notes of the lecture by Michael Krivelevich

Thomas Rast

Spring Semester 2010

Acknowledgements Many thanks to Michael Krivelevich for his corrections and suggestions for these notes, and to Luca Gugelmann for taking notes while I was absent.

Disclaimer Despite our best efforts there are probably still some typos and mistakes left.

Corrections and other feedback would be greatly appreciated and can be sent to `trast@inf.ethz.ch`. Please always state the version you were looking at. If you want to provide patches but do not have a login for the mitschriften.ethz.ch subversion repository, you can also use git to clone

`git://git.thomasrast.ch/topics_in_random_graphs.git`

Version You are looking at version 30a5dd2 of this document.

Contents

Contents **iii**

1 Introduction **1**

1.1 Basic inequalities and tools 1
1.2 Basic models: $G(n,p)$, $G(n,m)$, random graph process 4
1.3 Staged exposure 6
1.4 Monotonicity 6
1.5 Reminder from Graph Theory 7
1.6 Three illustrative examples 8
1.6.1 Lower bounds for Ramsey numbers 8
1.6.2 Graphs with high girth and high chromatic number . 9
1.6.3 Hadwiger's Conjecture 10
1.7 Asymptotic equivalence between $G(n,p)$ and $G(n,m)$ 14

2 Random Regular Graphs **17**

2.1 Preliminaries 17
2.2 Configurations 19

3 Long paths and Hamiltonicity **25**

3.1 Long paths 25
3.2 Hamiltonicity 29
3.2.1 Combinatorial background 30
3.2.2 Pósa's rotation-extension technique 30
3.2.3 Pósa's lemma 31
3.2.4 Hamiltonicity of $G(n,p)$ 33

4 Random Graph Processes and Hitting Times **39**

4.1 Hitting Time of Connectivity 40

5 Coloring Random Graphs **43**

5.1 Graph Theoretic Background 43
5.2 Elementary Facts About Coloring 43
5.3 Greedy Coloring 44
5.4 Coloring $G(n, \frac{1}{2})$ 44
5.4.1 Lower bound for $\chi(G(n, \frac{1}{2}))$ 44
5.4.2 Greedy's Performance 45
5.4.3 Lower bounding $\chi_g(G)$ 47
5.4.4 Chromatic Number of $G(n, 1/2)$ 48
5.5 List Coloring of Random Graphs 53
5.5.1 Combinatorial Background: List Coloring 53
5.5.2 Choosability in Random Graphs 55

Bibliography **59**

Index **61**

Chapter 1

Introduction

Tentative plan:

- models of random graphs and random graph processes
- random regular graphs
- long paths and Hamilton cycles in random graphs
- coloring problems in random graphs
- sharp thresholds
- eigenvalues of random graphs and algorithmic applications
- pseudo-random graphs

Not to be covered:

- phase transition
- appearance of a copy of a fixed graph H

Assume:

- working familiarity with basic notions of graph theory
- knowledge of basic notions in probability, linear algebra

1.1 Basic inequalities and tools

Theorem 1.1 *For all x: $1 + x \leq e^x$.*

Proof: $f(x) = e^x - 1 - x$ attains its minimum at $x = 0$. As a corollary: For all $x \geq 0$, $1 - x \leq e^{-x}$.

Proposition 1.2 *For* $1 \leq k \leq n$,

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k .$$

Proof Lower bound:

$$\begin{aligned}\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)\cdot(n-k+1)}{k(k-1)\cdot 1} \\ = \frac{n}{k}\cdot\frac{n-1}{k-1}\cdot\frac{n-k+1}{1} \geq \left(\frac{n}{k}\right)^k .\end{aligned}$$

Upper bound: we will prove

$$\sum_{i=0}^{k}\binom{n}{i} \leq \left(\frac{en}{k}\right)^k .$$

Note that for all x, $(1+x)^n = \sum_{i=0}^{n}\binom{n}{i}x^i$ hence if we assume $x \geq 0$, then by the above inequality,

$$(1+x)^n \geq \sum_{i=0}^{n}\binom{n}{i}x^i \implies \frac{(1+x)^n}{x^k} \geq \sum_{i=0}^{k}\binom{n}{i}x^{i-k} .$$

So for $0 < x \leq 1$,

$$\sum_{i=0}^{k}\binom{n}{i} \leq \frac{(1+x)^n}{x^k} .$$

Choose $x = k/n$, which is ok because $1 \leq k \leq n$; then

$$\sum_{i=0}^{k}\binom{n}{i} \leq \frac{(1+\frac{k}{n})^n}{(\frac{k}{n})^k} \leq \frac{e^{\frac{k}{n}n}}{(\frac{k}{n})^k} = \left(\frac{en}{k}\right)^k .$$

□

Theorem 1.3 (Stirling formula)

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}(\frac{n}{e})^n} = 1 .$$

Theorem 1.4 (Markov inequality) *Take a random variable* $X \geq 0$, *for which* $\mathbb{E}[X]$ *exists. Then for* $t > 0$,

$$\Pr[X > t] \leq \frac{\mathbb{E}[x]}{t} .$$

As usual

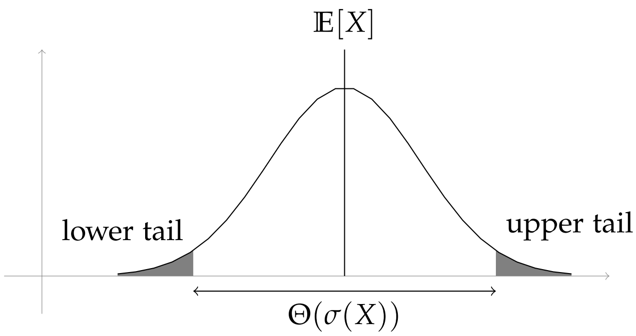
$$\mathrm{Var}[X] = \mathbb{E}\left[(X - E[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}^2[X] \geq 0 ,$$

and $\sigma(X) = \sqrt{\mathrm{Var}[X]}$.

Theorem 1.5 (Chebyshev inequality) *Let X be a random variable for which $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ exist. Then for $t > 0$,*

$$\Pr[|X - \mathbb{E}[X]| > t] \le \frac{\mathrm{Var}[X]}{t^2}\,.$$

Typical picture:



Definition 1.6 (Binomial Random Variable) *Take $X = X_1 + \cdots + X_n$, where each X_i is i.i.d.* Bernoulli(p) *distributed, i.e.* $\Pr[X_i = 1] = p$, $\Pr[X_i = 0] = 1 - p$. *Then X is a* binomial random variable *and we write $X \sim \mathrm{Bin}(n, p)$.*

Note that because of linearity and independence

$$\mathbb{E}[X] = np, \quad \mathrm{Var}[X] = np(1-p)\,.$$

Theorem 1.7 (Chernoff-type bounds on Bin) *For $X \sim \mathrm{Bin}(n, p)$ and $a > 0$, we have the following bounds for lower and upper tail, resp.:*

$$\Pr[X \le np - a] \le e^{-\frac{a^2}{2np}}\,,$$
$$\Pr[X \ge np + a] \le e^{-\frac{a^2}{2np} + \frac{a^3}{2(np)^3}}\,.$$

Asymptotics We will always assume $n \to \infty$ (but is finite!). Let $\Omega = (\Omega_n)$ be a sequence of probability spaces and $A = (A_n)$ a sequence of events with $A_n \subseteq \Omega_n$.

We say that A holds *with high probability* (w.h.p.) in Ω if

$$\lim_{n \to \infty} \Pr_{\Omega_n}[A_n] = 1.$$

For $f, g \colon \mathbb{N} \to \mathbb{R}^+$, we write

$$\begin{aligned}
f = o(g) &\iff \lim_{n\to\infty} f(n)/g(n) = 0\,,\\
f = O(g) &\iff \exists C > 0 \text{ such that } f(n) \leq Cg(n) \forall n\,,\\
f = \Omega(g) &\iff g = O(f)\,,\\
f = \omega(g) &\iff g = o(f)\,,\\
f = \Theta(g) &\iff f = O(g) \text{ and } f = \Omega(g)\,.
\end{aligned}$$

1.2 Basic models: $G(n, p)$, $G(n, m)$, random graph process

Write $[n] := \{1, \ldots, n\}$ and $N = \binom{n}{2}$.

Definition 1.8 ($G(n,p)$) *Take $V = [n]$: note that this means the vertices are labelled! For each pair $1 \leq i \leq j \leq n$, we let*

$$\Pr[\{i, j\} \in E] = p = p(n)\,,$$

independently of the other pairs. Then $G(V, E)$ is the binomial random graph *$G(n, p)$.*

Informally, we have N independent coins and each has probability p of showing head. For a fixed graph G,

$$\Pr[G] = p^{|E(G)|}(1-p)^{N-|E(G)|}\,.$$

In fact, the two definitions are equivalent.

Example 1.9 *Let $p = 1/3$ and*

$$G = \begin{array}{ccc} 1 & \text{---} & 3 \\ | & / & | \\ 2 & & 4 \end{array}\,,$$

Then $\Pr[G] = (\frac{1}{3})^4(\frac{2}{3})^2$.

Edges in $G(n, p)$ appear independently, so we work with product probability spaces.

Special case: if $p = \frac{1}{2}$, then $\forall G = (V, E)$, we have

$$\Pr[G] = (\tfrac{1}{2})^{|E(G)|}(\tfrac{1}{2})^{N-|\mathbb{E}(G)|}\,,$$

So all graphs are equiprobable, hence a typical graph in $G(n, \frac{1}{2})$ has property $P \Leftrightarrow$ a typical graph with vertex set $[n]$ has property P.

Definition 1.10 ($G(n,m)$, Erdős-Rényi) *Take the ground set*

$$\Omega = \{G = (V,E),\ V = [n],\ |E| = m\}\,.$$

Let all graphs be equiprobable:

$$\Pr[G] = \frac{1}{|\Omega|} = \frac{1}{\binom{N}{m}}\,.$$

Comparing $G(n,p)$ **to** $G(n,m)$ We expect $G(n,p) \approx G(n,m)$ if we choose $p(n)$ and $m(n)$ such that $m = \binom{n}{2}p$.

Random graph process The idea is to evolve/grow G gradually from the empty graph to the complete graph. Formally:

Definition 1.11 (Random graph process) *Take a permutation $\sigma = (e_1, \ldots, e_N)$ of the edges of K_n. Define*

$$\begin{aligned} G_0 &= ([n], \emptyset)\,, \\ G_i &= ([n], \{e_1, \ldots, e_i\}) \quad \forall 1 \le i \le N\,. \end{aligned}$$

Then $\tilde{G} = \tilde{G}(\sigma) = (G_i)_{i=0}^{N}$ is a graph process.

If you choose $\sigma \in S_N$ uniformly at random, then $\tilde{G}(\sigma)$ is called a random graph process.

Informally:

1. Start with $G_0 = ([n], \emptyset)$.
2. At step i ($1 \le i \le N$): choose a random missing edge e_i uniformly from $E(K_n) \setminus E(G_{i-1})$. Set $G_i \coloneqq G_{i-1} + \{e_i\}$.

Obviously $|E(G_i)| = i$. We can take a "snapshot" of $\tilde{G}$:

Proposition 1.12 *Let $\tilde{G} = (G_i)$ be a random graph process. Then $G_i \sim G(n,i)$.*

Proof $\forall G = ([n], E)$ with $|E| = i$, just staring at the evolution gives

$$\Pr[G_i = G] = \frac{i!(N-i)!}{N!} = \frac{1}{\binom{N}{i}}\,.$$

We have the same probability for every fixed G, so G_i must induce the distribution $G(n,i)$. □

Hence random graph processes encode/contain $G(n,m)$.

1.3 Staged exposure

Proposition 1.13 (Staged exposure in $G(n,p)$) *Suppose* $0 \leq p, p_1, \ldots, p_k \leq 1$ *satisfy*

$$1 - p = \prod_{i=1}^{k} (1 - p_i) .$$

Then the distributions $G(n,p)$ *and* $\bigcup_{i=1}^{k} G(n, p_i)$ *are identical.*

Proof Let $G_1 \sim G(n,p)$ and $G_2 \sim \bigcup_{i=1}^{k} G(n, p_i)$. Observe that in both G_1 and G_2, every edge $e = \{i, j\}$ appears independently. Moreover, $\Pr[e \notin G_1] = 1 - p$ and

$$\Pr[e \notin G_2] = \prod_{i=1}^{k} (1 - p_i) = 1 - p .$$

□

1.4 Monotonicity

A *graph property* is just a subset of graphs that are said to satisfy the property.

Definition 1.14 *A graph property P is called* monotone (increasing) *if*

$$G \in P \text{ and } H \supseteq G \implies H \in P .$$

Examples include: Hamiltonicity, connectivity, containment of a copy of a fixed graph H.

Proposition 1.15 *Let P be a monotone graph property,* $0 \leq p_1 \leq p_2 \leq 1$ *and* $0 \leq m_1 \leq m_2 \leq N$*. Then:*

(i) $\Pr[G(n, p_1) \in P] \leq \Pr[G(n, p_2) \in P]$.

(ii) $\Pr[G(n, m_1) \in P] \leq \Pr[G(n, m_2) \in P]$.

Proof (ii) Consider a random graph process $\tilde{G} = (G_i)$, and let $G_{m_1} \sim G(n, m_1)$, $G_{m_2} \sim G(n, m_2)$. Then the event "$G_{m_1} \in P$" is contained in the event "$G_{m_2} \in P$" (by monotonicity of P). From this we immediately have $\Pr[G(n, m_1) \in P] \leq \Pr[G(n, m_2) \in P]$.

(i) Take $G_2 \sim G(n, p_2)$, $G_1 \sim G_(n, p_1)$ and $G_0 \sim G(n, p_0)$ where $(1 - p_1)(1 - p_0) = 1 - p_2$ (with $p_0 \geq 1$. Then we can represent G_2 as $G_1 \cup G_0$ and then

$$\{G_1 \text{ has } P\} \subseteq \{G_2 \text{ has } P\} \implies \Pr[G_1 \text{ has } P] \leq \Pr[G_2 \text{ has } P] .$$

□

1.5 Reminder from Graph Theory

For a graph $G = (V, E)$, the *complement* $\bar{G}$ is defined via

$$V(\bar{G}) = V(G), \quad E(\bar{G}) = \binom{V(G)}{2} \setminus E(G)\,.$$

We say that a subset $V_0 \subseteq V(G)$ is an *independent* (or *stable*) set in G if V_0 spans no edges of G. The maximum size of an independent set in G is its *independence number* $\alpha(G)$.

$V_0 \subseteq V(G)$ is a *clique* in G if V_0 spans a complete graph in G, i.e., $\forall u, v \in V_0\colon \{u, v\} \in E(G)$. The maximum size of a clique in G is its *clique number* $\omega(G)$.

Note that V_0 is an independent set in G iff it is a clique in $\bar{G}$, hence $\alpha(G) = \omega(\bar{G})$.

A function $f\colon V \to [k]$ is a *k-coloring* of G if for every edge $\{u, v\} \in E(G)$, $f(u) \neq f(v)$. G is *k-colorable* if it admits a k-coloring.

Observe: if f is a k-coloring of G, then $\forall 1 \leq i < k$,

$$f^{-1}(i) = \{v \in V \mid f(v) = i\} \subseteq V$$

is an independent set. So G is k-colorable iff $V(G)$ can be partitioned as $V = V_1 \cup \cdots \cup V_k$ into k independent sets V_i.

We denote by $\chi(G)$ the *chromatic number of* G: the minimum k for which G is k-colorable.

For example: $\chi(K_n) = n$.

Proposition 1.16 *Let $G = (V, E)$ be a graph. Then*

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

Proof If $V = V_1 \cup \cdots \cup V_k$ is an optimal coloring of G, i.e., $k = \chi(G)$, then $|V_i| \leq \alpha(G)$, so

$$|V| = \sum_{i=1}^{k} |V_i| \leq k \cdot \alpha(G) = \chi(G) \cdot \alpha(G)\,.$$

□

1.6 Three illustrative examples

1.6.1 Lower bounds for Ramsey numbers

Definition 1.17 *Given integers $k, \ell \geq 2$, the* Ramsey number $R(k, \ell)$ *is the smallest n such that every red-blue coloring of the edges of K_n contains a red copy of K_k or a blue copy of K_ℓ.*

Theorem 1.18 (Ramsey, 1930) $R(k, \ell) < \infty$.

We will use the following equivalent definition:

Definition 1.19 *$R(k, \ell)$ is the minimal n such that every graph of size $|V| = n$ satisfies $\omega(G) \geq k$ or $\alpha(G) \geq \ell$ (or both).*

They are equivalent because a red-blue coloring of $E(K_n)$ induces a red graph G and its blue complement $\bar{G}$. A red copy of K_k then corresponds to a clique of size k in G, and a blue copy of K_ℓ corresponds to an independent set of size ℓ in G.

Theorem 1.20 (Erdős, 1947) *Lower bounds for diagonal Ramsey numbers $R(k, k)$:*

$$R(k, k) \geq 2^{k/2}$$

for k large enough.

Proof Need to prove: there is a $G = (V, E)$ with $|V| \geq 2^{k/2}$ such that $\omega(G) < k$ and $\alpha(G) < k$. Set $n = \lceil 2^{k/2} \rceil$, and consider $G(n, \frac{1}{2})$.

Let X be the number of k-cliques in G (a random variable). Then

$$\begin{aligned}\mathbb{E}[X] = \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \leq \left(\frac{en}{k}\right)^k 2^{\frac{-k(k-1)}{2}} &= \left(\frac{en}{k}2^{-\frac{k-1}{2}}\right)^k \\ &\leq \left(\frac{e2^{k/2}}{k}2^{-\frac{k}{2}+\frac{1}{2}}\right)^k = \left(\frac{e\sqrt{2}}{k}\right)^k = o(1)\,,\end{aligned}$$

where in the last step we assumed k to be large enough, i.e., $k \to \infty$. By Markov $\Pr[X \geq 1] = o(1)$.

Now let Y be the number of k-independent sets in G. Then by the same computation

$$\mathbb{E}[Y] = \binom{n}{k} 2^{-\binom{k}{2}} = o(1)\,,$$

so also $\Pr[Y \geq 1] = o(1)$.

From this we get

$$\Pr[X = 0 \text{ and } Y = 0] = 1 - o(1)\,,$$

which is a positive probability, so there must be at least one graph G with $|V| = n$ and $\omega(G), \alpha(G) < k$. $\square$

Problems (posed by Erdős):

1. Does the limit
$$\lim_{k\to\infty} \sqrt[k]{R(k,k)}$$
exist? (This carries a prize of \$100.)

2. Compute the above limit (\$250).

Current state of knowledge is

$$\sqrt{2} \leq \sqrt[k]{R(k,k)} \leq 4\,,$$

where the lower bound is from [Erd47] and the upper bound from [ES35].

1.6.2 Graphs with high girth and high chromatic number

Definition 1.21 *The* girth *of a graph G, denoted by* $\mathrm{girth}(G)$*, is the length of a shortest cycle in G. If G is a forest, we set* $\mathrm{girth}(G) = \infty$.

Theorem 1.22 (Erdős, 1959) *For all integers $k, \ell \geq 3$, there is a graph G with* $\mathrm{girth}(G) > \ell$ *and $\chi(G) > k$.*

Proof Fix a constant θ, where $0 < \theta < 1/\ell$, and consider a random graph $G(n,p)$, where $p = n^{-1+\theta}$.

Let X be the number of cycles of length at most ℓ in G. We compute

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=3}^{\ell} (\text{number of } i\text{-cycles in } K_n) \cdot p^i \\
&= \sum_{i=3}^{\ell} \frac{n(n-1)\cdots(n-i+1)}{2\cdot i} \cdot p^i \\
&\leq \sum_{i=3}^{\ell} n^i p^i \leq \sum_{i=3}^{\ell} n^{\theta i} = O(n^{\theta\ell}) = o(n)\,.
\end{aligned}$$

Hence by Markov, $\Pr[X \geq n/2] = o(1)$.

Bounding $\alpha(G)$: Set $t = \lceil \frac{3\ln n}{p} \rceil$, then

$$\begin{aligned}
\Pr[\alpha(G) \geq t] &\leq \binom{n}{t}(1-p)^{\binom{t}{2}} \leq \left(\frac{en}{t}\right)^t e^{-\frac{1}{2}pt(t-1)} \\
&= \left(\frac{en}{t} e^{-\frac{1}{2}p(t-1)}\right)^t \leq \left(en \cdot e^{-1.4\ln n}\right)^t = o(1)\,.
\end{aligned}$$

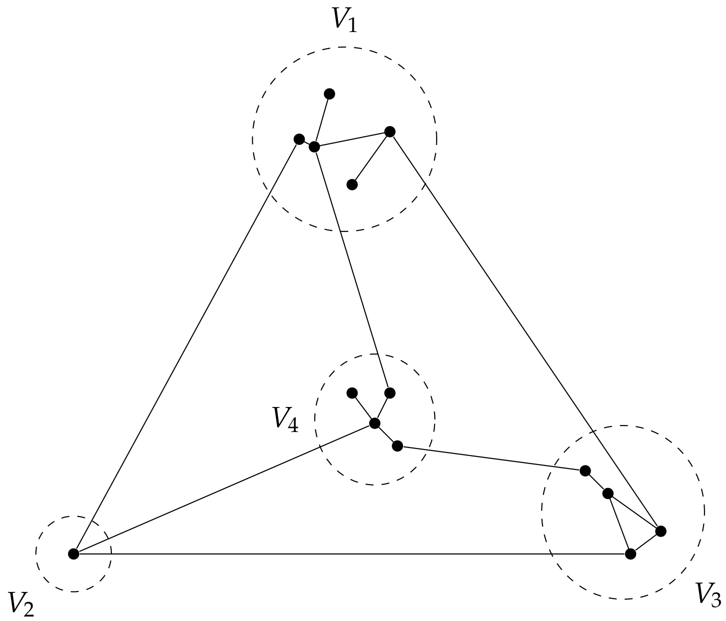


Figure 1.1: Example of a K_4 minor

Taking the two halves together, we get

$$\Pr[\alpha(G) \geq t \text{ or } X \geq n/2] = o(1)\,,$$

so there exists a graph G with $|V| = n$ such that $\alpha(G) < t$ and $\chi(G) \leq n/2$.

For every cycle of length $\leq \ell$ in G, delete an arbitrary vertex. We get an induced subgraph $G' \subseteq G$, with $|V(G')| \geq n/2$, girth$(G') > \ell$, and $\alpha(G') \leq \alpha(G) < t$. Finally,

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{n/2}{\left\lceil \frac{3 \ln n}{p} \right\rceil} \geq n^{\theta/2} > k\,.$$

So G' has all the properties we want. □

1.6.3 Hadwiger's Conjecture

Definition 1.23 *A graph $H = ([k], F)$ is a* minor *of G if $V(G)$ contains k nonempty disjoint subsets $V_1, \ldots, V_k$, such that:*

(i) $G[V_i]$ is connected,

(ii) for every $f = \{i, j\} \in F$, there is an edge between V_i and V_j in G.

See Figure 1.1 for an example. An equivalent way to put it is:

Definition 1.24 *H is a* minor *of G if H can be obtained from G by a sequence of the following operations:*

(i) deleting a vertex,

(ii) deleting an edge,

(iii) contracting an edge: replacing an edge $e = \{u, v\}$ by a new vertex uv, and connecting uv to all vertices in $N(u) \cup N(v) \setminus \{u, v\}$ (and deleting parallel edges).

Definition 1.25 *The* Hadwiger number *of G is the maximum k such that K_k is a minor of G.*

Conjecture 1.26 (Hadwiger, 1943) *For all graphs, $h(G) \geq \chi(G)$.*

Very little is known:

$k = 2$: obviously trivial.

$k = 3$: trivial because every cycle can be contracted to K_3.

$k = 4$: was proven by Hadwiger himself.

$k = 5$: equivalent to the four-color theorem (Wagner, 1937).

$k = 6$: proven by Robertson, Seymour, and Thomas.

What if we give up on answering the question for all graphs, and weaken it to a probabilistic statement: does Hadwiger's conjecture hold for almost every graph? Formally, let $G \sim G(n, \frac{1}{2})$. What about Hadwiger's conjecture for such G?

Theorem 1.27 (Bollobás, Catlin and Erdős, 1980) *Hadwiger's conjecture holds w.h.p. for $G \sim G(n, \frac{1}{2})$.*

Proof We will prove: if $G \sim G(n, \frac{1}{2})$, then w.h.p.

(i) $h(G) \geq \frac{n}{6\sqrt{\log_2 n}}$.

(ii) $\chi(G) \leq \frac{6n}{\log_2 n}$.

(We ignore rounding issues; they can be fixed.)

Part (i) We represent $G = G_1 \cup G_2$, where $G_1, G_2 \sim G(n, p)$ such that $\frac{1}{2} = (1-p)^2$. From there $p \geq \frac{1}{4}$, which will be enough.

Claim *W.h.p. $G_1 \sim G(n, p)$ has a path on $n/2$ vertices.*

We use the following algorithm:

1. Choose v_1 arbitrarily, and set $i := 1$.

2. As long as $i < n/2$, find a neighbor u of v_i outside $\{v_1, \ldots, v_i\}$. If there is no such neighbor, declare failure.

   Otherwise set $v_{i+1} := u$ and update $i := i + 1$.

Analyzing the algorithm:

$$v_1 \qquad\qquad v_{i-1} \quad v_i \qquad\qquad n-i$$

At v_i, we have explored only edges from v_i to $\{v_1, \ldots, v_{i-1}\}$, so

$$\begin{aligned}&\Pr[\text{no edge in } G(n,p) \text{ between } v_i \text{ and } V \setminus \{v_1, \ldots, v_{i-1}\}] \\ &\qquad = \Pr[\mathrm{Bin}(n-i,p) = 0] \leq (1-p)^{n-i} \leq (1-p)^{n/2}\end{aligned}$$

hence

$$\Pr[\text{fail}] \leq \sum_{i=1}^{n/2} (1-p)^{n-i} \leq \frac{n}{2}(1-p)^{n/2} = o(1)\,.$$

This proves the claim.

By the end of phase 1, w.h.p. we have created a path P on $n/2$ vertices.

Phase 2: split P into $s = \frac{n}{6\sqrt{log_2 n}}$ disjoint paths $P_1, \ldots, P_s$, with $|V(P_i)| = \frac{n/2}{s} = 3\sqrt{log_2 n}$.

Expose $G_2 \sim G(n,p)$ $(p \geq \frac{1}{4})$:

$$\begin{aligned}&\Pr[\exists 1 \leq i \neq j \leq s \text{ s.t. } G_2 \text{ has no edge between } V(P_i) \text{ and } V(P_j)] \\ &\qquad \leq \binom{s}{2}(1-p)^{(3\sqrt{\log_2 n})^2} \leq n^2 \left(\frac{3}{4}\right)^{9\log_2 n} = o(1)\,.\end{aligned}$$

Both stages succeed w.h.p. and between them create a clique minor of size s on sets $V(P_1), \ldots, V(P_s)$. Therefore w.h.p. $h(G) \geq s = \frac{n}{6\sqrt{\log_2 n}}$.

Remark: Actually, w.h.p. $h(G(n, \frac{1}{2})) = (1 + o(1))\frac{n}{\sqrt{\log_2 n}}$.

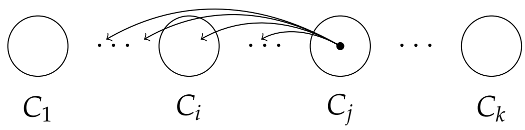
Part (ii): We use the following:

Definition 1.28 *A sequence of k subsets $C_1, \ldots, C_k$ is a* rigid k-coloring *of G if*

(i) $V = C_1 \uplus \cdots \uplus C_k$,

(ii) each C_i is independent,

(iii) $\forall 1 \leq i < j \leq k$, $\forall v \in C_j$: *v has a neighbor in C_i.*

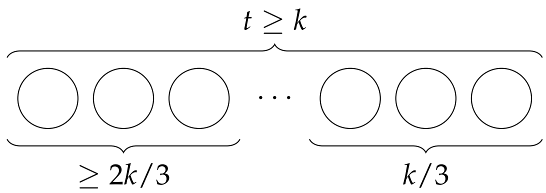


Claim *If $\chi(G) = k$, then G admits a rigid k-coloring.*

Start with any k-coloring $C_1, \ldots, C_k$, and as long as there exist $1 \leq i < j \leq k$ and $v \in C_j$ s.t. v has no neighbor in C_i, move v over to C_i. The final collection $(C_1^*, \ldots, C_k^*)$ is a valid k-coloring of G, which is rigid and with no empty color classes (as $\chi(G) \geq k$). This proves the claim.

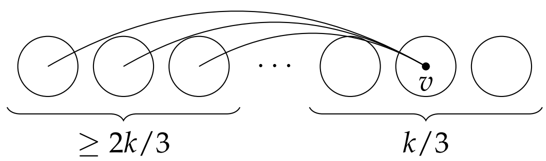
Back to $G(n, \frac{1}{2})$: Let $k = \frac{6n}{\log_2 n}$. Assume that $\chi(G) = t \geq k$. By the claim there is a rigid t-coloring $(C_1, \ldots, C_t)$ of G.

Then for the last $k/3$ color classes C_j, every $v \in C_j$ has a neighbor in C_i for all $i < j$.



Take a vertex v_j in each of the last $k/3$ color classes C_j.

Between the first $t - k/3 \geq 2k/3$ color classes, there are at least $k/3$ color classes of cardinality at most $3n/k$ each (as otherwise the largest $k/3$ color classes between them are of size $> 3n/k$ each, totaling in size $> n$, a contradiction).



Now we estimate the probability for this to happen, by fixing a coloring and looking at the probability that edges between v and C_i exist for all last v and small color classes C_i:

$$\Pr[\chi(G) \geq k] = \Pr[\exists \text{ rigid } \chi(G)\text{-coloring of } G] \leq k^n \left(1 - 2^{-3n/k}\right)^{\frac{k}{3} \cdot \frac{k}{3}}$$
$$\leq k^n \cdot e^{-2^{-3n/k} \cdot \frac{k^2}{9}} \leq k^n \cdot e^{-\frac{1}{\sqrt{n}} \frac{cn^2}{\log_2 n}} = o(1)\,. \quad \square$$

1.7 Asymptotic equivalence between $G(n, p)$ and $G(n, m)$

Denote $N = \binom{n}{2}$ and $q = 1 - p$.

Informally: $G(n, p)$ and $G(n, m)$ are "similar" if $m = Np(1 + o(1))$.

Of course this is only an intuition, and it cannot be true literally:

$$\Pr[|E(G(n, m))| = m] = 1\,,$$
$$\Pr[|E(G(n, p))| = m] = \frac{1}{\Omega(\sqrt{Npq})}\,.$$

Still, we get some results.

Proposition 1.29 *Let $P = (P_n)_{n\geq 1}$ (P_n a set of graphs on $[n]$) be an arbitrary graph property. Let $p = p(n)$ be a sequence of real numbers with $0 \leq p(n) \leq 1$. Let further $0 \leq a \leq 1$.*

If for every sequence $m = m(n)$ satisfying $m = Np + O(\sqrt{Npq})$ we have

$$\lim_{n\to\infty} \Pr[G(n, m) \in P] = a \tag{1.1}$$

then also

$$\lim_{n\to\infty} \Pr[G(n, p) \in P] = a\,. \tag{1.2}$$

Proof Choose large enough constant $c > 0$. Denote

$$M = M(c) = \{0 \leq m \leq N : |m - Np| \leq c\sqrt{Npq}\}\,,$$

then by Chebyshev

$$\Pr[|E(G(n, p))| \notin M] \leq \frac{1}{c^2}\,.$$

Also, denote

$$m_* = \arg\min_{m\in M} \Pr[G(n, m) \in P]\,,$$
$$m^* = \arg\max_{m\in M} \Pr[G(n, m) \in P]\,.$$

Then, by the law of total probability

$$\begin{aligned}
\Pr[G(n, p) \in P] &= \sum_{m=0}^{M} \Pr[|E(G(n, p))| = m] \cdot \Pr[G(n, m) \in P] \\
&\geq \sum_{m\in M} \Pr[|E(G(n, p))| = m] \cdot \Pr[G(n, m) \in P] \\
&\geq \sum_{m\in M} \Pr[|E(G(n, p))| = m] \cdot \Pr[G(n, m_*) \in P] \\
&= \Pr[G(n, m_*) \in P] \cdot \Pr[|E(G(n, p))| \in M] \\
&= \Pr[G(n, m_*) \in P] \left(1 - c^{-2}\right)\,,
\end{aligned}$$

hence

$$\begin{aligned}&\liminf_{n\to\infty} \Pr[G(n,p) \in P] \geq a(1-c^{-2})\\ \Longrightarrow\ &\lim_{n\to\infty} \Pr[G(n,p) \in P] \geq a\,.\end{aligned}$$

For the upper bound, we treat M and the rest of the interval separately:

$$\begin{aligned}\Pr[G(n,p) \in P] &= \sum_{m=0}^{M} \Pr[|E(G(n,p))| = m] \cdot \Pr[G(n,m) \in P]\\ &\leq c^{-2} + \sum_{m \in M} \Pr[|E(G(n,p))| = m] \cdot \Pr[G(n,m) \in P]\\ &\leq c^{-2} + \Pr[G(n,m_*) \in P]\,,\end{aligned}$$

therefore

$$\liminf_{n\to\infty} \Pr[G(n,p) \in P] \leq a + c^{-2}\,.$$

Taking c large enough, we get

$$\lim_{n\to\infty} \Pr[G(n,p) \in P] = a\,. \qquad \square$$

For the direction $G(n,p) \rightsquigarrow G(n,m)$, we have to be a bit more careful. We need monotonicity; consider the counterexample of P being the property of having exactly m edges. We also need a bit of slack.

Proposition 1.30 *Let P be a* monotone *graph property. Let $0 \leq m = m(n) \leq N$, and $0 \leq a \leq 1$. If for every sequence $p = p(n)$ s.t.*

$$p = \frac{m}{N} + O\left(\sqrt{\frac{m(N-m)}{N^3}}\right)$$

it holds that $\lim_{n\to\infty} \Pr[G(n,p) \in P] = a$, then

$$\lim_{n\to\infty} \Pr[G(n,m) \in P] = a\,.$$

Chapter 2

Random Regular Graphs

2.1 Preliminaries

Notation:

1. For an even m,

$$(m-1)!! := (m-1)(m-3)\cdots 3\cdot 1$$

By Stirling we have

$$\begin{aligned}(m-1)!! &= \frac{m(m-1)(m-2)\cdots 3\cdot 2\cdot 1}{m(m-1)\cdots 4\cdot 2} \\ &= \frac{m!}{2^{m/2}(m/2)!} = (1+o(1))\frac{\sqrt{2}m^{m/2}}{e^{m/2}}\,.\end{aligned}$$

2. *Falling factorial*: for $x \in \mathbb{R}$ and an integer $k \geq 1$, define

$$(x)_k := x(x-1)(x-2)\cdots(x-k+1)\,.$$

Use:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{(n)_k}{k!}\,.$$

3. For a random variable X and an integer $k \geq 1$, the *k-th factorial moment* of X is

$$\mathbb{E}[(X)_k] = \mathbb{E}[X(X-1)\cdots(X-k+1)]\,.$$

Definition 2.1 *A random variable X is said to have* Poisson with parameter λ *distribution, written $X \sim \mathrm{Poi}(\lambda)$, if X takes non-negative integer values and for every $k \geq 0$,*

$$\Pr[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}\,.$$

Note that $\mathrm{Bin}(n, p) \to \mathrm{Poi}(\lambda, p)$ if $\lambda = np$.

We have

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda\,.$$

More generally, if $X \sim \mathrm{Poi}(\lambda)$, then the k-th factorial moment of X is λ^k.

Definition 2.2 (Convergence in distribution) *Let (X_n) be a sequence of random variables, Z another r.v., and let them take integer values only. We say that X_n* converges in distribution *to Z, written $X_n \xrightarrow{d} Z$, if for every integer k,*

$$\lim_{n\to\infty} \Pr[X_n = k] = \Pr[Z = k]\,.$$

This can be extended to vectors of random variables by requiring convergence in every component. Finally, it can be extended to infinite vectors by requiring that every finite subvector converges.

Definition 2.3 *Let $n, r \geq 1$ be integers. Denote by $G_{n,r}$ the set of all r-regular graphs with vertex set $[n]$. Endow $G_{n,r}$ with the uniform distribution: $\Pr[G] = 1/|G_{n,r}|$ for every $G \in G_{n,r}$. This defines the* probability space of r-regular graphs on $[n]$.

Obvious necessary condition:

$$nr = \sum_{v \in V(G)} d(v) = 2|E|\,,$$

so we assume nr is even.

Questions to address:

1. How many graphs $L_{n,r} = |G_{n,r}|$ are there?
2. What are the properties of a typical graph $G \in G_{n,r}$?
3. How to sample $G \in G_{n,r}$?

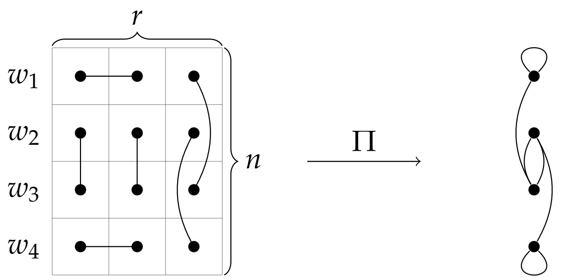


Figure 2.1: Example configuration and its projection

2.2 Configurations

Approach invented by [BC78] and independently [Bol80].

Define $W = [n] \times [r]$. Elements of W are called *cells*. We can represent $W = \bigcup_{i=1}^{n} W_i$, $|W_i| = r$; cells of W_i are the cells of vertex i.

A *configuration* F on W is a partition of W into $nr/2$ pairs. (In graph terms, it is a perfect matching on W.)

There are

$$(nr-1)(nr-3)\cdots 3\cdot 1 = (nr-1)!!$$

configurations:

1. Each permutation σ on W produces a partition as follows: $(\sigma(1), \sigma(2))$, $(\sigma(3), \sigma(4))$, … In this way we get every configuration F. Moreover, each F is obtained

   $$(nr/2)!2^{nr/2}$$

   times (ordering of the elements and orientation of the pairs), so the total number is $(nr-1)!!$.

2. Fix an arbitrary order on the cells of W.

   - Match the first cell of W: $(nr-1)$ options,

   - Match the first currently unmatched cell: $(nr-3)$ options, etc.

   Altogether we get $(nr-1)!!$.

From configurations to (multi)graphs Given a configuration F on W, project it into a multigraph $G = \Pi(F)$ in the following way: let $V = [n]$. For every pair $((u_i, x_i), (v_i, y_i)) \in F$ (where $1 \le u_i, v_i \le n$ and $1 \le x_i, y_i \le r$) we put an edge (u_i, v_i) into E (and forget about x_i and y_i). See Figure 2.1.

The result $G = \Pi(F)$ is an r-regular *multigraph* (can contain loops and multiple edges) on the vertex set $[n]$. Loops contribute 2 to the degree of their vertex.

Claim *Every r-regular graph G on $[n]$ is obtained from exactly $(r!)^n$ configurations F.*

Proof Need to match, for every $i \in [n]$, the r edges entering i with r cells of W_i: can be done in $r!$ ways per vertex, altogether in $(r!)^n$ ways. Thus $|\Pi^{-1}(G)| = (r!)^n$. $\square$

Turn the space of $(nr-1)!!$ configurations on W into a probability space by putting the uniform measure $\Pr[F] = 1/(nr-1)!!$. Then the uniform probability measure on all configurations induces the probability measure on r-regular multigraphs on $[n]$ through the projection operator. Denote the latter probability space by $G^*_{n,r}$. This $G^*_{n,r}$ is *different* from $G_{n,r}$, as in particular $G^*_{n,r}$ contains multigraphs. However, if we condition on not having loops or multi-edges, then we get a random element $G \in G_{n,r}$.

Call a configuration *simple* if $\Pi(F)$ has no loops or multiple edges. Denote the event that a random configuration is simple by SIMPLE. Then

$$L_{n,r} = |G_{n,r}| = \frac{(nr-1)!! \cdot \Pr[\textsc{simple}]}{(r!)^n} .$$

The probability of SIMPLE is non-negligible, so computing it will be key.

Lemma 2.4 *Let $1 \le k \le nr/2$. Let $E_0 = \{e_1, \ldots, e_k\}$ be a fixed set of k disjoint pairs in W. Then*

$$p_k := \Pr[E_0 \subseteq F] = \frac{1}{(nr-1)(nr-3)\cdots(nr-2k+1)} .$$

Proof Let $e_i = ((u_i, x_i), (v_i, y_i))$ and $E_0 = \{e_1, \ldots, e_k\}$. Then

$$\Pr[e_1 \in F] = 1/(nr-1) ,$$

by symmetry: the probability that (u_i, x_i) is matched to (v_i, y_i) is exactly $1/(nr-1)$. Next,

$$\Pr[e_2 \in F \mid e_1 \in F] = 1/(nr-3) ,$$

because there are only $(nr-3)$ choices of (v_2, y_2) left to match (u_2, x_2) with. Continue in this way, and eventually we get

$$\Pr[E_0 \subseteq F] = \frac{1}{(nr-1)(nr-3)\cdots(nr-2k+1)} . \qquad \square$$

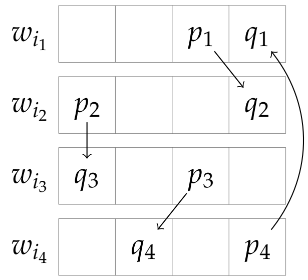


Figure 2.2: Enumerating cycles via ordered pairs

Remark 2.5 For a constant k,

$$p_k = \frac{1}{(nr-1)(nr-3)\cdots(nr-2k+1)} \sim \frac{1}{(rn)^k}\,.$$

We will look at cycle lengths in $G^*_{n,r}$.

Theorem 2.6 *For $k = 1, 2, \ldots$ denote $\lambda_k := (r-1)^k/2k$. Also, let $Z^*_k \sim \mathrm{Poi}(\lambda_k)$ be independent. Finally, let Z_k denote the number of k-cycles in $G^*_{n,r}$. Then the vector of random variables $(Z_k)_{k\geq 1}$ converges in distribution to the vector $(Z^*_k)_{k\geq 1}$.*

Note that the Z_k are *not* independent, they just converge in distribution to the *independent* Z^*_k.

Proof We will use the method of moments. We will prove that the joint factorial moments of (Z_k) converge to the corresponding joint factorial moments of (Z^*_k).

Let us first compute the expected values of Z_k. Denote by a_k the number of possible k-cycles in W, i.e., the number of families of k pairs in W that produce a k-cycle when projected by Π. We can generate all possible k-cycles in W as follows:

1. Choose the *ordered* set of vertices of a k-cycle in G in $n(n-1)\cdots(n-k+1) = (n)_k$ ways.
2. For each vertex i of the cycle choose an ordered pair (p_i, q_i) where $1 \leq p_i \neq q_i \leq r$, using p_i for the outgoing edge and q_i for the incoming edge (Figure 2.2). There are $r(r-1)$ ways to do this for every vertex, thus $(r(r-1))^k$ altogether.

We get $(n)_k(r(r-1))^k$ ways. We counted ordered cycles and hence overcounted by 2 for the choice of direction and k for the choice of the initial vertex. Therefore

$$a_k = \frac{(n)_k(r(r-1))^k}{2k}\,.$$

Each possible cycle appears with probability p_k, so

$$\mathbb{E}[Z_k] = a_k \cdot p_k = \frac{(n)_k (r(r-1))^k}{2k} \frac{1}{(nr-1)(nr-3)\cdots(nr-2k+1)}$$

hence

$$\mathbb{E}[Z_k] \to \frac{(r-1)^k}{2k} = \lambda_k = \mathbb{E}[Z_k^*]\,.$$

This accounts for the first moment.

Remark: For every fixed graph H, the expected number of copies of H in $G_{n,r}^*$ is $O(n^{|V(H)|}n^{-|E(H)|})$. In particular, if $|E(H)| > |V(H)|$ then the expected number of copies of H in $G_{n,r}^*$ is $O(n^{-1})$ and by Markov we do not expect them to appear at all.

Proceed to higher factorial moments of $(Z_k)_{k\geq 1}$. Let's estimate the second factorial moments

$$\mathbb{E}[(Z_k)_2] = \mathbb{E}[Z_k(Z_k - 1)]\,.$$

Notice that $(Z_k)_2$ is the number of ordered pairs of distinct k-cycles in $G_{n,r}^*$.

We represent $(Z_k)_2 = Y' + Y''$ where Y' is the number of ordered pairs of *disjoint* k-cycles and Y'' accounts for k-cycles that share at least one vertex.

In order to estimate $\mathbb{E}[Y'']$ we go over all possible ways H to create two k-cycles sharing a vertex. In each of them $|E(V)| > |V(H)|$ and there is a bounded (by some function of k) number of them. Hence, by the above remark, each H is expected to appear $O(n^{-1})$ times, and therefore $\mathbb{E}[Y''] = O(n^{-1})$.

Denote by a_{2k} the number of sets of $2k$ pairs in W projecting to 2 disjoint k-cycles. Then

$$a_{2k} = \frac{n(n-1)\cdots(n-2k+1)(r(r-1))^{2k}}{(2k)^2} \sim \frac{n^{2k}(r(r-1))^{2k}}{(2k)^2}\,,$$

hence

$$\mathbb{E}[Y'] \sim \frac{(n)_{2k}(r(r-1))^{2k}}{(2k)^2} p_{2k} \sim \frac{n^{2k}(r(r-1))^{2k}}{(2k)^2} \frac{1}{(rn)^{2k}} = \frac{(r-1)^{2k}}{(2k)^2} = \lambda_k^2\,.$$

It follows that $\mathbb{E}[(Z_k)_2] = \lambda_k^2 + O(n^{-1})$.

In general, using similar arguments, one can prove: for $t_1, \ldots, t_m \geq 0$ integers,

$$\mathbb{E}[(Z_1)_{t_1} \cdots (Z_m)_{t_m}] \sim \lambda_1^{t_1} \cdots \lambda_m^{t_m}\,.$$

Now we cite the following probabilistic statement:

Let $(Z_n^{(1)}, \dots, Z_n^{(m)})_n$ be a sequence of random variables. Assume that there exist $\lambda_1, \dots, \lambda_m \geq 0$ such that for all integers $t_1, \dots, t_m \geq 0$,

$$\lim_{n\to\infty} \mathbb{E}[Z_{t_1}^{(1)} \cdots Z_{t_m}^{(m)}] = \lambda_1^{t_1} \cdots \lambda_m^{t_m} .$$

Then the vector $(Z_n^{(1)}, \dots, Z_n^{(m)})$ converges in distribution to $(Z^{(i)*})_{i=1}^m$ where $Z^{(i)^*} \sim \mathrm{Poi}(\lambda_i)$, and the $Z^{(i)*}$ are independent. □

Corollary 2.7 *The probability that a configuration is simple is a constant if r is constant:*

$$\begin{aligned}\Pr[\textsc{simple}] = \Pr[Z_1 = Z_2 = 0] \sim \Pr[Z_1^* = Z_2^* = 0] = e^{-\lambda_1} e^{-\lambda_2} \\ = e^{-(r-1)/2-(r-1)^2/4} = e^{-(r^2-1)/4} .\end{aligned}$$

Therefore

$$L_{n,r} = |G_{n,r}| \sim \frac{(rn-1)!!e^{-(r^2-1)/4}}{(r!)^n} \sim \sqrt{2}e^{-(r^2-1)/4} \left(\frac{r^{r/2}e^{-r/2}}{r!}\right)^n n^{nr/2} .$$

And probably more importantly:

Corollary 2.8 *If $G_{n,r}^*$ has property P w.h.p., then $G_{n,r}$ has P w.h.p.*

Proof Use the definition of conditional probability:

$$\begin{aligned}\Pr[G_{n,r} \notin P] = \Pr[G_{n,r}^* \notin P \mid \textsc{simple}] = \frac{\Pr[G_{n,r}^* \notin P \wedge \textsc{simple}]}{\Pr[\textsc{simple}]} \\ \leq \frac{\Pr[G_{n,r}^* \notin P]}{\Pr[\textsc{simple}]} = \frac{o(1)}{e^{-(r^2-1)/4}} = o(1) .\end{aligned}$$ □

Remark 2.9 Our estimate on $\Pr[\textsc{simple}]$ allows us to derive and analyze the following algorithm for generating a random r-regular graph with vertex set $[n]$:

1. Generate a random configuration F,
2. Project to get $G = \Pi(F)$,
3. If G has no loops or multiple edges, output G; otherwise restart.

Each round succeeds with probability $\Pr[\textsc{simple}] \sim e^{-(r^2-1)/4}$. Therefore we expect to need about $e^{(r^2-1)/4}$ rounds to generate a random r-regular graph.

Similar techniques can also be used for $r = r(n)$ slowly growing to infinity, but then the probability of SIMPLE is negligible, so the last few statements do not hold.

Chapter 3

Long paths and Hamiltonicity

3.1 Long paths

Take $G \sim G(n, p)$. What is the typical behaviour of a longest path in G?

What to expect:

- For $p = (1 - \varepsilon)/n$, with $\varepsilon > 0$, all connected components are $O(\log n)$ big w.h.p. Therefore the longest path is also $O(\log n)$ w.h.p. (and you can indeed obtain such a path).
- For $p \geq (1 + \varepsilon)/n$, w.h.p. G has a connected component with linear size. We may hope to get a path of linear length w.h.p.
- For $p \geq (\log n + \omega(1))/n$, w.h.p. G has no isolated vertices. We may hope to get a Hamilton path or cycle.

Theorem 3.1 (Ajtai, Komlós, Szemerédi [AKS81]) *There is a function $\alpha = \alpha(c) = (1, \infty) \to (0, 1)$ such that $\lim_{c \to \infty} \alpha(c) = 1$ such that a random graph $G \sim G(n, c/n)$ has a path of length at least $\alpha(c)n$.*

We will instead prove a weaker result by W. Fernandez de la Vega [dlV79] which gives a similar result but with the domain of α being $(4 \log 2, \infty)$.

Theorem 3.2 (W. Fernandez de la Vega) *Let $p(n) = \theta/n$. Then w.h.p. a random graph $G(n, p)$ has a path of length at least*

$$\left(1 - \frac{4 \log 2}{\theta}\right) n \,.$$

Proof Represent $G \sim G(n, p)$ as $G = G_1 \cup G_2$ where $G_1, G_2 \sim G(n, r)$ are independent. The value of r is determined by $(1 - p) = (1 - r)^2$, implying $r \geq p/2 = \theta/2n$. Call G_1 the red graph and G_2 the blue graph.

Informal description: Try to advance using red edges first. If you can't advance using red edges, rewind the current path to the last vertex from which blue edges haven't been tried. If this fails, give up the current path and start from scratch.

Formal description: During the course of the algorithm, maintain the triple (P_k, U_k, B_k) where:

- P_k is the current path from vertex y_k to vertex y_k.
- $U_k \subseteq V \setminus V(P_k)$ is the set of untried vertices.
- $B_k \subseteq V$ is the set of blue vertices, i.e., the vertices in which we have exposed red edges.

Denote $u_k = |U_k|$.

We will maintain the following properties:

- If $y_k \notin B_k$, then the red edges between y_k and U_k have not been exposed.
- If $y_k \in B_k$, then the blue edges between y_k and U_k have not been exposed.

Initialization: Choose an arbitrary x_0, and define

$$P_0 = (x_0)\,, \quad y_0 = (x_0)\,, \quad U_0 = V \setminus \{x_0\}\,, \quad B_0 = \emptyset\,.$$

As the algorithm advances, the set U_k shrinks, while the set B_k grows: $U_{k+1} \subseteq U_k$ and $B_{k+1} \supseteq B_k$.

At a generic step k we distinguish between the following cases:

Case 1, $y_k \notin B_k$: If there is a red edge y_k–U_k, say edge $(y_k, y_{k+1}) \in E(G_1)$, then extend the current path by adding edge (y_k, y_{k+1}), i.e., update:

$$P_{k+1} = x_k P_k y_k y_{k+1}\,, \quad U_{k+1} = U_k \setminus \{y_{k+1}\}\,, \quad B_{k+1} = B_k\,.$$

If there is no such edge, put y_k into B_k:

$$P_{k+1} = P_k\,, \quad U_{k+1} = U_k\,, \quad B_{k+1} = B_k \cup \{y_k\}\,.$$

Case 2, $y_k \in B_k$ and $V(P_k) \setminus B_k \neq \emptyset$: If there is a blue edge y_k–U_k, say edge $(y_k, y_{k+1}) \in E(G_2)$, then update:

$$P_{k+1} = x_k P_k y_k y_{k+1}\,, \quad U_{k+1} = U_k \setminus \{y_{k+1}\}\,, \quad B_{k+1} = B_k\,.$$

If there is no such edge, let y_{k+1} be the last vertex in P_k that is not in B_k, shorten P_k and recolor y_{k+1} blue:

$$P_{k+1} = x_k P_k y_{k+1}\,, \quad U_{k+1} = U_k\,, \quad B_{k+1} = B_k \cup \{y_{k+1}\}\,.$$

Case 3, $V(P_k) \subseteq B_k$: For technical reasons (to become apparent later) stay put with probability $(1-r)^{u_k}$, i.e.,

$$P_{k+1} = P_k\,, \quad U_{k+1} = U_k\,, \quad B_{k+1} = B_k\,.$$

With probability $1-(1-r)^{u_k}$, check for a blue edge between y_k and U_k. If there is such an edge, extend P_k as in Case 2. Otherwise give up the current path P_k and restart by choosing $u \in U_k$ and setting

$$P_{k+1} = (u)\,, \quad U_{k+1} = U_k - \{u\}\,, \quad B_{k+1} = B_k\,.$$

Let us analyze the above algorithm. Denote $b_k = |B_k|$. Observe that under no circumstances we both shrink U_k and extend B_k in the same step. Therefore

$$(n - u_{k+1}) + b_{k+1} \leq (n - u_k) + b_k + 1\,.$$

Since in the beginning we have $n - u_0 = 1$ and $b_0 = 0$, we conclude that

$$(n - u_k) + b_k \leq k + 1\,. \tag{3.1}$$

Also, $V \setminus V(P_k) \setminus U_k \subseteq B_k$. Hence using (3.1),

$$\begin{aligned} |V(P_k)| - 1 \geq n - u_k - b_k - 1 \geq n - u_k - (k + 1 - (n - u_k)) - 1 \\ = 2(n - u_k) - k - 2\,. \end{aligned} \tag{3.2}$$

In the (very unlikely) case when we use red edges successfully to extend the current path in k rounds, we get $n - u_k = k + 1$, i.e.,

$$|V(P - k)| - 1 \geq 2(k+1) - k - 2 = k\,.$$

This is very unlikely, but still we prove that w.h.p. the RHS of (3.2) is large for some k.

Let us turn to the sequence (u_k). Observe that in all three cases, $u_{k+1} = u_k$ or $u_{k+1} = u_k - 1$, where

$$\begin{aligned} \Pr[u_{k+1} = u_k] \quad &= (1-r)^{u_k}\,, \\ \Pr[u_{k+1} = u_k - 1] &= 1 - (1-r)^{u_k}\,. \end{aligned} \tag{3.3}$$

Therefore the sequence (u_k) forms a Markov chain governed by (3.3). Define the r.v. X_i as

$$X_i = \max\{k - \ell \mid u_k = u_\ell = i\}\,.$$

Then the sequence (u_k) spends exactly time $X_i + 1$ at state i. Furthermore define

$$Y_i = \sum_{i=j+1}^{n-1} (X_i + 1)\,.$$

which is the total time spent by the sequence u_k to reach value j.

Background: suppose we conduct independent experiments where each experiment succeeds with probability p. Then the number of unsuccessful experiments before the first success is a geometric random variable with parameter p.

Formally, a *geometric r.v.* with parameter p is a r.v. X taking non-negative integer values $k = 0, 1, \ldots$ such that

$$\Pr[X = k] = (1-p)^k p$$

and one can show that

$$\mathbb{E}[X] = \frac{1-p}{p}, \quad \mathrm{Var}[X] = \frac{1-p}{p^2}.$$

Going back to the proof, X_i is geometrically distributed with parameter $1 - (1-r)^i$. Hence

$$\mathbb{E}[X_i] = \frac{(1-r)^i}{1-(1-r)^i}, \quad \mathrm{Var}[X_i] = \frac{(1-r)^i}{(1-(1-r)^i)^2}.$$

Therefore, since the X_i's are independent, we get

$$\mathbb{E}[Y_j] = \sum_{i=j+1}^{n-1} 1 + \mathbb{E}[X_i] = \sum_{i=j+1}^{n-1} 1 + \frac{(1-r)^i}{1-(1-r)^i}.$$

Also,

$$\mathrm{Var}[Y_j] = \sum_{i=j+1}^{n-1} \frac{(1-r)^i}{(1-(1-r)^i)^2}.$$

Choose $j = \lceil \frac{\log 2}{r} \rceil$ (note that this means $j = \Theta(n)$). Then

$$(1-r)^j \leq e^{-rj} \leq \tfrac{1}{2}.$$

Hence the standard deviation is on the order of $\sqrt{n}$:

$$\mathrm{Var}[Y_j] \leq \sum_{i=j+1}^{n-1} \frac{1/2}{(1-1/2)^2} \leq 2n.$$

On the other hand, the expectation is linear:

$$\mathbb{E}[Y_j] = \sum_{i=j+1}^{n-1} \frac{1}{1-(1-r)^i} \leq \sum_{i=j+1}^{n-1} \frac{1}{1-e^{-ri}} \leq \int_j^{n-1} \frac{dx}{1-e^{-rx}} = \frac{\log(e^{rn}-1)}{r}.$$

By Chebyshev we derive that w.h.p.

$$Y_j \leq \frac{\log(e^{rn}-1)}{r} + \sqrt{n}\omega(1)\,.$$

Therefore the algorithm w.h.p. finds a path of length at least

$$2\left(n - \left\lceil \frac{\log 2}{r} \right\rceil\right) - \frac{\log(e^{rn}-1)}{2} - \sqrt{n}\omega(1) \geq n - 2\frac{\log 2}{r} \geq n - \frac{4\log 2}{\theta} n\,.$$

□

Remarks:

1. We took $p = \theta/n$ and assumed $\theta > 0$ to be constant. In fact the same proof goes through even if $\theta = \theta(n)$ increases "slightly" with n, implying in particular that if $p \gg 1/n$ then w.h.p. the $G(n,p)$ has a path of length $(1-o(1))n$.

2. Observe that the algorithm analyzed never reversed the direction of any edge. Therefore the algorithm and its analysis also produce the following result.

   Theorem 3.3 *Let $\vec{G}(n,p)$ be a* random directed graph, *where each ordered pair (u,v) is a directed edge independently with probability p. Let $p = \theta/n$. Then w.h.p. $\bar{G}(n,p)$ has a directed path of length at least $(1 - \frac{4\log 2}{\theta})n$.*

3. Since the property of containing a path of length ℓ is monotone, we can derive the corresponding result for the model $G(n,m)$:

   Theorem 3.4 *Consider $G(n,m)$ with $m = \theta n/2$. Then w.h.p. $G \sim G(n,m)$ has a path of length at least $(1 - \frac{4\log 2}{\theta})n$.*

3.2 Hamiltonicity

Definition 3.5 (Hamiltonicity) *Let G be a graph on n vertices.*

(i) A path *P in G is called Hamilton if it has $n-1$ edges.*

(ii) A cycle *C in G is called Hamilton if it has n edges.*

(iii) The graph *G itself is called Hamiltonian if it has a Hamilton cycle.*

One ultimate goal is to establish the threshold for Hamiltonicity in the random graph $G(n,p)$ and to prove the following result:

Theorem 3.6 *The threshold for Hamiltonicity is at $p = (\log n + \log\log n)/n$:*

1. *Let*
$$p(n) = \frac{\log n + \log\log n - \omega(1)}{n},$$
where $\omega(1) \to \infty$. Then w.h.p. $G(n,p)$ is not Hamiltonian.

2. *Let*
$$p(n) = \frac{\log n + \log\log n + \omega(1)}{n},$$
where $\omega(1) \to \infty$. Then w.h.p. $G(n,p)$ is Hamiltonian.

Note that the first part hinges on proving that not all vertices have degree at least 2, which of course is a necessary condition, and is thus fairly easy.

3.2.1 Combinatorial background

Notation: Given a graph G, let $\ell(G)$ be the length of a longest path in G.

Definition 3.7 *Let $G = (V,E)$ be a graph and $e = (u,v) \notin E$. If adding e to G results in a graph $G' = G + e$ such that $\ell(G') > \ell(G)$ or G' has a Hamilton cycle, then e is called a* booster.

Observe that if G is already Hamiltonian, then *every* edge $e \notin E(G)$ is a booster. More importantly, starting from *any* graph G and adding a sequence of n boosters creates a Hamiltonian graph (observe that the set of boosters may change with every addition).

Claim *Let G be a connected graph. Let $P = (v_0, \ldots, v_\ell)$ be a longest path in G. If $e = (v_0, v_\ell) \notin E(G)$ then e is a booster in G.*

Proof Consider $G' = G + e$. Obviously, G' contains a cycle $C = (v_0, \ldots, v_\ell, v_0)$. If C is Hamilton, e is obviously a booster. Otherwise, $V(C) \subsetneq V(G) = V(G')$ and therefore there is $u \in V(G) \setminus V(C)$. Furthermore G is connected, and hence so is G'. Therefore G' contains a path P' from u to C. Let w be the last vertex of P' before hitting C, and let $(w, v_i) \in E(P')$. Then G contains a path Q as follows:

$$Q := (w, v_i, v_{i-1}, \ldots, v_0, v_\ell, v_{\ell-1}, \ldots, v_{i+1}).$$

Observe that $V(P) \subsetneq V(Q)$ because $w \in V(Q) \setminus V(P)$. Hence Q is strictly longer than P, implying $\ell(G') > \ell(G)$. □

3.2.2 Pósa's rotation-extension technique

Developed by Pósa in 1976 [Pós76] (in his proof that $G(n, C\log n/n)$ is w.h.p. Hamiltonian, for some large enough $C > 0$). Suppose $P = (v_0, \ldots, v_\ell)$ is a

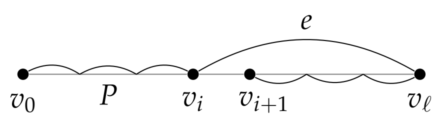


Figure 3.1: Pósa rotation

longest path in G. For every edge $e \in E(G)$ containing v_ℓ, the other endpoint of e is on P (otherwise you could extend the path).

If $e = (v_0, v_\ell)$ then we get a cycle C with $V(C) = V(P)$. If $e = (v_i, v_\ell)$ with $1 \leq i \leq \ell - 2$, we can *rotate* P at v_i by adding edge e, deleting (v_i, v_{i+1}) to get a new path P' (see Figure 3.1):

$$P' := (v_0, \ldots, v_i, v_\ell, v_{\ell-1}, \ldots, v_{i+1}) .$$

Observe that $V(P') = V(P)$, and therefore P' is also a longest path in G, but ending at v_{i+1} instead of v_ℓ. We can then rotate P' to get a new longest path etc.

Suppose Q is obtained from P by a sequence of rotations with a fixed starting point v_0. Let v be the endpoint of Q. If $(v_0, v) \in E(G)$, then G has a cycle C' with $V(C') = V(P)$. If $(v, u) \in E(G)$ for some $u \notin V(P)$, then we can append (v, u) to Q and get a longer path (*extension* step).

3.2.3 Pósa's lemma

Let $G = (V, E)$ be a graph, with longest path $P = (v_0, \ldots, v_\ell)$. Notation:

1. For $U \subseteq V(G)$, denote by $N(U)$ the *external neighborhood* of U in G,

$$N(U) := \{v \in V \setminus U \mid v \text{ has a neighbor in } U\} .$$

2. Suppose $R \subseteq V(P)$. Denote by R^- and R^+ the set of vertices of P which are neighbors of R to the left and right, respectively, relative to P.

For example, if $P = (v_0, \ldots, v_6)$ and $R = \{v_2, v_3, v_6\}$, then $R^- = \{v_1, v_2, v_5\}$ and $R^+ = \{v_3, v_4\}$. Observe that $|R^-| \leq |R|$ and $|R^+| \leq |R|$. Moreover, $|R^+| = |R| - 1$ if $v_\ell \in R$.

Lemma 3.8 (Pósa) *Let G and P as before. Let R be the set of endpoints of paths Q obtained from P by a sequence of rotations with a fixed starting point v_0. Then*

$$N(R) \subseteq R^- \cup R^+ .$$

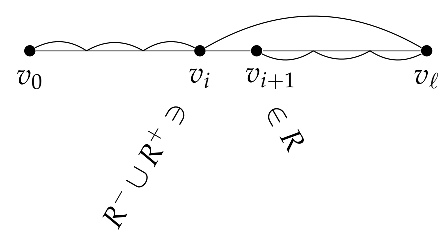


Figure 3.2: Rotation (situation in proof)

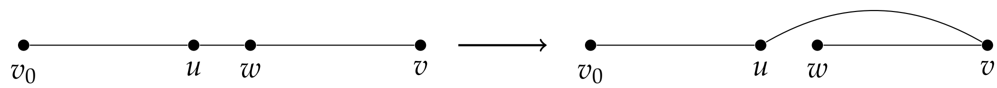


Figure 3.3: u, v rotation

This is useful, because if every small set U of G expands itself, then R has to be large, as

$$|N(R)| \leq |R^-| + |R^+| \leq 2|R|\,.$$

Proof Let $v \in R$, and let $u \in V(G) \setminus (R \cup R^- \cup R^+)$. We need to prove $(u, v) \notin E(G)$.

If $u \notin V(P)$, consider a path Q ending at v and obtained from P by a sequence of rotations with v_0 as a fixed starting point. Then $(u, v) \notin E(G)$ as otherwise one can append (u, v) to Q and get a longer path, contradicting the choice of P.

If $u \in V(P) \setminus (R \cup R^- \cup R^+)$ (as in Figure 3.2), then u has the *exact same* neighbors along every path P' obtained from P by a sequence of rotations (as breaking an edge during a rotation puts one of its endpoints in R and the other in $R^- \cup R^+$).

Consider now the path Q as before. If $(u, v) \in E(G)$, we can rotate Q at u (Figure 3.3) to get a new path R' ending at w such that w is a neighbor of u along P. Such a rotation puts u into $R^- \cup R^+$, which is a contradiction. □

Definition 3.9 *Let k be a positive integer and let $t > 0$ be real. A graph $G = (V, E)$ is called a (k, t)-*expander *if $|N(U)| > t|U|$ for every $U \subseteq V$ with $|U| \leq k$.*

Theorem 3.10 *Let G be a $(k, 2)$-expander. Then G has a path of length at least $3k - 1$. If G is in addition connected and non-Hamiltonian, then G has at least $(k+1)^2/2$ boosters.*

Proof Let $P = (v_0, \ldots, v_\ell)$ be a longest path in G, and let R be the set of endpoints of paths Q obtained from P by a sequence of rotations with a fixed starting point v_0.

Then $|R^-| \leq |R|$ and $|R^+| \leq |R| - 1$ (as $v_\ell \in R$). By Pósa's lemma, $N(R) \subseteq R^- \cup R^+$, implying

$$|N(R)| \leq |R^-| + |R^+| \leq 2|R| - 1 .$$

By our assumption this implies that $|R| > k$. Let $R_0 \subseteq R$ with $|R_0| = k$. Then $N(R_0) \subseteq R \cup R^- \cup R^+$, implying $N(R_0) \subseteq V(P)$. By our assumption $|N(R_0)| \geq 2|R_0| = 2k$. But R_0 and $N(R_0)$ are both contained in $V(P)$, hence

$$|V(P)| \geq k + 2k = 3k .$$

implying that P has at least $3k - 1$ edges, as required.

For the second part, assume G is connected and non-Hamiltonian. If $v \in R$, then $(v_0, v) \notin E(G)$ (as otherwise we would get a cycle C with $V(C) = V(P)$ and would proceed as before). For the same reason, (v_0, v) is a booster. It follows that v_0 participates in $\geq (k+1)$ boosters (v_0, v) (for $v \in R$).

Now, for every $v \in R$, there is a longest path $P(v)$ starting at v. We can now rotate $P(v)$ while keeping v as a fixed starting point, to get a set of endpoints $R(v)$, satisfying $|R(V)| \geq k + 1$. Every pair (u, v) with $u \in R(v)$ is a non-edge of G and is a booster as before. Altogether we get $(k+1)^2$ pairs. Each pair is counted at most twice, therefore we get at least $(k+1)^2/2$ boosters.□

3.2.4 Hamiltonicity of $G(n, p)$

The goal of this section is to prove Theorem 3.6.

The lower bound follows immediately from the following proposition: if $\delta(G) < 2$ (where δ is the minimum degree as usual), then G is not Hamiltonian.

Proposition 3.11 *Let $k \geq 1$ be a fixed integer.*

(i) If

$$p(n) = \frac{\log n + (k-1) \log\log n - \omega(n)}{n}$$

where $\omega(n) \to \infty$, then w.h.p. $\delta(G(n, p)) \leq k - 1$.

(ii) If

$$p(n) = \frac{\log n + (k-1) \log\log n + \omega(n)}{n}$$

where $\omega(n) \to \infty$, then w.h.p. $\delta(G(n, p)) \geq k$.

Proof Homework. □

Let $G \sim G(n,p)$ and represent G as $G = G_1 \cup G_2$, where $G_i \sim G(n,p_i)$, $i = 1,2$. Set $p_2 = \frac{c}{n}$ for $c > 0$ large enough. From $(1-p_1)(1-p_2) = 1-p$ we get $p_1 \geq p - p_2$ and therefore

$$p_1 = \frac{\log n + \log\log n + \omega_1(n)}{n},$$

where $\omega_1(n) \to \infty$.

Lemma 3.12 *Let $p(n) = \frac{\log n + \log\log n + \omega_1(n)}{n}$. Then with high probability the random graph $G(n,p)$ is an $(n/4,2)$-expander.*

To prove this lemma we first define the subset Small $\subseteq [n]$ by

$$\textsc{Small} = \{v \in [n] : d_G(v) \leq \log^{7/8} n\}.$$

The lemma is an immediate consequence of the following two propositions.

Proposition 3.13 *Let $p = p(n)$ be as above. Then with high probability $G(n,p)$ has the following property:*

(a) $\delta(G) \geq 2$.

(b) No vertex of Small *lies on a cycle of length ≤ 4. G does not contain a path between two vertices of* Small *of length at most 4.*

(c) G has an edge between every two disjoint vertex subsets A, B of sizes $|A|, |B| \geq \frac{n}{\log^{1/2} n}$.

(d) Every vertex set $V_0 \subseteq [n]$ of size $|V_0| \leq \frac{2n}{\log^{3/8} n}$ spans at most $3|V_0|\log^{5/8} n$ edges.

Proposition 3.14 *Let G be a graph with vertex set $[n]$ satisfying the properties (a)-(d) above. Then G is an $(n/4,2)$-expander, for large enough n.*

Proof (of Proposition 3.13)

(a) Apply Proposition 3.11 with $k = 2$.

(Remark: Property (a) is in fact the bottleneck of the proof. Properties (b)-(d) hold with high probability for somewhat smaller values of $p(n)$, for example for $p(n) \geq \frac{\log n}{n}$.)

(b) Intuition first. For a fixed $v \in [n]$, its degree $d(v)$ in $G(n,p)$ is $\text{Bin}(n-1,p)$ distributed. Note that for every constant $c > 0$ and for every $k \leq$

$\log^{1-\varepsilon} n$ we have

$$\begin{aligned}
\Pr[\mathrm{Bin}(n-c,p) \le k] &\le (k+1)\Pr[\mathrm{Bin}(n-c,p) = k] \\
&= (k+1)\binom{n-c}{k} p^k (1-p)^{n-c-k} \\
&\le \log n (np)^k e^{-pn} e^{(k+c)p} \\
&\le \log n (2\log n)^k \frac{1}{n} \le n^{-0.9} \,.
\end{aligned}$$

where the first inequality is due to the monotonicity of Bin. In particular this implies that

$$\mathbb{E}\left[|\textsc{Small}|\right] \le n n^{-0.9} = n^{0.1}$$

and so by Markov we have w.h.p. $|\textsc{Small}| \le n^{0.2}$.

Now we prove (b) formally. Let's prove that with high probability Small is an independent set. For $u \ne v \in [n]$ let $A_{u,v}$ be the event

$$A_{u,v} = \text{“}u, v \in \textsc{Small}, (u,v) \in E(G)\text{”} \,.$$

We have that

$$\Pr[A_{u,v}] = p\big(\Pr[\mathrm{Bin}(n-2,p) \le \log^{7/8} n - 1]\big)^2 \le p n^{-0.9\cdot 2} \le n^{-2.7}$$

Hence by the union bound

$$\Pr[\textsc{Small} \text{ is not ind.}] = \Pr\Big[\bigcup_{u\ne v} A_{u,v}\Big] \le \binom{n}{2} n^{-2.7} = o(1) \,.$$

Let's prove that with high probability every two vertices in Small do not have a common neighbor. For $u, v, w \in [n]$ distinct denote by $A_{u,v,w}$ the event

$$A_{u,v,w} = \text{“}u, v \in \textsc{Small}; (u,w), (v,w) \in E(G)\text{”} \,.$$

Here we have that

$$\begin{aligned}
\Pr[A_{u,v,w}] = \ & p^2\Big(p\big(\Pr[\mathrm{Bin}(n-3,p) \le \log^{7/8} n - 2]\big)^2 \\
& \quad + \big((1-p)\Pr[\mathrm{Bin}(n-3,p) \le \log^{7/8} n - 1]\big)^2\Big) \\
\le \ & p^2 n^{-0.9\cdot 2} \le n^{-3.7} \,.
\end{aligned}$$

Then by the union bound $\Pr[\exists u, v, w\; A_{u,v,w}] = o(1)$. The remaining cases are treated similarly.

(c) For this case we have that

$$\begin{aligned}
&\Pr[\text{(c) does not hold}] \\
&\quad = \Pr\Big[\exists A, B \text{ disjoint}, |A|, |B| = \frac{n}{\log^{1/2} n}, e_G(A,B) = 0\Big] \\
&\quad \le \binom{n}{\frac{n}{\log^{1/2} n}}^2 (1-p)^{n^2/\log n} \\
&\quad \le (e \log^{1/2} n)^{\frac{2n}{\log^{1/2} n}} e^{-\frac{pn^2}{\log n}} = o(1)\,.
\end{aligned}$$

(d) Here we have

$$\begin{aligned}
&\Pr\Big[\exists V_0 \subseteq [n], |V_0| \le \frac{2n}{\log^{3/8} n}, e(V_0) \ge 3|V_0| \log^{5/8} n\Big] \\
&\qquad \le \sum_{k \le \frac{2n}{\log^{3/8} n}} \binom{n}{k} \Pr[\mathrm{Bin}(\tbinom{k}{2}, p) \ge 3k \log^{5/8} n] \\
&\qquad \le \sum_{k \le \frac{2n}{\log^{3/8} n}} \binom{n}{k} \binom{\binom{k}{2}}{3k \log^{5/8} n} p^{3k \log^{5/8} n} \\
&\qquad \le \sum_{k \le \frac{2n}{\log^{3/8} n}} \left(\frac{en}{k} \Big(\frac{ekp}{6 \log^{5/8} n} \Big)^{3 \log^{5/8} n} \right)^k \\
&\qquad = \sum_{k \le \frac{2n}{\log^{3/8} n}} a_k
\end{aligned}$$

where

$$a_k = \left(\frac{en}{k} \Big(\frac{ekp}{6 \log^{5/8} n} \Big)^{3 \log^{5/8} n} \right)^k .$$

If $k \le \sqrt{n}$, then

$$a_k \le \left(enn^{-\frac{1}{3} 3 \log^{5/8} n} \right)^k = o(1/n)\,.$$

If on the other hand $\sqrt{n} \le k \le \frac{2n}{\log^{3/8} n}$, then

$$\begin{aligned}
a_k &= \left(\frac{en}{k} \frac{ekp}{6 \log^{5/8} n} \Big(\frac{ekp}{6 \log^{5/8} n} \Big)^{3 \log^{5/8} n - 1} \right)^k \\
&\le \left(\log n \Big(\frac{ekp}{6 \log^{5/8} n} \Big)^{2 \log^{5/8} n} \right)^k \\
&\le \left(\log n 0.95^{2 \log^{3/8} n} \right)^{\sqrt{n}} = o(1/n)\,.
\end{aligned}$$

Therefore (d) holds with high probability. □

Proof (of Proposition 3.14) We need to prove that if G has properties (a)-(d) then for all $U \subseteq [n]$, $|U| \leq n/4$ we have that $|N(U)| \geq 2|U|$.

If $|U| \geq \frac{n}{\log^{1/2} n}$ we have by property (c) that

$$|N(U)| \geq n - |U| - \frac{n}{\log^{1/2} n} \geq \frac{n}{2} \geq 2|U|\,.$$

Assume now that $|U| \leq \frac{n}{\log^{1/2} n}$. Let $A = U \cap \textsc{Small}$, and $B = U \setminus \textsc{Small} = U - A$. Then

$$N(U) = N(A \cup B) = N(A) - \big(B \cap N(A)\big) + \big(N(B) - A \cup N(N)\big)$$

so we obtain

$$|N(U)| \geq |N(A)| - |B| + |N(B) - A \cup N(A)|\,.$$

Recall that by Property (a), $d(u) \geq 2$ for every $v \in [n]$. Also, by Property (b), $\textsc{Small}$ is an independent set, and the neighborhoods of the vertices in $\textsc{Small}$ are pairwise disjoint. Hence

$$|N(A)| = \sum_{v \in A} d(v) \geq 2|A|\,.$$

Now let us estimate $|N(B)|$. Since $B \cap \textsc{Small} = \emptyset$, we have $d(v) \geq \log^{7/8} n$ for every $v \in B$. It thus follows that the set $B \cup N(B)$ spans all edges touching B, whose number is at least $(|B| \log^{7/8} n)/2$ (every edge is counted at most twice). We claim that $|N(B)| \geq |B| \log^{1/8} n$. Assume otherwise. Then

$$|B \cup N(B)| \leq |B| + |B| \log^{1/8} n \leq \frac{2n}{\log^{3/8} n}\,.$$

This is because we assumed that $|U| \leq \frac{n}{\log^{1/2} n}$, and hence Property (d) applies to $B \cup N(B)$. Thus, $B \cup N(B)$ has at most $|B| + |B| \log^{1/8} n$ vertices and yet spans at least $(|B| \log^{7/8} n)/2$ edges. We obtain

$$\frac{e(B \cup N(B))}{|B \cup N(B)|} \geq \frac{|B| \log^{7/8} n}{|B|(1 + \log^{1/8} n)} = \Omega(\log^{3/4} n)$$

thus contradicting Property (d).

Also observe that by Property (b), every vertex $b \in B$ has at most one neighbor in $A \cup N(A)$. Hence $|N(B) - A \cup N(A)| \geq |N(B)| - |B|$.

Putting everything together, we get

$$\begin{aligned}|N(U)| &\geq |N(A)| - |B| + |N(B)| - |B| \\ &\geq 2|A| + |B| \log^{1/8} n - 2|B| \geq 2|A| + 2|B| = 2|U|\,.\end{aligned} \qquad \square$$

The second part of Theorem 3.6 will be established by proving the following statement.

Lemma 3.15 *Let G_1 be an $(n/4, 2)$-expander with vertex set $[n]$. Let furthermore $G_2 \sim G(n, p_2)$ with $p_2 = 80/n$. Then with high probability (over the choice of G_2) the graph $G = G_1 \cup G_2$ is Hamiltonian.*

Proof First observe that G_1 is necessarily connected. Indeed, let C be a connected component of G_1. Then $N(C) = \emptyset$, implying that $|C| > n/4$. Choose $V_0 \subseteq C$ with $|V_0| = n/4$. Then $|N(V_0)| \geq n/2$, and $V_0 \cup N(V_0) \subseteq C$, implying $|C| \geq 3n/4$. So there is no room for more than one connected component.

Represent

$$G_2 = \bigcup_{j=1}^{2n} G_{2,j}$$

where $G_{2,j} \sim G(n, \rho)$ and $\rho = \rho(n)$ satisfies $(1-\rho)^{2n} = (1-p_2)$, implying $\rho \geq p_2/(2n) = 40/n^2$. For $0 \leq j \leq 2n$ denote

$$H_j = G_1 \cup \bigcup_{k=1}^{j} G_{2,k}\,.$$

We say that round j is *successful* if H_{j-1} is Hamiltonian or $E(G_{2,j})$ hits the set of boosters of H_{j-1}.

Observe that if at least n rounds are successful, then the final graph $G_1 \cup G_2$ is necessarily Hamiltonian. Let us consider round j. Either H_{j-1} is Hamiltonian, or, due to the fact that $H_{j-1} \supseteq G_1$ and G_1 is a connected $(n/4, 2)$-expander, H_{j-1} has at least $(n/4)^2/2 = n^2/32$ boosters.

In either case,

$$\begin{aligned}\Pr[\text{round } j \text{ is successful}] &\geq 1 - (1-\rho)^{n^2/32} \geq 1 - e^{-\rho n^2/32} \\ &\geq 1 - e^{-\frac{40}{n^2}\frac{n^2}{32}} = 1 - e^{-5/4} \geq \frac{2}{3}\,.\end{aligned}$$

Let X be the random variable counting the number of successful rounds. Then X stochastically dominates the binomial r.v. $\mathrm{Bin}(2n, 2/3)$. Hence w.h.p. (in fact with exponentially high probability) $X \geq n$. □

Proof (Proof of Theorem 3.6) Represent $G = G_1 \cup G_2$, with $G_i \sim G(n, p_i)$ as usual, and

$$p_1 = \frac{\log n + \log\log n + \omega_1(n)}{n}\,, \quad p_2 = \frac{c}{n}\,.$$

Then by Proposition 3.14 and Lemma 3.15 the result follows. □

Chapter 4

Random Graph Processes and Hitting Times

Recall from Section 1.2:

Definition 4.1 (Random graph process) *Take a permutation $\sigma = (e_1, \ldots, e_N)$ of the edges of K_n. Define*

$$G_0 = ([n], \emptyset)\,,$$
$$G_i = ([n], \{e_1, \ldots, e_i\}) \quad \forall 1 \le i \le N\,.$$

Then $\tilde{G} = \tilde{G}(\sigma) = (G_i)_{i=0}^{N}$ is a graph process.

If you choose $\sigma \in S_N$ uniformly at random, then $\tilde{G}(\sigma)$ is called a random graph process.

Informally:

1. Start with $G_0 = ([n], \emptyset)$.
2. At step i ($1 \le i \le N$): choose a random missing edge e_i uniformly from $E(K_n) \setminus E(G_{i-1})$. Set $G_i := G_{i-1} + \{e_i\}$.

Generating the random graph process $\tilde{G}$ and taking a *snapshot* at time m (i.e. looking at the graph G_m) induces the probability space $G(n, m)$.

Definition 4.2 (Hitting time) *Let P be a monotone (increasing) graph property. Let us assume that $\bar{K}_n \notin P$ (the empty graph does not satisfy the property) and $K_n \in P$. The* hitting time *of property P w.r.t. a graph process $\tilde{G}$, denoted by $\tau(\tilde{G}, P)$, is defined by*

$$\tau(\tilde{G}, P) = \min\{i \mid G_i \in P\}\,.$$

Since P is monotone, $G_i \in P$ for every $i \ge \tau(\tilde{G}, P)$.

If $\tilde{G}$ is a random graph process, $\tau(\tilde{G}, P)$ becomes a random variable.

4.1 Hitting Time of Connectivity

Define

$$\tau_1(\tilde{G}) := \min\{i \mid \delta(G_i) \geq 1\}\,,$$
$$\tau_c(\tilde{G}) := \min\{i \mid G_i \text{ is connected}\}\,.$$

Obviously $\tau_1(\tilde{G}) \leq \tau_c(\tilde{G})$.

Theorem 4.3 *A typical random graph process $\tilde{G}$ becomes connected* exactly *at the moment the last isolated vertex disappears. Formally, for a random graph process $\tilde{G}$, w.h.p.*

$$\tau_1(\tilde{G}) = \tau_c(\tilde{G})\,.$$

Corollary 4.4 *Consider the random graph $G(n,m)$. Then*

(a) If $m = \frac{1}{2}n(\log n - \omega(n))$ (with $\omega(n) \to \infty$), then w.h.p. $G(n,m)$ is not *connected.*

(b) If $m = \frac{1}{2}n(\log n + \omega(n))$ (with $\omega(n) \to \infty$), then w.h.p. $G(n,m)$ is *connected.*

Proof For the first part, recall that for such $m = m(n)$, w.h.p. $G(n,m)$ does have isolated vertices. Hence w.h.p. it is not connected.

For the second part, recall that for such $m = m(n)$, w.h.p. $G(n,m)$ has *no* isolated vertices. Take a random graph process $\tilde{G}$ and run it till m as above. Then w.h.p. $\tau_1(\tilde{G}) \leq m$. By Theorem 4.3, $\tau_1(\tilde{G}) = \tau_c(\tilde{G})$, implying that $G(n,m)$ is connected w.h.p. □

Corollary 4.5 *Let $G \sim G(n,p)$.*

(a) If $p = (\log n - \omega(n))/n$, then w.h.p. $G(n,p)$ is not *connected.*

(b) If $p = (\log n + \omega(n))/n$, then w.h.p. $G(n,p)$ is *connected.*

Proof Follows from the asymptotic equivalence of $G(n,p)$ and $G(n,m)$. □

Proof (Proof of Theorem 4.3) Given a graph $G = ([n], E)$, define the vertex subset Small as

$$\text{Small}(G) = \{v \in [n] \mid d_G(v) \leq \log^{7/8} n\}\,.$$

Define the following three graph properties:

(A1) G has no isolated vertices.

(A2) Small is an independent set in G.

(A3) For every partition $V = V_1 \cup V_2$ with $|V_1|, |V_2| \geq \log^{7/8} n$, the graph G has an edge between V_1 and V_2.

Then the proof will follow from the following lemmas. $\square$

Lemma 4.6 *Let G be a graph on n vertices having properties (A1), (A2) and (A3). Then G is connected.*

Proof Observe that if v is a vertex of G and C is a connected components of G including v, then $N(v) \subseteq C$, and for every neighbor u of v, also $N(u) \subseteq C$.

Let C be a connected component of G, and let $v \in C$. If $v \in \textsc{Small}$, then by property (A1), v has a neighbor u, and by property (A2) $u \notin \textsc{Small}$, implying $|C| \geq |N(u)| > \log^{7/8} n$. On the other hand, if $v \notin \textsc{Small}$, then $|N(v)| \geq \log^{7/8} n$, again implying $|C| \geq |N(u)| > \log^{7/8} n$.

Now, if G is disconnected, then there is a partition $V = V_1 \cup V_2$ such that both V_1 and V_2 contain some connected components of G in full, and G has no edge between V_1 and V_2. But then $|V_1|, |V_2| \geq \log^{7/8} n$, thus contradicting property (A3). $\square$

Lemma 4.7 *W.h.p. the random graph process $\tilde{G}$ is such that G_{τ_1} has properties (A1), (A2) and (A3).*

Proof (A1) follows by definition of τ_1. To prove (A2) and (A3), define

$$\begin{aligned} m_1 &:= \tfrac{1}{2} n \log n - \tfrac{1}{2} n \log\log\log n \,, \\ m_2 &:= \tfrac{1}{2} n \log n + \tfrac{1}{2} n \log\log\log n \,, \\ p_1 &:= \frac{m_1}{\binom{n}{2}} \geq \frac{\log n - \log\log\log n}{n} \,. \end{aligned}$$

The lemma then follows from the following propositions. $\square$

Proposition 4.8 *W.h.p.,*

$$m_1 \leq \tau_1 \leq m_2 \,.$$

Proof Follows from previously cited results about isolated vertices. $\square$

Corollary 4.9 *W.h.p.,*

$$\textsc{Small}(G_{m_1}) \supseteq \textsc{Small}(G_{\tau_1}) \supseteq \textsc{Small}(G_{m_2}) \,.$$

Proposition 4.10 *W.h.p. $\textsc{Small}(G_{m_1})$ is an independent set.*

Proof Follows from Proposition 3.13. $\square$

Proposition 4.11 *W.h.p. no edge of $\tilde{G}$ falls into $\textsc{Small}(G_{m_1})$ between m_1 and m_2.*

Proof We have previously observed that for every $v \in [n]$, we can bound $\Pr[v \in \textsc{Small}] \leq n^{-0.9}$, implying $\mathbb{E}[|\textsc{Small}|] \leq n^{0.1}$. By Markov, w.h.p. $|\textsc{Small}(G_{m_1})| \leq n^{0.2}$. Then the probability of putting some edge inside $\textsc{Small}(G_{m_1})$ between m_1 and m_2 can be estimated as follows:

$$(m_2 - m_1)\frac{\binom{|\textsc{Small}(G_{m_1})|}{2}}{\binom{n}{2} - m_2} = o(1)\,.$$

Hence w.h.p. $\textsc{Small}(G_{\tau_1})$ is independent as well. $\square$

This proves (A2), so it remains to show (A3). Since w.h.p. $\tau_1 \geq m_1$, it is enough to prove that $G_{m_1} \sim G(n, m_1)$ has property (A3). By the asymptotic equivalence of $G(n,p)$ and $G(n,m)$ it is enough to prove it for $G(n, p_1)$.

Proposition 4.12 *W.h.p. $G(n, p_1)$ has the property (A3).*

Proof

$$\begin{aligned}
\Pr[\text{(A3) is violated in } G(n,p_1)] &\leq \sum_{i=\log^{7/8} n}^{n/2} \binom{n}{i}(1-p_1)^{i(n-i)} \\
&\leq \sum_{i=\log^{7/8} n}^{n/2} \left(\frac{en}{i}e^{-p_1(n-i)}\right)^i = \sum_{i=\log^{7/8} n}^{n/2} g(i)\,,
\end{aligned}$$

where

$$g(i) = \left(\frac{en}{i}e^{-p_1(n-i)}\right)^i .$$

If $i \leq n^{2/3}$, then

$$\begin{aligned}
g(i) &\leq \left(\frac{en}{i}e^{-(\log n - \log\log\log n)(1-i/n)}\right)^i \leq \left(\frac{en}{i}e^{-\log n + 2\log\log\log n}\right)^i \\
&\leq \left(\frac{en}{\log^{7/8} n}\frac{(\log\log n)^2}{n}\right)^i \leq \left(\frac{1}{\log^{1/2} n}\right)^i .
\end{aligned}$$

If $n^{2/3} \leq i \leq n/2$, then

$$\begin{aligned}
g(i) &\leq \left(\frac{en}{i}e^{-(\log n - \log\log\log n)(1-i/n)}\right)^i \leq \left(\frac{en}{i}e^{-\frac{1}{2}\log n + \frac{1}{2}\log\log\log n}\right)^i \\
&\leq \left(en^{1/3}\frac{(\log\log n)^{1/2}}{n^{1/2}}\right)^i = o(1/n)\,.
\end{aligned}$$

So altogether, $\Pr[\text{(A3) does not hold}] = o(1)$. $\square$

Chapter 5

Coloring Random Graphs

5.1 Graph Theoretic Background

Let $G = (V, E)$ be a graph.

Definition 5.1 *A set $I \subset V$ is* independent *(or* stable*) if I spans no edges of G. The* independence number *of G, denoted by $\alpha(G)$, is the largest size of an independent set in G.*

*A partition $V = C_1 \cup \cdots \cup C_k$ is a k-*coloring *of G if each color class C_i is an independent set. Equivalently, a function $f : V \to [k]$ is a k-coloring if for every edge $e = (u, v) \in E$, we have $f(u) \neq f(v)$.*

*G is called k-*colorable *if it admits a k-coloring. The* chromatic number $\chi(G)$ *of G is the smallest k for which G is k-colorable.*

Examples:

1. Let $G = K_n$, then $\alpha(G) = 1$ and $\chi(G) = n$.
2. Let $G = K_{m,n}$, then $\alpha(G) = \max(m, n)$ and $\chi(G) = 2$.

5.2 Elementary Facts About Coloring

1. If $V(G) = V_1 \dot\cup V_2$, then
$$\chi(G) \leq \chi(G[V_1]) + \chi(G[V_2]) \leq \chi(G[V_1]) + |V_2| .$$
2. $\chi(G) \geq |V(G)|/\alpha(G)$: Let $V = C_1 \cup \cdots \cup C_k$ such that $k = \chi(G)$ and each C_i is independent. Then $|V| = \sum_{i=1}^{k} |C_i| \leq k\alpha(G)$.

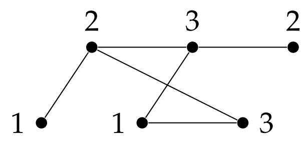


Figure 5.1: Greedy coloring example (vertices ordered left-to-right)

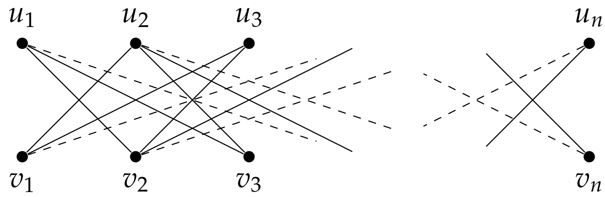


Figure 5.2: Greedy performs arbitrarily bad in this construction

5.3 Greedy Coloring

Take $G = (V, E)$ with $V = [n]$. Let $\sigma \in S_n$ be a permutation of $[n]$. The *greedy coloring* of G according to σ is defined by setting, for $i \leq i \leq n$,

$$c(\sigma(i)) = \min \{j \geq 1 \mid j \notin \{c(u) \mid u \in N(\sigma(i))\}\} .$$

In general the greedy algorithm performs reasonably well (see Figure 5.1 but note that this graph could be colored with only two colors). However, there are some bad cases: Take $G = K_{n,n} - M$ where M is a perfect matching consisting of edges $e_1 = (u_1, v_1), \ldots, e_n = (u_n, v_n)$ (see Figure 5.2). Let the permutation σ alternate between the parts of the bipartition in the order $(u_1, v_1, u_2, \ldots, u_n, v_n)$. Then the greedy algorithm uses n colors while $\chi(G) = 2$.

Remark 5.2 For every graph G, there is a permutation σ of $V(G)$ such that the greedy algorithm on G according to σ uses exactly $\chi(G)$ colors.

5.4 Coloring $G(n, \frac{1}{2})$

$p = \frac{1}{2}$ is chosen for illustrative purposes; similar results available for other values of $p = p(n)$.

5.4.1 Lower bound for $\chi(G(n, \frac{1}{2}))$

Define

$$f(k) = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^{\binom{k}{2}} .$$

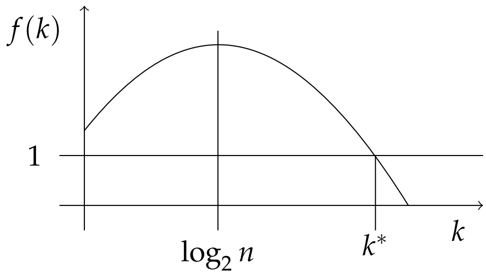


Figure 5.3: Expected number of independent sets in $G(n, \frac{1}{2})$

Clearly, $f(k)$ is the expected number of independent sets of size k in $G(n, \frac{1}{2})$. Calculate

$$\frac{f(k+1)}{f(k)} = \frac{\binom{n}{k+1} \cdot 2^{-\binom{k+1}{2}}}{\binom{n}{k} \cdot 2^{-\binom{k}{2}}} = \frac{n-k}{k+1} 2^{-k} .$$

The maximum is attained roughly at $\log_2 n$ (Figure 5.3).

Let $k^* = \max\{k \mid f(k) \geq 1\}$. It is easy to get

$$k^* = 2\log_2 n - 2\log_2 \log_2 n + \Theta(1) .$$

For $k \approx k^*$, $f(k+1)/f(k) = \tilde{\Theta}(1/n)$.

Claim *W.h.p. in $G(n, \frac{1}{2})$, we have $\alpha(G) < k^* + 2$.*

Proof Since $f(k^* + 1) < 1$ by the definition of k^*, we get that

$$f(k^* + 2) = f(k^* + 1)\tilde{\Theta}(1/n) = o(1) .$$

Therefore the expected number of independent sets of size $k^* + 2$ in $G(n, \frac{1}{2})$ is $o(1)$. By Markov, w.h.p. G has *no* independent sets of size $k^* + 2$. $\square$

Corollary 5.3 *W.h.p. in $G(n, \frac{1}{2})$,*

$$\chi(G) \geq \frac{n}{2\log_2 n - 2\log_2 \log_2 n + \Theta(1)} .$$

5.4.2 Greedy's Performance

Look at the performance of the greedy algorithm on $G(n, 1/2)$. Let σ be the identity permutation. Let $\chi_g(G)$ denote the random variable counting the number of colors used by the greedy algorithm on $G(n, \frac{1}{2})$ according to σ.

Observe that one can expose/generate $G(n, \frac{1}{2})$ as follows: for each $2 \leq i \leq n$, each pair (j, i) (where $1 \leq j \leq i-1$) is an edge of G with probability $\frac{1}{2}$. Thus the greedy algorithm is easy to analyze on random graphs.

Theorem 5.4 ([GM]) *W.h.p.,*

$$\chi_g(G(n, \tfrac{1}{2})) \leq (1 + o(1))\frac{n}{\log_2 n}.$$

Proof Set

$$t = \left\lceil \frac{n}{\log_2 n - 3\log_2 \log_2 n} \right\rceil .$$

We will prove that w.h.p. in $G(n, \frac{1}{2})$, we have $\chi_g(G) \leq t$.

for $t+1 \leq i \leq n$, let A_i be the event "i is the first vertex colored by color $t+1$ by the greedy algorithm". Then

$$\text{“}\chi_g(G) > t\text{”} = \bigcup_{i=t+1}^{n} A_i \,,$$

hence

$$\Pr[\chi_g(G) > t] = \Pr\left[\bigcup_{i=t+1}^{n} A_i\right] = \sum_{i=t+1}^{n} \Pr[A_i]\,.$$

Consider A_{i_0} with $t+1 \leq i_0 \leq n$. In order for A_{i_0} to happen, the following should occur:

(i) The greedy algorithm has used all of t colors at vertices $1, \ldots, i_0 - 1$.

(ii) i_0 has a neighbor in each of the t color classes created so far.

Condition on a coloring $C_1, \ldots, C_t$ of the first $i_0 - 1$ vertices; observe that $C_j \neq \emptyset$ (for $1 \leq j \leq t$) by (i). Expose the edges from i_0 to $[i_0 - 1]$. Then the probability that i_0 has a neighbor in each C_j ($1 \leq j \leq t$) is

$$\prod_{j=1}^{t}\left(1 - 2^{-|C_j|}\right) .$$

The function $g(x) = 1 - 2^{-x}$ is concave, so

$$\begin{aligned}
\prod_{j=1}^{t} \left(1 - 2^{-|C_j|}\right) &\leq \left(1 - 2^{-\frac{1}{t}\sum_{j=1}^{t}|C_j|}\right)^t \\
&= \left(1 - 2^{-\frac{i_0 - 1}{t}}\right)^t \\
&\leq \left(1 - 2^{-\frac{n}{t}}\right)^t \\
&\leq \exp\left(-t2^{-n/t}\right) \\
&= \exp\left(-\left\lceil \frac{n}{\log_2 n - 3\log_2\log_2 n} \right\rceil 2^{-\log_2 n + 3\log_2\log_2 n}\right) \\
&\leq \exp\left(-(1 - o(1))\log_2^2 n\right) \\
&= o\left(\frac{1}{n}\right)
\end{aligned}$$

Lifting the conditioning on the particular coloring $(C_1, \ldots, C_t)$, we get

$$\Pr[\chi_g(G) > t] = o(1) .$$

□

The greedy algorithm is extremely robust when applied to $G(n, \frac{1}{2})$, as witnessed by the following result.

Theorem 5.5 ([McD79])

$$\Pr\left[\chi_g(G) \geq \left(1 + \frac{5\log_2\log_2 n}{\log_2 n}\right)\frac{n}{\log_2 n}\right] = o(n^{-n}) .$$

Since $|S_n| = n! = o(n^n)$, we conclude that w.h.p. $G \sim G(n, \frac{1}{2})$ is such that the greedy algorithm uses at most

$$(1 + o(1))\frac{n}{\log_2 n}$$

colors for *any* permutation σ of G's vertices.

5.4.3 Lower bounding $\chi_g(G)$

Theorem 5.6 *W.h.p. in $G(n, \frac{1}{2})$ all color classes produced by the greedy algorithm are of size at most $\log_2 n + 2\sqrt{\log_2 n}$.*

Corollary 5.7 *W.h.p. the greedy algorithm, when applied to $G(n, \frac{1}{2})$, uses*

$$\chi_g(G) \geq \frac{n}{\log_2 n + 2\sqrt{\log_2 n}}$$

colors.

Proof (Proof of Theorem 5.6) Denote $r_1 = \log_2 n$ and $r_2 = 2\sqrt{\log_2 n}$. If C_i gets $r_1 + r_2$ vertices, we first put r_1 vertices $u_1, \dots, u_{r_1}$ into C_i, and then add another r_2 vertices $v_1, \dots, v_{r_2}$. We can assume

$$u_1 < u_2 < \cdots < u_{r_1} < v_1 < \cdots < v_{r_2} .$$

Let us condition on $(u_1, \dots, u_{r_1})$. For a given sequence $(v_1, \dots, v_{r_2})$, the probability that $v_1, \dots, v_{r_2}$ gets added to $\{u_1, \dots, u_{r_1}\}$ is at most

$$2^{-r_1} \cdot 2^{-(r_1+1)} \dots 2^{-(r_1+r_2-1)}$$

(each of the v_i cannot have any neighbors among the u_i and from v_1 to v_{i-1}). We get a total probability of

$$2^{-r_1 \cdot r_2 - \binom{r_2}{2}} .$$

Altogether there are less than $(n)_{r_2}$ choices of $(v_1, \dots, v_{r_2})$. Therefore the probability that at least r_2 vertices get added to $\{u_1, \dots, u_{r_1}\}$ is at most

$$\begin{aligned}(n)_{r_2} \cdot 2^{-r_1 r_2 - \binom{r_2}{2}} \le n^{r_2} \cdot 2^{-r_1 r_2 - \frac{r_2(r_2-1)}{2}} &= \left(n \cdot 2^{r_1} + \frac{r_2 - 1}{2}\right)^{r_2} \\ &= 2^{-\binom{r_2}{2}} = 2^{-(1-o(1))2\log_2 n} \ll \frac{1}{n} .\end{aligned}$$

Lifting the conditioning on $(u_1, \dots, u_{r_1})$, we get

$$\Pr[|C_i| \ge r_1 + r_2] = o(1/n)$$

hence

$$\Pr[\exists C_i \colon |C_i| \ge r_1 + r_2] = o(1/n) .$$

□

5.4.4 Chromatic Number of $G(n, 1/2)$

Theorem 5.8 (Bollobás [Bol88]) *Let $G \sim G(n, \frac{1}{2})$, then w.h.p.*

$$\chi(G) = (1 + o(1)) \frac{n}{2\log_2 n}$$

Proof (Upper bound) We know: the expected number of independent sets of size k in $G(n, \frac{1}{2})$ is

$$f(k, n) = \binom{n}{k} 2^{-\binom{k}{2}} .$$

Let

$$k^* = \max\{k \mid f(k, n) \ge 1\}$$

then

$$k^* = 2\log_2 n - 2\log_2 \log_2 n + \Theta(1) .$$

W.h.p. $\alpha(G(n, \frac{1}{2})) \leq k^* + 1$, thus

$$\chi(G(n, \tfrac{1}{2})) \geq \frac{n}{k^* + 1} = \frac{(1 + o(1))n}{2 \log_2 n} . \qquad \square$$

For the lower bound we need to do some more work.

Janson Inequalities

Setting: Finite ground set Ω. Form a random subset R of Ω as follows: for all $r \in \Omega$, $\Pr[r \in R] = p_r$ independently from each other. (Typical example: $\Omega = E(K_n)$, $R = E(G(n, p))$ or $R = E(K_n) = E(G(n, p))$.)

Let I be indexes, and take a family $\{A_i\}_{i \in I}$ of subsets $A_i \subseteq \Omega$. We want to find out how many of the A_i's fall into R, in particular, to estimate the probability that *none* of them fall into R.

For each A_i define the corresponding indicator r.v.

$$X_i = \begin{cases} 1 & A_i \subseteq R\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Then

$$\Pr[X_i = 1] = \mathbb{E}[X_i] = \prod_{r \in A_i} p_r\,.$$

Now let X count the number of A_i's that fall entirely into R:

$$X = \sum_{i \in I} X_i\,.$$

Denote

$$\mu = \mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[X_i] = \sum_{i \in I} \prod_{r \in A_i} p_r\,.$$

If the A_i's are pairwise disjoint, the probabilities are independent and thus

$$\Pr[X = 0] = \prod_{i \in I} \Pr[X_i = 0] = \prod_{i \in I} \Big(1 - \prod_{r \in A_i} p_r\Big)$$

If $\prod_{r \in A_i} p_r = o(1)$, then we can approximate the latter expression by

$$\leq \prod_{i \in I} \exp\Big(- \prod_{r \in A_i} p_r\Big) = e^{-\mu}\,,$$

the so called *Poisson paradigm*: if X is a non-negative integer r.v. with $\mathbb{E}[X] = \mu$, then $\Pr[X = 0] \approx e^{-\mu}$. (This is exact for the Poisson distribution.)

Usually the A_i's intersect and therefore some dependencies should be taken into account. Write $i \sim j$ if $A_i \cap A_j \neq \emptyset$ (in this case X_i and X_j may be dependent). Now define

$$\Delta = \sum_{i \sim j} \Pr[(A_i \subseteq R) \wedge (A_j \subseteq R)]$$

where the summation runs over ordered pairs.

Theorem 5.9 (Janson's inequality [JŁR00]) *With the above notation,*

$$\Pr[X = 0] \leq e^{-\mu + \frac{\Delta}{2}} .$$

This is essentially useless if $\Delta \geq 2\mu$, so we have another form.

Theorem 5.10 (Extended/generalized Janson's inequality) *If in addition $\Delta \geq \mu$, then*

$$\Pr[X = 0] \leq e^{\frac{-\mu^2}{2\Delta}} .$$

For proof see, e.g., [AS00, Chapter 8]. Back to colorings:

Proposition 5.11 *Let $1 \leq k \leq m \leq n$ be integers. Let G be a graph on n vertices in which every m vertices span an independent set of size at least k. Then*

$$\chi(G) \leq \frac{n}{k} + m .$$

Proof Coloring by excavation. Start with $G' := G$ and $i := 1$. Proceed in two phases:

1. As long as $|V(G')| \geq m$: find an independent set C_i in G' of size $|C_i| = k$, color C_i by a new color, put it aside. Update $G' := G' - C_i$ and $i := i + 1$.
2. Color each of the remaining vertices in a new separate color.

Then phase 1 is repeated at most n/k times and thus uses at most n/k colors. Phase 2 uses m colors. Altogether we get the claim. □

Let $m = n/\log_2^2 n$. We will prove that w.h.p. every subset of size m of $G(n, \frac{1}{2})$ spans an independent set of size $(1 - o(1))2\log_2 n$.

Lemma 5.12 *Let $G \sim G(m, \frac{1}{2})$. Write*

$$k^* = k^*(m) = \max\left\{ k \,\middle|\, \binom{m}{k} 2^{-\binom{k}{2}} \geq 1 \right\} .$$

Set $k = k^* - 3$, then

$$\Pr[\alpha(G) < k] = \exp\left\{ -\Omega\left(\frac{m^2}{k^4}\right) \right\} .$$

Proof Notice first that

$$f(k, m) = \binom{m}{k} 2^{-\binom{k}{2}} \geq m^{3+o(1)} ,$$

since $k^* = 2\log_2 m(1 - o(1))$, $k = k^* - 3$, and

$$\frac{f(k_1 + 1, m)}{f(k_1, m)} = m^{-1+o(1)}$$

for $k_1 = (1 + o(1))k^*$. We will apply the extended Janson inequality. Set $\Omega = E(K_m) = \binom{[m]}{2}$ and let R be the set of non-edges of $G(m, \frac{1}{2})$. Then $\Pr[r \in R] = \frac{1}{2}$ for all $r \in \Omega$. For a subset $S \subseteq [m]$ with $|S| = k$, let $A_S = \binom{S}{2}$ be the set of $\binom{k}{2}$ pairs inside S. Denote

$$X_S = \begin{cases} 1 & A_S \subseteq R , \\ 0 & \text{otherwise} . \end{cases}$$

Then $X_S = 1$ iff S is an independent set in $G(m, \frac{1}{2})$. Let X count the number of independent sets of size k in $G(n, \frac{1}{2})$:

$$X = \sum_{\substack{S \subseteq [m] \\ |S| = k}} X_S .$$

Then

$$\Pr[\alpha(G(m, \tfrac{1}{2})) < k] = \Pr[X = 0] .$$

We have

$$\mathbb{E}[X] = \mu = \sum_{\substack{S \subseteq [m] \\ |S| = k}} \mathbb{E}[X_S] = \binom{m}{k} 2^{-\binom{k}{2}} \geq m^{3+o(1)} .$$

(We cannot hope that the Poisson principle applies here because with probability 2^{-m^2}, the graph is complete.)

We write $S \sim S'$ if $2 \leq |S \cup S'| \leq k - 1$ with $S, S' \subseteq [m]$ and $|S| = |S'| = k$.

Then

$$\begin{aligned}
\Delta &= \sum_{S\sim S'} \Pr[(X_S = 1) \wedge (X_{S'} = 1)] \\
&= \sum_{\substack{S\subseteq[m]\\|S|=k}} \sum_{i=2}^{k-1} \binom{k}{i}\binom{m-k}{k-i}\left(\frac{1}{2}\right)^{2\binom{k}{2}-\binom{i}{2}} \\
&= \binom{m}{k} \sum_{i=2}^{k-1} \binom{k}{i}\binom{m-k}{k-i}\left(\frac{1}{2}\right)^{2\binom{k}{2}-\binom{i}{2}} \\
&= \binom{m}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \sum_{i=2}^{k-1} \binom{k}{i}\binom{m-k}{k-i}\left(\frac{1}{2}\right)^{\binom{k}{2}-\binom{i}{2}} \\
&= \mu \sum_{i=2}^{k-1} g(i)\,,
\end{aligned}$$

where

$$g(i) = \binom{k}{i}\binom{m-k}{k-i}\left(\frac{1}{2}\right)^{\binom{k}{2}-\binom{i}{2}}\,.$$

Informally, looking at the ratio $g(i+1)/g(i)$ we find that g has a decreasing and an increasing part. It is easy to see that

$$\sum_{i=2}^{k-1} g(i) \le (1+o(1))(g(2)+g(k-1))\,.$$

Now

$$\begin{aligned}
g(2) &= \binom{k}{2}\binom{m-k}{k-2}\left(\frac{1}{2}\right)^{\binom{k}{2}-1} = (1+o(1))k^2\binom{m-k}{k-2}\left(\frac{1}{2}\right)^{\binom{k}{2}} \\
&= (1+o(1))k^2\frac{\binom{m-k}{k-2}}{\binom{m}{k}}\binom{m}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}} \\
&= (1+o(1))k^2\Theta\left(\frac{k^2}{m^2}\right)\mu \\
&= \Theta\left(\frac{k^4}{m^2}\mu\right)\,.
\end{aligned}$$

On the other hand

$$g(k-1) = \binom{k}{k-1}\binom{m-k}{1}\left(\frac{1}{2}\right)^{\binom{k}{2}-\binom{k-1}{2}} = k(m-k)2^{-(k-1)} = \tilde{O}(1/m)\,,$$

in particular $g(2) \gg g(k-1)$.

Altogether

$$\Delta = (1 + o(1))\mu g(2) = \Theta\left(\frac{k^4}{m^2}\mu^2\right) .$$

Observe that

$$\frac{\Delta}{\mu} = \Theta\left(\frac{k^4}{m^2}\mu\right) \gg 1 ,$$

hence we can apply the extended Janson inequality. It gives

$$\Pr[\alpha(G(n, \tfrac{1}{2})) < k] = \Pr[X = 0] \le \exp\left(-\frac{\mu^2}{2\Delta}\right) = \exp\left(-\Theta\left(\frac{m^2}{k^4}\right)\right) . \qquad \square$$

Let us go back to $G(n, \frac{1}{2})$.

Proof (Proof of Theorem 5.8, lower bound) Fix $V_0 \subseteq [n]$ with $|V_0| = m$. Then the subgraph $G[V_0]$ is distributed *exactly* $G(m, \frac{1}{2})$. Hence by Lemma 5.12

$$\Pr[\alpha(G[V_0]) < k] = \exp\left(-\Omega\left(\frac{m^2}{k^4}\right)\right) .$$

Applying the union bound, we get that

$$\begin{aligned}\Pr\left[\exists V_0 \subseteq [n],\, |V_0| = m \colon \alpha(G[V_0]) < k\right] &\le \binom{n}{m} \exp\left(-\Omega\left(\frac{m^2}{k^4}\right)\right) \\ &\le 2^n \exp\left(-\frac{cn^2}{\log_2^8 n}\right) = o(1) .\end{aligned}$$

Therefore w.h.p. $G(n, \frac{1}{2})$ satisfies the conditions of Proposition 5.11 for $m = n/\log_2^2 n$ and $k = k^*(m) - 3$, and we get

$$\begin{aligned}\chi(G(n, \tfrac{1}{2})) \le \frac{n}{k} + m &= \frac{n}{(1 + o(1))2\log_2 m} + m \\ &= \frac{n}{(1 + o(1))2\log_2 n} + \frac{n}{\log_2^2 n} = \frac{n(1 + o(1))}{2\log_2 n} . \qquad \square\end{aligned}$$

5.5 List Coloring of Random Graphs

5.5.1 Combinatorial Background: List Coloring

Introduced independently in [Viz76] and [ERT79].

Recall: G is k-colorable if $\exists c \colon V \to \{1, \ldots, k\}$ such that $c(u) \ne c(v)$ for every $(u, v) \in E(G)$. Note that the list of colors $\{1, \ldots, k\}$ is the same for every vertex. Generalization:

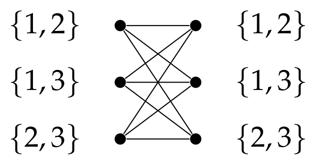


Figure 5.4: $K_{3,3}$ with lists proving choice number > 2

Definition 5.13 *Let $\mathcal{S} = \{S(v) \mid v \in V\}$ be a family of color lists ($S(v) \subseteq \mathbb{Z}$ for all $v \in V$). $G = (V, E)$ is* $\mathcal{S}$-choosable *if there exists $c \colon V \to \mathbb{Z}$ such that*

(i) $c(v) \in S(v)$ for every vertex $v \in V$, and

(ii) $c(u) \neq c(v)$ for every edge $(u, v) \in E$.

Example 5.14 *$S(v) = \{1, \ldots, k\}$ for all vertices, $\mathcal{S} = \{S(v) \mid v \in V\}$ gives: G is $\mathcal{S}$-choosable $\iff$ G is k-colorable.*

Definition 5.15 *G is* k-choosable *if G is $\mathcal{S}$-choosable for every family $\mathcal{S}$ of color lists satisfying $|S(v)| = k$ for all $v \in v$.*

Definition 5.16 *The* choice number *(or list chromatic number) of G, denoted by* $\operatorname{ch} G$ *or $\chi_\ell(G)$, is the least k for which G is k-choosable.*

Claim $\operatorname{ch}(G) \geq \chi(G)$ *for all graphs G.*

Proof Let $\operatorname{ch}(G) = k$. Then by definition G is $\mathcal{S}$-choosable from every family $\mathcal{S}$ of color lists satisfying $|S(v)| \geq k \ \forall v \in V$. Set $S(v) := \{1, \ldots, k\}$ for all $v \in V$ and $\mathcal{S} = \{S(v) \mid v \in V\}$. Then G is $\mathcal{S}$-choosable meaning that it is also k-colorable. $\square$

Claim *Let $\mathcal{S} = \{S(v) \mid v \in V\}$ be a family of color lists for a graph $G = (V, E)$. Assume $S(u) \cap S(v) = \emptyset$ for every pair of vertices $u \neq v$. Then G is $\mathcal{S}$-choosable.*

Proof Any choice function satisfying $c(v) \in S(v)$ for all $v \in V$ will do. $\square$

Does it happen that $\operatorname{ch}(G) \gg \chi(G)$? Yes, but as we will see, very rarely.

$\operatorname{ch}(G)$ can be larger than $\chi(G)$:

Example 5.17 *Let $G = K_{3,3}$ with lists as in Figure 5.4. Look at one half of the bipartition. It cannot be colored with only one color, since no color is available for all vertices. So one must use at least two colors. But then clearly the colors used for the two halves must overlap, so there is a monochromatic edge, so the choice number is at least* 3. *Hence* $\operatorname{ch}(G) > 2 = \chi(G)$.

The above example can easily be generalized:

Proposition 5.18 *Let $n = \binom{2k-1}{k}$. Then* $\operatorname{ch}(K_{n,n}) > k$.

Proof There are exactly $n = \binom{2k-1}{k}$ size k subsets of $[2k-1]$. Denote the sides of $G = K_{n,n}$ by A and B with $|A| = |B| = n$. Fix two arbitrary bijections f and g between the k-subsets of $[2k-1]$ and the vertices of A and B (respectively). Form a family $\mathcal{S} = \{S(v) \mid v \in V\}$ by setting $S \equiv f$ on A and $S \equiv g$ on B. We will prove that G is not $\mathcal{S}$-choosable.

Let $c\colon A \cup B \to [2k-1]$ be a choice function satisfying $c(v) \in S(v)$ for $v \in A \cup B$. Denote

$$T_A = \{c(a) \mid a \in A\}\,, \quad T_B = \{c(b) \mid b \in B\}\,.$$

Observe that $|T_A| \geq k$: if $|T_A| \leq k-1$, then $|[2k-1] \setminus T_A| \geq k$. Then there is a k-subset $S \subseteq [2k-1]$ completely missed by T_A (i.e., $S \cap T_A = \emptyset$). But S is a color list of some vertex, so $S \cap T_A \neq \emptyset$, a contradiction. Similarly, $|T_B| \geq k$.

But now since T_A and T_B are both subsets of $[2k-1]$, we have $T_A \cap T_B \neq \emptyset$. Let $i \in T_A \cap T_B$ and let $a \in A$, $b \in B$ be such that $c(a) = c(b) = i$. But then (a, b) is a monochromatic edge of $K_{n,n}$. □

Conclusion: $\mathrm{ch}(K_{n,n}) \geq (\frac{1}{2} - o(1)) \log_2 n$.

In fact [ERT79] proved that $\mathrm{ch}(K_{n,n}) = (1 + o(1)) \log_2 n$ (they connected between choosability of complete bipartite graphs and the so called property B of hypergraphs).

Furthermore, [Alo00] proved in 2000 the following result:

Theorem 5.19 *Let G be a graph of average degree d. Then*

$$\mathrm{ch}(G) \geq (\tfrac{1}{2} - o(1)) \log_2 d\,.$$

The choice number thus grows with the graph density. This stands in striking contrast to the usual chromatic number: $\chi(K_{n,n}) = 2$ but $K_{n,n}$ is n-regular.

5.5.2 Choosability in Random Graphs

Theorem 5.20 (J. Kahn, appeared in [Alo93]) *Let $G \sim G(n, \frac{1}{2})$. Then w.h.p.*

$$\mathrm{ch}(G) = (1 + o(1))\chi(G)\,.$$

The proof is based on the following lemma:

Lemma 5.21 *Let $k \leq m \leq n$ be positive integers. Let $G = (V, E)$ be a graph on n vertices, in which every subset of m vertices spans an independent set of size k. Then*

$$\mathrm{ch}(G) \leq \frac{n}{k} + m\,.$$

(Compare to Proposition 5.11, in which the chromatic number is bounded.)

Proof Let $\mathcal{S} = \{S(v) \mid v \in V\}$ be a family of color lists satisfying $|S(V)| \geq n/k + m$ for all $v \in V$. We will prove that G is $\mathcal{S}$-choosable.

We color from lists in two stages. Start with $G_0 := G$ and $i := 0$.

1. As long as there is a color c appearing on the lists of at least m vertices of G_i: let
$$U = \{v \in V \mid c \in S(v)\} .$$
By assumption U spans an independent set I of size k. Color all vertices in I with color c. Delete c from all remaining color lists, and update $G_{i+1} := G_i - I$ and $i := i + 1$.

2. Denote by G^* the output of the first stage. Observe that the iteration of stage 1 was performed at most n/k times. Therefore the length of the color list of each remaining vertex of G^* is at least m. On the other hand each color appears on fewer than m lists of remaining vertices. Now we assign colors to vertices such that:

   (i) each vertex of G^* gets a color from its current list;

   (ii) each color gets assigned to at most one vertex.

   Define an auxiliary bipartite graph $\Gamma = (X \cup Y, F)$ as follows:
$$X = G^* , \quad Y = \bigcup_{v \in V(G^*)} S(v) .$$
Connect $y \in Y$ to $v \in X$ by an edge in F if $y \in S(v)$. We are looking for a matching in Γ saturating side X. Observe:

   (i) $d_\Gamma(v) = |S(v)| \geq m$ for all $v \in X$.

   (ii) For $y \in Y$, $d_\Gamma(y)$ is the number of appearances of y in the lists of G^*, i.e., $< m$.

   Then Hall's condition applies to X, and therefore Γ contains a matching M of size $|M| = |X| = |V(G^*)|$. Coloring the vertices of G^* according to M produces a valid coloring of G^* and thus completes a valid coloring of G. □

Proof (Proof of Theorem 5.20) Set $m = n/\log_2^2 n$ and $k = k^*(m) - 3 = (1 + o(1))2\log_2 n$. We have proved that w.h.p. in $G(n, \frac{1}{2})$, every m vertices span an independent set of size k. Then w.h.p. when $G \sim G(n, \frac{1}{2})$,
$$\mathrm{ch}(G) \leq \frac{n}{k} + m = (1 + o(1))\frac{n}{2\log_2 n} .$$

We have also observed that w.h.p. in $G(n, \frac{1}{2})$,

$$\chi(G) \geq (1 + o(1))\frac{n}{2\log_2 n}.$$

Hence w.h.p.

$$\mathrm{ch}(G) = (1 + o(1))\chi(G)\,. \qquad \square$$

[KSVW03] proved that in $G(n, p)$ with $p \geq n^{-\frac{1}{3}+\varepsilon}$ for some $\varepsilon > 0$, then w.h.p.

$$\mathrm{ch}(G) = (1 + o(1))\chi(G)\,.$$

In fact they conjecture that the bound on p is just an artifact of the proof.

Conjecture 5.22 *Consider $G \sim G(n, p)$ for $p = p(n)$. Then w.h.p.*

$$\mathrm{ch}(G) = (1 + o(1))\chi(G)$$

where the $o(1)$ term tends to 0 as $np \to \infty$.

Bibliography

[AKS81] M. Ajtai, J. Komlós, and E. Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), no. 1, 1–12.

[Alo93] N. Alon, *Restricted colorings of graphs*, Surveys in combinatorics **1993** (1993), 1–33.

[Alo00] ———, *Degrees and choice numbers*, Random Structures and Algorithms **16** (2000), no. 4, 364–368.

[AS00] N. Alon and J.H. Spencer, *The probabilistic method*, Wiley-Interscience, 2000.

[BC78] E.A. Bender and E.R. Canfield, *The asymptotic number of labeled graphs with given degree sequences*, Journal of Combinatorial Theory, Series A **24** (1978), no. 3, 296–307.

[Bol80] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin **1** (1980), no. 4, 311–316.

[Bol88] ———, *The chromatic number of random graphs*, Combinatorica **8** (1988), no. 1, 49–55.

[Bol01] ———, *Random graphs*, Cambridge Univ Press, 2001.

[dlV79] W. Fernandez de la Vega, *Long paths in random graphs*, Studia Sci. Math. Hungar **14** (1979), 335–340.

[Erd47] P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc **53** (1947), 292–294.

[ERT79] P. Erdős, A.L. Rubin, and H. Taylor, *Choosability in graphs*, Congr. Numer **26** (1979), 125–157.

[ES35] P. Erdős and G. Szekeres, *A combinatorial theorem in geometry*, Compositio Math **2** (1935), 463–470.

[GM] G.R. Grimmett and C.J.H. McDiarmid, *On colouring random graphs*, Mathematical Proceedings of the Cambridge Philosophical Society, no. 77, 313–324.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Citeseer, 2000.

[KSVW03] M. Krivelevich, B. Sudakov, V.H. Vu, and N.C. Wormald, *On the probability of independent sets in random graphs*, Random Structures and Algorithms **22** (2003), no. 1, 1–14.

[McD79] C. McDiarmid, *Determining the chromatic number of a graph*, SIAM J. Comput. **8** (1979), no. 1, 1–14.

[Pós76] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics **14** (1976), no. 4, 359–364.

[Viz76] V.G. Vizing, *Coloring the vertices of a graph in prescribed colors*, Diskret. Anal **29** (1976), 3–10.

Index

B

binomial distribution . . . 3
booster . . . 30

C

Chebyshev . . . 3
Chernoff
  on Bin . . . 3
chromatic number . . . 7, 43
clique number . . . 7
configuration . . . 19
convergence in distribution . . . 18

F

factorial-like products . . . 17

G

girth . . . 9
$G(n, m)$ . . . 5
  asymptotic eq. to $G(n, p)$ . . 14
$G(n, p)$ . . . 4
$G(n, r)$ . . . 18
graph
  minor . . . 10
  property . . . 6
greedy coloring . . . 44
  of $G(n, \frac{1}{2})$ . . . 46

H

Hadwiger number . . . 11
Hamiltonicity . . . 29
hitting time . . . 39
  of connectivity . . . 40

I

independence number . . . 7, 43

M

Markov . . . 2
monotonicity . . . 6

P

Pósa rotation . . . 30
Poisson distribution . . . 18

R

Ramsey number . . . 8
random graph process . . . 5, 39
random regular graph . . . 18

S

SIMPLE . . . 20, 23
staged exposure . . . 6
Stirling . . . 2

W

with high probability . . . 3