

## Random graphs and the probabilistic method

The material in this handout is taken from: Noga Alon and Joel H. Spencer, *The Probabilistic Method (3rd edition)*, John Wiley & Sons, 2008. Reinhard Diestel, *Graph Theory (5th edition)*, Springer, 2017.

A generative model for graphs is a probability distribution over all graphs of a certain order. One of the simplest such models is the one proposed by Paul Erdős and Alfred Rényi around 1960. The Erdős-Rényi random graph  $G = (V, E)$  of parameters  $n \in \mathbb{N}$  and  $p \in [0, 1]$  is such that the events  $(i, j) \in E$  for every  $1 \leq i < j \leq n$  are independent and have probability  $p$ . We use  $\mathcal{G}(n, p)$  to denote the resulting probability distribution over graphs of order  $n$ . If  $0 < p < 1$ , then  $\mathcal{G}(n, p)$  assigns a nonzero probability to every graph of order  $n$ . In particular, for any given  $G = (V, E)$  we have that

$$\mathbb{P}(G) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$$

where the probability is computed with respect to the distribution  $\mathcal{G}(n, p)$ . Note that  $\mathcal{G}(n, \frac{1}{2})$  is the uniform distribution of all graphs of order  $n$ . The distribution of the number of edges follows a binomial distribution of parameters  $\binom{n}{2}$  and  $p$ ,

$$\mathbb{P}(|E| = k) = \sum_{E' \subseteq [V]^2: |E'|=k} \mathbb{P}(E') = \binom{\binom{n}{2}}{k} p^k (1-p)^{\binom{n}{2}-k}$$

We write  $G_n \sim \mathcal{G}(n, p)$  to denote the random graph distributed according to  $\mathcal{G}(n, p)$ .

We use the Erdős-Rényi model to prove properties about graphs via the so-called probabilistic method. In order to prove that there exist graphs with certain properties, we show that the Erdős-Rényi model generates the desired graphs with probability strictly bigger than zero.

In the following, we repeatedly use two simple facts from probability theory. The first one is known as the union bound: for any finite collection  $E_1, \dots, E_N$  of events,

$$\mathbb{P}(E_1 \cup \dots \cup E_N) \leq \sum_{i=1}^N \mathbb{P}(E_i)$$

The second one is Markov's inequality: for all nonnegative random variables  $X$  and all  $a > 0$ ,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

We start with a simple application of the probabilistic method to show that there exist graphs of any order  $n \geq 4$  that have neither a large clique nor a large independent set, where large means at least  $2 \log_2 n$ .

**Fact 1** For all  $n \geq 4$  and all  $k \geq 2 \log_2 n$ , there exists a graph  $G$  of order  $n$  such that  $\alpha(G) < k$  and  $\omega(G) < k$ .

PROOF. Let  $G \sim \mathcal{G}(n, \frac{1}{2})$ . The probability that an arbitrary  $U \subset V$  of size  $k$  is an independent set is  $2^{-\binom{k}{2}}$ . Therefore,

$$\begin{aligned}
\mathbb{P}(\alpha(G) \geq k) &= \mathbb{P}(\exists U \subset V : |U| = k, U \text{ is independent in } G) \\
&\leq \sum_{U \subset V : |U|=k} 2^{-\binom{k}{2}} && \text{(using the union bound)} \\
&= \binom{n}{k} 2^{-\binom{k}{2}} \\
&\leq \left(\frac{n}{2}\right)^k 2^{-\frac{k(k-1)}{2}} && \text{(using } \binom{n}{k} \leq \left(\frac{n}{2}\right)^k \text{ for } 4 \leq k \leq n) \\
&= 2^{k(\log_2 n - 1) - \frac{k(k-1)}{2}} \\
&\leq 2^{\frac{k^2}{2} - k - \frac{k(k-1)}{2}} && \text{(using } k \geq 2 \log_2 n) \\
&\leq 2^{-\frac{k}{2}} < \frac{1}{2} && \text{(because } k \geq 4)
\end{aligned}$$

Likewise, the probability that an arbitrary  $U \subset V$  of size  $k$  is a clique is also  $2^{-\binom{k}{2}}$ . Therefore, with an identical proof, we can prove that  $\mathbb{P}(\omega(G) \geq k) < \frac{1}{2}$ . This implies

$$\mathbb{P}(\alpha(G) < k \text{ and } \omega(G) < k) = 1 - \mathbb{P}(\alpha(G) \geq k \text{ or } \omega(G) \geq k) \geq 1 - \mathbb{P}(\alpha(G) \geq k) - \mathbb{P}(\omega(G) \geq k) > 0$$

where we used the union bound. Since  $\mathbb{P}(\alpha(G) < k \text{ and } \omega(G) < k) > 0$ , we conclude there exists  $G$  of order  $n$  such that  $\alpha(G) < k$  and  $\omega(G) < k$ .  $\square$

We now use the probabilistic method to prove a harder result. Namely, that we can find graphs that lack short cycles and simultaneously have a large chromatic number. Recall that a high chromatic number requires a small independence number, which in turn is favored by a high density. On the other hand, a high density also favors the presence of short cycles. As we see next, the trick to obtain the right density is to choose  $p$  slightly larger than  $\frac{1}{n}$ .

**Theorem 2 (Erdős, 1959)** *For every integer  $k \geq 3$  there exists a graph  $G$  such that  $g(G) > k$  and  $\chi(G) > k$ .*

We start by proving an auxiliary lemma on the expected number of cycles of a given length.

**Lemma 3** *The expected number of cycles of length  $k$  in  $G \sim \mathcal{G}(n, p)$  is*

$$\frac{n(n-1) \cdots (n-k+1)}{2k} p^k$$

PROOF. Let  $\mathcal{C}_k$  be the set of all cycles of length  $k$  on  $n$  vertices. Since each cycle is specified by a sequence of  $k$  distinct vertices, and there are  $n(n-1) \cdots (n-k+1)$  ways of choosing this sequence,  $|\mathcal{C}_k| = n(n-1) \cdots (n-k+1)/(2k)$ , where we divide by  $2k$  because there are exactly  $2k$  sequences that correspond to the same cycle. Since a cycle is also specified by a sequence of  $k$  edges,  $\mathbb{P}(G \text{ contains } C) = p^k$  for any given  $C \in \mathcal{C}_k$ . Finally, let  $N_k(G)$  be the number of cycles of length  $k$  in  $G$ .

$$\mathbb{E}[N_k(G)] = \sum_{C \in \mathcal{C}_k} \mathbb{P}(G \text{ contains } C) = \sum_{C \in \mathcal{C}_k} p^k = |\mathcal{C}_k| p^k$$

concluding the proof.  $\square$

We are now ready to prove Erdős theorem.

PROOF OF THEOREM 2. Fix  $\varepsilon > 0$  with  $0 < \varepsilon < \frac{1}{k}$ , and let  $p = n^{\varepsilon-1}$ . Let  $N_{\leq k}(G)$  be the number of cycles of length at most  $k$  in  $G$ . By Lemma 3 we have

$$E[N_{\leq k}(G)] = \sum_{i=3}^k \frac{n(n-1)\cdots(n-i+1)}{2^i} p^i \leq \frac{1}{2} \sum_{i=3}^k (np)^i \leq \frac{k-2}{2} (np)^k \quad (1)$$

where we used  $(np)^i \leq (np)^k$  because  $np = n^\varepsilon \geq 1$ . Now,

$$\begin{aligned} \mathbb{P}(N_{\leq k}(G) \geq n/2) &\leq \frac{E[N_{\leq k}(G)]}{n/2} && \text{(by Markov's inequality)} \\ &\leq (k-2)n^{k-1}p^k && \text{(by (1))} \\ &= (k-2)n^{k-1}n^{(\varepsilon-1)k} \\ &= (k-2)n^{k\varepsilon-1} \\ &< \frac{1}{2} && \text{(for } n \text{ large since } k\varepsilon - 1 < 0 \text{ due to } \varepsilon < 1/k) \end{aligned}$$

Now, using the argument in the proof of Fact 1,

$$\begin{aligned} \mathbb{P}(\alpha(G) \geq n/(2k)) &\leq \binom{n}{n/(2k)} (1-p)^{\binom{n/(2k)}{2}} \\ &= \binom{n}{r} (1-p)^{\binom{r}{2}} && \text{(letting } r = n/(2k)) \\ &\leq 2^n e^{-p\binom{r}{2}} && \text{(using } \binom{n}{r} < 2^n \text{ and } 1-p \leq e^{-p}) \\ &\leq 2^n e^{-pr^2/4} && \text{(using } \binom{r}{2} \geq \frac{r^2}{4} \text{ for } r \geq 2) \\ &= 2^n e^{-pn^2/(16k^2)} \\ &\leq 2^n e^{-n} && \text{(as } pn = n^\varepsilon \geq 16k^2 \text{ for } n \text{ large enough)} \\ &< \frac{1}{2} && \text{(for } n \text{ large enough)} \end{aligned}$$

Combining these two results, we conclude that for  $n$  large enough with respect to  $k$ ,

$$\mathbb{P}(N_{\leq k}(G) < n/2, \alpha(G) < n/(2k)) > 0$$

Then, there exists a graph  $G$  of order  $n$  with fewer than  $n/2$  cycles of length at most  $k$  and  $\alpha(G) < n/(2k)$ . From each of those cycles delete a vertex and let  $H = (V_H, E_H)$  be the graph obtained. Then  $|V_H| \geq n/2$  and  $g(H) > k$ . By definition of  $G$ , and using the fact that  $\chi(H)\alpha(H) \geq |V_H|$  which holds for any graph,

$$\chi(H) \geq \frac{|V_H|}{\alpha(H)} \geq \frac{n/2}{\alpha(G)} > k$$

where we used the fact that  $\alpha(H) \leq \alpha(G)$  since  $H$  is obtained by deleting vertices from  $G$  (see Exercise 1). This concludes the proof.  $\square$

**Properties that asymptotically holds for almost all graphs.** A predicate over the set of all graphs of order  $n$  is a  $\{0, 1\}$ -valued function  $\mathcal{P}_n$ . For instance,  $\mathcal{P}_n(G_n) = \mathbb{I}\{\omega(G_n) \leq f(n)\}$  for some function  $f : \mathbb{N} \rightarrow \mathbb{N}$ . Let  $\mathcal{P}_1, \mathcal{P}_2, \dots$  be predicates. Then we say that  $\mathcal{P}_n$  *asymptotically holds for almost all  $G_n$*  if

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_n(G_n) = 1) = 1$$

where the probability is with respect to  $G_n \sim \mathcal{G}(n, \frac{1}{2})$ .

In Fact 1 we proved that if  $k \geq 2 \log_2 n$ , then it becomes easy to find graphs  $G$  of order  $n$  such that  $\alpha(G) < k$  and  $\omega(G) < k$ . It turns out that both predicates  $\omega(G_n) = 2 \log_2 n$  and  $\alpha(G_n) = 2 \log_2 n$  asymptotically hold for almost every  $G_n \sim \mathcal{G}(n, \frac{1}{2})$ . In other words, almost all graphs  $G$  of large order  $n$  have  $\alpha(G) = \omega(G) = 2 \log_2 n$ .

We only prove the lower bound. Let  $k = 2 \log_2 n$ . From the proof of Fact 1,

$$\mathbb{P}(\alpha(G) \geq k) \leq 2^{-\frac{k}{2}} = \frac{1}{n} \rightarrow 0 \quad (\text{for } n \rightarrow \infty)$$

and the exactly same result also holds for the clique number. The upper bound is an immediate consequence of the following result, whose proof is omitted,

$$\mathbb{P}(\omega(G) < 2 \log_2 n) \leq 2^{-n^{2+o(1)}} \quad (2)$$

We now study  $\chi(G_n)$  for  $G_n \sim \mathcal{G}(n, \frac{1}{2})$ . Since  $\chi(G) \geq n/\alpha(G)$ , we already know that  $\chi(G_n) \geq \frac{n}{2 \log_2 n}$  asymptotically holds for almost all  $G_n \sim \mathcal{G}(n, \frac{1}{2})$ .

**Theorem 4**  $\chi(G_n) = \frac{n}{2 \log_2 n}$  asymptotically holds for  $G_n \sim \mathcal{G}(n, \frac{1}{2})$ .

**PROOF.** For  $G = (V, E)$  recall that  $G[S]$  is the graph induced by  $S \subseteq V$ . For any fixed such  $S$  of size  $m$ , if  $G \sim \mathcal{G}(n, \frac{1}{2})$ , then  $G[S] \sim \mathcal{G}(m, \frac{1}{2})$ . In order to prove  $\chi(G_n) \leq \frac{n}{2 \log_2 n}$ , we show that for  $n \rightarrow \infty$  any  $S \subseteq V$  of size  $m = n/(\log_2 n)^2$  contains an independent set of size at least  $2 \log_2 m$ .

$$\begin{aligned} & \mathbb{P}(\exists S \subseteq V : |S| = m, \alpha(G[S]) < 2 \log_2 m) \\ & \leq \binom{n}{m} 2^{-m^{2+o(1)}} && (\text{using (2) for each } G[S]) \\ & < 2^{n-m^{2+o(1)}} && (\text{because } \binom{n}{m} < 2^n) \\ & = 2^{n - \left(\frac{n}{(\log_2 n)^2}\right)^{2+o(1)}} && (\text{for } m = n/(\log_2 n)^2) \\ & = 2^{\frac{n}{(\log_2 n)^{4+o(1)}} \left( (\log_2 n)^{4+o(1)} - n^{1+o(1)} \right)} \\ & \rightarrow 0 && (\text{for } n \rightarrow \infty) \end{aligned}$$

This implies that for  $n$  large, in almost all graphs  $G \sim \mathcal{G}(n, \frac{1}{2})$  we can find  $n/m$  independent sets of size at least  $2 \log_2 m$ . In order to upper bound the chromatic number of such a graph  $G$ , we can assign the same color to all the vertices in the same independent set. Since  $m$  might not divide  $n$ , we are left with less than  $m$  vertices, which we can color with less than  $m$  additional colors. Therefore,

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

concluding the proof. □

**Acknowledgements.** Thanks to Andrea Rovati for flagging mistakes and typos in earlier versions of this handout.

### Exercises

1. For any  $G = (V, E)$  and for any  $S \subseteq V$ , let  $H = G[S]$ . Prove that  $\alpha(H) \leq \alpha(G)$ .