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 $G_n = (V_n, E_n)$ of parameters $n \in \mathbb{N}$ and $p \in [0, 1]$ is such that the events $(i, j) \in E_n$ for every $1 \le i < j \le n$ 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 $G = (V, E)$ of order n 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 $\mathcal{G}\left(n, \frac{1}{2}\right)$ is the uniform distribution over all graphs of order n. The distribution of the number of edges of G_n follows a binomial distribution of parameters $\binom{n}{2}$ and p,

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

We write $G_n \sim \mathcal{G}(n, p)$ to denote the random graph $G_n = (V_n, E_n)$ distributed according to $\mathcal{G}(n, p)$. Note that $\mathbb{E}\big[|E_n|\big] = p\binom{n}{k} = p\frac{n(n-1)}{2}$ which is $\Theta(n^2)$ when $p = \Theta(1)$.

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,

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

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

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[X]}{a}$$

We start with a simple application of the probabilistic method to show that there exist graphs of any order $n \ge 4$ 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 $k \geq \lceil 2 \log_2 n \rceil$, there exists a graph G of order n such that $\alpha(G) < k$ and $\omega(G) < k$.*

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

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

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 $\mathbb{P}\big(\omega(G) \geq k\big) \leq \frac{1}{4}$. This implies

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

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

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. $\qquad\square$

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

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

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,

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

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. $\square$

Asymptotic properties. Given $G_n \sim \mathcal{G}\big(n, \frac{1}{2}\big)$, a function $\mu : G_n \mapsto \mu(G_n)$, and a function $f : \mathbb{N} \to \mathbb{R}$, we say that $\mu(G_n) \xrightarrow{p} f(n)$ if

$$\lim_{n\to\infty} \mathbb{P}\Big(\mu(G_n) = \big(1 + o(1)\big)f(n)\Big) = 1 \qquad \text{where } o(1) \to 0 \text{ for } n \to \infty$$

To understand this definition, recall that $g = \big(1 + o(1)\big)f$ means that for all $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n(\varepsilon)$ we have $(1 - \varepsilon)f(n) \leq g(n) \leq (1 + \varepsilon)f(n)$. Therefore, we can define $\varepsilon(n) = \min\{\varepsilon > 0 : (\forall n' \geq n)\,(1 - \varepsilon)f(n') \leq g(n') \leq (1 + \varepsilon)f(n')\}$ such that $\lim_{n\to\infty} \varepsilon(n) = 0$. So $g(n) \xrightarrow{p} f(n)$ means that the probability of $(\exists n' \geq n)\,\big|f(n') - g(n')\big| > \varepsilon(n)$ goes to zero as $n \to \infty$.

In Fact 1 we proved that if $k \geq \lceil 2\log_2 n \rceil$, then one can find graphs G_n such that $\alpha(G_n) < k$ and $\omega(G_n) < k$. It turns out that $\omega(G_n) \xrightarrow{p} 2\log_2 n$ and $\alpha(G_n) \xrightarrow{p} 2\log_2 n$. More precisely, there exists a sequence $k_1, k_2, \ldots$ such that

$$k_n = \big(1 + o(1)\big)2\log_2 n \qquad \text{and} \qquad \lim_{n\to\infty} \mathbb{P}\big(\omega(G_n) \in \{k_n, k_n + 1\}\big) = 1$$

and the exactly same result also holds for the independence number. To get some intuition on why this is true, note that the expected number of cliques of size k in G_n is $f(n, k) = \binom{n}{k}2^{-\binom{k}{2}}$. Using the standard bounds

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

we obtain $f(n, k) = 2^{k\left(\log_2 n - \frac{k}{2} - \log_2 k + \Theta(1)\right)}$. Hence, it is easy to see that

$$\lim_{n,k\to\infty} f(n, k) = \begin{cases} 0 & \text{if } k > 2\log_2 n \\ \infty & \text{if } k < 2\log_2 n \end{cases}$$

In other words, the expected number of cliques of size k goes to zero for $k > 2\log_2 n$ and to infinity for $k < 2\log_2 n$

We now study $\chi(G_n)$. We already know that $\chi(G_n) \geq n/\alpha(G_n)$ for every G_n. Let

$$k_n = \max\{k \in \mathbb{N} : f(n, k) \geq 1\} \tag{2}$$

Note that $k_n = \big(1 + o(1)\big)2\log_2 n$. To prove $\chi(G_n) \leq \big(1 + o(1)\big)\frac{n}{\alpha(G_n)}$ with probability one for $n \to \infty$, we use the following lemma (without proof).

Lemma 4 $\mathbb{P}\big(\omega(G_n) < k_n - 4\big) < 2^{-n^{2+o(1)}}$.

Clearly, the same result holds with $\alpha(G_n)$ in place of $\omega(G_n)$.

Theorem 5 $\chi(G_n) \xrightarrow{p} \frac{n}{2\log_2 n}$.

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 $G[S] \sim \mathcal{G}(m, \frac{1}{2})$. We now show that for $n \to \infty$ any $S \subseteq V$ of size $m = n/(\log_2 n)^2$ contains an independent set of size at least k_m defined in (2).

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

This implies that for n large, almost all graphs are such that any subset of m vertices contains an independent set of size at least k_m. For any such graph, we can repeat the following procedure until less than m vertices remain: find an independent set of size k_m, assign the same color to its vertices, and remove it from the graph. Once the procedure is over, we are left with less than m vertices, which we can color with less than m additional colors. Therefore,

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

because $m = n/(\log_2 n)^2$ and $k_m = \big(1 + o(1)\big) 2 \log_2 n - 2 \log_2 \log_2 n = \big(1 + o(1)\big) 2 \log_2 n$. $\qquad \square$

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 $\alpha(H) \leq \alpha(G)$.