Graph Theory

Random graphs and the probabilistic method

Instructor: *Nicolò Cesa-Bianchi* version of June 17, 2020

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 $\mathcal{G}\big(n, \frac{1}{2}\big)$ is the uniform distribution of all graphs of order n. The distribution of the number of edges follows a binomial distribution of parameters n and p,

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

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.

We start with a simple application of the probabilistic method to show that there exist graphs that have neither a large clique nor a large independent set. We use the following bound on the binomial coefficients,

$$\binom{n}{k} \leq \left(\frac{n}{2}\right)^k \qquad 4 \leq k \leq n \tag{1}$$

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 $G \sim \mathcal{G}\big(n, \frac{1}{2}\big)$. The probability that an arbitrary $U \subset V$ of size k is an independent

set is $2^{-\binom{k}{2}}$. Therefore,

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

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

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

Since $\mathbb{P}\big(\alpha(G) < k,\, \omega(G) < k\big) > 0$, we conclude there exists G of order n such that $\alpha(G) < k$ and $\omega(G) < k$. □

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 large value of p in $\mathcal{G}(n,p)$. On the other hand, p large also favors the presence of short cycles. As we see next, the trick is to choose p slightly larger than $\frac{1}{n}$.

Theorem 2 (Erdős, 1959) *For every integer k 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, $\mathbb{P}\big(G \text{ contains } C\big) = p^k$ for any given $C \in \mathcal{C}_k$. Finally, let $N_k(G)$ be the number of cycles of length k in G.

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

concluding the proof. □

We are now ready to prove Erdős theorem.

Proof of Theorem 2. Assume $k \geq 3$, 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

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

where we used $(np)^i \leq (np)^k$ because $np = n^{\varepsilon} \geq 1$. Using Markov's inequality $\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a$ for all random variables $X \geq 0$ and all $a > 0$,

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

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

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

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

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

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 be the graph obtained. Then $|H| \geq n/2$ 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,

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

where we used the fact that $\alpha(H) \leq \alpha(G)$ since H is obtained by deleting vertices from G (prove this fact). This concludes the proof. □

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, $\mathcal{P}_n(G_n) = \mathbb{I}\{\omega(G_n) \le f(n)\}$ for some function $f : \mathbb{N} \to \mathbb{N}$. Let $\mathcal{P}_1, \mathcal{P}_2, \ldots$ be predicates and $p_1, p_2, \ldots \in (0,1)$ be a sequence of edge probabilities. Then we say that $\mathcal{P}_n$ *asymptotically holds for almost all* G_n if

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

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

An isomorphism $\varphi : V \to V'$ between $G = (V, E)$ and $G = (V', E')$ is a bijection that respects neighborhoods: $(i, j) \in E$ if and only if $\big(\varphi(i), \varphi(j)\big) \in E'$ for all $i, j \in V$.

Given $i, j \in \mathbb{N}$, let $\mathcal{P}_{i,j}(G)$ denote the predicate that G contains, for any disjoint vertex sets U, W with $|U| \le i$ and $|W| \le j$, a vertex $v \notin U \cup W$ that is adjacent to all the vertices in U but to none in W.

Fact 4 *For every graph $H = (V_H, E_H)$, if $G = (V_G, E_G)$ satisfies $\mathcal{P}_{i,j}$ for $i = \Delta(H)$ and $j = |V_H| - 1$, then G contains an induced copy of H.*

The proof is based on a simple gadget to sequentially build an isomorphism φ between H and some subgraph of G: consider an enumeration $1, \ldots, k$ of the vertices in V_H. For each $i = 1, \ldots, k$ map $i \in V_H$ to a vertex $\varphi(i) \in V_G \setminus \varphi\big(\{1, \ldots, i-1\}\big)$ such that any previously mapped vertex $j < i$ is adjacent to i if and only if $\varphi(j)$ is adjacent to $\varphi(i)$. It the process ends after all vertices of H are mapped, it is easy to verify that the resulting φ is an isomorphism between V_H and $\varphi(V_H) \subseteq V_G$.
Proof. Since G satisfies $\mathcal{P}_{i,j}$, each time we need to map $i \in V_H$ we can find $\varphi(i) \in V_G$ by choosing $U, W \subseteq V_G$ such that $U \equiv \{\varphi(j) : j \in N(i),\, j = 1, \ldots, i-1\}$ and $W \equiv \big\{\varphi(1), \ldots, \varphi(i-1)\big\} \setminus U$. Note that $|U| \le \Delta(H)$ and $|W| \le |V_H| - 1$. □

The next result shows that each property $\mathcal{P}_{i,j}$ asymptotically holds for almost all graphs.

Lemma 5 *For every constant $p \in (0,1)$ and $i, j \in \mathbb{N}$, $\mathcal{P}_{i,j}$ asymptotically holds for almost all $G_n \sim \mathcal{G}(n, p)$.*

Proof. Let $G_n = (V, E)$. For fixed $U, W \in V$ and $v \in V \setminus (U \cup W)$, the probability that v is adjacent to all the vertices in U but to none in W is $p^{|U|}(1-p)^{|W|}$. Therefore, the probability that no suitable v exists for these U and W, is

$$\big(1 - p^{|U|}(1-p)^{|W|}\big)^{n-|U|-|W|} \le \big(1 - p^i(1-p)^j\big)^{n-i-j}$$

(for $n \ge i + j$), since the corresponding events are independent for different v.

$$\begin{aligned}
1 - \mathbb{P}\big(\mathcal{P}_{i,j}\big) &= \mathbb{P}\big(\exists\, U, W \subset V \,:\, U \cap W = \emptyset,\, |U| \le i,\, |W| \le j, \text{ no suitable } v \text{ exists for } U, W\big) \\
&\le \sum_{\substack{U,W \subset V\,:\, U\cap W = \emptyset \\ |U| \le i,\, |W| \le j}} \mathbb{P}\big(\text{no suitable } v \text{ exists for } U, W\big) \\
&\le n^{i+j}\big(1 - p^i(1-p)^j\big)^{n-i-j} \to 0 \qquad \text{(for } n \to \infty\text{)}
\end{aligned}$$

because $1 - p^i(1-p)^j < 1$. □

Corollary 6 *For every constant $p \in (0,1)$ and $k \in \mathbb{N}$, almost all graphs $G_n \sim \mathcal{G}(n,p)$ are asymptotically k-connected.*

PROOF. By Lemma 5, it is enough to show that every graph satisfying $\mathcal{P}_{2,k-1}$ is k-connected. But this is easy: any graph satisfying $\mathcal{P}_{2,k-1}$ has order at least $2+k-1 = k+2$, and if W is a set of fewer than k vertices, then by definition of $\mathcal{P}_{2,k-1}$ any other two vertices x, y have a common neighbour $v \notin W$. Thus W does not separate x from y. □

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 almost all graphs G have $\alpha(G) = \omega(G) = 2\log_2 n$ when $n \to \infty$.

Let $k = 2\log_2 n$. From the proof of Fact 1,

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

and the exactly same result also holds for the clique number. One can also show (proof omitted) that

$$\mathbb{P}\big(\omega(G) < 2\log_2 n\big) \leq 2^{-n^{2+o(1)}} \tag{3}$$

This implies the desired result.

Corollary 7 *Both predicates $\omega(G_n) = 2\log_2 n$ and $\alpha(G_n) = 2\log_2 n$ asymptotically hold for almost every $G_n \sim \mathcal{G}(n, \frac{1}{2})$.*

We now study $\chi(G_n)$ for $G_n \sim \mathcal{G}(n, \frac{1}{2})$. Since $\chi(G) \geq n/\alpha(G)$, we 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 8 *$\chi(G_n) = \frac{n}{2\log_2 n}$ asymptotically holds for $G_n \sim \mathcal{G}(n, \frac{1}{2})$.*

PROOF. For $G = (V, E)$ let $G_{|S}$ be 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 $n \to \infty$ any $S \subseteq V$ of size $m = n/(\log_2 n)^2$ contains an independent set of size at least $2\log_2 m$.

$$\begin{aligned}
\mathbb{P}\big(\exists S \subseteq V \,:\, |S| = m,\, \alpha(G_{|S}) < 2\log_2 m\big) &\leq \binom{n}{m} 2^{-m^{2+o(1)}} && \text{(using (3) for each } G_{|S}\text{)} \\
&< 2^{n - m^{2+o(1)}} && \text{(because } \tbinom{n}{m} < 2^n\text{)} \\
&\leq 2^{n - \left(\frac{n}{(\log n)^2}\right)^{2+o(1)}} && \text{(for } m = n/(\log_2 n)^2\text{)} \\
&\to 0 && \text{(for } n \to \infty\text{)}
\end{aligned}$$

Now note that $\log_2 m = \log_2 n - 2\log_2\log_2 n = (1+o(1))\log_2 n$. Hence, as long as we can find a subset of m vertices, we can take out an independent set of size $k = (1+o(1))\log_2 n$ and assign it the same color. At the end, we are left with less than m vertices, which we can color with less than m additional colors. Therefore,

$$\chi(G_n) \leq \left\lceil \frac{n-m}{k} \right\rceil + m \leq \frac{n}{k} + m = (1+o(1))\frac{n}{2\log_2 n} + \frac{n}{(\log_2 n)^2} = (1+o(1))\frac{n}{2\log_2 n}$$

concluding the proof. □

The Rado graph. The Rado graph R is an infinite graph constructed inductively as follows. Let $R_0 = K_1$. For all $n \in \mathbb{N}$, let R_{n+1} be obtained from $R_n = (V_n, E_n)$ by adding for every subset $U \subseteq V_n$ (including the empty set and V_n) a new vertex v such that $N(v) \equiv U$ (in particular, the new vertices form an independent set in R_{n+1}). Then $R = \bigcup_{n \geq 1} R_n$. Note that $V_{n+1} = |V_n| + 2^{|V_n|}$.

R has the property $\mathcal{P}^*$. That is, R has the property $\mathcal{P}_{i,j}$ for all $i, j \in \mathbb{N}$. Indeed, for all finite and disjoint $U, W \subseteq \mathbb{N}$ with, say, $|U| = i$ and $|W| = j$, let n be the smallest integer such that $U, W \subseteq V_n$. The there exists $v \in V_{n+1}$ whose neighborhood is exactly U.

Because R has the property $\mathcal{P}^*$, Fact 4 implies that R contains an induced copy of any countable graph G. To see this, note that our gadget to sequentially construct an isomorphism works even when the graph to copy has a countable number of vertices. Indeed, every time we need to map $i \in \mathbb{N}$, the required sets U and W are finite. We now prove that the Rado graph is unique.

Theorem 9 (Unicity of the Rado graph) *The Rado graph is the unique graph (up to isomorphisms) having the property $\mathcal{P}^*$.*

Proof. Let $R = (V, E)$ and $R' = (V', E')$ be two graphs satisfying the property $\mathcal{P}^*$, each one with a fixed vertex enumeration. We construct a bijection $\varphi : V \to V'$ in an infinite sequence of steps, defining $\varphi(v)$ for a new vertex $v \in V$ at each step. At every odd step we look at the first vertex v in the enumeration of V for which $\varphi(v)$ has not yet been defined. Let U be the set of those among its neighbours $u \in V$ for which $\varphi(u)$ has already been defined. This is a finite set. Using the property $\mathcal{P}^*$ for R', find a vertex $v' \in V'$ outside the image of φ (which is a finite set), so that v' is adjacent in R' to all the vertices in $\varphi(U)$ but to no other vertex in the image of φ. Put $\varphi(v) = v'$. At even steps in the process we do the same thing with the roles of R and R' interchanged: we look at the first vertex v' in the enumeration of V' that does not yet lie in the image of φ, and set $\varphi(v) = v'$ for a new vertex v that matches the adjacencies and non-adjacencies of v' among the vertices for which φ (respectively, φ^{-1}) has already been defined. By our minimum choices of v and v', the bijection gets defined on all of V and all of V', and it is an isomorphism by construction. □

Finally, we show that the Rado graph is isomorphic to the infinite Erdős-Rényi graph. Recall that $\aleph_0$ is the cardinality of $\mathbb{N}$.

Theorem 10 *For every constant $p \in (0, 1)$, $G \sim \mathcal{G}(\aleph_0, p)$ is isomorphic to the Rado graph with probability* 1.

Proof. Given fixed disjoint finite sets $U, W \subseteq \mathbb{N}$, the probability that a vertex $v \notin U \cup W$ does not satisfy $\mathcal{P}^*$ for U and W (i.e., it is not joined to all of U or is joined to some vertex in W) is $r = 1 - p^{|U|}(1-p)^{|W|} < 1$ depending only on $|U|$ and $|W|$. The probability that none of k given vertices v satisfies $\mathcal{P}^*$ for U and W is r^k, which tends to 0 as $k \to \infty$. Hence the probability that all the (infinitely many) vertices outside $U \cup W$ fail to witness $\mathcal{P}^*$ for these sets U and W is 0. Now there are only countably many choices for U and W as above. So the probability that $\mathcal{P}^*$ fails for any sets U and W is still 0. Therefore G satisfies $\mathcal{P}^*$ with probability 1. By Theorem 9, this means that G is isomorphic to R. □