Probability makes counting (sometimes) easy

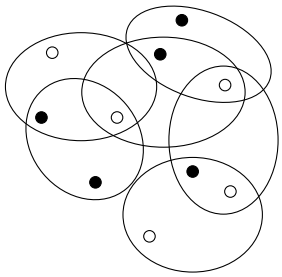
Chapter 40

Just as we started this book with the first papers of Paul Erdős in number theory, we close it by discussing what will possibly be considered his most lasting legacy — the introduction, together with Alfred Rényi, of the *probabilistic method*. Stated in the simplest way it says:

> *If, in a given set of objects, the probability that an object does not have a certain property is less than* 1*, then there must exist an object with this property.*

Thus we have an *existence* result. It may be (and often is) very difficult to find this object, but we know that it exists. We present here three examples (of increasing sophistication) of this probabilistic method due to Erdős, and end with a particularly elegant recent application.

As a warm-up, consider a family $\mathcal{F}$ of subsets A_i, all of size $d \geq 2$, of a finite ground-set X. We say that $\mathcal{F}$ is 2-*colorable* if there exists a coloring of X with two colors such that in every set A_i both colors appear. It is immediate that not every family can be colored in this way. As an example, take *all* subsets of size d of a $(2d-1)$-set X. Then no matter how we 2-color X, there must be d elements which are colored alike. On the other hand, it is equally clear that every subfamily of a 2-colorable family of d-sets is itself 2-colorable. Hence we are interested in the *smallest* number $m = m(d)$ for which a family with m sets exists which is not 2-colorable. Phrased differently, $m(d)$ is the largest number which guarantees that *every* family with less than $m(d)$ sets is 2-colorable.



A 2-colored family of 3-sets

Theorem 1. *Every family of at most* 2^{d-1} d*-sets is* 2*-colorable, that is,* $m(d) > 2^{d-1}$.

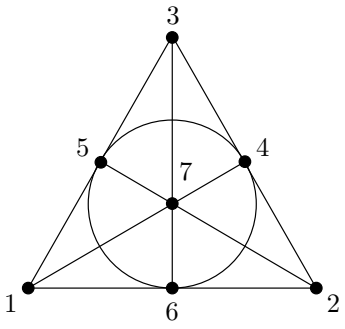
■ **Proof.** Suppose $\mathcal{F}$ is a family of d-sets with at most 2^{d-1} sets. Color X randomly with two colors, all colorings being equally likely. For each set $A \in \mathcal{F}$ let E_A be the event that all elements of A are colored alike. Since there are precisely two such colorings, we have

$$\text{Prob}(E_A) \;=\; (\tfrac{1}{2})^{d-1},$$

and hence with $m = |\mathcal{F}| \leq 2^{d-1}$ (note that the events E_A are not disjoint)

$$\text{Prob}(\bigcup_{A\in\mathcal{F}} E_A) \;<\; \sum_{A\in\mathcal{F}} \text{Prob}(E_A) \;=\; m\,(\tfrac{1}{2})^{d-1} \;\leq\; 1.$$

We conclude that there exists some 2-coloring of X without a unicolored d-set from $\mathcal{F}$, and this is just our condition of 2-colorability. □

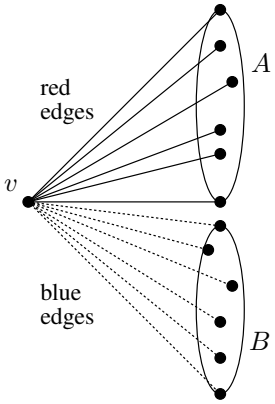


An upper bound for $m(d)$, roughly equal to $d^2 2^d$, was also established by Erdős, again using the probabilistic method, this time taking random sets and a fixed coloring. As for exact values, only the first two $m(2) = 3$, $m(3) = 7$ are known. Of course, $m(2) = 3$ is realized by the graph K_3, while the Fano configuration yields $m(3) \leq 7$. Here $\mathcal{F}$ consists of the seven 3-sets of the figure (including the circle set $\{4, 5, 6\}$). The reader may find it fun to show that $\mathcal{F}$ needs 3 colors. To prove that all families of six 3-sets are 2-colorable, and hence $m(3) = 7$, requires a little more care.

Our next example is the classic in the field — Ramsey numbers. Consider the complete graph K_N on N vertices. We say that K_N has property (m, n) if, no matter how we color the edges of K_N red and blue, there is always a complete subgraph on m vertices with all edges colored red or a complete subgraph on n vertices with all edges colored blue. It is clear that if K_N has property (m, n), then so does every K_s with $s \geq N$. So, as in the first example, we ask for the *smallest* number N (if it exists) with this property — and this is the ***Ramsey number*** $R(m, n)$.

As a start, we certainly have $R(m, 2) = m$ because either all of the edges of K_m are red or there is a blue edge, resulting in a blue K_2. By symmetry, we have $R(2, n) = n$. Now, suppose $R(m-1, n)$ and $R(m, n-1)$ exist. We then prove that $R(m, n)$ exists and that

$$R(m, n) \;\leq\; R(m-1, n) \;+\; R(m, n-1). \tag{1}$$



Suppose $N = R(m-1, n) + R(m, n-1)$, and consider an arbitrary red-blue coloring of K_N. For a vertex v, let A be the set of vertices joined to v by a red edge, and B the vertices joined by a blue edge.

Since $|A| + |B| = N - 1$, we find that either $|A| \geq R(m-1, n)$ or $|B| \geq R(m, n-1)$. Suppose $|A| \geq R(m-1, n)$, the other case being analogous. Then by the definition of $R(m-1, n)$, there either exists in A a subset A_R of size $m-1$ all of whose edges are colored red which together with v yields a red K_m, or there is a subset A_B of size n with all edges colored blue. We infer that K_N satisfies the (m, n)-property and Claim (1) follows.

Combining (1) with the starting values $R(m, 2) = m$ and $R(2, n) = n$, we obtain from the familiar recursion for binomial coefficients

$$R(m, n) \;\leq\; \binom{m+n-2}{m-1}, \tag{2}$$

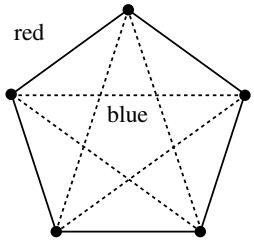
and, in particular

$$R(k, k) \;\leq\; \binom{2k-2}{k-1} \;=\; \binom{2k-3}{k-1} + \binom{2k-3}{k-2} \;\leq\; 2^{2k-3}.$$

Now what we are really interested in is a lower bound for $R(k, k)$. This amounts to proving for an as-large-as-possible $N < R(k, k)$ that there *exists* a coloring of the edges such that no red or blue K_k results. And this is where the probabilistic method comes into play.

Theorem 2. *For all $k \geq 2$, the following lower bound holds for the Ramsey numbers:*

$$R(k,k) \;\geq\; 2^{\frac{k}{2}}.$$



■ **Proof.** We have $R(2,2) = 2$. From (2) we know $R(3,3) \leq 6$, and the pentagon colored as in the figure shows $R(3,3) = 6$.

Now let us assume $k \geq 4$. Suppose $N < 2^{\frac{k}{2}}$, and consider all red-blue colorings, where we color each edge independently red or blue with probability $\frac{1}{2}$. Thus all colorings are equally likely with probability $2^{-\binom{N}{2}}$. Let A be a set of vertices of size k. The probability of the event $A_{\scriptscriptstyle R}$ that the edges in A are all colored red is then $2^{-\binom{k}{2}}$. Hence it follows that the probability $p_{\scriptscriptstyle R}$ for *some* k-set to be colored all red is bounded by

$$p_{\scriptscriptstyle R} \;=\; \text{Prob}\Big(\bigcup_{|A|=k} A_{\scriptscriptstyle R}\Big) \;\leq\; \sum_{|A|=k} \text{Prob}(A_{\scriptscriptstyle R}) \;=\; \binom{N}{k} 2^{-\binom{k}{2}}.$$

Now with $N < 2^{\frac{k}{2}}$ and $k \geq 4$, using $\binom{N}{k} \leq \frac{N^k}{2^{k-1}}$ for $k \geq 2$ (see page 12), we have

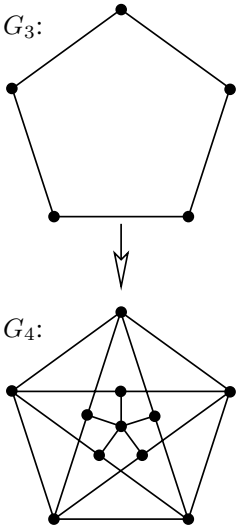
$$\binom{N}{k} 2^{-\binom{k}{2}} \;\leq\; \frac{N^k}{2^{k-1}} 2^{-\binom{k}{2}} \;<\; 2^{\frac{k^2}{2} - \binom{k}{2} - k + 1} \;=\; 2^{-\frac{k}{2}+1} \leq \frac{1}{2}.$$

Hence $p_{\scriptscriptstyle R} < \frac{1}{2}$, and by symmetry $p_{\scriptscriptstyle B} < \frac{1}{2}$ for the probability of some k vertices with all edges between them colored blue. We conclude that $p_{\scriptscriptstyle R} + p_{\scriptscriptstyle B} < 1$ for $N < 2^{\frac{k}{2}}$, so there *must* be a coloring with no red or blue K_k, which means that K_N does not have property (k,k). □

Of course, there is quite a gap between the lower and the upper bound for $R(k,k)$. Still, as simple as this Book Proof is, no lower bound with a better exponent has been found for general k in the more than 50 years since Erdős' result. In fact, no one has been able to prove a lower bound of the form $R(k,k) > 2^{(\frac{1}{2}+\varepsilon)k}$ nor an upper bound of the form $R(k,k) < 2^{(2-\varepsilon)k}$ for a fixed $\varepsilon > 0$.

Our third result is another beautiful illustration of the probabilistic method. Consider a graph G on n vertices and its chromatic number $\chi(G)$. If $\chi(G)$ is high, that is, if we need many colors, then we might suspect that G contains a large complete subgraph. However, this is far from the truth. Already in the fourties Blanche Descartes constructed graphs with arbitrarily high chromatic number and no triangles, that is, with every cycle having length at least 4, and so did several others (see the box on the next page).

However, in these examples there were many cycles of length 4. Can we do even better? Can we stipulate that there are no cycles of small length and still have arbitrarily high chromatic number? Yes we can! To make matters precise, let us call the length of a shortest cycle in G the *girth* $\gamma(G)$ of G; then we have the following theorem, first proved by Paul Erdős.



Constructing the Mycielski graph

Triangle-free graphs with high chromatic number

Here is a sequence of triangle-free graphs $G_3, G_4, \ldots$ with

$$\chi(G_n) \;=\; n.$$

Start with $G_3 = C_5$, the 5-cycle; thus $\chi(G_3) = 3$. Suppose we have already constructed G_n on the vertex set V. The new graph G_{n+1} has the vertex set $V \cup V' \cup \{z\}$, where the vertices $v' \in V'$ correspond bijectively to $v \in V$, and z is a single other vertex. The edges of G_{n+1} fall into 3 classes: First, we take all edges of G_n; secondly every vertex v' is joined to precisely the neighbors of v in G_n; thirdly z is joined to all $v' \in V'$. Hence from $G_3 = C_5$ we obtain as G_4 the so-called *Mycielski graph*.

Clearly, G_{n+1} is again triangle-free. To prove $\chi(G_{n+1}) = n + 1$ we use induction on n. Take any n-coloring of G_n and consider a color class C. There must exist a vertex $v \in C$ which is adjacent to at least one vertex of every other color class; otherwise we could distribute the vertices of C onto the $n - 1$ other color classes, resulting in $\chi(G_n) \leq n - 1$. But now it is clear that v' (the vertex in V' corresponding to v) must receive the same color as v in this n-coloring. So, all n colors appear in V', and we need a new color for z.

Theorem 3. *For every $k \geq 2$, there exists a graph G with chromatic number $\chi(G) > k$ and girth $\gamma(G) > k$.*

The strategy is similar to that of the previous proofs: We consider a certain probability space on graphs and go on to show that the probability for $\chi(G) \leq k$ is smaller than $\frac{1}{2}$, and similarly the probability for $\gamma(G) \leq k$ is smaller than $\frac{1}{2}$. Consequently, there must exist a graph with the desired properties.

■ **Proof.** Let $V = \{v_1, v_2, \ldots, v_n\}$ be the vertex set, and p a fixed number between 0 and 1, to be carefully chosen later. Our probability space $\mathcal{G}(n, p)$ consists of all graphs on V where the individual edges appear with probability p, independently of each other. In other words, we are talking about a Bernoulli experiment where we throw in each edge with probability p. As an example, the probability $\text{Prob}(K_n)$ for the complete graph is $\text{Prob}(K_n) = p^{\binom{n}{2}}$. In general, we have $\text{Prob}(H) = p^m(1-p)^{\binom{n}{2}-m}$ if the graph H on V has precisely m edges.

Let us first look at the chromatic number $\chi(G)$. By $\alpha = \alpha(G)$ we denote the *independence number*, that is, the size of a largest independent set in G. Since in a coloring with $\chi = \chi(G)$ colors all color classes are independent (and hence of size $\leq \alpha$), we infer $\chi\alpha \geq n$. Therefore if α is small as compared to n, then χ must be large, which is what we want.

Suppose $2 \leq r \leq n$. The probability that a fixed r-set in V is independent

is $(1-p)^{\binom{r}{2}}$, and we conclude by the same argument as in Theorem 2

$$\begin{aligned}\text{Prob}(\alpha \geq r) &\leq \binom{n}{r}(1-p)^{\binom{r}{2}} \\ &\leq n^r(1-p)^{\binom{r}{2}} = (n(1-p)^{\frac{r-1}{2}})^r \leq (ne^{-p(r-1)/2})^r,\end{aligned}$$

since $1-p \leq e^{-p}$ for all p.

Given any fixed $k > 0$ we now choose $p := n^{-\frac{k}{k+1}}$, and proceed to show that for n large enough,

$$\text{Prob}\Big(\alpha \geq \frac{n}{2k}\Big) < \frac{1}{2}. \tag{3}$$

Indeed, since $n^{\frac{1}{k+1}}$ grows faster than $\log n$, we have $n^{\frac{1}{k+1}} \geq 6k\log n$ for large enough n, and thus $p \geq 6k\frac{\log n}{n}$. For $r := \lceil \frac{n}{2k} \rceil$ this gives $pr \geq 3\log n$, and thus

$$ne^{-p(r-1)/2} = ne^{-\frac{pr}{2}}e^{\frac{p}{2}} \leq ne^{-\frac{3}{2}\log n}e^{\frac{1}{2}} = n^{-\frac{1}{2}}e^{\frac{1}{2}} = (\tfrac{e}{n})^{\frac{1}{2}},$$

which converges to 0 as n goes to infinity. Hence (3) holds for all $n \geq n_1$.

Now we look at the second parameter, $\gamma(G)$. For the given k we want to show that there are not too many cycles of length $\leq k$. Let i be between 3 and k, and $A \subseteq V$ a fixed i-set. The number of possible i-cycles on A is clearly the number of cyclic permutations of A divided by 2 (since we may traverse the cycle in either direction), and thus equal to $\frac{(i-1)!}{2}$. The total number of possible i-cycles is therefore $\binom{n}{i}\frac{(i-1)!}{2}$, and every such cycle C appears with probability p^i. Let X be the random variable which counts the number of cycles of length $\leq k$. In order to estimate X we use two simple but beautiful tools. The first is linearity of expectation, and the second is Markov's inequality for nonnegative random variables, which says

$$\text{Prob}(X \geq a) \leq \frac{EX}{a},$$

where EX is the expected value of X. See the appendix to Chapter 15 for both tools.

Let X_C be the indicator random variable of the cycle C of, say, length i. That is, we set $X_C = 1$ or 0 depending on whether C appears in the graph or not; hence $EX_C = p^i$. Since X counts the number of all cycles of length $\leq k$ we have $X = \sum X_C$, and hence by linearity

$$EX = \sum_{i=3}^{k}\binom{n}{i}\frac{(i-1)!}{2}p^i \leq \frac{1}{2}\sum_{i=3}^{k}n^ip^i \leq \frac{1}{2}(k-2)n^kp^k$$

where the last inequality holds because of $np = n^{\frac{1}{k+1}} \geq 1$. Applying now Markov's inequality with $a = \frac{n}{2}$, we obtain

$$\text{Prob}(X \geq \tfrac{n}{2}) \leq \frac{EX}{n/2} \leq (k-2)\frac{(np)^k}{n} = (k-2)n^{-\frac{1}{k+1}}.$$

Since the right-hand side goes to 0 with n going to infinity, we infer that $p(X \geq \frac{n}{2}) < \frac{1}{2}$ for $n \geq n_2$.

Now we are almost home. Our analysis tells us that for $n \geq \max(n_1, n_2)$ there exists a graph H on n vertices with $\alpha(H) < \frac{n}{2k}$ and fewer than $\frac{n}{2}$ cycles of length $\leq k$. Delete one vertex from each of these cycles, and let G be the resulting graph. Then $\gamma(G) > k$ holds at any rate. Since G contains more than $\frac{n}{2}$ vertices and satisfies $\alpha(G) \leq \alpha(H) < \frac{n}{2k}$, we find

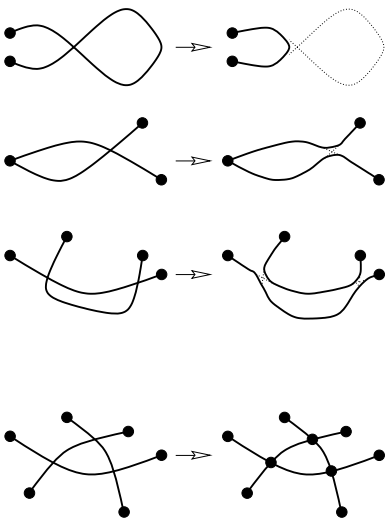
$$\chi(G) \;\geq\; \frac{n/2}{\alpha(G)} \;\geq\; \frac{n}{2\alpha(H)} \;>\; \frac{n}{n/k} \;=\; k,$$

and the proof is finished. □

Explicit constructions of graphs with high girth and chromatic number (of huge size) are known. (In contrast, one does not know how to construct red/blue colorings with no large monochromatic cliques, whose existence is given by Theorem 2.) What remains striking about the Erdős proof is that it proves the existence of relatively small graphs with high chromatic number and girth.

To end our excursion into the probabilistic world let us discuss an important result in geometric graph theory (which again goes back to Paul Erdős) whose stunning Book Proof is of recent vintage.

Consider a simple graph $G = G(V, E)$ with n vertices and m edges. We want to embed G into the plane just as we did for planar graphs. Now, we know from Chapter 12 — as a consequence of Euler's formula — that a simple planar graph G has at most $3n - 6$ edges. Hence if m is greater than $3n - 6$, there must be crossings of edges. The *crossing number* $\mathrm{cr}(G)$ is then naturally defined: It is the smallest number of crossings among all drawings of G, where crossings of more than two edges in one point are not allowed. Thus $\mathrm{cr}(G) = 0$ if and only if G is planar.



In such a minimal drawing the following three situations are ruled out:

- No edge can cross itself.
- Edges with a common endvertex cannot cross.
- No two edges cross twice.

This is because in either of these cases, we can construct a different drawing of the same graph with fewer crossings, using the operations that are indicated in our figure. So, from now on we assume that any drawing observes these rules.

Suppose that G is drawn in the plane with $\mathrm{cr}(G)$ crossings. We can immediately derive a lower bound on the number of crossings. Consider the following graph H: The vertices of H are those of G together with all crossing points, and the edges are all pieces of the original edges as we go along from crossing point to crossing point.

The new graph H is now plane and simple (this follows from our three assumptions!). The number of vertices in H is $n + \mathrm{cr}(G)$ and the number

of edges is $m + 2\text{cr}(G)$, since every new vertex has degree 4. Invoking the bound on the number of edges for plane graphs we thus find

$$m + 2\,\text{cr}(G) \;\leq\; 3(n + \text{cr}(G)) - 6,$$

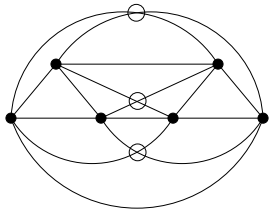
that is,

$$\text{cr}(G) \;\geq\; m - 3n + 6. \tag{4}$$

As an example, for the complete graph K_6 we compute

$$\text{cr}(K_6) \geq 15 - 18 + 6 = 3$$

and, in fact, there is an drawing with just 3 crossings.



The bound (4) is good enough when m is linear in n, but when m is larger compared to n, then the picture changes, and this is our theorem.

Theorem 4. *Let G be a simple graph with n vertices and m edges, where $m \geq 4n$. Then*

$$\text{cr}(G) \;\geq\; \frac{1}{64}\frac{m^3}{n^2}.$$

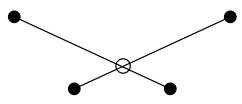
The history of this result, called the *crossing lemma*, is quite interesting. It was conjectured by Erdős and Guy in 1973 (with $\frac{1}{64}$ replaced by some constant c). The first proofs were given by Leighton in 1982 (with $\frac{1}{100}$ instead of $\frac{1}{64}$) and independently by Ajtai, Chvátal, Newborn and Szemerédi. The crossing lemma was hardly known (in fact, many people thought of it as a conjecture long after the original proofs), until László Székely demonstrated its usefulness in a beautiful paper, applying it to a variety of hitherto hard geometric extremal problems. The proof which we now present arose from e-mail conversations between Bernard Chazelle, Micha Sharir and Emo Welzl, and it belongs without doubt in The Book.

■ **Proof.** Consider a minimal drawing of G, and let p be a number between 0 and 1 (to be chosen later). Now we generate a subgraph of G, by selecting the vertices of G to lie in the subgraph with probability p, independently from each other. The induced subgraph that we obtain that way will be called G_p.

Let n_p, m_p, X_p be the random variables counting the number of vertices, of edges, and of crossings in G_p. Since $\text{cr}(G) - m + 3n \geq 0$ holds by (4) for *any* graph, we certainly have

$$E(X_p - m_p + 3n_p) \;\geq\; 0.$$

Now we proceed to compute the individual expectations $E(n_p)$, $E(m_p)$ and $E(X_p)$. Clearly, $E(n_p) = pn$ and $E(m_p) = p^2 m$, since an edge appears in G_p if and only if both its endvertices do. And finally, $E(X_p) = p^4\text{cr}(G)$, since a crossing is present in G_p if and only if all four (distinct!) vertices involved are there.



By linearity of expectation we thus find

$$0 \;\leq\; E(X_p) \;-\; E(m_p) \;+\; 3E(n_p) \;=\; p^4\mathrm{cr}(G) - p^2 m + 3pn,$$

which is

$$\mathrm{cr}(G) \;\geq\; \frac{p^2 m - 3pn}{p^4} \;=\; \frac{m}{p^2} - \frac{3n}{p^3}. \tag{5}$$

Here comes the punch line: Set $p := \frac{4n}{m}$ (which is at most 1 by our assumption), then (5) becomes

$$\mathrm{cr}(G) \;\geq\; \frac{1}{64}\left[\frac{4m}{(n/m)^2} - \frac{3n}{(n/m)^3}\right] \;=\; \frac{1}{64}\frac{m^3}{n^2},$$

and this is it. □

Paul Erdős would have loved to see this proof.

References

[1] M. AJTAI, V. CHVÁTAL, M. NEWBORN & E. SZEMERÉDI: *Crossing-free subgraphs,* Annals of Discrete Math. **12** (1982), 9-12.

[2] N. ALON & J. SPENCER: *The Probabilistic Method,* Third edition, Wiley-Interscience 2008.

[3] P. ERDŐS: *Some remarks on the theory of graphs,* Bulletin Amer. Math. Soc. **53** (1947), 292-294.

[4] P. ERDŐS: *Graph theory and probability,* Canadian J. Math. **11** (1959), 34-38.

[5] P. ERDŐS: *On a combinatorial problem I,* Nordisk Math. Tidskrift **11** (1963), 5-10.

[6] P. ERDŐS & R. K. GUY: *Crossing number problems,* Amer. Math. Monthly **80** (1973), 52-58.

[7] P. ERDŐS & A. RÉNYI: *On the evolution of random graphs,* Magyar Tud. Akad. Mat. Kut. Int. Közl. **5** (1960), 17-61.

[8] T. LEIGHTON: *Complexity Issues in VLSI,* MIT Press, Cambridge MA 1983.

[9] L. A. SZÉKELY: *Crossing numbers and hard Erdős problems in discrete geometry,* Combinatorics, Probability, and Computing **6** (1997), 353-358.