8.5. Random Graphs

In its simplest form, the probabilistic method is used to prove the existence of desired combinatorial objects without constructing them. An appropriate probability model is defined on a large class of objects. The occurrence of the desired structure is an event. If this event has positive probability, then some object with the desired structure exists. Designing the model and applying probabilistic and asymptotic techniques may involve considerable art.

We discuss these methods in the context of random graphs. The study of random graphs is itself motivated by the modeling of physical properties and by the analysis of algorithms in computer science.

8.5.1. Example. *Melting points.* The behavior of random graphs suggests a mathematical explanation for melting points. Think of a solid as a three-dimensional grid of molecules, with neighboring molecules joined by bonds. For example, consider the graph $P_l \square P_m \square P_n$, with bonds corresponding to edges.

Adding energy excites molecules and breaks bonds. We assume that bonds break at random as we raise the temperature (energy level). Each temperature corresponds to some fraction of bonds broken. While the graph remains largely connected, the material seems solid. Breaking off small pieces doesn't change this, but when all the components are small the global nature of the material changes. Small components of molecules float freely, like a liquid or gas.

Mathematically, there is a threshold for the number of bonds to be broken (in terms of the size of the grid) such that almost every way of breaking somewhat fewer bonds leaves a giant component, and almost every way of breaking somewhat more bonds leaves all components being tiny. Just below the threshold temperature the material will almost certainly be a solid, and just above it the material will almost certainly not be a solid. ∎

8.5.2. Example. *Analysis of algorithms.* Worst-case complexity is the maximum running time for an algorithm over all inputs of size n (see Appendix B). For difficult problems, we may seek an algorithm that takes many steps on a few bizarre graphs while running quickly on most graphs. We need a way to describe the usefulness of such algorithms.

The answer is **probabilistic analysis**. We assume a probability distribution on the inputs and study the expected running time with respect to this distribution. Choosing a realistic distribution can be difficult. In practice, we choose a probability distribution that makes the analysis feasible. We cannot define a probability distribution over infinitely many graphs, so we define a distribution on the graphs of each order. This is consistent with viewing the expected running time as a function of the input size. ∎

Erdős and Rényi [1959] introduced random graphs. The subject developed rapidly in the 1980s, with books by Bollobás [1985], by Palmer [1985], and by Alon and Spencer [1992] (the last treats broader combinatorial applications of probabilistic methods). The book Janson–Łuczak–Ruciński [2000] emphasizes later developments.

More sophisticated probabilistic techniques than we can present here are now being applied to random graphs. We describe the basic techniques and suggest the flavor of the subject, with no attempt at exhaustive treatment.

EXISTENCE AND EXPECTATION

We begin by showing how probabilistic methods can prove existence statements. Suppose we want to prove that an object with some desired property exists. We define a probability space where occurrence of the desired property is an event A. If A has positive probability, then the desired object exists.

8.5.3. Definition. A discrete **probability space** or **probability model** is a finite or countable set S together with nonnegative weights on the elements that sum to 1. **An event** is a subset of S. The **probability** $P(A)$ of an event A is the sum of the weights of the elements of A. Events A and B are **independent** if $P(A \cap B) = P(A)P(B)$.

Erdős popularized the probabilistic method in 1947 by using it to prove lower bounds on Ramsey numbers (Definition 8.3.6). We phrased this combinatorially in Theorem 8.3.12; here we present the same proof in probabilistic language. It uses the observation that $P(\bigcup_i A_i) \le \sum_i P(A_i)$. Note that **in this section, all graphs are simple.**

8.5.4. Theorem. (Erdős [1947]) If $\binom{n}{p}2^{1-\binom{p}{2}} < 1$, then $R(p, p) > n$.

Proof: It suffices to show that when $\binom{n}{p}2^{1-\binom{p}{2}} < 1$ there is an n-vertex graph G with $\omega(G) < p$ and $\alpha(G) < p$. We define a probability model on graphs with vertex set $[n]$ by letting each edge appear independently with probability .5. If the probability of the event Q = "no p-clique or independent p-set" is positive, then the desired graph exists.

Each possible p-clique occurs with probability $2^{-\binom{p}{2}}$, since obtaining the complete graph requires obtaining all its edges, and they occur independently. Hence the probability of having at least one p-clique is bounded by $\binom{n}{p}2^{-\binom{p}{2}}$. The same bound holds for independent p-sets. Hence the probability of "not Q" is bounded by $\binom{n}{p}2^{1-\binom{p}{2}}$, and the given inequality guarantees that $P(Q) > 0$. ■

8.5.5. Remark. Existence arguments can be used as probabilistic construction algorithms. The probability that a random 1024-vertex graph has a 10-clique or independent 10-set is less than $2^{11}/20!$. If the first random one doesn't work,

generate another; the probability of continued failure is the *product* of these small numbers and soon becomes incomprehensibly small. ∎

The lower bound in Theorem 8.5.4 is roughly $\sqrt{2}^k$; the inductive upper bound in Theorem 8.3.11 is roughly 4^k. The gap is large. More sophisticated probabilistic methods have achieved only small improvements in the lower bound. Nevertheless, the constructive bounds are much weaker, so this is a triumph for the probabilistic method. The proof is essentially just a counting argument. Many probabilistic arguments with finite sample spaces can be rephrased as weighted counting arguments, but the proofs are simpler in the language of probability.

The introduction of random variables adds considerable power. We assign values to the elements of our probability space.† We have already used the comparison between the average and maximum values of a random variable to prove inequalities.

8.5.6. Definition. A **random variable** is a function assigning a real number to each element of a probability space. We use $X = k$ to denote the event consisting of all elements where the variable X has the value k.

The **expectation** $E(X)$ of a random variable X is the weighted average $\sum_k kP(X = k)$. The **pigeonhole property** of the expectation is the statement that there exists an element of the probability space for which the value of X is as large as (or as small as) $E(X)$.

Applying the pigeonhole property requires a value or bound for $E(X)$. Often the computation applies the **linearity of expectation** to an expression for X in terms of simpler random variables. For our purposes, we generally restrict our attention to probability models on finite sets and sum only finitely many random variables. Analogous results hold in continuous probability spaces.

8.5.7. Lemma. (Linearity property) If X and the finite set $\{X_i\}$ are random variables on the same space and $X = \sum X_i$, then $E(X) = \sum E(X_i)$. Also $E(cX) = cE(X)$ for $c \in \mathbb{R}$.

Proof: In a discrete probability space, each element contributes the same amount to each side of the desired equations. ∎

We often apply Lemma 8.5.7 to random variables that count substructures. Such a random variable is a sum of variables indicating whether one of the possible things being counted actually occurs. These **indicator variables** take values in $\{0, 1\}$ (they are also called 0, 1-variables). The expectation of an indicator variable is the probability that it equals 1. These properties facilitate what was perhaps the first use of the probabilistic method.

†We consider only discrete probability spaces, but analogous concepts hold for continuous probability spaces.

8.5.8. Theorem. (Szele [1943]) Some n-vertex tournament has at least $n!/2^{n-1}$ Hamiltonian paths.

Proof: Generate tournaments on $[n]$ randomly by choosing $i \to j$ or $j \to i$ with equal probability for each pair $\{i, j\}$. Let X be the number of Hamiltonian paths; X is the sum of $n!$ indicator variables for the possible Hamiltonian paths. Each Hamiltonian path occurs with probability $1/2^{n-1}$, so $E(X) = n!/2^{n-1}$. In some tournament, X is at least as large as the expectation. ∎

This simple bound using expectation gives almost the right answer for the maximum number of Hamiltonian paths in an n-vertex tournament; Alon [1990] proved that it is at most $n!/(2 - o(1))^n$. When almost all instances have a value near the extreme, probabilistic arguments are especially effective.

Many inequalities can be interpreted as statements about the expected value of a random variable. This often yields a shorter proof than combinatorial methods. Exercise 3.1.42 requests a combinatorial proof of the next result.

8.5.9. Theorem. (Caro [1979], Wei [1981]) $\alpha(G) \geq \sum_{v \in V(G)} \frac{1}{d(v)+1}$ for every graph G.

Proof: (Alon–Spencer [1992, p81]) Given an ordering of the vertices of G, the set of vertices that appear before all their neighbors form an independent set. When the ordering is chosen uniformaly at random, the probability that v appears before all its neighbors is $1/(d(v)+1)$. Thus the right side of the inequality is the expected size of the independent set formed by choosing the vertices appearing before their neighbors in a random vertex ordering. ∎

When a randomly generated object is close to having a desired property, a slight alteration may produce it. This technique is called the **deletion method**, the **alteration method**, or the **two-step method**. Ramsey numbers furnish a classical application (Exercise 16). We provide two others.

Recall that $S \subseteq V(G)$ is a *dominating set* in G if every vertex outside S has a neighbor in S (Definition 3.1.26). When G is k-regular, every vertex dominates $k+1$ vertices (including itself), so every dominating set has at least $n(G)/(k+1)$ vertices. The alteration method yields a dominating set close to that bound in every graph with minimum degree k. The argument, like many involving these techniques, uses the fundamental inequality $1 - p < e^{-p}$ (Exercise 2).

8.5.10. Theorem. (Alon [1990]) Every n-vertex graph with minimum degree $k > 1$ has a dominating set of size at most $n\frac{1+\ln(k+1)}{k+1}$.

Proof: In such a graph G, select a random set $S \subseteq V(G)$ by including each vertex independently with probability $p = \ln(k+1)/(k+1)$. Given S, let T be the set of vertices outside S having no neighbor in S; adding T to S yields a dominating set. We seek the expected size of $S \cup T$.

Since each vertex appears in S with probability p, linearity yields $E(|S|) = np$. The random variable $|T|$ is the sum of n indicator variables for whether individual vertices belong to T. We have $v \in T$ if and only if v and its neighbors

all fail to be in S. This has probability bounded by $(1-p)^{k+1}$, since v has degree at least k. Since $(1-p)^{k+1} < e^{-p(k+1)}$, we have $E(|S|+|T|) \le np + ne^{-p(k+1)} = n\frac{1+\ln(k+1)}{k+1}$. The pigeonhole property of the expectation completes the proof. ∎

This easy bound yields almost the smallest s_k such that every graph G with minimum degree k has a dominating set of size at most $s_k n(G)$ (Alon [1990]). A greedy algorithm proves the same result constructively (Theorem 3.1.30).

A striking and famous application of the deletion method is the existence of graphs with large girth and chromatic number. Explicit constructions came much later (Lovász [1968a], Nešetřil–Rödl [1979], Kriz [1989]). We present a simplification of the original proof (Alon–Spencer [1992, p35]). It uses a property of the expectation that we will prove in Lemma 8.5.17.

8.5.11. Theorem. (Erdős [1959]) Given $m \ge 3$ and $g \ge 3$, there exists a graph with girth at least g and chromatic number at least m.

Proof: We generate graphs with vertex set $[n]$ by letting each pair be an edge with probability p, independently. A graph with no large independent set has large chromatic number, since $\chi(G) \ge n(G)/\alpha(G)$. We therefore choose p large enough to make large independent sets unlikely. We also choose p small enough to make the expected number of short cycles (length less than g) small. Given a graph satisfying both conditions, we can delete a vertex from each short cycle to obtain the desired graph.

To make it unlikely that we generate more than $n/2$ short cycles, we let $p = n^{t-1}$, where $t < 1/g$. Each of the possible cycles of length j occurs with probability p^j. Since there are $n_{(j)}/(2j)$ of these for each j, the total number X of cycles of length less than g has expectation

$$E(X) = \sum_{i=3}^{g-1} n_{(i)} p^i/(2i) \le \sum_{i=3}^{g-1} n^{ti}/(2i).$$

Since $tg < 1$, this implies that $E(X)/n \to 0$ as $n \to \infty$. In Markov's Inequality we will complete the details of concluding from this that $P(X \ge n/2) \to 0$ as $n \to \infty$. For n large enough, $P(X \ge n/2) < 1/2$.

Since $\alpha(G)$ cannot grow when we delete vertices, at least $(n - -X)/\alpha(G)$ independent sets are needed to color the vertices remaining when we delete a vertex of each cycle. If $X < n/2$ and $\alpha(G) \le n/(2k)$, then at least k colors are needed for the graph remaining. With $r = \lceil 3 \ln n/p \rceil$, we have

$$P(\alpha(G) \ge r) \le \binom{n}{r}(1-p)^{\binom{r}{2}} < [ne^{-p(r-1)/2}]^r.$$

This approaches 0 as n grows.

Since $r = \lceil 3n^{1-t} \ln n \rceil$ and k is fixed, we can choose n large enough to obtain $r < n/(2k)$. If we also choose n large enough so that $P(X \ge n/2) < 1/2$ and $P(\alpha(G) \ge r) < 1/2$, then there will exist an n-vertex graph G such that $\alpha(G) \le n/(2k)$ and such that G has fewer than $n/2$ cycles of length less than g. We delete a vertex from each short cycle and retain a graph with girth at least g and chromatic number at least k. ∎

PROPERTIES OF ALMOST ALL GRAPHS

We have proposed studying properties that "almost always" hold. This phrase has meaning in the context of a probability model.

8.5.12. Definition. Given a sequence of probability spaces, let q_n be the probability that property Q holds in the nth space. Property Q **almost always** holds if $\lim_{n\to\infty} q_n = 1$.

For us, the nth space is a probability distribution over n-vertex graphs. When property Q almost always holds, we say "almost every graph has property Q". Making all graphs with vertex set $[n]$ equally likely is equivalent to letting each vertex pair appear as an edge with probability $1/2$. Models where edges arise independently with the same probability are the most common for random graphs because they lead to the simplest computations. We allow this probability to depend on n.

8.5.13. Definition. Model A: Given n and $p = p(n)$, generate graphs with vertex set $[n]$ by letting each pair be an edge with probability p, independently. Each graph with m edges has probability $p^m(1-p)^{\binom{n}{2}-m}$. The random variable G^p denotes a graph drawn from this probability space. "*The* random graph" means Model A with $p = 1/2$, which makes all graphs with vertex set $[n]$ equally likely.

Computations are much simpler for graphs with a fixed vertex set ("labeled" graphs) than for random isomorphism classes. Since inputs to algorithms are graphs with specified vertex sets, this model is consistent with applications.

We often measure running times of algorithms in terms of the number of vertices and number of edges; hence we may want to control the number of edges. This suggests a model in which the n-vertex labeled graphs with m edges are equally likely. (We use m to count edges in this section because the number $e = 2.71828...$ plays an important role in asymptotic arguments.)

8.5.14. Definition. Model B: Given n and $m = m(n)$, let each graph with vertex set $[n]$ and m edges occur with probability $\binom{N}{m}^{-1}$, where $N = \binom{n}{2}$. The random variable G^m denotes a graph generated in this way.

These two are the most common of many models studied. Model B seems more pertinent for applications. We ask questions like "as a function of n, how many edges are needed to make a graph almost surely connected?" In Model A we would say, "as a function of n, what edge probability is needed to make a graph almost surely connected?" Unfortunately, calculations needed to answer such questions are messier in Model B than in Model A.

Fortunately, Model B is accurately described by Model A when n is large and $p = m/\binom{n}{2}$, because the actual number of edges generated in Model A is almost always very close to the resulting expectation m. The correspondence

is valid for most properties of interest. The proof of this requires detailed use of the binomial distribution for the number of edges. A graph property Q is **convex** if G satisfies Q whenever $F \subseteq G \subseteq H$ and F, H satisfy Q.

8.5.15. Theorem. (Bollobás [1985, p34-35]) If Q is convex and $p(1-p)\binom{n}{2} \to \infty$, then almost every G^p satisfies Q if and only if, for every fixed x, almost every G^m satisfies Q, where $m = \lfloor p\binom{n}{2} + x[p(1-p)\binom{n}{2}]^{1/2} \rfloor$. ∎

Theorem 8.5.15 justifies restricting our attention to Model A. It also motivates letting p be a function of n; to study graphs with a linear number of edges, we must let p vanish at a rate like c/n, where c is constant. Constant p yields dense graphs.

Proving $P(Q) \to 1$ is usually much easier than computing $P(Q)$; this distinction is important. Exact computation of probabilities is difficult, unnecessary, and avoided wherever possible. Instead we use asymptotic analysis, which rests on limits. We write $a_n \to L$ for $\lim_{n\to\infty} a_n = L$. To compare growth rates of sequences, we use "big O" and "little o" notation (see Appendix B for definitions). We write $a_n = b_n(1 + o(1))$ when $\langle a \rangle$ and $\langle b \rangle$ differ by a sequence that grows more slowly than $\langle b \rangle$; equivalently, $a_n/b_n \to 1$. When $a_n/b_n \to 1$, we say that a_n is **asymptotic** to b_n, written $a_n \sim b_n$.

We use asymptotic statements to discard lower-order terms that don't affect whether $\lim_{n\to\infty} P(Q) = 1$. Computing $P(Q)$ first and then proving that the formula tends to 1 is harder and is unnecessary. We need only show that $P(\neg Q)$ is *bounded* by something tending to 0. Many asymptotic arguments are "sloppy" in this sense; we don't care how loose the bound is as long as it tends to 0. Experience refines our intuition about what can be discarded safely.

8.5.16. Theorem. (Gilbert [1959]) When p is constant, almost every G^p is connected.

Proof: We can make G disconnected by picking a vertex partition into two sets and forbidding edges between the two sets. Occurrence of edges within the sets is irrelevant. We bound the probability q_n that G^p is disconnected by summing $P([S, \overline{S}] = \emptyset)$ over all bipartitions $S, \overline{S}$. Graphs with many components are counted many times. When $|S| = k$, there are $k(n-k)$ possible edges in $[S, \overline{S}]$. Each has probability $1-p$ of not appearing, independently, so $P([S, \overline{S}] = \emptyset) = (1-p)^{k(n-k)}$. Considering all S generates each partition from each side, so $q_n \le \frac{1}{2}\sum_{k=1}^{n-1}\binom{n}{k}(1-p)^{k(n-k)}$.

This formula is symmetric in k and $n-k$; hence q_n is bounded by $\sum_{k=1}^{\lfloor n/2 \rfloor}\binom{n}{k}(1-p)^{k(n-k)}$. We loosen the bound to simplify it. Using $\binom{n}{k} < n^k$ and $(1-p)^{n-k} \le (1-p)^{n/2}$ (for $k \le n/2$) yields $q_n < \sum_{k=1}^{\lfloor n/2 \rfloor}(n(1-p)^{n/2})^k$. For large enough n, we have $n(1-p)^{n/2} < 1$. This makes our bound the initial portion of a convergent geometric series. We obtain $q_n < x/(1-x)$, where $x = n(1-p)^{n/2}$. Since $n(1-p)^{n/2} \to 0$ when p is constant, our bound on q_n approaches 0 as $n \to \infty$. ∎

We avoid struggling with probability formulas by introducing integer-valued random variables and techniques involving expectation. If X is a nonnegative random variable such that $X = 0$ when G^p has property Q, then $E(X) \to 0$ implies that almost every G^p satisfies Q. This is a special case of the following lemma. We prove it only for integer variables, but it also holds for continuous variables.

8.5.17. Lemma. (Markov's Inequality) If X takes only nonnegative values, then $P(X \geq t) \leq E(X)/t$. In particular, if X is integer-valued, then $E(X) \to 0$ implies $P(X = 0) \to 1$.

Proof: $E(X) = \sum_{k\geq 0} kp_k \geq \sum_{k\geq t} kp_k \geq t \sum_{k\geq t} p_k = tP(X \geq t)$. ∎

For connectedness, we can define $X(G^p)$ by $X = 1$ if G is disconnected and $X = 0$ otherwise. The expectation of an indicator variable is the probability that it equals 1. We proved $P(X = 1) \to 0$ (when p is constant) to prove that almost every G^p is connected. With a different random variable we can simplify the proof and strengthen the result. We still want G to satisfy Q if $X = 0$ (in order to apply Markov's Inequality), but we don't need $(X = 0) \Leftrightarrow (G$ satisfies $Q)$. We use a sum X of many indicator variables, such that G satisfies Q if $X = 0$. The linearity of expectation and convenience of $E(X_i) = P(X_i = 1)$ for the indicator variables simplify the task of proving $E(X) \to 0$.

8.5.18. Theorem. If p is constant, then almost every G_p has diameter 2 (and hence is connected).

Proof: Let $X(G^p)$ be the number of unordered vertex pairs with no common neighbor. If there are none, then G_p is connected and has diameter 2. By Markov's Inequality, we need only show $E(X) \to 0$. We express X as the sum of $\binom{n}{2}$ indicator variables $X_{i,j}$, one for each vertex pair $\{v_i, v_j\}$, where $X_{i,j} = 1$ if and only if v_i, v_j have no common neighbor.

When $X_{i,j} = 1$, the $n-2$ other vertices fail to have edges to both of these, so $P(X_{i,j} = 1) = (1-p^2)^{n-2}$ and $E(X) = \binom{n}{2}(1-p^2)^{n-2}$. When p is fixed, $E(X) \to 0$, and hence almost every G_p has diameter 2. ∎

The intuition behind this argument, made precise by Markov's Inequality, is that if we expect almost no bad pairs, then almost every graph has none. The summation disappears, and for the limit we need only know that $(1 - p^2)^{n-2}$ tends to 0 faster than any polynomial function of n.

THRESHOLD FUNCTIONS

Roughly speaking, random graphs with constant edge probability are connected because they have many more edges than needed to be connected. To improve Theorem 8.5.18, we want to make $p(n)$ as small as possible to have

almost every G^p connected. We need the notion of a threshold probability function. By the relationship between Model A and Model B, a threshold edge probability also yields a threshold number of edges.

8.5.19. Definition. A **monotone property** is a graph property preserved by addition of edges. A **threshold probability function** for a monotone property Q is a function $t(n)$ such that $p(n)/t(n) \to 0$ implies that almost no G^p satisfies Q, and $p(n)/t(n) \to \infty$ implies that almost every G^p satisfies Q. **Threshold edge function** is defined similarly for Model B.

This is a broad notion of threshold function; it allows a property to have many threshold functions. A threshold function $t(n)$ is "sharper" when the "almost surely" behavior occurs when $p(n)/t(n)$ approaches nonzero constants. Still sharper is a threshold $t(n)$ such that this behavior occurs when $p(n)$ differs from $t(n)$ by the subtraction or addition of a lower-order term.

Markov's Inequality does half the job of deriving a threshold function. If $X = 0$ implies property Q and we prove that $E(X) \to 0$, then $P(Q) \to 1$. We obtain candidates for threshold functions by determining which functions $p(n)$ yield $E(X) \to 0$. Often we obtain $p(n)$ such that $E(X) \to 0$ or $E(X) \to \infty$, depending on the value of a parameter c. The property $E(X) \to \infty$ suggests that $P(X = 0) \to 0$, but this does not always follow. For example, $E(X) \to \infty$ when $P(X = 0) = .5$ and $P(X = n) = .5$. To obtain $P(X = 0) \to 0$, we must prevent the probability from spreading out like this.

8.5.20. Definition. The rth **moment** of X is the expectation of X^r. The **variance** of X, written $Var(X)$, is the quantity $E[(X - E(X))^2]$. The **standard deviation** of X is the square root of $Var(X)$.

8.5.21. Lemma. (Second Moment Method) If X is a random variable, then $P(X = 0) \leq \frac{E(X^2)-E(X)^2}{E(X)^2}$. In particular, $P(X = 0) \to 0$ when $\frac{E(X^2)}{E(X)^2} \to 1$.

Proof: Applied to the variable $(X - E(X))^2$ and the value t^2, Markov's Inequality yields $P[(X - E(X))^2 \geq t^2] \leq E[(X - E(X))^2]/t^2$. We rewrite this as $P[|X - E(X)| \geq t] \leq Var(X)/t^2$ (Chebyshev's Inequality). Since

$$E[(X - E(X))^2] = E[X^2 - 2XE(X) + (E(X))^2] = E(X^2) - (E(X))^2,$$

Chebyshev's Inequality becomes $P[|X - E(X)| \geq t] \leq (E(X^2) - E(X)^2)/t^2$. Since $X = 0$ only when $|X - E(X)| \geq E(X)$, setting $t = E(X)$ completes the proof. ∎

Intuitively, if the mean grows and the standard deviation grows more slowly, then all the probability is pulled away from 0, and $P(X = 0) \to 0$ results. We illustrate the method by considering the disappearance of isolated vertices. Since a connected graph has no isolated vertices, a threshold for connectedness must be at least as large as a threshold for disappearance of isolated vertices. The computations for the latter are simpler, because we can express this condition using a sum of identically distributed indicator variables with

easily computed expectations. In fact, both properties have the same threshold, since it happens that at the threshold almost every graph consists of one huge component plus isolated vertices.

8.5.22. Theorem. In Model A, $\ln n/n$ (natural logarithm) is a threshold probability function for the disappearance of isolated vertices (that is, $\delta(G) \geq 1$). (The corresponding threshold in Model B is $\frac{1}{2}n \ln n$.)

Proof: Let X be the number of isolated vertices, with X_i indicating whether vertex i is isolated. Then $E(X) = \sum E(X_i) = n(1-p)^{n-1}$. We study the asymptotic behavior of $E(X)$ in terms of $p(n)$. Since

$$(1-p)^n = e^{n\ln(1-p)} = e^{-np}e^{-np^2[1/2+p/3+\cdots]},$$

our expression for $E(X)$ simplifies asymptotically if $np^2 \to 0$. This is equivalent to $p \in o(1/\sqrt{n})$ and implies $(1-p)^n \sim e^{-np}$ and $(1-p)^{-1} \sim 1$, yielding $E(X) \sim ne^{-np}$. To simplify further, set $p = c\ln n/n$ to obtain $ne^{-np} = n^{1-c}$, where c may depend on n. Constant c yields $p \in o(1/\sqrt{n})$, as we needed earlier. When $c > 1$, we have $E(X) \sim n^{1-c} \to 0$, which proves one side of the threshold.

When $c < 1$, we have $E(X) \to \infty$ and use the second moment method. We need only show that $E(X^2) \sim E(X)^2$. This uses another helpful property of indicator variables: $X_i^2 = X_i$. Thus,

$$E(X^2) = \textstyle\sum_{i=1}^{n} E(X_i^2) + \sum_{i\neq j} E(X_iX_j) = E(X) + n(n-1)E(X_iX_j).$$

The indicator variable X_iX_j has value 1 only when v_i and v_j are both isolated, which forbids $2(n-2)+1$ edges. Thus $E(X_iX_j) = (1-p)^{2n-3}$. Again $(1-p)^n \sim e^{-np}$, so $E(X_iX_j) \sim e^{-2np}$, and

$$E(X^2) \sim E(X) + n(n-1)e^{-2np} \sim E(X) + E(X)^2.$$

Since $E(X) \to \infty$, this yields $E(X^2) \sim E(X)^2$. ∎

Theorem 8.5.22 is stronger than required by the definition of threshold function. The threshold is sharper: we guarantee or forbid isolated vertices when the ratio of $p(n)$ to $\ln n/n$ approaches a nonzero constant, not 0 or ∞.

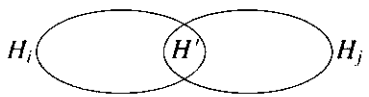
In fact, yet sharper information is known about the threshold for isolated vertices. When $p = \lg n/n + x/n$ and X counts the isolated vertices, $P(X = k) \sim e^{-\mu}\mu^k/k!$, where $\mu = e^{-x}$. (Readers may recognize this limiting distribution as the **Poisson distribution**.) For $k = 0$, we have $P(X = 0) \sim e^{-\mu}$. Thus this additive term in p describes the movement through the threshold from almost always isolated vertices to almost never isolated vertices. Many such sharp thresholds are known, but the techniques for deriving the asymptotic Poisson distribution are beyond our scope here.

Next we derive a threshold function for the appearance of fixed subgraphs. A graph is **balanced** if the average vertex degree in every induced subgraph is no larger than the average degree of the entire graph. All regular graphs and all forests are balanced.

8.5.23. Theorem. If H is a balanced graph with k vertices and l edges, then $p = n^{-k/l}$ is a threshold function in Model A for the appearance of H as a subgraph of almost every G^p.

Proof: Let X be the number of copies of H in G^p; X is the sum of indicator variables for the possible copies of H in K_n. There are $n(n-1)\cdots(n-k+1)$ ways to map $V(H)$ into $[n]$. Each copy of H arises A times, where A is the number of automorphisms of H. We thus have $\frac{1}{A}\prod_{j=0}^{k-1}(n-j)$ variables X_i. Since a copy of H occurs when its edges occur, $P(X_i = 1) = p^l$. Because k is fixed, $E(X) \sim n^k p^l / A$.

Setting $p(n) = c_n n^{-k/l}$ yields $E(X) \sim c_n^l / A$. Hence $c_n \to 0$ yields $E(X) \to 0$, and $c_n \to \infty$ yields $E(X) \to \infty$. It remains only to obtain $E(X^2) \sim E(X)^2$ when $c_n \to \infty$. Again $E(X^2) = E(X) + \sum_{i \neq j} E(X_i X_j)$. The summands are not equal; $E(X_i X_j)$ depends on $H' = H_i \cap H_j$. We group the terms by the choice of $H' \subseteq H$. When H' has r vertices and s edges, the number of edges needed to create H_i and H_j is $2l - s$, so $E(X_i X_j) = p^{2l-s}$.



To specify pairs i, j such that $H' = H_i \cap H_j$, we choose r vertices for H', $k-r$ vertices for each of $H_i - H'$ and $H_j - H'$, and an extension of H' to each of those sets. The number of ways to choose the vertex sets is $\frac{n!}{r!(k-r)!(k-r)!(n-2k+r)!}$, which is asymptotic to $n^{2k-r}/[r!(k-r)!^2]$. The number M of ways to extend H' to obtain copies of H in both specified k-sets depends only on H and H'; it is independent of n and p. Let $\alpha_{H'}$ be the constant $M/[r!(k-r)!^2]$. The contribution to $\sum E(X_i X_j)$ from pairs i, j such that $H_i \cap H_j = H'$ is asymptotic to $\alpha_{H'} n^{2k-r} p^{2l-s}$; we call this $E_{H'}$.

When $r = s = 0$, we have $M = (k!/A)^2$. Hence $\alpha_{H'} \sim n^{2k} p^{2l} / A^2 \sim E(X)^2$ when H' is the "null graph". This is the total contribution to $\sum E(X_i X_j)$ for all i, j with H_i, H_j disjoint and is asymptotic to $E(X)^2$. The proof is completed by showing that the total contribution from all other choices of H' has lower order. We have $E_{H'} \sim \alpha_{H'} A^2 E(X)^2 n^{-r} p^{-s}$. Since $2s/r$ is the average degree of H', the hypothesis that H is balanced yields $2r/s \geq 2k/l$, or $pn^{r/s} \geq pn^{k/l} \to \infty$ when $c_n \to \infty$. Since $pn^{r/s} \to \infty$ is equivalent to $n^{-r} p^{-s} \to 0$, we obtain $E_{H'} \in o(E(X)^2)$ for $H' \neq \emptyset$. Since the number of possible subgraphs H' is bounded (by an expression involving the constants k and l), this implies that $E(X^2) \sim E(X) + E_\emptyset \sim E(X)^2$. ∎

This result generalizes for all H. The ratio $d(H) = e(H)/n(H)$ is the **density** of H, and $\rho(H) = \max_{F \subseteq H} d(F)$ is the **maximum density**. These are equal precisely when H is balanced, and then $p = n^{-1/\rho(H)}$ is the threshold for appearance of H. Every graph H has a balanced subgraph F such that $d(F) = \rho(H)$. When $pn^{\rho(H)} \to 0$, almost every G^p has no copy of F; hence it also has no copy of H. In fact, $p = n^{-1/\rho(H)}$ is always a threshold function for the appearance of H (Exercise 25).

EVOLUTION AND GRAPH PARAMETERS

In the subtitle to his book, Palmer [1985] tells us that random graphs involve the study of

> "THRESHOLD FUNCTIONS, which facilitate the careful study of the structure of a graph as it grows, and specifically reveal the mysterious circumstances surrounding the abrupt appearance of the UNIQUE GIANT COMPONENT, which systematically absorbs its neighbors, devouring the larger first and ruthlessly continuing until the last ISOLATED VERTICES have been sucked up, whereupon the Giant is suddenly brought under control by a SPANNING CYCLE."

The evolutionary viewpoint generates random graphs with m edges in a way that yields the same probability space as Model B but makes intuitive reasoning easier. Almost everything suggested about random graphs by intuition or experimentation is true. The evolutionary viewpoint develops this intuition.

Generating m edges simultaneously or one-by-one yields the same probability distribution, making the graphs with m edges equally likely. By studying the likely effect of a new edge on the present structure, we can make intuitive hypotheses about the properties of the graph at any stage. A *stage* of evolution is a range of values for $m(n)$ (or $p(n)$) in which the structural description of the typical graph doesn't change much. We have studied the basic techniques for verifying these descriptions, but the computations can be difficult. Hence we will only describe the stages using the evolutionary intuition.

We remark first that a constant multiple of almost nothing is almost nothing. Therefore, when each of $A_1, \ldots, A_r$ happens almost always (r is fixed), it follows that almost always they all happen.

Beginning with many vertices and no edges, each new edge is likely to be isolated. The random graph is a matching until a substantial fraction of the vertices are involved in edges. The thresholds $p \sim cn^{-k/(k-1)}$ for appearance of fixed subtrees generalize this. Let $t_k(n) = n^{-k/(k-1)}$. If $p/t_k \to \infty$ but $p/t_{k+1} \to 0$, then every fixed subtree on k vertices appears, but none on $k+1$ vertices appears. (The statements about individual trees become statements about all trees of that order.) Furthermore, this p is also below the threshold for appearance of fixed cycles (density 1, length bounded by k), so G^p is a forest of trees of order at most k, and every tree on k vertices appears as a component.

Intuitively, the random graph has no cycles in this stage of evolution because when there is no large component a random added edge is much more likely to join two components than to lie in one component. To make the intuition precise, we let X be the number of cycles in G^p and compute

$$E(X) = \sum_{k=3}^{n} \binom{n}{k} \tfrac{1}{2}(k-1)!p^k < \sum_{k=3}^{n} (np)^k/2k.$$

If $pn \to 0$, then $E(X) \to 0$.

The next major stage of evolution is $p = c/n$ with $0 < c < 1$. With X counting cycles, we can no longer say that $E(X) \sim \sum_{k=3}^{n} (np)^k/2k$, because when

k is a substantial fraction of n the ratio $n^k/(n)_k$ does not approach 1. We must break $E(X)$ into two sums, and the arguments become more difficult. When $pn \to c$, we find that $E(X)$ approaches a constant c', and the number of cycles in G^p is asymptotically Poisson distributed. With cycles in a few components and all components small, we still expect the next edge to join two components or create a cycle in a component that doesn't have one. In this range, the size of the largest component is about $\log n$, there are many components, each having at most one cycle. Most vertices still belong to acyclic components.

When c reaches and passes 1, the structure of G^p changes radically. This is called the **double jump** because the structure of G^p is significantly different for $c < 1$, $c \sim 1$, and $c > 1$. At $pn = 1$, the second moment method guarantees that almost every G^p has a cycle. Also, the order of the largest component jumps from $\log n$ to $n^{2/3}$. For $pn = c > 1$, the number of vertices outside the "giant component" becomes $o(n)$. Also G^p is likely to have some cycle with three crossing chords and be nonplanar.

Next, let p approach $c \ln n/n$. With $c < 1$, we have proved that almost every G^p has isolated vertices. With $c > 1$, these disappear. As we add edges to a disconnected graph, the edges may go within a component or connect two components. When the components are all small, added edges will almost surely join components. Eventually, this results in the creation of a giant component. At this point, added edges are likely to lie within the giant component or to join it to one of the small components. Of the small components, those most likely to receive such edges are the larger ones. In other words, as c passes through 1 the last remaining small components swallowed by the giant component are isolated vertices. This explains intuitively why the threshold for connectedness is the same as the threshold for the disappearance of isolated vertices. With $c > 1$, suddenly almost every G^p also has a spanning cycle. Minimum degree k (and the appearance of the Hamiltonian cycle when $k = 2$) has a threshold that involves a lower-order term: $\ln n/n + (k-1)\ln\ln n/n$.

The last stages of evolution are those where $pn/\ln n \to \infty$ but $p = o(1)$, and then finally $p = c$; this brings us back to where we began our study.

When $p = c\log n/n$ with $c \to \infty$, we leave the domain of sparse graphs. The evolutionary viewpoint becomes less valuable, and we study properties of the random graph. We pay less attention to probability threshold functions and concentrate on the likely value of graph parameters, especially when p is constant. Given a parameter μ, we want to show that $\mu(G^p) \sim f(n)$ for almost every G^p. We can view this as a threshold when $\mu(G^p)$ is almost always between $(1-\epsilon)f(n)$ and $(1+\epsilon)f(n)$, for each $\epsilon > 0$. If $\mu(G^p)$ is almost always between $f(n) - \epsilon g(n)$ and $f(n) + \epsilon g(n)$, where $g(n) = o(f(n))$, then we have a stronger statement, written as $\mu(G^p) \in f(n)(1 + o(1))$.

Some properties that are true for almost all graphs occur in no known examples! For the known lower bound on Ramsey numbers, there is still no construction of an infinite class of graphs such that $\alpha(G) < \log_{\sqrt{2}}(n(G))$ and $\omega(G) < \log_{\sqrt{2}}(n(G))$, even though almost all graphs have this property.

Properties of the random graph can lead to a fast algorithm that solves a

difficult problem on almost all inputs. For example, after stating two results about vertex degrees in random graphs, we show how to use properties of the degree sequence to design a fast algorithm to test isomorphism "almost always". In the literature of random graphs, ω_n denotes a function that is unbounded but grows arbitrarily slowly.

8.5.24. Theorem. (Erdős–Rényi [1966]) If $p = \omega_n \log n/n$ and $\epsilon > 0$ is fixed, then almost every G^p satisfies

$$(1-\epsilon)pn < \delta(G^p) \leq \Delta(G^p) \leq (1+\epsilon)pn. \qquad \blacksquare$$

Most vertices have degree near the average, but there is still considerable variation. Bollobás [1982] showed that for $p \leq 1/2$, the vertex of maximum degree is unique in almost every G^p if and only if $pn/\log n \to \infty$. When we complete evolution by returning to the realm of constant edge probability, more detailed results are known about the degree distribution. There will almost always be some vertices with isolated high degrees before the degrees begin to bunch up. Bollobás determined how many distinct degrees can be guaranteed.

8.5.25. Theorem. (Bollobás [1981b]) In Model A with p fixed and $t \in o(n/\log n)^{1/4}$, almost every G^p has different degrees for its t vertices of highest degree. If $t \notin o(n/\log n)^{1/4}$, then almost every G^p has $d_i = d_{i+1}$ for some $i < t$. $\blacksquare$

We apply this result to isomorphism testing. No polynomial-time algorithm is known for this problem, but Babai–Erdős–Selkow [1980] used the degree results for the random graph to develop a fast algorithm that almost always works. We define a set **H** that contains almost all graphs and show that isomorphism with a graph in **H** can be tested quickly.

The testing is done by a *canonical labeling algorithm*, which accepts and labels a graph in a canonical way if it belongs to **H**. The desired property is that when vertices are labeled as $v_1, \ldots, v_n$ in one graph and $w_1, \ldots, w_n$ in another, only the bijection mapping v_i to w_i is a possible isomorphism. Isomorphism can then be tested by comparing the adjacency matrices under this labeling.

8.5.26. Corollary. (Babai–Erdős–Selkow [1980]) There is a quadratic algorithm that tests isomorphism for almost all pairs of graphs.

Proof: Given a graph G on n vertices, presented by its adjacency matrix, compute and sort the vertex degrees, labeling the vertices in decreasing order of degree. Fix $r = \lfloor 3 \lg n \rfloor$. If $d(v_i) = d(v_{i+1})$ for any $i < r$, reject G. Using $p = 1/2$ in Theorem 8.5.25 implies that almost every graph successfully passes this test.

Let $U = \{v_1, \ldots, v_r\}$. With $r = \lfloor 3 \lg n \rfloor$, there are about n^3 distinct subsets of the vertices of U. Since only $n-r$ vertices remain outside U, there is a chance that they can be distinguished by their neighborhoods in U. The set **H** will be all the graphs reaching this stage for which this holds: the vertices of $V - U$ have distinct neighborhoods in U. To test this in $O(n^2)$ time and complete the