labeling, for each $x \in V - U$ encode $N(x) \cap U$ as a binary r-tuple. Evaluate these as binary integers, and sort them! These steps take $O(n \log n)$ time. Relabel the vertices v_{r+1} to v_n as $w_{r+1}, \ldots, w_n$ in decreasing order of these values. If two consecutive values are the same, reject G.

If G has passed this far, then G has no nontrivial automorphisms. A graph isomorphic to G has only one isomorphism to G, given by applying the canonical labeling algorithm to it. The last stage, if both graphs pass canonical labeling, is to compare the adjacency matrices with rows and columns indexed by the canonical labeling. The graphs are isomorphic if and only if the matrices are now identical. This comparison takes $O(n^2)$ time.

We must show that for almost every G^p, the adjacency vectors within a specified set of r vertices are distinct for the remaining vertices. If $p \leq 1/2$, then the probability for any pair x, y that x, y have the same adjacencies in U is bounded approximately by $(1-p)^r$. We say approximately because U is not chosen at random; choosing U as the set of vertices of highest degree impairs randomness, increasing the probability of a specified edge incident to these vertices. Nevertheless, it doesn't change by much, and the expected number of pairs of vertices outside U with identical adjacencies in U is bounded by $O(\binom{n-r}{2}(1-p)^r)$. Given our choice of r, we can bound the base 2 logarithm of this by $2 \lg n - 3 \lg b \lg n$, where $b = 1/(1-p) \geq 2$ (if $p \leq 1/2$). This tends to $-\infty$, so almost all graphs have distinct adjacency vectors in this set. ■

The probability of rejection in this labeling algorithm is bounded by $n^{-1/7}$ for sufficiently large n. Later improvements led to an algorithm running in time $O(n^2)$ with rejection probability c^{-n} (Babai–Kučera [1979]).

CONNECTIVITY, CLIQUES, AND COLORING

Studying the "typical behavior" of a random structure often involves studying probability distributions of its parameters. Here we consider connectivity, cliques, and colorings for random graphs.

For random graphs, naive algorithms may become good. For example, finding a maximum clique is NP-hard. If we know that almost every graph has clique number about $2 \lg n$, then we can test all vertex subsets up to size $3 \lg n$ for being cliques. If $\omega(G) < 3 \lg n$, then this computes $\omega(G)$, since every set of size $\omega(G) + 1$ is not a clique. If $\omega(G) \geq 3 \lg n$, then the algorithm fails to compute $\omega(G)$, but this rarely happens. There are too many subsets of size $2 \lg n$ for this to be a polynomial-time algorithm, but it's close, and it illustrates one way in which the properties of random graphs can be used algorithmically.

Some NP-hard problems are trivial for random graphs. Although $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$ for every simple graph G (Vizing [1964]), deciding between these values is NP-hard (Holyer [1981]). Vizing proved that $\chi'(G) = \Delta(G) + 1$ only when G has at least 3 vertices of maximum degree. Thus Erdős and Wilson [1977], who noted the uniqueness of the vertex of maximum degree when $p = 1/2$, also observed that $\chi'(G) = \Delta(G)$ for the random graph.

For sparse graphs and constant k, the thresholds for connectivity k and minimum degree k are the same. Does this also hold for constant edge probability? Theorem 8.5.18 can be generalized and strengthened to show that if $k \in o(n/\log n)$ and p is fixed, then almost every G^p has k common neighbors for every vertex pair and hence is k-connected (Exercise 33). Improving this requires other methods; Bollobás [1981b] showed for constant p that almost every G^p has connectivity equal to minimum degree.

What about clique number? For fixed k, Theorem 8.5.23 yields a probability threshold for the appearance of a k-clique, but for constant p the clique number grows with n. Determining the clique number is NP-complete, but for a random graph we can guess the correct value with high probability without looking at the graph! Amazingly, for fixed p almost every G^p has one of two possible values for the clique number (as a function of n), and for each $k \in \mathbb{N}$ there is a range of n where the clique number almost always equals k. The approach is to find bounds on $r(n)$ such that almost every G^p has an r-clique and almost none has an $r+1$-clique.

8.5.27. Theorem. (Matula [1972]) For fixed $p = 1/b$ and fixed $\epsilon > 0$, almost every G^p has clique number between $\lfloor d - \epsilon \rfloor$ and $\lfloor d + \epsilon \rfloor$, where $d = 2\log_b n - 2\log_b \log_b n + 1 + 2\log_b(e/2)$.

Proof: (sketch) If X_r is the number of r-cliques, then $E(X_r) = \binom{n}{r} p^{\binom{r}{2}}$. Since $r! \sim (r/e)^r \sqrt{2\pi r}$ (Stirling's approximation), also $E(X_r) \sim (2\pi r)^{-1/2} (enr^{-1} p^{(r-1)/2})^r$. If $r \to \infty$ and $(enr^{-1}p^{(r-1)/2}) \le 1$, then we expect that $E(X_r) \to 0$. To determine $r(n)$ such that this holds, take logarithms (base b) in the inequality and solve for r to find

$$r \ge 2\log_b n - 2\log_b r + 1 + 2\log_b e.$$

This is approximately equivalent to $r \ge d(n)$ as defined above. More precisely, if $r > d + \epsilon$, then almost every G^p has no clique of size r.

The lower bound comes from careful application of the second moment method, as in Theorem 8.5.23, but the dependence of r on n makes the analysis more difficult. The expectation of X_r^2 sums the probability of common occurrence for all ordered pairs of r-cliques. This probability depends only on the number of common vertices, so

$$E(X_r^2) = \binom{n}{r} \sum_{k=0}^{r} \binom{r}{k} \binom{n-r}{r-k} p^{2\binom{r}{2} - \binom{k}{2}}.$$

We want to show that the term for $k = 0$ (disjoint cliques) dominates. Let $E(X_r^2)/E(X_r)^2 = \alpha_n + \beta_n$, where $\alpha_n = \binom{n}{r}^{-1} \binom{n-r}{r}$ and $\beta_n = \binom{n}{r}^{-1} \sum_{k=1}^{r} \binom{r}{k} \binom{n-r}{r-k} b^{\binom{k}{2}}$. We seek $\alpha_n \sim 1$ and $\beta_n \to 0$. When $r \sim 2\log_b n$, an asymptotic formula for $\binom{a}{k}/\binom{b}{k}$ leads to $\alpha_n \sim e^{-r^2/(n-r)} \to 1$. The discussion of β_n is more difficult; see Palmer [1985, p75-80]. ■

Our study of graph parameters can be applied to measure the strength of conditions for Hamiltonian cycles (Palmer [1985, p81-85]). A theorem proves nothing if its hypotheses are never satisfied; this suggests saying that such a theorem has strength 0. A theorem is strong if the conclusion is satisfied only when the hypothesis is satisfied; then the hypotheses cannot be weakened. Define the **strength** of a theorem to be the probability that its hypotheses are satisfied divided by the probability that its conclusion is satisfied.

Consider sufficient conditions for Hamiltonian cycles. Since $p = \log n/n$ is a threshold for a Hamiltonian cycle, almost every G^p is Hamiltonian when p is fixed. Dirac [1952b] showed that G is Hamiltonian when every vertex degree is at least $n/2$ (Theorem 7.2.8). When $p > 1/2$, this condition holds for almost every G^p; when $p \leq 1/2$, it almost never holds. Hence the asymptotic strength of Dirac's Theorem is 0 when p is a constant at most $1/2$. The same fate befalls the other degree conditions of Section 7.2.

Meanwhile, Chvátal and Erdős [1972] proved that G is Hamiltonian whenever its connectivity exceeds its independence number (Theorem 7.2.19). Our thresholds for these parameters imply that this result is strong for every constant $p > 0$. We know that $\alpha(G^p) < 2(1+\epsilon)\log_b n$ almost always, and we know that $\kappa(G^p) \geq k$ almost always (when $k = o(n/\log n)$). Hence $\kappa > \alpha$ for almost every G^p, and the asymptotic strength of the theorem is 1.

Finally, we consider chromatic number for constant p. Since $1 - p$ is also constant, we can apply the results on clique number: Almost every G^p has no stable set with more than $(1+o(1))2\log_b n$ vertices, where $b = 1/(1-p)$. Hence $\chi(G^p) \geq (1/2 + o(1))n/\log_b n$ almost always. Achieving this bound requires finding many disjoint stable sets with near-maximum sizes. For a decade, the best result was an algorithmic guarantee of a coloring with at most twice the number of colors in the lower bound.

Bollobás [1988] proved that the lower bound is achievable, by using another probabilistic technique that guarantees finding enough large stable sets. He proved that, in almost every G^p, *every* set having at least $n/(\log_b n)^2$ vertices contains a clique of order at least $2\log_b n - 5\log_b \log_b n$. This allows stable sets of near-maximum size to be extracted until too few vertices remain to cause trouble; the remainder can be given distinct colors.

Before developing Bollobás' approach, we present the earlier result for its algorithmic interest; the greedy algorithm uses at most $(1+\epsilon)n/\log_b n$ colors on almost every G^p. Thus it "almost always works" as an approximation algorithm in the same sense that our earlier isomorphism algorithm almost always works. Garey and Johnson [1976] showed there is no fast algorithm that uses at most twice the optimum number of colors on *every* graph unless P = NP. Bollobás' proof does not yield a fast algorithm for coloring almost every graph with an asymptotically optimal number of colors; it is an existence proof only.

8.5.28. Theorem. (Grimmett–McDiarmid [1975]) Given edge probability p, let $b = 1/(1-p)$. For constant p and constant $\epsilon > 0$, almost every G^p satisfies

$$(1/2 - \epsilon)n/\log_b n \leq \chi(G^p) \leq (1+\epsilon)n/\log_b n.$$

Proof: The lower bound follows using stable sets as suggested above. For the upper bound, we show that the greedy coloring of $v_1, \dots, v_n$ in order uses at most $f(n) = (1+\epsilon)n/\log_b n$ colors on almost every G^p (for simplicity, choose ϵ so that $f(n)$ is an integer). Within the set of n-vertex graphs using more colors, let $\mathbf{B}_m$ be the set such that v_m is the first vertex to use color $f_n + 1$. We prove that $\sum_{m=1}^{n} P(\mathbf{B}_m) \to 0$ as $n \to \infty$.

Given G, let $G_m = G[\{v_1, \dots, v_{m-1}\}]$. Before color $f_n + 1$ is used, color f_n must be used, so for each $G \in \mathbf{B}_m$ the greedy coloring of G_m uses f_n colors. Let k_i be the number of times color i appears in this coloring. To require use of color $f_n + 1$, v_{m+1} must have at least one neighbor of each color $1, \dots; f_n$. Given the numbers $\{k_i\}$, the probability of this is $\prod_{i=1}^{f(n)}[1-(1-p)^{k_i}]$.

Bollobás and Erdős [1976] simplified the subsequent computations involving this bound by observing that the bound is maximized when the k_i's are all equal (Exercise 8.3.37). Thus

$$\prod_{i=1}^{f(n)}[1-(1-p)^{k_i}] \le [1-(1-p)^{(m-1)/f}]^f < [1-(1-p)^{n/f}]^f.$$

Given G_m, we have $b_n = [1-(1-p)^{n/f(n)}]^{f(n)}$ as a bound on the probability that the full graph G belongs to $\mathbf{B}_m$. Since this holds for each G_m, we conclude that $P(\mathbf{B}_m) < b_n$. This holds for all m, so $\sum_{m=1}^{n} P(\mathbf{B}_m) < nb_n$.

Using $(1-p)^{-x} < e^{-x}$, we obtain $nb_n < ne^{-f(1-p)^{n/f}}$. Substituting $f_n = cn/\log_b n$ yields $(1-p)^{n/f} = n^{-1/c}$. The logarithm of the bound becomes $\log n - cn^{1-1/c}/\log_b n$. This tends to $-\infty$ for $c > 1$, so the probability that the greedy algorithm uses more than $f(n)$ colors is bounded by a function tending to 0. ■

The order of growth of $\chi(G)$ sheds light on other famous problems in graph theory. Hajós conjectured that every r-chromatic graph contains a subdivision of K_r (see Remark 5.2.21). This was disproved by Catlin [1979] (Exercise 5.2.40). Erdős and Fajtlowicz [1981] observed that the chromatic number of G^p almost always grows like $\Theta(n/\log n)$). On the other hand, the largest r such that G^p contains a subdivision of K_r grows like $\Theta(\sqrt{n})$). Thus the chromatic number is almost always much larger, and Hajós' Conjecture is almost always very false.

In contrast, almost every G^p has a subgraph contractible to K_r when $r \in \Theta(n/\sqrt{\log n})$. Thus almost every graph satisfies the weaker conjecture of Hadwiger (Remark 5.2.21), which states that every r-chromatic graph has a subgraph contractible to K_r.

MARTINGALES

Advanced techniques in probability lead to elegant results on combinatorial structures without the drudgery involved in second moment and higher moment computations. The theory aims to develop paradigms that can be applied without repeating computational details.

Some of these methods employ lists of related random variables. The resulting stochastic process displays more consistent and predictable global behavior than the individual random variables do.

In the classical random walk on a line, at each step there is probability p of moving one unit to the left, probability p of moving one unit to the right, and probability $1-2p$ of not moving. No matter what the earlier history of the walk has been, the expected position after t steps equals the actual position after $t-1$ steps. This is the defining property of a martingale.

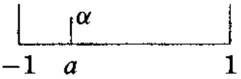
8.5.29. Definition. A **martingale** is a list of random variables $X_0, \dots, X_n$ such that the expectation of X_i, given the values of $X_0, \dots, X_{i-1}$, equals X_{i-1}.

The expected position of the random walk after n steps is at the origin. Less obvious is that the walk is highly unlikely to be very far from the origin, as a function of n. We shall see that this follows from its inability to move more than one unit in each step.

Martingales can make it easy to show that a random variable is highly concentrated around its expected value. When the technique applies, it makes the detailed computation in the Second Moment Method unnecessary. The hard work is accomplished by Azuma's Inequality, also called the Martingale Tail Inequality. This inequality states that if successive random variables in a martingale always differ by at most 1, then the probability that $X_n - X_0$ exceeds $\lambda\sqrt{n}$ is bounded by $e^{-\lambda^2/2}$. We first prove two lemmas. These statements hold for continuous random variables, but again we consider only discrete variables.

8.5.30. Lemma. Let Y be a random variable such that $E(Y) = 0$ and $|Y| \le 1$. If f is a convex function on $[-1, 1]$, then $E(f(Y)) \le \frac{1}{2}[f(-1) + f(1)]$. In particular, $E(e^{tY}) \le \frac{1}{2}(e^t + e^{-t})$ for all $t > 0$.

Proof: When Y takes only the values ± 1, each with probability .5, we have $E(f(Y)) = \frac{1}{2}[f(-1)+f(1)]$. For other distributions, pushing probability "out to the edges" increases $E(f(Y))$. For discrete variables, we can use induction on the number of values with nonzero probability. Convexity implies that $f(a) \le \frac{1-a}{2} f(-1) + \frac{a+1}{2} f(1)$. If $P(Y = a) = \alpha$, then we can decrease the probability at a to 0, increase $P(Y = -1)$ by $\alpha\frac{1-a}{2}$ and increase $P(Y = 1)$ by $\alpha\frac{a+1}{2}$ to obtain a new variable Y' with the same expectation. By the convexity inequality and the induction hypothesis, $E(f(Y)) \le E(f(Y')) \le \frac{1}{2}[f(-1) + f(1)]$. ■



8.5.31. Definition. For events A and B, the **conditional probability** of A *given* B is obtained by treating the event B as the full probability space, which means normalizing by $P(B)$. Thus we define $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$.

When Y, X are random variables, we write $Y|X$ for "Y *given* X". This defines a random variables for each value of X; we treat X as a constant i and normalize the resulting distribution for Y by $P(X = i)$.

For Azuma's Inequality, we use expectation of conditional variables. For each i, we compute the expected value of Y when restricted to the sample points where $X = i$. The expectation $E(E(Y|X))$ is the expectation of $E(Y|X = i)$ over the choices for i, which occur with probability $P(X = i)$. The result is an expectation over the entire sample space. It removes the effect of conditioning, and we obtain $E(E(Y|X)) = E(Y)$.

8.5.32. Lemma. $E(E(Y|X)) = E(Y)$.

Proof: Let $p_{i,j} = P(X = i \text{ and } Y = j)$. Since $E(Y|X = i) = \frac{\Sigma_j j p_{i,j}}{P(X=i)}$,

$$E(E(Y|X)) = \sum_i E(Y|X = i)P(X = i) = \sum_i \sum_j j p_{i,j} = E(Y). \qquad \blacksquare$$

8.5.33. Theorem. (Azuma's Inequality) If $X_0, \ldots, X_n$ is a martingale with $|X_i - X_{i-1}| \le 1$, then $P(X_n - X_0 \ge \lambda\sqrt{n}) \le e^{-\lambda^2/2}$.

Proof: By translation, we may assume that $X_0 = 0$. For $t > 0$, we have $X_n \ge \lambda\sqrt{n}$ if and only if $e^{tX_n} \ge e^{t\lambda\sqrt{n}}$, and hence $P(X_n \ge \lambda\sqrt{n}) = P(e^{tX_n} \ge e^{t\lambda\sqrt{n}})$. Applied to e^{tX_n}, Markov's Inequality yields $P(e^{tX_n} \ge e^{t\lambda\sqrt{n}}) \le E(e^{tX_n})/e^{\lambda t\sqrt{n}}$. This bound holds for each $t > 0$, and later we will choose t to minimize the bound.

First we prove by induction on n that $E(e^{tX_n}) \le \frac{1}{2}(e^t + e^{-t})$. We introduce X_{n-1} to condition on it. Lemma 8.5.32 yields

$$E(e^{tX_n}) = E(e^{tX_{n-1}}e^{t(X_n - X_{n-1})}) = E(E(e^{tX_{n-1}}e^{t(X_n - X_{n-1})}|X_{n-1})).$$

When we condition on X_{n-1}, the value of X_{n-1} is constant for the inner expectation. Hence we can remove $e^{tX_{n-1}}$ from the inner expectation to obtain $E(e^{tX_n}) = E(e^{tX_{n-1}}E(e^{tY}|X_{n-1}))$, where $Y = X_n - X_{n-1}$. Because $\{X_n\}$ is a martingale, $E(Y) = 0$, and by hypothesis $|Y| \le 1$. Hence Lemma 8.5.30 applies, yielding $E(e^{tY}|X_{n-1}) \le \frac{1}{2}(e^t + e^{-t})$. This itself is now a constant, yielding $E(e^{tX_n}) = \frac{1}{2}(e^t + e^{-t})E(e^{tX_{n-1}})$. The induction hypothesis completes the proof.

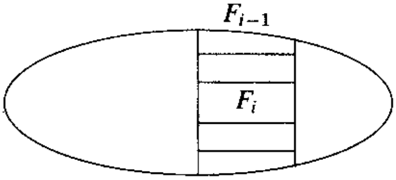
We weaken the bound to a more useful form by observing that $\frac{1}{2}(e^t + e^{-t}) \le e^{t^2/2}$. This holds because the left side is $\sum t^{2k}/(2k)!$ and the right side is $\sum t^{2k}/(2^k k!)$. Hence our original probability is bounded by $e^{nt^2/2 - \lambda t\sqrt{n}}$ for each $t > 0$. We obtain the best bound by minimizing over t. The exponent is quadratic; we minimize it by choosing t to solve $tn - \lambda\sqrt{n} = 0$, or $t = \lambda/\sqrt{n}$. The resulting bound is $e^{-\lambda^2/2}$. $\blacksquare$

Azuma's Inequality is one-sided; it bounds the probability that X_n is much larger than X_0. Since the conditions are symmetric in sign, applying the inequality to $\{-X_i\}$ yields the same inequality for the other tail, in which X_n is much smaller than X_0.

8.5.34. Example. *The pragmatic gambler.* A gambler can bet up to n times, where n is fixed. Each time he bets, he wins or loses 1 with equal probability.

His goal is winning $\lambda\sqrt{n}$, so he stops if he reaches that value. Letting X_i be his winnings after i games, we have $X_i = X_{i-1}$ if $X_{i-1} \ge \lambda\sqrt{n}$, and otherwise $X_i = X_{i-1} \pm 1$, each with probability .5. Hence $\{X_i\}$ is a martingale that changes by at most 1 at each step, and Azuma's Inequality applies. The probability that the gambler will earn $\lambda\sqrt{n}$ is bounded by $e^{-\lambda^2/2}$. If $\lambda = 1$, then there may be a reasonable chance of success, but if $\lambda = 10$, then there is little hope. ■

In combinatorial applications, we consider a special type of martingale. We have an underlying probability space, and X_0 is the expectation of a random variable X. The variable X_n is the value of X at one sample point. We define a martingale $X_0, \dots, X_n$ that describes a gradual process of learning more about the final value $X_n = X$.



8.5.35. Lemma. Let X be a random variable defined on a probability space. Let $F_0 \supseteq F_1 \supseteq \cdots \supseteq F_n$ be a chain of subsets of the space, where F_0 is the full space, F_n is a single outcome, and F_i is a random variable that is a block in a partition of F_{i-1}. The probability of choosing F_i within F_{i-1} is proportional to its probability in the underlying space. If $X_i = E(X|F_i)$, then the list $X_0, \dots, X_n$ is a martingale.

Proof: We must prove that $E(X_i|X_0, \dots, X_{i-1}) = X_{i-1}$. In a particular instance of the process, the list of values is the outcome of a particular sequence of restrictions. Each sequence of restrictions that generates the given values $X_0, \dots, X_{i-1}$ reaches some F_{i-1} such that $E(X|F_{i-1})$ has the given value of X_{i-1}. For every such F_{i-1}, we can take the expectation of X_i over the possible values of F_i. In each case, we obtain X_{i-1}, so the desired formula holds regardless of which F_{i-1} generated the list $X_0, \dots, X_{i-1}$.

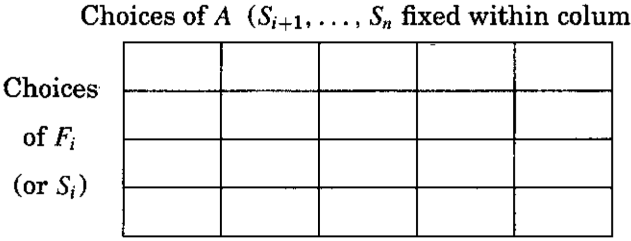
We thus condition on a fixed choice of F_{i-1} to compute $E(X_i|X_0, \dots, X_{i-1})$. Within F_{i-1}, Lemma 8.5.32 yields $E(X_i) = E(E(X|F_i)) = E(X)$. This is the expectation within the event F_{i-1} (treated as a probability space), so all of these expresssions are conditioncd on F_{i-1}, and the final expression is actually $E(X|F_{i-1}) = X_{i-1}$. ■

Such martingales, which we call **restriction martingales**, arise when we gradually discover a randomly generated object. Here F_i is the subset of the probability space where the object is confined after i steps (F for "inFormation"). In coin-flipping, the sample points are list of length n, and F_i may be the knowledge of the first i values. In random graphs, F_i may be the subgraph induced by the vertices $\{v_1, \dots, v_i\}$, or F_i may be the knowledge of which among the first i edges are present.

To apply Azuma's Inequality, we need to bound $|X_i - X_{i-1}|$. The knowledge of which edges arise incident to a fixed vertex v_i can change the chromatic number by at most 1, since $\chi(G - v_i)$ equals $\chi(G)$ or $\chi(G) - 1$. From this we can conclude that $|X_i - X_{i-1}| \leq 1$ in the restriction martingale defined by revealing vertices one by one.

8.5.36. Lemma. Consider a random structure specified by independent steps $S_1, \dots, S_n$. Let F_i be the knowledge of $S_1, \dots, S_i$, and let $X_0, \dots, X_n$ be the corresponding restriction martingale for a random variable X. Let A be the knowledge of S_j for all $j \neq i$, with S_i unknown. If for each such A the values of X on points in A differ by at most 1, then $|X_i - X_{i-1}| \leq 1$ for all i (and hence $P(X - E(X)) \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}$).

Proof: Consider a particular instance of F_{i-1}, with $X_{i-1} = E(X|F_{i-1})$ given. We arrange the points of F_{i-1} in the cells of a grid. For all these points, the outcomes of $S_1, \dots, S_{i-1}$ are the same. Each row is a choice for F_i: a block in the partition of F_{i-1}. Each column is an A in which $S_{i+1}, \dots, S_n$ are fixed and only S_i varies. By hypothesis, in each column the maximum and minimum values of X differ by at most 1. Let m_s, M_s be the minimum and maximum of X in column s.



Because S_i and $S_{i+1}, \dots, S_n$ are specified independently, the probability of the outcome in row r and column s is $q_r p_s$, where q_r is the probability that S_i yields this row and p_s is the probability that $S_{i+1}, \dots, S_n$ yields this column. The computation of X_i is the expectation across a single row:

$$\sum m_s p_s \leq E(X|F_i) \leq \sum M_s p_s \leq 1 + \sum m_s p_s.$$

Since these upper and lower bounds are independent of the row index, taking the expectation over the entire grid to compute X_{i-1} yields the same inequalities. Hence X_{i-1} and X_i are confined to a single interval of length 1 and differ by at most 1. Therefore, Azuma's Inequality applies. ∎

When the conditions of Lemma 8.5.36 hold, we conclude immediately that the value of X is highly concentrated around its mean.

8.5.37. Example. *Chromatic number of random graphs.* Fix n, and consider Model A with edge probability p. Suppose we reveal the random n-vertex graph one vertex at a time. At stage i, we learn the edges from v_i to the previous vertices; this is S_i, and Model A specifies the outcomes of the S_i's independently. The event A in which all but S_i are specified is the subgraph $G - v_i$ of the random graph G plus the knowledge of edges from v_i to *later* vertices. Since $\chi(G - v_i) \le \chi(G) \le \chi(G - v_i) + 1$, the value of X differs by at most one over all possibilities in A. The hypotheses of Lemma 8.5.36 hold. Using both tails, we conclude that

$$P(|\chi(G) - E(\chi(G))|) \ \ge\ \lambda\sqrt{n}) \ \le\ 2e^{-\lambda^2/2}. \qquad \blacksquare$$

The result of Example 8.5.37 says nothing about the value of $E(\chi(G))$. To approximate this we again use Azuma's Inequality. With constant edge probability p, we know that the clique number of G^p is almost always within 1 of $d = 2\log_b n - 2\log_b \log_b n + 1 + 2\log_b(e/2)$, where $b = 1/p$. The same result holds for stable sets using the base $c = 1/(1-p)$ for the logarithm. To show that the chromatic number of G^p is close to $n/(2\log_c n)$, Bollobás showed that it is possible to extract stable sets of almost the maximum size until the number of vertices remaining is too small to matter.

8.5.38. Theorem. (Bollobás [1988]) For almost every G^p with constant $p = 1 - 1/c$, every induced subgraph of order at least $m = \lceil n/\log_c^2 n\rceil$ has a stable set of size at least $r = 2\log_c n - 5\log_c \log_c n$.

Proof: (sketch) We use r-**staset**, by analogy with r-clique, to mean a stable set of size r. Let S be a set of m vertices. We bound the probability that S has no r-staset by $e^{-dm^{1+\epsilon}}$ for some d, ϵ. This in turn bounds the probability that there exists an m-set with no r-staset by $\binom{n}{m}e^{-dm^{1+\epsilon}} < 2^n e^{-dm^{1+\epsilon}}$. Since $n = m^{1+o(1)}$, this bound goes to 0, and the first moment method implies that almost every G^p has no bad m-set.

It suffices to study the subgraph G induced by $[m]$. Let X be the maximum number of pairwise pair-disjoint r-stasets in this subgraph, where *pair-disjoint* means they share at most one vertex. We will show that $X \ge 1$ almost always. To do this, it suffices to show that (1) X is highly concentrated around its mean, and (2) $E(X)$ is bigger than something large (and growing).

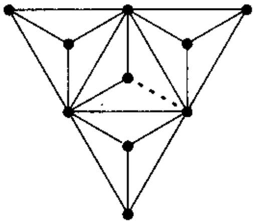
We invoke Azuma's Inequality for (1). Consider the restriction martingale for X that results from revealing G *one edge-slot at a time*. At each step, we learn whether one additional pair of vertices induces an edge. We have $X_0 = E(X)$ and $X_{\binom{m}{2}} = X$. The status of one edge slot changes the value of X by at most 1, so Lemma 8.5.36 applies, and $P(X - E(X)) \le -\lambda\binom{m}{2}^{1/2}) \le e^{-\lambda^2/2}$. With $\lambda = E(X)/\binom{m}{2}^{1/2}$, we have

$$P(X = 0) \ =\ P(X - E(X) \le -E(X)) \ \le\ e^{-E(X)^2/(m^2 - m)}.$$

Hence it suffices to show that $E(X)/m \to \infty$.

To prove this, we consider another random variable $\hat{X}$, the number of r-stasets in G that have no pair in common with *any* other r-staset. Such a collection forms a pairwise pair-disjoint collection of r-stasets, so $X \ge \hat{X}$. We introduced X because the restriction martingale for $\hat{X}$ does not satisfy

$|\hat{X}_i - \hat{X}_{i-1}| \le 1$. In the drawing of $\overline{G}$ in the figure below, for example, we have $r = 4$ and seek 4-cliques; if the last (dotted) edge is present in $\overline{G}$ (absent in G), then $\hat{X} = 0$, but if it is absent from $\overline{G}$ (present in G), then $\hat{X} = 3$.



It is easier to compute $E(\hat{X})$ than $E(X)$. Expressing $\hat{X}$ as the sum of $\binom{m}{r}$ indicator variables, we obtain $E(\hat{X})$ as $\binom{m}{r}$ times the‹probability that $[r]$ induces an r-staset that is pair-disjoint from all others. This is $(1-p)^{\binom{r}{2}}$ times the conditional probability that $[r]$ does not conflict with other r-stasets, given the event Z that $[r]$ is in fact independent. Let Y be the number of other r-stasets overlapping $[r]$ in at least two elements. By Markov's Inequality, $E(Y|Z) \to 0$ implies $P(Y = 0|Z) \to 1$. Since each set counted shares at least two vertices with $[r]$, we have

$$E(Y|Z) = \textstyle\sum_{i \ge 2, r-1} \binom{r}{i}\binom{m-r}{r-i}(1-p)^{\binom{r}{2}-\binom{i}{2}}.$$

As $m \to \infty$, this tends to 0; this follows from the expression for r in terms of m. Hence $E(\hat{X})$ is asymptotic to $\binom{m}{r}(1-p)^{\binom{r}{2}}$. The expression for r in terms of m yields $E(\hat{X}) \in \Omega(m^{5/3})$. Thus $E(X)/m \to \infty$, which completes the proof. ■

8.5.39. Corollary. (Bollobás [1988]) For constant edge probability $p = 1 - 1/c$, almost every G^p satisfies

$$(1+\epsilon)n/(2\log_c n) \le \chi(G^p) \le (1+\epsilon')n/(2\log_c n),$$

where $\epsilon = \log_c \log_c n / \log_c n$ and $\epsilon' = 5\log_c \log_c n / \log_c n$.

Proof: The lower bound holds because almost every G^p has no stable set larger than $2\log_c n - 2\log_c \log_c n$. The upper bound follows from Theorem 8.5.38, because we can almost always select stable sets of size $2\log_c n - 5\log_c \log_c n$ until we have only $n/\lg_c^2 n$ vertices left. Since $n/\lg_c^2 n \in o(n/\log_c n)$, we can complete the coloring by using distinct new colors on the remaining vertices. ■

EXERCISES

8.5.1. (−) *Expectation.*

a) Compute the expected number of fixed points in a random permutation of $[n]$.

b) Determine the expected number of vertices of degree k in a random n-vertex graph with edge probability p.

8.5.2. (−) Prove that $1 - p < e^{-p}$ for $p > 0$.

8.5.3. (−) Determine the expected number of monochromatic triangles in a random 2-coloring of $E(K_6)$.

8.5.4. (−) Prove that some 2-coloring of the edges of $K_{m,n}$ has at least $\binom{m}{r}\binom{n}{s}2^{1-rs}$ monochromatic copies of $K_{r,s}$.

8.5.5. (−) The statement "$f(G_n) \le (1+\epsilon)n$" means that for all $\epsilon > 0$, the inequality holds for sufficiently large n. The statement "$f(G_n) \le n + o(n)$" means that $f(G_n)/n \to 1$ as $n \to \infty$. Prove that these two statements are equivalent.

8.5.6. Compute explicitly the probability that the Hamiltonian closure of a random graph with vertex set [5] is complete.

8.5.7. Let G be a graph with p vertices, q edges, and automorphism group of size s. Let $n = (sk^{q-1})^{1/p}$. Prove that some k-coloring of $E(K_n)$ has no monochromatic copy of G. (Chvátal–Harary [1973])

8.5.8. (!) a) Use a random partition of the vertices to prove that every graph has a bipartite subgraph with at least half its edges.

b) Use equipartitions of the vertices to improve part (a): if G has m edges and n vertices, then G has a bipartite subgraph with at least $m\frac{\lceil n/2\rceil}{2\lceil n/2\rceil - 1}$ edges.

8.5.9. An army of computers is configured as a complete k-ary tree with leaves at distance l from the root. At a fixed time, each node is working with probability p, independently of other nodes. When a node is not working, the entire subtree below it is inaccessible. What is the expected number of nodes accessible from the root?

8.5.10. Let G be a matching of size n. Select a set of k vertices at random. Compute the expected number of edges induced by the selected vertices.

8.5.11. Consider a drawing in the plane of a simple graph G with n vertices and m edges, where $m \ge 4n$. Let H be a random induced subdrawing, generated by letting each vertex be retained with probability p, independently. Let Y be the number of edge crossings in H. Let $X = Y - [e(H) - (3n(H) - 6)]$. Use expectations to prove that $3n + p^3\nu(G) - pm > 0$, and conclude that $\nu(G) \ge m^3/[64n^2]$, where $\nu(G)$ is the minimum number of crossings in a drawing of G. (Comment: This is an alternative proof of Theorem 6.3.16.)

8.5.12. Given a random permutation of the vertices of a simple graph G, orient each edge toward the vertex with higher index in the permutation. Compute the expected number of sink vertices (outdegree 0) in the resulting orientation. In terms of $n(G)$, determine the minimum and maximum values of this expectation. Prove that the probably of having only one sink is at most $e(G)/\binom{n(G)}{2}$. (Jeurissen [1997])

8.5.13. (!) A **hypergraph** consists of a collection of vertices and a collection of edges; if the vertex set is V, then the edges are subsets of V. The **chromatic number** $\chi(H)$ of a hypergraph H is the minimum number of colors needed to label the vertices so that no edge is monochromatic. A hypergraph is k-**uniform** if its edges all have size k.

a) Prove that every k-uniform hypergraph with fewer than 2^{k-1} edges is 2-colorable. (Erdős [1963])

b) Use part (a) to prove that if each vertex of an n-vertex bipartite graph has a list of more than $1 + \lg n$ usable colors, then a proper coloring can be chosen from the lists.

8.5.14. (!) Use the deletion method to prove that a graph with n vertices and average degree $d \ge 1$ has an independent set with at least $n/(2d)$ vertices. (Hint: Choose a random subset by including each vertex independently with a probability p to be chosen later. Compute the expected number of edges induced.)

8.5.15. The maximum size of an n-vertex graph not containing H is $ex(n; H)$. Use the deletion method to prove that $ex(n; C_k) \in \Omega(n^{1+1/(k-1)})$. (Comment: One can also show that $ex(n; C_k) \in O(n^{1+2/k})$ by considering the average degree.) (Bondy–Simonovits)

8.5.16. (!) For $n \in \mathbb{N}$, prove that $R(k,k) > n - \binom{n}{k}2^{1-\binom{k}{2}}$. Use this to conclude that $R(k,k) > (1/e)(1-o(1))k2^{k/2}$.

8.5.17. For natural numbers n, t, let $m = n - \binom{n}{t}^2 2^{1-t^2}$. Prove that there is a 2-coloring of the edges of $K_{m,m}$ with no monochromatic copy of $K_{t,t}$.

8.5.18. (+) *Off-diagonal Ramsey numbers.* Suppose that $0 < p < 1$.

a) Prove that if $\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}} < 1$, then $R(k,l) > n$.

b) Prove that $R(k,l) > n - \binom{n}{k}p^{\binom{k}{2}} - \binom{n}{l}(1-p)^{\binom{l}{2}}$ for all $n \in \mathbb{N}$.

c) Choose n and p in part (b) to prove that $R(3,k) > k^{3/2-o(1)}$. What lower bound on $R(3,k)$ can be obtained from part (a)? (Spencer [1977])

8.5.19. Let H be a graph. For constant p, prove that almost every G^p contains H as an induced subgraph.

8.5.20. a) Fix k, s, t, p. Prove that almost every G^p has the following property: for every choice of disjoint vertex sets S, T of sizes s, t, there are at least k vertices that are adjacent to every vertex of S and no vertex of T. (Blass–Harary [1979])

b) Conclude that almost every G^p is k-connected.

c) Apply the same argument to random tournaments: almost every one has the property that for every choice of disjoint vertex sets S, T of sizes s, t, there are at least k vertices with edges to every vertex of S and from every vertex of T.

8.5.21. A random labeled tournament is generated by orienting each edge v_iv_j as $v_i \to v_j$ or $v_j \to v_i$ independently with probability $1/2$.

a) Prove that almost every tournament is strongly connected.

b) In a tournament, a "king" is a vertex such that every other vertex can be reached from it by a path of length at most 2. It is known that every tournament contains a king. Is it true that in almost every tournament every vertex is a king? (Palmer [1985])

8.5.22. Find a threshold probability function for the property that at least half the possible edges of a graph are present. How sharp is the threshold?

8.5.23. For $p = 1/n$ and fixed $\epsilon > 0$, show that almost every G^p has no component with more than $(1+\epsilon)n/2$ vertices. (Hint: Instead of trying to bound the probability directly, show that it is bounded by the probability of another event, which tends to 0.)

8.5.24. Determine the smallest connected simple graph that is not balanced.

8.5.25. Extend the second moment argument of Theorem 8.5.23 to prove that $n^{-1/\rho(H)}$ is a threshold function for the appearance of H as a subgraph of G^p, where $\rho(G) = \max_{G \subseteq H} e(G)/n(G)$. (Bollobás [1981a], Ruciński–Vince [1985])

8.5.26. Let Q be the following graph property: for every choice of disjoint vertex sets S, T of size $c\lg n$, there is an edge with endpoints in S and T. Prove that almost every graph has property Q if $c > 2$. (Comment: This implies that the random graph has bandwidth at least $n - 2\log n$.)

8.5.27. Prove that if $k = \lg n - (2+\epsilon)\lg\lg n$, then almost every n-vertex tournament has the property that every set of k vertices has a common successor.

8.5.28. A tournament is **transitive** if it has a vertex ordering $u_1, \ldots, u_n$ such that $u_i \to u_j$ if and only if $i < j$. Prove that every tournament has a transitive subtournament with $\lg n$ vertices, and almost every tournament has no transitive subtournament with more than $2\lg n + c$ vertices if c is a constant greater than 1.

8.5.29. (!) *The Coupon Collector.*

a) Consider repetitions of an experiment with independent success probability p. Prove that the expected number of the trial on which the first success occurs is $1/p$.

b) Every box of a certain type of candy contains one of n prizes, each with probability $1/n$. Receiving the grand prize requiries obtaining each of these prizes at least once. Prove that the expected number of the box on which the last prize is obtained is $n\sum_{i=1}^{n} 1/i$.

c) Prove that $m(n) = n\ln n + (k-1)n\ln\ln n$ is a threshold function for the number of boxes needed to obtain at least k copies of each prize. (Hint: Prove that when $p = o(1)$ and k is constant, the probability of at most k successes in m trials with success probability p is asymptotic to the probability of exactly k successes.)

8.5.30. Prove that the length of the longest run in a list of n random heads and tails is $(1+o(1))\lg n$. In other words, for $\epsilon > 0$, almost no list has at least $(1+\epsilon)\lg n$ consecutive identical flips, and almost every list has at least $(1-\epsilon)\lg n$ consecutive identical flips.

8.5.31. With $p = (1-\epsilon)\log n/n$, find a large m such that almost every graph has at least m isolated vertices. What $m(n)$ results from Chebyshev's Inequality?

8.5.32. Given a graph G, say that a k-set S is *bad* if G has no vertex v such that $S \subseteq N(v)$. For fixed p, how large can k be so that almost every G^p has no bad k-set? How slowly can k grow so that almost every G^p has a bad k-set?

8.5.33. By examining common neighbors, prove that if p is fixed and $k = o(n/\log n)$, then almost every G^p is k-connected.

8.5.34. (!) With $p = (1-\epsilon)\log n/n$, how large can m be such that almost every graph has at least m isolated vertices? (Hint: Use Chebyshev's Inequality.)

8.5.35. A t-**interval** is a subset of $\mathbb{R}$ that is the union of at most t intervals. The **interval number** of a graph G is the minimum t such that G is an intersection graph of t-intervals (each vertex is assigned a set that is the union of at most t intervals). Prove that almost all graphs (edge probability 1/2) have interval number at least $(1-o(1))n/(4\lg n)$. (Hint: Compare the number of representations with the number of simple graphs. Comment: Scheinerman [1990] showed that almost all graphs have interval number $(1+o(1))n/(2\lg n)$.) (Erdős–West [1985])

8.5.36. (!) *Threshold for perfect matching in a random bipartite graph.* Let G be a random subgraph of $K_{n,n}$ with partite sets A, B, generated by independent edge probability $p = (1+\epsilon)\ln n/n$, where ϵ is a nonzero constant. Call S a *violated set* if $|N(S)| < |S|$.

a) Prove that if $\epsilon < 0$, then almost every G has no perfect matching.

b) Let S be a minimal violated set. Prove that $|N(S)| = |S| - 1$ and that $G[S \cup N(S)]$ is connected.

c) Suppose that G has no perfect matching. Prove that A or B contains a violated set with at most $\lceil n/2\rceil$ elements.

d) For $r, s \geq 1$, the number of spanning trees of $K_{r,s}$ is $r^{s-1}s^{r-1}$. Use this, part (b), part (c), and Markov's Inequality to prove that if $\epsilon > 0$, then G almost surely has a perfect matching. (Hint: A summation in the bound on the expected number of minimal violated sets can be bounded by a geometric series.)

8.5.37. Suppose that $0 < p < 1$ and that $k_1, \dots, k_r$ are nonnegative integers summing to m. Prove that $\prod_{i=1}^{r}[1 - (1-p)^{k_i}] \le [1 - (1-p)^{m/r}]^r$.

8.5.38. *Tail inequality for binomial distribution.* Let $X = \sum X_i'$, where each X_i' is an indicator variable with success probability $P(X_i' = 1) = .5$, so $E(X) = n/2$. Applying Markov's Inequality to the random variable $Z = (X - E(X))^2$ yields $P(|Z| \ge t) \le Var(X)/t^2$. Setting $t = \alpha\sqrt{n}$ yields a bound on the tail probability: $P(|X - np| \ge \alpha\sqrt{n}) \le 1/(2\alpha^2)$. Use Azuma's Inequality to prove the stronger bound that $P(|X - np| > \alpha\sqrt{n}) < 2e^{-2\alpha^2}$. (Hint: Let $Y_i' = X_i' - .5$. Let F_i be the knowledge of $Y_1', \dots, Y_i'$, and let $Y_i = E(Y|F_i)$.)

8.5.39. *Bin-packing.* Let the numbers $S = \{a_1, \dots, a_n\}$ be drawn uniformly and independently from the interval $[0, 1]$. The numbers must be placed in bins, each having capacity 1. Let X be the number of bins needed. Use Lemma 8.5.36 to prove that $P(|X - E(X)|) \ge \lambda\sqrt{n}) \le 2e^{-\lambda^2/2}$.

8.5.40. (!) *Azuma's Inequality and the Traveling Salesman Problem.*

a) Prove the generalization of Azuma's Inequality to general martingales: If $E(X_i) = X_{i-1}$ and $|X_i - X_{i-1}| \le c_i$, then $P(X_n - X_0) \ge \lambda\sqrt{\sum c_i^2}) \le e^{-\lambda^2/2}$.

b) Let Y be the distance from a given point z in the unit square to the nearest of n points chosen uniformly and independently in the unit square. Prove that $E(Y) < c/\sqrt{n}$, for some constant c. (Hint: For a nonnegative continuous random variable Y, $E(Y) = \int_0^\infty P(Y \ge y)dy$, which can be verified using integration by parts. In order to bound this integral, use (somewhere) the inequality $1 - a < e^{-a}$ and the definite integral $\int_0^\infty e^{-t^2}dt = \sqrt{\pi}/2$.)

c) Apply parts (a) and (b) to prove that the smallest length of a polygon bounding a random set of n points in the unit square is highly concentrated around its expectation. In particular, the probability that this deviates from the expected tour length by more than $\lambda c\sqrt{\ln n}$ is bounded by $2e^{-\lambda^2/2}$, for some appropriate c. (Hint: For the martingale in which X_i is the expected length of the tour when the first i points are known, prove that $|X_i - X_{i-1}| < c(n-i)^{-1/2}$. Lemma 8.5.36 does not apply directly.)

8.6. Eigenvalues of Graphs

Techniques from group theory and linear algebra assist in studying the structure and enumeration of graphs.

From linear algebra, we have seen hints of vector spaces and determinants. In a graph G with edges $e_1, \dots, e_m$, the **incidence vector** for a set $F \subseteq E(G)$ has coordinates $a_i = 1$ when $e_i \in F$ and $a_i = 0$ when $e_i \notin F$. Let **C** be the set of incidence vectors of even subgraphs (those with all vertex degrees even), and let **B** be the set of incidence vectors of edge cuts. Because these sets are closed under binary vector addition, **C** and **B** are vector spaces (Exercises 1–2), called the **cycle space** and **bond space** of G. Since an even subgraph and an edge cut share an even number of edges, **C** and **B** are orthogonal. This is closely related