Probabilistic Method

Lectures delivered by Jacob Fox
Notes by Holden Lee

Spring 2011, MIT

Last updated Fri. 4/29/2011

Contents

Introduction

Jacob Fox taught a course (18.997) on Probabilistic Method at MIT in Spring 2011. These are my "live-TEXed" notes from the course. The template is borrowed from Akhil Mathew.

Please email corrections to `holden1@mit.edu`.

Lecture 1
Tue. 2/1/2011

§1 Ramsey Numbers

What is the probabilistic method? Come up with probability space; show something exists with positive probability.

Three theorems due to Erdös.

Definition 1.1: A **Ramsey number** $R(k,l)$ is the minimum n such that every red-blue edge-coloring of the complete graph K_n on n vertices contains a red K_k or a blue K_l.

Theorem 1.2 (Ramsey's Theorem)**:** $R(k,l)$ is finite for all k,l.

If there are a lot of people at a party, you can find a large number that are friends or a large number that are not friends.

Probabilistic method gives a lower bound.

Proposition 1.3: If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$. Thus $R(k,k) \geq \left\lfloor 2^{\frac{k}{2}} \right\rfloor$ for $k \geq 3$.

Proof. Consider a random coloring of K_n where each edge is colored independently red or blue with probability $\frac{1}{2}$.

For any set S of k vertices, let A_S be the even that S induces a monochromatic K_k. The probability is

$$P(A_S) = 2^{1-\binom{k}{2}}$$

because there are $\binom{k}{2}$ edges between vertices of S, and they could either be all red or all blue. There are $\binom{n}{k}$ choices for S. Thus (since $P(A \vee B) = P(A)+P(B)-P(A \wedge B) \leq P(A)+P(B)$),

$$P(\text{at least one } A_s \text{ occurs}) \leq \binom{n}{k}2^{1-\binom{k}{2}} < 1.$$

Thus with positive probability no event A_S occurs, and there exists a 2-coloring of K_n with no monochromatic K_k. This means $R(k,k) > n$.

If $n = \left\lfloor 2^{\frac{k}{2}} \right\rfloor$ and $k \geq 3$,

$$2^{1-\binom{k}{2}} \leq \frac{2^{1+\frac{k}{2}}}{k!}\frac{n^k}{2^{k^2/2}} < 1.$$

□

Question: Can we construct a 2-edge-coloring of K_n without a monochromatic $K_{2\log n}$?

Silly answer: Yes; try all $2^{\binom{n}{2}}$ colorings.

Better answer: Still wide open (can it be done in polynomial time?).

But... a random coloring almost surely works, since $\frac{2^{1+\frac{k}{2}}}{k!} \to 0$ as $k \to \infty$. (Verification?)

Probabilistic method shows coloring exists but does not tell us how to find one!

§2 Tournaments

Definition 1.4: A **tournament** is an oriented complete graph. (Each edge has an orientation from one vertex to another.) One way to interpret this is that the vertices represent players and there is an edge from x to y if x beats y. Say a tournament T (with at least k vertices) has property S_k if for every k vertices, there is another vertex which beats them all. In other words, there is no small set of winners.

Theorem 1.5 (Erdös)**:** If $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there exists a tournament T on n vertices with property S_k. (For sufficiently large n the inequality holds.)

Proof. Consider a random tournament on n vertices $V = \{1, \ldots, n\}$, each edge having probability $\frac{1}{2}$ of going either way. For every $K \subseteq V$, $|K| = k$, let A_K be the event that there is no vertex which beats K. The probability that a fixed vertex outside of K beats all of K is 2^{-k}, and there are $n-k$ vertices outside of K, so

$$P(A_K) = (1-2^{-k})^{n-k}.$$

Since there are $\binom{n}{k}$ possibilities for K,

$$P(\text{any } A_K \text{ occurs}) = \sum_K P(A_K) = \binom{n}{k}(1-2^{-k})^{n-k} < 1.$$

Hence with positive probability, no A_K occurs and there exists a tournament T on n vertices with property S_K. □

Let $f(k)$ be the minimum n such that there exists T on n vertices with property S_K. Using the bound $\binom{n}{k} > \left(\frac{en}{k}\right)^k$ and $1 - x < e^{-x} \implies (1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, this gives

$$f(k) \leq (\ln 2 + o(1))k^2 2^k.$$

Szekeres showed that $f(k) \geq ck2^k$ (so this bound is pretty good).

§3 Sum-free subsets

Definition 1.6: A subset A of an abelian group is **sum-free** if there do not exist a_1, a_2, a_3 such that $a_1 + a_2 = a_3$.

Theorem 1.7: Every set $B = \{b_1, \ldots, b_n\}$ of n nonzero integers contains a sum-free subset A with $|A| > \frac{n}{3}$.

Proof. Let $p = 3k+2$ be a prime with $p > 2\max\{|b_i| : 1 \leq i \leq n\}$. Let $C = \{k+1, \ldots, 2k+1\} \subseteq \mathbb{Z}/p$; C is sum-free subset containing more than $\frac{1}{3}$ of the nonzero residues modulo $\mathbb{Z}/p$:

$$\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

Choose $x \in \{1, \ldots, p-1\}$ uniformly at random. Let $d_i \equiv xb_i \pmod{p}$ with $0 \leq d_i < p$. As x ranges from $1, \ldots, p-1$, d_i ranges from 1 to $p-1$ as well. Then

$$P(d_i \in C) = \frac{|C|}{p-1}.$$

Adding up the probabilities (expected values add linearly),

$$\mathbb{E}(\#d_i \in C) \geq \frac{n}{3}.$$

Therefore there exists x and $A \subseteq B$ with $|A| > \frac{n}{3}$ such that $xa \pmod{p} \in C$ for all $a \in A$. Now A is sum-free: indeed if $a_1 + a_2 = a_3$ with $a_1, a_2, a_3 \in A$, then $xa_1 + xa_2 \equiv xa_3 \pmod{p}$ with $xa_i \in C$, contradiciton the fact that C is sum-free. □

A history...

$$\begin{aligned} \text{Erdös } |A| &\geq \frac{n}{3} \\ \text{Alon, Kleitman } |A| &\geq \frac{n+1}{3} \\ \text{Bourgain } |A| &\geq \frac{n+2}{3} \end{aligned}$$

It isn't known where $\frac{1}{3}$ is the best constant. The current best construction is $\frac{11n}{28}$.

Lecture 2
Thu. 2/3/2011

First problem set: 1.1, 2, 4*, 6*, 8, 10, due Feb. 24 (latex please)

§1 Dominating sets

Definition 2.1: A set $U \subseteq V$ is **dominating** in a graph $G = (V, E)$ if every vertex $v \in V - U$ is has at least one neighbor in U.

Theorem 2.2: Let $G = (V, E)$ be a graph on n vertices with minimum degree $\delta > 1$. Then G has a dominating set of size at most $\frac{n(1+\ln(\delta+1))}{\delta+1} \sim \frac{n \ln \delta}{\delta}$ (as $\delta \to \infty$).

Proof. New technique: Let probability be arbitrary and choose it later on.

Fix $p \in [0, 1]$. Pick randomly and independently each vertex with probability p. Set X to be the set of picked vertices. Then

$$\mathbb{E}(|X|) = pn. \tag{1}$$

Let Y be the set in vertices in $V - X$ with no neighbors in X. Then $U = X \cup Y$ is dominating.

The probability that a vertex is in Y is (since there is probability p that a given vertex is in X; we care about the vertex and its neighbors)

$$P(v \in Y) = (1-p)^{\deg(v)+1} \leq (1-p)^{\delta+1}.$$

Hence

$$\mathbb{E}(|Y|) \leq n(1-p)^{\delta+1}. \tag{2}$$

Then by linearity of expectation with (1) and (2),

$$\mathbb{E}(|U|) = \mathbb{E}(|X| + |Y|) = \mathbb{E}(|X|) + \mathbb{E}(|Y|) \leq pn + (1-p)^{\delta+1}n.$$

This works for any $p \in [0,1]$ so we can choose p to make this expression small. We use $1-p \leq e^{-p}$ (these are actually close to each other for p small). Thus there exists $U = X \cup Y$ which is dominating and $|U| \leq pn + e^{-p(\delta+1)}n$. Taking $p = \frac{\ln(\delta+1)}{\delta+1}$, we get

$$|U| \leq \frac{n(1+\ln(\delta+1))}{\delta+1}.$$

(Note that actually calculating the minimum and plugging it in is messier.) □

This proof reveals four important ideas.

1. Linearity of expectation:
$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i).$$
Note that independence is not required.

2. Alteration principle: X wasn't enough; we had to alter it a bit. (See Chapter 3)

3. Optimized p at the end of the proof.

4. Asymptotic estimates. (The actual minimum above is $p = 1 - (\delta+1)^{\frac{1}{\delta}}$, but this is difficult to work with, and we prefer a clean bound.)

§2 Hypergraph coloring

Definition 2.3: A **hypergraph** is $H = (V, E)$ consists of a set of vertices V, and a set of edges E, where an edge is a subset of vertices. H is n-**uniform** if every edge has exactly n vetices. (A 2-uniform hypergraph is simply a graph.)

H is 2-colorable (has property B) if there exists a 2-coloring of V with no monochromatic edge. Let $m(n)$ be the minimum number of edges of a n-uniform hypergraph without property B.

Note $m(n) \leq \binom{2n-1}{n} \approx 4^n$ because we can let $|V| = 2n-1$ and let the edges be all n-subsets.

Proposition 2.4: $m(n) \leq 2^{n-1}$

Proof. Suppose H is a hypergraph with $|E| < 2^{n-1}$ edges. Color V randomly with 2 colors. For each $e \in E$,

$$P(e \text{ is monochromatic}) = 2 \cdot 2^{-n}$$
$$P(\text{at least one edge is monochromatic}) \leq |E| \cdot 2^{1-n} < 1$$

Thus there exists a 2-coloring without a monochromatic edge, i.e. H has property B. □

The upper bound also uses the probabilistic method.

Theorem 2.5: $m(n) = O(n^2 2^n)$.

Proof. Fix V with v vertices, where v is even. Pick edges at random. Let χ be a coloring of V with a points in the first color and $b = v - a$ points in the second color. Let S be a random subset of V with $|S| = n$.

Then

$$P(S \text{ is monochromatic under } \chi) = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}} \geq \frac{2\binom{v/2}{n}}{\binom{v}{n}} =: p$$

where the last inequality follows from convexity (Jensen). Let $S_1, \ldots, S_m$ be chosen uniformly at random with replacement. Let A_χ be the event that no S_i is monochromatic under χ. Then $\#\chi = 2^v$. We have

$$P(A_\chi) \leq (1-p)^m$$
$$P\left(\bigvee_x A_x\right) \leq 2^v (1-p)^m < 1 \qquad \text{for } m = \left\lceil \frac{v \ln 2}{p} \right\rceil$$

Thus there exists H with at most m edges such that every 2-coloring gives monochromatic edges. (Take S_i to be the edges such that $\bigvee_x A_x$ does not hold.) Picking v to minimize

$$m = \left\lceil \frac{v \ln 2}{p} \right\rceil = \left\lceil \frac{v(\ln 2)\binom{v}{n}}{2\binom{\frac{v}{2}}{n}} \right\rceil$$

we get $O(n^2 2^n)$. □

§3 Erdös-Ko-Rado Theorem

Definition 2.6: A family $\mathcal{F}$ of sets is **intersecting** if for any $A, B \in \mathcal{F}$, $A \cap B \neq \phi$.

Theorem 2.7 (Erdös-Ko-Rado)**:** Suppose $n \geq 2k$ and $\mathcal{F}$ is an intersecting family of k-subsets of a n-set. Then $|F| \leq \binom{n-1}{k-1}$, and this bound is attainable.

Proof. Create an "obstruction" and copy it a lot.

Lemma 2.8: For $0 \leq s \leq n-1$, set $A_s = \{s, s+1, \ldots, s+k-1\} \subseteq \mathbb{Z}/n\mathbb{Z}$. Then $\mathcal{F}$ contains at most k of the sets A_s.

Proof. Fix $A_s \in \mathcal{F}$. All other A_t which intersect A_s can be partitioned into disjoint pairs $\{A_{s-i}, A_{s+k-1}\}$. (They are disjoint since $n \geq 2k$.) There are $k-1$ such pairs. □

Pick a permutation σ of $\{0, \ldots, n-1\}$ and $i \in \{0, \ldots, n-1\}$ at random, uniformly and independently. Set $A = \{\sigma(i), \ldots, \sigma(i+k-1)\}$ (addition modulo n). For any fixed σ, by Lemma 2.8,

$$P(A \in \mathcal{F}) \leq \frac{k}{n}.$$

Thus this holds for σ chosen randomly. But

$$P(A \in \mathcal{F}) = \frac{|F|}{\binom{n}{k}}.$$

Hence

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}.$$

This is attainable by choosing all k-element subsets containing a fixed element. □

Lecture 3
Tue. 2/8/2011

§1 Linearity of Expectation

Let $X_1, \ldots, X_n$ be random variables and $X = cX_1 + \cdots + cX_n$. Then

$$\mathbb{E}(X) = c_1\mathbb{E}(X_1) + \cdots + c_n\mathbb{E}(X_n).$$

(Independence is not required.)

Example 3.1: Let σ be a random permutation of $\{1, \ldots, n\}$. What is the expected value of the number of fixed points? (i is a fixed point if $\sigma(i) = i$)

Solution. Let X be the number of fixed points. Then $X = X_1 + \ldots + X_n$ where X_i is the indicator random variable for the event $\sigma(i) = i$. (1 if the event happens, 0 if it doesn't) Now $\mathbb{E}(X_i) = P(X_i) = \frac{1}{n}$. Hence

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = n \cdot \frac{1}{n} = 1.$$

In applications, we often use that there is a point in the probability space for which $X \geq \mathbb{E}(X)$ and a point where $X \leq \mathbb{E}(X)$.

Theorem 3.2: There exists a tournament T with n players and at least $\frac{n!}{2^{n-1}}$ hamiltonian paths. (A hamiltonian path is a sequence of vertices $v_1, \ldots, v_n$ such that v_iv_{i+1} is a directed edge.)

Proof. Pick a random tournament, each edge has probability $\frac{1}{2}$ going either way. Each permutation $\sigma(1), \ldots, \sigma(n)$ of the n players has probability $2^{-(n-1)}$ has probability of forming a hamiltonian path because there are $n-1$ edges between them. Let X_σ be the indicator random variable for σ forming a hamiltonian path and $X = \sum_{\sigma \in S_n} X_\sigma$ be the number of hamiltonian paths. Then $P(X_\sigma) = \frac{1}{2^{n-1}}$ and the expected number of hamiltonian paths is

$$\mathbb{E}(X) = \sum_{\sigma \in S_n} \mathbb{E}(X_\sigma) = n! 2^{-(n-1)}.$$

□

Remark 3.3: The bound is roughly best possible.

Every tournament has a hamiltonian path. If T is transitive there is exactly one hamiltonian path.

§2 Max-cut

Definition 3.4: The **max-cut** of a graph G is the maximum number of edges of a bipartite subgraph of G. (Not necessarily induced subgraph)

Theorem 3.5: If G contains e edges then G contains a bipartite subgraph with at least $\frac{e}{2}$ edges.

Proof. Let $T \subseteq V$ be a random subset. Let $B = V - T$. Let H be the bipartite subgraph consisting of edges between T and B ("crossing edges"). For any edge e of G, $P(e \in H) = \frac{1}{2}$ (consider where its endpoints lie). Hence letting X_e be the indicator for $e \in H$,

$$\mathbb{E}(\text{number of edges of } H) = \sum_{e \in E} \mathbb{E}(X_e) = \frac{e}{2}.$$

□

Note greedy algorithm also works. Or use extremal principle; if some vertex has more edges going to the same group, then move it to the other group to increase the number of edges of H.

We can improve this by picking T uniformly at random from all n-subsets: (Picking a better probability space can improve the bound.)

Theorem 3.6: If G has $2n$ vertices and e edges then G has a bipartite subgraph with at least $\frac{n}{2^{n-1}} e$ edges.

Proof. The probability that $e \in H$ is $\frac{n}{2n-1}$. (Given $e = xy$ and x is in T or B, there is $\frac{n}{2n-1}$ probability that y is too.) The rest of the proof is the same. □

Algorithmic question: compute or estimate max-cut. Goemans and Williamson give a .878... (polynomial-time) approximation algorithm for max cut. Khot showed that assuming the Unique Games Conjecture[1] this is the best possible.

[1] `http://en.wikipedia.org/wiki/Unique_games_conjecture`

§3 Ramsey multiplicity

Question: How many monochromatic K_a are there in every 2-edge-coloring of K_n?

Theorem 3.7: There exists a 2-coloring of the edges of K_n with at most $\binom{n}{a}2^{1-\binom{a}{2}}$ monochromatic K_a.

Proof. Take a random 2-coloring of the edges of K_n. Each K_a has probability $2^{1-\binom{k}{2}}$ of being monochromatic. Since $\#K_a = \binom{n}{a}$,

$$\mathbb{E}(\text{number of monochromatic} K_a) = \sum_{a-\text{cliques}} 2^{1-\binom{a}{2}} = \binom{n}{a}2^{1-\binom{a}{2}}.$$

Thus there exists a coloring with at most $\binom{n}{a}2^{1\binom{a}{s}}$ momochromatic K_a. □

Remark 3.8: How good is this bound? Erdös conjectures this in 1962; Goodman probed it true for $a = 3$ in 1959 (AMM) The theroem is false for $a \geq 4$ (thin $H = K_a$).

If H has e edges random bound gives that there exists a coloring with the fraction of monochromatic H at most 2^{n-1}.

$H = K_a, a^2 \ln a$, one color class is disjoint union of $K_{n/(a-1)}$.

Sidoranko conjectures that the random bound is tirght if H is bipartite.

Lecture 4
Thu. 2/10/11

§1 Balancing vectors

Theorem 4.1: Let $v_1, \ldots, v_n \in \mathbb{R}^n$, $|v_i| = 1$ for $1 \leq i \leq n$. Then there exist $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ with

$$|\varepsilon_1 v_1 + \cdots \varepsilon_n v_n| \leq \sqrt{n}$$

and also $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ with

$$|\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n| \geq \sqrt{n}.$$

Note this is tight because letting v_i be the ith standard basis vector we get all vertices of $[-1, 1]^n$.

Proof. Pick $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ with probability $\frac{1}{2}$ uniforomly and independently at random. Let $X = |\varepsilon_1 v_1 + \cdots + \varepsilon_n v_n|^2$. Then

$$X = \sum_{1 \leq i,j \leq n} \varepsilon_i \varepsilon_j v_i v_j.$$

$$\mathbb{E}(X) = \sum_{1 \leq i,j \leq n} v_i v_j \mathbb{E}(\varepsilon_i \varepsilon_j).$$

If $i \neq j$ then by independence $\mathbb{E}(\varepsilon_i\varepsilon_j) = \mathbb{E}(\varepsilon_i)\mathbb{E}(\varepsilon_j) = 0$ and if $i = j$ then $\mathbb{E}(\varepsilon_i\varepsilon_j) = \varepsilon_i\varepsilon_j = 1$. Hence

$$\mathbb{E}(X) = \sum_{i=1}^{n} v_i v_j = n.$$

There exist $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ with $X \leq \mathbb{E}(X)$ and $\varepsilon_1, \ldots, \varepsilon_n = \pm 1$ with $X \leq \mathbb{E}(X)$. □

This can be proved without probability.

Proof. Pick $\varepsilon_1, \ldots, \varepsilon_n$ sequentially such that

$$w_i = \varepsilon_1 v_1 + \cdots + \varepsilon_i v_i$$

satisfies $|w_i| \leq \sqrt{i}$ for all i, $1 \leq i \leq n$. Once $\varepsilon_1, \ldots, \varepsilon_i$ have been chosen pick ε_{i+1} such that $\varepsilon_{i+1}v_{i+1}$ and w_i make an obtuse (or right) angle. Then

$$|w_{i+1}|^2 \leq |w_i|^2 + |\varepsilon_{i+1}v_{i+1}|^2 \leq i + 1.$$

□

§2 Unbalancing lights

Suppose we have a $n \times n$ array of lights; each can be on or off. We want as many on as possible, but we can only flip all of the lights in a row or all the lights in a column in one step. How many lights can be turn on?

Theorem 4.2: Let $a_{ij} = \pm 1$ for $1 \leq i, j \leq n$. Then there exist $x_i, y_j = \pm 1$ for $1 \leq i, j \leq n$ so that

$$\sum_{1 \leq i,j \leq n} a_{ij} x_i y_j \geq \left(\sqrt{\frac{2}{\pi}} + o(1) \right) n^{\frac{3}{2}}.$$

(The a_{ij} are the initial states of the lights; x_i is 1 if row i is flipped and 0 otherwise; y_j is 1 if row j is flipped and 0 otherwise.)

Proof. Forget the x_i. Let $y_1, \ldots, y_n$ be selected uniformly at random. Let

$$R_i = \sum_{j=1}^{n} a_{ij} y_j, \quad R = \sum_{i=1}^{n} |R_i|.$$

Once y_j have been chosen, we can choose x_i so the total is R, by choosing x_i so $x_i R_i$ is positive. Regardless of a_{ij}, R_i has distribution S_n, the sum of n random variables from $\{-1, 1\}$. Now

$$\mathbb{E}(|R_i|) = \mathbb{E}(|S_n|) = \sqrt{n}(\mathbb{E}(|N|) + o(1)) = \left(\sqrt{\frac{2}{\pi}} + o(1) \right) \sqrt{n}$$

where N is the standard normal distribution. (Alternatively, use $E(|S_n|) = n2^{1-n}\binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$ and Stirling's approximation.) Then

$$\mathbb{E}(R) = \sum_{i=1}^{n} \mathbb{E}(|R_i|) = n\left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{\frac{1}{2}}$$

so there exist $y_1, \ldots, y_n = \pm 1$ for which $R \geq \mathbb{E}(R)$. As mentioned, pick x_i so that they are the same sign as R_i. □

§3 Alterations

Theorem 4.3 (Markov's inequality)**:** Let X be a nonnegative random variable. Then

$$P(X > \alpha \mathbb{E}(X)) \leq \frac{1}{\alpha}.$$

Note for $\alpha = 1$ this tells us nothing, but for α large the probability is small.

Proof. Substitute $t = \alpha E(X)$ in

$$\mathbb{E}(X) > tP(X > t).$$

□

Definition 4.4: The **girth** of G is the length of the shortest cycle. (Think of G looking locally like a tree, sparse.)

The **chromatic number** $\chi(G)$ is the minimum number of colors needed to properly color the vertices of G (i.e. no two adjacent vertices are the same color). (Think of G as globally dense.)

The **independence number** $\alpha(G)$ is the maximum size of an independent set, a set of pairwise nonadjacent vertices.

Theorem 4.5 (Erdös)**:** For all k, l, there exists a graph G^* with $\text{girth}(G^*) > l$ and $\chi(G^*) > k$.

Proof. Pick $\theta \in (0, \frac{1}{l})$ and let $p = n^{\theta - 1}$. Let $G = G(n, p)$ be a random graph with n vertices where each edge is picked with probability p independent of the other vertices. Idea: pick p just right—small enough so that there aren't a lot of small cycles, and large enough so that the chromatic number is large.

We can't avoid short cycles in a random graph G^* but we can make their number small. There are $\frac{1}{2i}n^{\underline{i}}$ possible i-cycles ($n^{\underline{i}}$ ways of choosing the set of vertices; divide by $2i$ because rotations and reversals are the same.

$$\mathbb{E}(X) = \sum_{i=3}^{n} \frac{n^{\underline{i}}}{2i} p^i \leq \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

We later delete some vertices to get rid of the cycles—alteration. (We won't have to delete too many.) By Markov's Inequality,

$$P\left(X \geq \frac{n}{2}\right) = o(1). \tag{3}$$

Chromatic number is hard to bound directly, so we instead bound $\alpha(G)$ from above and note

$$\chi(G) \geq \frac{n}{\alpha(G)}$$

because a coloring with $\chi(G)$ colors partitions the vertices into independent sets, each of which has size at most $\alpha(G)$, giving $n \geq \chi(G)\alpha(G)$. Set $x = \left\lceil \frac{3}{p} \ln n \right\rceil$. Then since there are $\binom{n}{x}$ sets of size x and probability $(1-p)^{\binom{x}{2}}$ that a given x-set is independent, using the union bound gives

$$P(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} \leq \left(ne^{-\frac{p(x-1)}{2}}\right)^x = o(1). \tag{4}$$

Let n be sufficient large so that both probabilities (3) and (4) are less than $\frac{1}{2}$. Pick G with $X < \frac{n}{2}$ and $\alpha(G) < x$. Delete one vertex from each cycle of length at most l, and let G^* be the resulting graph. Now

$$\text{girth}(G) > l.$$

Now $X < \frac{n}{2}$ so G^* has $n^* \geq \frac{n}{2}$ vertices. Now G^* is an induced subgraph of G, so any independent set in G^* is also independent in G, and $\alpha(G^*) \leq \alpha(G)$. Hence (recall $p = n^{\theta-1}$)

$$\chi(G^*) \geq \frac{n^*}{\alpha(G^*)} \geq \frac{n/2}{x} \geq \frac{n^\theta}{(6+\varepsilon)\ln n}.$$

For n sufficiently large this is greater than k. □

Lecture 5
Tue. 2/15/11

§1 Alterations: Ramsey Numbers

Recall that

$$\binom{n}{k}2^{\binom{k}{2}} \implies R(k,k) > n. \tag{5}$$

We prove a different bound.

Theorem 5.1: For any n,

$$R(k,k) > n - \binom{n}{k}2^{\binom{k}{2}} \tag{6}$$

If we choose n appropriately then this will be a better bound.

Proof. Consider a random 2-coloring of K_n. Let X be the number of monochromatic K_k. Then

$$\mathbb{E}(X) = \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Fix a coloring with $X \leq \mathbb{E}(X)$. Delete a vertex for each monochromatic K_k. The resulting coloring has no monochromatic K_k and has at least

$$n - X \geq n - \binom{n}{k} 2^{1-\binom{k}{2}}$$

vertices. □

Alteration idea: first get coloring with not too many monochromatic cliques and then delete vertices to get rid of them.

Bound (5) gives

$$R(k,k) > \frac{1}{e\sqrt{2}}(1+o(1))k2^{\frac{k}{2}}$$

while bound (6) gives

$$R(k,k) > \frac{1}{e}(1+o(1))k2^{\frac{k}{2}}.$$

The Lovasz Local Lemma in Chapter 5 will give

$$R(k,k) > \frac{\sqrt{2}}{e}(1+o(1))k2^{\frac{k}{2}}.$$

The best known upper bound is $R(k,k) \leq (4+o(1))^k$. (Improvements are small but are good examples of the method.)

§2 Independent Sets

If we have a graph with few edges, we expect large independent sets.

Theorem 5.2: Let G have n vertices and $\frac{dn}{2}$ edges, $d \geq 1$ (so the average degree of the graph is at least d). Then $\alpha(G) \geq \frac{n}{2d}$. (Recall $\alpha(G)$ is the independence number.)

Proof. Let $S \subseteq V$ be a random set defined where $P(v \in S) = p$ for any v and these events are independent of each other. We will delete vertices to make it independent.

Let $X = |S|$. Then

$$\mathbb{E}(X) = pn.$$

Let Y be the number of edges in $G[S]$ (the induced subgraph of G with vertices of S). Then each edge has probability p^2 of lying in $G[S]$ so

$$\mathbb{E}(Y) = p^2 \cdot \frac{dn}{2}.$$

Then

$$\mathbb{E}(X-Y) = \mathbb{E}(X) - \mathbb{E}(Y) = pn - p^2 \cdot \frac{dn}{2}.$$

Letting $p = \frac{1}{d}$ maximizes this expression. Then

$$\mathbb{E}(X-Y) = \frac{n}{2d}.$$

Fix G for which $X - Y \leq \mathbb{E}(X-Y)$. Delete from S a vertex from each edge. At least $X-Y$ vertices remain, and it is an independent set. □

Turan's Theorem will give a tighter bound.

§3 Combinatorial Geometry

Let S be a set of n points in a closed unit square U. Let $T(S)$ be the minimal area among all triangles with vertices in S. Let

$$T(n) = \max_{|S|=n} T(S).$$

Heilbroon conjectured that $T(n) = O\left(\frac{1}{n^2}\right)$, but this was disproved by KPS with a probabilistic method giving $T(n) = \Omega\left(\frac{\ln n}{n^2}\right)$ (complicated).

Theorem 5.3: $T(n) \geq \frac{1}{100n^2}$.

Proof. Let P, Q, R be independent and uniformly selected from U, and let $\mu = [PQR]$ be the area of $\triangle PQR$. We bound $P([PQR] \leq \varepsilon)$. Let $x = |PQ|$. Now

$$P(b \leq x \leq b + \Delta b) \leq \pi(b+\Delta b)^2 - \pi b^2$$

so as $\Delta b \to 0$,

$$P(b \leq x \leq b + db) \leq 2\pi b\, db.$$

Given $d(P,Q) = b$, we bound $P(\mu \leq \varepsilon)$. The distance of R from PQ must be $h \leq \frac{2\varepsilon}{b}$; thus R is in a strip of width $\frac{4\varepsilon}{b}$ and of length at most $\sqrt{2}$, so given $|PQ| = b$,

$$P(\mu \leq \varepsilon) \leq \sqrt{2} \cdot \frac{4\varepsilon}{b}.$$

Hence

$$P(\mu \leq \varepsilon) \leq \int_0^{\sqrt{2}} \sqrt{2} \cdot \frac{4\varepsilon}{b} \cdot 2\pi b\, db = 16\pi\varepsilon.$$

Let $P_1, \ldots, P_{2n}$ be selected uniformly and independently at random from U. Let X be the number of triangles P_1, P_2, P_3 with area less than $\frac{1}{100n^2}$ "bad triangles". Then

$$\mathbb{E}(X) \leq \binom{2n}{3} 16\pi \cdot \frac{1}{100} n^2 < n.$$

Delete point from each bad triangle. The resulting set will have greater than n points and satisfies the equations. □

An explicit example (Erdös) gives $T(n) \geq \frac{1}{2(n-1)^2}$ for n prime (but doesn't extend to better bounds). Consider $[0, n-1] \times [0, n-1]$ and points (x, y) where $0 \leq x \leq n-1$, $y \equiv x^2 \pmod{n}$ and $0 \leq y \leq n-1$. We claim this set works (after scaling by $\frac{1}{n-1}$). No three points are collinear: otherwise they are on a line $y = mx + b$, m rational with denominator less than n. But then $x^2 - mx - b$ would have 3 solutions in $\mathbb{Z}/n$, n prime, a contradiction. The area of every nontrivial lattice triangle is at least $\frac{1}{2}$. Contract by a factor of $n-1$.

Definition 5.4: Let C be a bounded measurable subset of $\mathbb{R}^d$ (with $\mu(C) > 0$). Let $B(x) = [0, x]^d$ be the d-dimensional cube of side length x. A **packing** of C in $B(x)$ is a family of mutually disjoint translates of C lying inside in $B(x)$. Let $f(x)$ be the size of the largest packing of C in $B(x)$. The packing constant is

$$\delta(C) = \mu(C) \lim_{x \to \infty} f(x) x^{-d},$$

i.e. the fraction of space that can be filled with copies of C.

For example, for C a sphere in $\mathbb{R}^3$, $\delta(C) = \frac{\pi}{3\sqrt{2}}$.

Theorem 5.5: Let X be a bounded, convex, centrally symmetric set around the origin. Then

$$\delta(C) \geq \frac{1}{2^{d+1}}.$$

Proof. Take random points x_i from $B(x)$. Consider $x_i + C$. Count the number that intersect. Now $(p + C) \cap (q + C)$ means that $p - q = c_2 - c_1 \in 2C$ (from convexity and symmetry). Now $[2C] = 2^d[C]$. Hence

$$P((p + C) \cap (q + C) \neq \phi) \leq \frac{[2C]}{x^d} = \frac{2^d[C]}{x^d}.$$

Now delete the problemsome points. □

Lecture 6
Thu. 2/17/11

§1 Second Moment Method

Definition 6.1: The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - E(X)^2$$

The standard deviation is

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Theorem 6.2 (Chebyshev's Inequality)**:** For any $\lambda > 0$,

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Proof.

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[(X-)^2] \geq \lambda^2\sigma^2 P(|X-\mu|^2 \geq \lambda^2\sigma^2).$$

This is Markov's inequality with $Z = (X-\mu)^2$, $\mathbb{E}(Z) = \sigma^2$, and λ^2. Now $Z \geq \lambda^2\sigma^2$ iff $|X-\mu| \geq \lambda\sigma$. □

Suppose $X = X_1 + \cdots + X_n$. Then

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)$$

where

$$\mathrm{Cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j).$$

In particular, if X_i, X_j are independent then

$$\mathrm{Cov}(X_i, X_j) = 0.$$

Suppose X_i is an indicator random variable, i.e. $X_i = 1$ if a certain even A_i occurs and 0 otherwise. If $X_i = 1$ with probability $P(A_i) = p$ then $\mathrm{Var}(X_i) = p(1-p) \leq p = \mathbb{E}(X_i)$. Hence

$$\mathrm{Var}(X) \leq \mathbb{E}(X) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j).$$

§2 Number Theory

Let $\nu(n)$ be the number of prime factors of n. We will show that almost all n have close to $\ln\ln n$ prime factors.

Theorem 6.3 (Hardy-Ramanujan)**:** Let $\omega(n) \to \infty$ arbitrarily slowly. Then the number of x in $\{1, \ldots, n\}$ with $|U(x) - \ln\ln x| > \omega(n)\sqrt{\ln\ln n}$ is $o(n)$. In other words, for x randomly chosen from $[1, n]$,

$$P(|\nu(x) - \ln\ln n| > \omega(n)\sqrt{\ln\ln n}) = o(1).$$

Proof. Let x be randomly chosen from 1 to n. For p prime, let X_p be the indicator random variable for the event $p|x$. Set $M = n^{\frac{1}{10}}$ and $X = \sum_{p\leq M, p \text{ prime}} X_p$. Note

$$x \leq \nu(x) < x + 10.$$

since a number has less than 10 prime factors greater than $n^{\frac{1}{10}}$. (We exclude the large primes because they will give a greater variance for X_p.) Now

$$\mathbb{E}(X_p) = \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor = \frac{1}{p} + O\left(\frac{1}{n}\right)$$

so

$$\mathbb{E}(X) = \sum_p \mathbb{E}(X_p) = \sum_p \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = \ln\ln M + O(1) = \ln\ln n + O(1)$$

using Mertens's estimate.

The rough idea is to show that for $p \neq q$, X_p and X_q are almost independent, so the covariances are small, and don't affect the variance of X much. I.e. $\mathbb{E}(X_p \wedge X_q) \approx \mathbb{E}(X_p)\mathbb{E}(X_q)$. We have

$$\operatorname{Var}(X_p) = \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{n}\right)$$

so

$$\sum_p \operatorname{Var}(X_p) = \sum_p \frac{1}{p} + O(1) = \ln\ln n + O(1). \tag{7}$$

Now $X_p X_q = 1$ iff $pq|X$, so

$$|\operatorname{Cov}(X_p, X_q)| = \left|\frac{\left\lfloor \frac{n}{pq} \right\rfloor}{n} - \frac{\left\lfloor \frac{n}{p} \right\rfloor}{n} \cdot \frac{\left\lfloor \frac{n}{q} \right\rfloor}{n}\right| \leq \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).$$

Hence

$$\left|\sum_{p\neq q} \operatorname{Cov}(X_p, X_q) \leq \frac{1}{n}\sum_{p\neq q}\left(\frac{1}{p} + \frac{1}{q}\right)\right| \leq \left|\frac{2M}{n} \sum_{\substack{p \leq M \\ p \text{ prime}}} \frac{1}{p}\right| = o(1). \tag{8}$$

since we counted each $\frac{1}{p}$ at most $2M$ times. Putting (7) and (8) together,

$$\begin{aligned}\operatorname{Var}(X) &= \sum_p \operatorname{Var}(X_p) + \sum_{p\neq q} \operatorname{Cov}(X_p, X_q) \\ &= \ln\ln n + O(1).\end{aligned}$$

Since $\sigma = \sqrt{\ln\ln n} + O(1)$, by Chebyshev's inequality

$$P(|X - \ln\ln n| \geq \lambda\sqrt{\ln\ln n}) < \lambda^{-2} + o(1).$$

and the same holds for $\nu(X)$. Letting $\lambda \to \infty$ gives the theorem. □

In fact, the distribution of $\nu(x)$ approaches a normal distribution with mean and variance $\ln\ln n$.

Theorem 6.4 (Erdös-Kac)**:** Fix $\lambda \in \mathbb{R}$. Then

$$\lim_{n\to\infty} |\{x : x \in [n], \nu(x) \geq \ln\ln n + \lambda\sqrt{\ln\ln n}\}| = \int_\lambda^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

For λ large, this is asymptotic to $\sqrt{\frac{2}{\pi}} e^{-\lambda^2/2}/\lambda \ll \lambda^{-2}$.

Proof. (Outline) Fix $s(n)$, $s(n) \to \infty$, and $s(n) = o(\sqrt{\ln \ln})$ such as $s(n) = \ln \ln \ln n$.[2] Set $M = n^{\frac{1}{n}}$ and

$$X = \sum_{\substack{p \leq M \\ p \text{ prime}}}$$

Then

$$\nu(x) - s(x) \leq X \leq \nu(x).$$

Let Y_p be independent random variables with $P(Y_p = 1) = \frac{1}{p}$ and $P(Y_p = 0) = 1 - \frac{1}{p}$. Y_p represents an idealized version of X_p. Set

$$\begin{aligned} \mu &= \mathbb{E}(Y) = \ln \ln n + o((\ln \ln n)^{\frac{1}{2}}) \\ \sigma^2 &= \text{Var}(Y) \sim \ln \ln n. \\ \tilde{Y} &= \frac{Y - \mu}{\sigma}. \end{aligned}$$

By the Central Limit Theorem, $\tilde{Y}$ approaches the standard normal distribution N and $\mathbb{E}(\tilde{Y}^k) \to \mathbb{E}(N^k)$. Let $\tilde{X} = \frac{X-\mu}{\sigma}$ and compare $\tilde{X}$ and $\tilde{Y}$. For distinct primes $p_1, \ldots, p_s$,

$$\mathbb{E}(X_{p_1} \cdots X_{p_s}) - \mathbb{E}(X_{p_1}) \cdots \mathbb{E}(X_{p_s}) = O\left(\frac{1}{n}\right). \tag{9}$$

Fix $k \in \mathbb{N}$. Compare $\mathbb{E}(\tilde{X}^k)$ with $\mathbb{E}(\tilde{Y}^k)$. Expanding, $\tilde{X}^k$ is a polynomial in X with coefficient $n^{o(1)}$. Expanding X^j, always reducing X_p^a for $a \geq 2$, each $X^j = (\sum X_p)^j$ gives $O(M^k)$ terms equal to $n^{o(1)}$ of the form $X_{p_1} \cdots X_{p_s}$. The same expansion applies to $\tilde{Y}^k$. As the corresponding terms have expectation within $O\left(\frac{1}{n}\right)$ by (9),

$$\mathbb{E}(\tilde{X}^k) - \mathbb{E}(\tilde{Y}^k) = o(1).$$

Thus each moment of $\hat{X}$ approaches that of the standard normal N, giving that (by a theorem from probability) $\hat{X}$ approaches the normal distribution. □

Lecture 7
Thu. 2/24/11

§1 Distinct sums

Definition 7.1: A set $x_1, \ldots, x_k \in \mathbb{N}$ has **distinct sums** if all sums

$$\sum_{i \in S} x_i, \quad S \subseteq [k]$$

are distinct.

[2] Drowning number theorists say log log log.

Let $f(n)$ be the maximum k such that there exists $\{x_1, \ldots, x_k\} \subseteq [n]$ with distinct sums. Since $\{2^i | i \leq \log_2 n\}$ has distinct sums,

$$f(n) \geq 1 + \lfloor \log_2 n \rfloor .$$

(There are actually sequences that do better, $f(n) \geq 3 + \lfloor \log_2 n \rfloor$ for large n.) Erdős asked the following: Determine if there is a constant C such that $f(n) \leq C + \log_2 n$.

If $k = f(n)$, all 2^k sums are distinct and at most kn, giving $2^{f(n)} \leq f(n)n$, and $f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$.

Theorem 7.2:

$$f(n) \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1).$$

Proof. Fix $\{x_1, \ldots, x_k\} \subseteq [n]$ with distinct sums. Let $\varepsilon_1, \ldots, \varepsilon_k$ be independent random variables with

$$P(\varepsilon_i = 0) = P(\varepsilon_i = 1) = \frac{1}{2}$$

and set $X = \varepsilon_1 x_1 + \cdots + \varepsilon_k x_k$. Then X is a random sum, and

$$\mu = \mathbb{E}(X) = \frac{x_1 + \cdots + x_k}{2}$$
$$\sigma^2 = \mathrm{Var}(X) = \frac{x_1^2 + \cdots + x_k^2}{4} \leq \frac{n^2 k}{4}$$

since the $X_i = \varepsilon_i x_i$ are independent random variables. ($\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \frac{X_i^2}{2} - \frac{X_i^2}{4}$.) By Chebyshev, (we want to show that a constant fraction of the sums lie in a small region around the mean, and use Pigeonhole to conclude that if there's too many, then two of them are equal)

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$
$$P\left(|X - \mu| \geq \lambda \frac{n\sqrt{k}}{2}\right) \leq \frac{1}{\lambda^2}$$

Then

$$1 - \frac{1}{\lambda^2} \leq P\left(|X - \mu| < \frac{\lambda n\sqrt{k}}{2}\right) \leq 2^{-k}(\lambda n\sqrt{k} + 1)$$

since at most $\lambda n\sqrt{k} + 1$ of the sums can be in the interval $\left(\mu - \frac{\lambda n\sqrt{k}}{2}, \mu + \frac{\lambda n\sqrt{k}}{2}\right)$ and each is chosen with probability 2^{-k}. Take $\lambda = \sqrt{3}$. Then the equation gives $2^k \leq Ck^{\frac{1}{2}}n$ for some constant C; take logs to get the answer. □

§2 Some bounds

Let X be a nonnegative integer-valued random variable (e.g. X counts something). We want to bound $P(X = 0)$ given $\mu = \mathbb{E}(X)$. If $\mu \leq 1$, then

$$P(X > 0) \leq \mathbb{E}(X) \implies P(X = 0) \geq 1 - \mathbb{E}(X).$$

If $\mathbb{E}(X) \to \infty$, then not necessarily $P(X = 0) \to 0$. But if the standard deviation is small relative to μ this is true.

Example 7.3: For example, X be the deaths due to nuclear war in the next year. Then $P(X > 0)$ is small but $\mathbb{E}(X)$ is large.

Theorem 7.4: Let X be a nonnegative integer-valued random variable. Then

$$P(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2}.$$

Proof. Let $\mu = \frac{\mu}{\sigma}$. Then Chebyshev's inequality gives

$$P(X = 0) \leq P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} = \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2}.$$

□

Corollary 7.5: If $\mathrm{Var}(X) = o(\mathbb{E}(X)^2)$ then $P(X > 0) \to 1$. In fact

$$X \sim \mathbb{E}(X)$$

almost surely. (For each $\varepsilon > 0$, $|X - \mathbb{E}(X)| < \varepsilon\mathbb{E}(X)$ goes to 0 as $X \to \infty$.)

Proof. Take $\lambda = \frac{\varepsilon\mu}{\sigma}$ and let $\varepsilon \to 0$. □

Lecture 8
Tue. 3/1/11

§1 More bounds

Suppose that $X = X_1 + \cdots + X_m$ where X_i is the indicator random variable for event A_i. Write $i \sim j$ if A_i and A_j are not independent. When $i \sim j$ and $i \neq j$,

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \leq \mathbb{E}(X_iX_j) = P(A_i \wedge A_j).$$

If A_i and A_j are independent then $\mathrm{Cov}(X_i, X_j) = 0$. Let

$$\Delta = \sum_{\substack{i \sim j \\ i \neq j}} P(A_i \wedge A_j).$$

Then

$$\begin{aligned}\mathrm{Var}(X) &= \sum_i \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j) \\ &\leq \mathbb{E}(X) + \Delta.\end{aligned}$$

Corollary 8.1: If $\mathbb{E}(X) \to \infty$ and $\Delta = o(\mathbb{E}(X)^2)$ then $X \sim \mathbb{E}(X)$ almost surely. In particular $X > 0$ almost surely.

Definition 8.2: $X_1, \ldots, X_m$ are **symmetric** if for every $i \neq j$, there is a measure preserving mapping of the underlying probability space that sends event A_i to A_j.

For example, A_i is the event that a certain triangle appears in the graph, the indices numbering all the triangles in K_n. If the X_i are symmetric, then

$$\sum_{i\sim j, i\neq j} P(A_i \wedge A_j) = \sum_i P(A_i) \sum_{j\sim i, j\neq i} P(A_j|A_i)$$

independent of i. Let

$$\Delta^* = \sum_{j\sim i, i\neq j} P(A_j|A_i).$$

(i is fixed in the sum.) Then

$$\Delta = \sum_{i=1}^{n} P(A_i)\Delta^* = \Delta^* \mathbb{E}(X).$$

Corollary 8.3: If $\mathbb{E}(X) \to \infty$ and $\Delta^* = o(\mathbb{E}(X))$ then $X \sim \mathbb{E}(X)$ almost surely; $X > 0$ almost surely.

§2 Random Graphs

Let $G(n,p)$ be the random graph on n vertices, each pair of vertices is an edge with probability p, independent of the other pairs. A **property** of graph is a family of graphs closed under isomorphism.

Definition 8.4: A fuction $r(n)$ is a threshold function for some property P if whenever $p = p(n) \ll r(n)$ then $G(n,p)$ does not satisfy P almost surely, and whenever $p = p(n) \gg r(n)$ then $G(n,p)$ satisfies P almost surely.

Let $\omega(G)$ be the class number of G.

Theorem 8.5: The property $\omega(G) \geq 4$ (i.e. G contains K_4) has threshold function $n^{-2/3}$.

Proof. For every 4-set S of vertices in $G(n,p)$ let A_S be the event "S forms a clique." Let X_S be the indicator random variable for A_S. The number of 4-cliques is $X := \sum_{S\subseteq V(G), |S|=4} X_S$. Note $\omega(G) \geq 4$ iff $X > 0$. Now

$$\begin{aligned}\mathbb{E}(X_S) &= P = P(A_S) = p^6 \\ \mathbb{E}(X) &= p^6 \binom{n}{4} \sim \frac{p^6 n^4}{24}.\end{aligned}$$

If $p = p(n) \ll n^{-2/3}$ then $\mathbb{E}(X) = o(1)$ and $X = 0$ almost surely.

Suppose $p = p(n) \gg n^{-2/3}$ so $\mathbb{E}(X) \to \infty$. By Corollary 8.3, we need to show $\Delta^* = \sum_{S \sim T, S \neq T} P(A_T | A_S) = o(\mathbb{E}(X))$. There are $O(n^2)$ sets T with $|S \cap T| = 2$ and for each of them $P(A_T|A_S) = p^5$, $O(n)$ sets with $|S \cap T| = 3$ and for each of them $P(A_T|A_S) = p^3$. (The sum is over S, T such that X_S, X_T are NOT independent.) Hence

$$\Delta^* = O(n^2 p^5) + O(np^3) = o(\mathbb{E}(X)).$$

Hence $X > 0$ almost surely as needed. □

Definition 8.6: Let H be a graph with v vertices and e edges. Define the **density** of H to be

$$\rho(H) = \frac{e}{v}.$$

H is **balanced** if $\rho(H') \leq \rho(H)$ for every subgraph H' of H.

Theorem 8.7 (Erdős and Rényi)**:** If H is balanced and A is the event that H is a subgraph of G then

$$p = n^{-\frac{1}{\rho(H)}}$$

is the threshold function for A.

If H is not balanced, $p = n^{-\frac{1}{\rho(H)}}$ is not the threshold function for A. The threshold function is

$$p = n^{-\frac{1}{\rho(H_1)}}$$

where H_1 is the subgraph with greatest density.

Proof. Same idea, but more involved.

For the second part, let H_1 be the subgraph of H with $\rho(H_1)$ maximum, so $\rho(H_1) > \rho(H)$. With this p, $\mathbb{E}(\#\text{ of copies of } H_1) = o(1)$, so there is no copy of H_1 (and hence no copy of H) almost surely. □

§3 Clique Number

Fix $p = \frac{1}{2}$. Consider $\omega(G)$. The expected number of cliques X is

$$\mathbb{E}(X) = f(k) = \binom{n}{k} 2^{-\binom{n}{k}}.$$

$f(k)$ drops under 1 at around $k = 2\log_2 n$. (Use the estimate $\binom{k}{2} \approx \frac{k^2}{2}$.)

Theorem 8.8: Let $k = k(n)$ satisfy $f(k) \to \infty$. Then almost surely $\omega(G) \geq k$.

Proof. For each k-set S let A_S be the event that S is a clique, and X_S be the indicator random variable for A_S. Then $X = \sum_S X_S$.

Examine Δ^*. Fix a k-set S. Then $T \sim S$ iff $|T \cap S| = i$ with $2 \le i \le k-1$. Now there are $\binom{k}{i}\binom{n-k}{k-i}$ sets of k vertices that intersect S in i vertices.

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}-\binom{k}{2}}.$$

$$\frac{\Delta^*}{\mathbb{E}(X)} = \sum_{i=2}^{k-1} \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}.$$

"Not hard" but time-consuming calculation (Stirling...) shows this is $o(1)$. hence $\omega(G) \ge k$ almost surely. □

Theorem 8.9: For all n there exists k such that

$$P(\omega(G) = k \text{ or } k+1) \to 1.$$

Proof. For $k \sim 2\log_2 n$,

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1} 2^{-k} = n^{-1+o(1)} = o(1).$$

For most n, $f(k) \to \infty$ but $f(k+1) = o(1)$. For those n, $\omega(G) = k$ almost surely. Else we get a two-point concentration. □

Lecture 9
Thu. 3/3/11

§1 Lovász Local Lemma

Consider the Ramsey number lower bound. Not only does there exist a 2-edge coloring of K_n with $n = 2^{\frac{k}{2}}$ without a monochromatic K_k but almost surely a random coloring has this property.

However, in many cases the probability is not large but we still need to show it is positive. The Lovász Local Lemma helps in this.

A trivial example with positive but small probability is when we have n mutually independent events that each hold with probability p, then they hold with probability p^n. Mutual independence can be generalized to rare dependencies. ("almost mutually independent"—each event is dependent only on a few other events.)

Definition 9.1: Let $A_1, \ldots, A_n$ be events in a probability space. A directed graph $D = (V, E)$ with $V = [n]$ is called a **dependency digraph** for the events $A_1, \ldots, A_n$ if for each i, $1 \le i \le n$, A_i is mutually independent of all the events $A_j, (i,j) \notin E$. In other words, A_i is independent of the event that any combination of those A_j's occur.

Lemma 9.2 (Lovász Local Lemma)**:** Suppose D is a dependency digraph for events $A_1, \ldots, A_n$ and there exist real numbers $x_1, \ldots, x_n$ with $0 \leq x_i < 1$ and $P(A_i) \leq x_i \prod_{(i,j)\in E}(1 - x_j)$ for each $i, 1 \leq i \leq n$. Then

$$P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n}(1 - x_i) > 0.$$

Thus with positive probability all the A_i occur.

Proof. First we show that for any $S \subset [n]$, $|S| = s < n$, and any $i \notin S$,

$$P\left(A_i | \bigwedge_{j\in S} A_n\right) \leq x_i.$$

We prove this by induction on s. The base case $s = 0$ holds. Now assume it holds for all $s' < s$. Let $S_1 = \{j \in S : (i, j) \in E\}$ and $S_2 = S \backslash S_1$. Let

$$A = A_i, \quad B = \bigwedge_{j\in S_1} \overline{A_j}, \quad C = \bigwedge_{j\in S_2} \overline{A_j}.$$

Then since A and C are independent,

$$P(A|B \wedge C) = \frac{P(A \wedge B|C)}{P(B|C)} \leq \frac{P(A|C)}{P(B|C)} = \frac{P(A)}{P(B|C)} \tag{10}$$

We try to show $P(B|C) \geq \prod_{(i,j)\in E}(1 - x_i)$. Suppose $S_1 = \{j_1, \ldots, j_r\}$. If $r = 0$ then $P(B|C) = 1$. Suppose $r \geq 1$. Then by the induction hypothesis,

$$P(\overline{A_{j_1}} \wedge \cdots \wedge \overline{A_{j_r}}|C) = \prod_{i=1}^{r}\left(1 - P\left(A_{j_i} | \bigwedge_{k=1}^{i-1} \overline{A_{j_k}} \wedge C\right)\right) \geq \prod_{i=1}^{r}(1 - x_{j_r}).$$

Substituting this and $P(A) \leq x_i \prod_{i=1}^{r}(1 - x_{j_r})$ in (10) gives the claim.

Now

$$P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) = \prod_{i=1}^{n}\left(1 - P\left(A_i | \bigwedge_{k=1}^{i-1} \overline{A_k}\right)\right) \geq (1 - x_1) \cdots (1 - x_n).$$

□

Lemma 9.3 (Symmetric case)**:** Let $A_1, \ldots, A_n$ be events in a probability space with each A_i mutually independent of all the other A_j but at most d (i.e. D has maximum outdegree at most d) and $P(A_i) \leq p$ for all $i, 1 \leq i \leq n$. If $ep(d + 1) \leq 1$ then $P\left(\bigwedge_{i=1}^{n} \overline{A_i}\right) > 0$. (Here $e = 2.71828...$)

Proof. The case $d = 0$ is trivial, so assume $d \geq 1$. Take $x_i = \frac{1}{d+1}$ for all i. Then

$$\left(1 - \frac{1}{d + 1}\right)^d > \frac{1}{e}$$

so

$$x_i \prod_{(i,j)\in E} (1-x_i) \geq \frac{1}{(d+1)e} \geq p \geq P(A_i).$$

Now apply the general version. □

Remark 9.4: We can replace mutual independence and $P(A_i) \leq x_i \prod_{(i,j)\in E}(1-x_i)$ by a weaker assumption: For each i and $S_2 \subseteq [n]\backslash\{j : (i,j) \in E\}$, $P\left(A_i | \bigwedge_{j\in S_2}\right) \leq x_i \prod_{(i,j)\in E}(1-x_i)$. ($E$ is some digraph, not dependency digraph.) (In the proof this was $P(A|C) = P(A)$.)

§2 Ramsey Numbers

Consider a random 2-edge-coloring of K_n. For each k-set S, let A_S be the event S is monochromatic. Then $P(A_S) = 2^{1-\binom{k}{2}}$. Each A_S is mutually independent to all A_T except at most $\binom{k}{2}\binom{n-2}{k-2}$. (Such T contain 2 edges of S and $k-2$ more edges.)

Proposition 9.5: If $e\binom{k}{2}\binom{n-2}{k-2}2^{1-\binom{k}{2}} \leq 1$, then $R(k,k) > n$. Thus

$$R(k,k) > \left(\frac{\sqrt{2}}{e} + o(1)\right) k 2^{\frac{k}{2}}.$$

Proof. Use the symmetric case of local lemma. The probability that there is no monochromatic clique is positive □

This is the best bound so far. We get bigger improvements for off-diagonal Ramsey numbers.

Now consider $R(k,3)$. The basic probabilistic method gives $\Omega(k)$. The alteration method gives $\Omega\left(\left(\frac{k}{\ln k}\right)^{\frac{3}{2}}\right)$. The Lovász Local Lemma gives $\Omega\left(\left(\frac{k}{\ln k}\right)^{2}\right)$. A pigeonhole argument gives the upper bound $\binom{k+1}{2}$. Later it was shown that actually $R(3,k) \sim \frac{k^2}{\ln k}$.[3]

Take a 2-edge-coloring K_n, each edge blue with probability p. Construct digraph D. For each 3-set T, let A_T be the event T is monochromatic blue and for each k-set S, let B_S be the event S is red. Now $P(A_T) = p^3$ and $P(B_S) = (1-p)^{\binom{k}{2}}$. Each A_T is mutually independent of all $A_{T'}$ except at most $3n$ and all B_S's but at most $\binom{n}{k}$. Each B_S is mutually independent of all $A_{T'}$ but at most $\binom{k}{2}(n-2)$ and all B'_S but at most $\binom{n}{k}$. If we can find $0 \leq p, x, y \leq 1$ with

$$p^3 \leq x(1-x)^{3n}(1-y)^{\binom{n}{k}}$$
$$(1-p)^{\binom{k}{2}} \leq y(1-x)^{\binom{k}{2}n}(1-y)^{\binom{n}{k}}$$

then $R(k,3) > n$. Take $p = c_1 n^{-\frac{1}{2}}$, $k = c_2 n^{\frac{1}{2}} \ln n$, $x = c_3 n^{-\frac{3}{2}}$, y so that $y^{\binom{n}{k}} = c_4$. Then use LLL.

[3]Differential equation and Rödl nibble. Randomly order $\binom{n}{2}$ edges. Put edges in graph as long as don't add triangles. Easy to describe but difficult to actually analyze. Differential equations control parameters—certain parameters depend on others. Show doesn't deviate much from ideal version.

Lecture 10
Tue. 3/8/11

§1 Local lemma on hypergraphs

Theorem 10.1: Let $H = (V, E)$ be a hypergraph in which each edge has at least k vertices, and suppose each edge intersects at most d other edges. If $e(d+1) \leq 2^{k-1}$, then H has property B (the vertices can be colored so there is no monochromatic edge).

Proof. Color each vertex v of H randomly and independently with blue or red with probability $\frac{1}{2}$. For each edge $f \in R$, let A_f be the event f is monochromatic. Then

$$P(A_f) \leq 2 \cdot \frac{1}{2^{|f|}} \leq 2^{1-k}.$$

Each A_f is mutually independent of all but at most d other A_g. The result follows from the symmetric case of LLL (the probability is positive when since $ep(d+1) \leq e2^{1-k}(d+1) \leq 1$). □

Corollary 10.2: If H is k-uniform and each vertex has degree at most $\frac{1}{k}(e^{-1}2^{k-1} - 1)$ then H has property B. (The degree of a vertex is the number of edges it's in.)

If H is k-regular and k-uniform if $k \geq 9$.

Proof. H being k-uniform means $d \leq e^{-1}2^{k-1}$. □

§2 Compactness arguments

For a k-coloring $c : \mathbb{R} \to [k]$ and a subset $T \subseteq \mathbb{R}$, T is **multicolored** if $c(T) = [k]$.

Theorem 10.3: Let $m, k \in \mathbb{N}$ such that $e(m(m-1)+1)k\left(1 - \frac{1}{k}\right)^m \leq 1$. Then for any set S of m real numbers, there is a k-coloring of the reals so that each translation $x + S, x \in \mathbb{R}$, is multicolored.

For example, this holds when $m > (3 + o(1))k \ln k$.

Proof. We first fix a finite set $X \subseteq \mathbb{R}$ and show there exists a k-coloring such that each translate $x \in X$ is multicolored.

Put $Y = \bigcup_{x \in X}(x + S)$ and choose $c : Y \to [k]$ uniformly at random. For each $x \in X$, let A_x be the event $x + S$ is not multicolored. Then (since the probability that a fixed color is missing is $\left(1 - \frac{1}{k}\right)^m$, and there are k colors)

$$P(A_x) < k\left(1 - \frac{1}{k}\right)^m$$

Each A_x is mutually independent of all $A_{x'}$ except those with

$$(x + S) \cap (x' + S) = \phi,$$

so $d \leq m(m-1)$. ($x + s_1 = x' + s_2$, there are d possibilities for s_1 and $d-1$ possibilities for $s_2 \neq s_1$.) Use the symmetic case of LLL.

Now we use a compactness argument to extend the result to reals. Discrete space with k points is compact. By Tychonoff's Theorem an arbitrary product of compact spaces is compact, so the space of all functions $f : \mathbb{R} \to [k]$ with the product topology is compact. For each $x \in \mathbb{R}$, let C_x be the set of all colorings such that $x + S$ is multicolored. Note C_x is closed. (Colorings in C_x can be described by their values at a finite number of points.) The intersection of any finite number of C_x is nonempty from above, so by compactness, $\bigcap_{x\in\mathbb{R}} C_x$ is nonempty. □

Definition 10.4: A family F of open unit balls in $\mathbb{R}^3$ is a k-**fold covering** if each $x \in \mathbb{R}^3$ is in at least k balls in F. F is **decomposable** if there exists a partition $F = F_1 \cup F_2$ so that each F_i is a covering of $\mathbb{R}^3$.

Theorem 10.5 (Mani-Levitska, Pach)**:** For all k, there exists a k-fold covering of $\mathbb{R}^3$ which is not decomposable.

In $\mathbb{R}^2$, every 33-fold covering of $\mathbb{R}^2$ is decomposable.

Theorem 10.6: Each k-fold covering $F = \{B_i\}_{i\in I}$ in which each point is in at most $t = c2^{\frac{k}{3}}$, $c = (2^{19}e)^{-\frac{1}{3}}$ balls is decomposable.

Proof. By our choice of t, $\frac{et^3 2^{18}}{2^{k-1}} \leq 1$. Define a hypergraph $H = (V(H), E(H))$ with $V(H) = F$. For each $x \in \mathbb{R}^3$, let E_x be the set of $B_i \in F$ that contain x. Let $E(H) = \{E_x : x \in \mathbb{R}^3\}$.

We claim that each E_x intersects less than $t^3 2^{18}$ other E_y. If $E_x \cap E_y$, then x and y are in intersecting balls, say $x \in B_i, y \in B_j$, and y is in an ball of radius 4 around x. B_j takes up $4^{-3} = 2^{-6}$ of the volume of this ball. Hence the number of such B_j is less than $m = 2^6 t$. (Each point is in at most t balls. Note strict inequality holds since balls cannot perfectly cover a larger ball.)

Any n balls in $\mathbb{R}^3$ partition $\mathbb{R}^3$ into at most n^3 regions (not counting the infinite portion), so we get less than $m^3 = 2^{18}t^3$. regions, i.e. edges.

Let L be a finite subhypergraph of H. Put in $d = t^3 2^{18}$, and use Theorem 10.1 to show that L has property B.

By the compactness argument, if every finite subhypergraph of H is 2-colorable, then H is 2-colorable. (Axioim of choice) □

Lecture 11
Thu. 3/10/11

§1 Chernoff bounds

First we give an estimate.

Lemma 11.1:

$$\binom{n}{k} \leq \frac{1}{e}\left(\frac{en}{k}\right)^k.$$

Proof.

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n^k}{k!}.$$

Thus it suffices to show that $k! \geq e\left(\frac{k}{e}\right)^k$. We show this by taking logs and using an integral estimate:

$$\begin{aligned}\ln k! &= \sum_{i=1}^{k} \ln i \\ &\geq \int_{x=1}^{k} \ln x \, dx \\ &= k \ln k - l + 1.\end{aligned}$$

Exponentiating gives $k! > e\left(\frac{k}{e}\right)^k$, as needed. □

Chernoff says that under certain circumstances, tail probabilities decay rapidly—exponentially. This is useful in random graphs, when we want a union bound over many events.

Consider $G(n,p)$, a random graph with n vertices, each edge picked with probability p. (When the probability is the below function and $n \to \infty$ the probabilities of the following go to 1)

1. $p < \frac{1}{n}$: All connected components have size $O(\ln n)$, and they are trees or unicyclic.
2. $p = \frac{1}{n}$: The largest component has size $\Theta(n^{\frac{2}{3}})$.
3. $p = \frac{1+\varepsilon}{n}$: The largest component has size asymptotic $\frac{1+\varepsilon}{n}$. [Phase transition]

Fixing a vertex v,

$$P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} < \left(\frac{ne}{k}\right)^k p^k (1-p)^{n-1-k}.$$

Hence using the union bound,

$$\begin{aligned}P(\text{there exists a vertex of degree } \geq k) &\leq n \sum_{i \geq k} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\ &\leq n \binom{n-1}{k} p^k < n \left(\frac{ne}{k}\right)^k p^k.\end{aligned}$$

At $p = \frac{1}{n}$, this is less than $n\left(\frac{e}{k}\right)^k$. Putting $k \sim \frac{\ln n}{\ln \ln n}$ shows that this is approximately the maximum degree.

Theorem 11.2 (Chernoff bound)**:** Let $X_i, 1 \leq i \leq n$ be mutually independent random variables with $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Set $S_n = X_1 + \cdots + X_n$. Let $a > 0$. Then

$$P(S_n > a) < e^{-a^2/2n}.$$

The theorem will also work if $P(X_i = 1 - p_i) = p_i$ and $P(X_i = -p_i) = 1 - p_i$.

Proof. We use Markov's inequality. If Y is a *nonnegative* random variable $\alpha > 0$, then

$$P(Y > \alpha \mathbb{E}(Y)) < \frac{1}{\alpha}.$$

Thus we exponentiate, making the random variables nonnegative, and turning sums into products. We have to take advantage of independence, which makes products behave nicely.

Fix n, a and let $\lambda > 0$ be arbitrary. For $1 \leq i \leq n$,

$$\mathbb{E}(x^{\lambda X_i}) = \frac{e^{\lambda} + e^{-\lambda}}{2} = \cosh \lambda \leq e^{\lambda^2/2}.$$

by looking at Taylor series. Since X_i are mutually independent, so are $e^{\lambda X_i}$. The the expected values multiply:

$$\mathbb{E}(e^{\lambda S_n}) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda X_i}) \leq e^{\lambda^2 n/2}.$$

Now $S_n > a$ iff $e^{\lambda S_n} > e^{\lambda a}$. By Markov's inequality,

$$P(S_n > a) = P(e^{\lambda S_n} > e^{\lambda a}) < \frac{\mathbb{E}(e^{\lambda S_n})}{e^{\lambda a}} \leq e^{\frac{\lambda^2 n}{2} - \lambda a}.$$

Picking $\lambda = \frac{a}{n}$, we get $P(S_n > a) < e^{-\frac{a^2}{2n}}$. □

Note by symmetry, $P(S_n < -a) < e^{-\frac{a^2}{2n}}$, so

$$P(|S_n| > a) < 2e^{-\frac{a^2}{2n}}.$$

Theorem 11.3: There exists a graph G on n vertices such that for every $U \subseteq V(G)$, with $|U| = u$,

$$\begin{aligned} \left| e(U) - \frac{1}{2}\binom{u}{2} \right| &\leq u^{\frac{3}{2}} \sqrt{\ln\left(\frac{en}{u}\right)} \\ &= O(n^{\frac{3}{2}}) \end{aligned} \tag{11}$$

Proof. Consider a random graph $G = G(n, \frac{1}{2})$. Consider a vertex subset U. Let $a_u = u^{\frac{3}{2}} \sqrt{\ln\left(\frac{en}{u}\right)}$. Then using Chernoff's bound with suitably translated and dilated random variables,

$$P\left(\left| e(U) - \frac{1}{2}\binom{u}{2} \right| > a_u \right) < 2e^{-\frac{(2a_u)^2}{2\binom{u}{2}}} < e^{-\frac{4a_u^2}{u^2}}.$$

Hence by the union bound,

$$\begin{aligned} P(\exists U \subseteq V(G) \text{ not satisfying (11)}) &< \sum_{u=1}^{n} \binom{n}{u} e^{-\frac{4a_u^2}{u^2}} \\ &< \sum_{u=1}^{n} \left(\frac{ne}{u}\right)^u e^{-\frac{4a_u^2}{u^2}} \\ &= \sum_{u=1}^{n} \left(\frac{ne}{u}\right)^{-3u} \\ &= o(1). \end{aligned}$$

□

If G has a clique of order k, then $\chi(G) \geq k$. The converse is not true, but the following holds.

Conjecture 11.4 (Hajós)**:** If $\chi(G) \geq k$, then G contains a subdivision of K_k. (This means that there is $H \subseteq G$ such that we can get to H from K_k by replacing edges by paths.)

(See Graph Theory, Diestel, 7.3.)

This is true for $k \leq 4$, false for $k \geq 7$. This is in fact very false. Consider $G = G(n, \frac{1}{2})$. Then $\chi(G) = (1 + o(1))\frac{n}{2\log_2 n}$ almost surely (we only need $\geq$ here), since the clique number closely concentrated at $2\log_2 n$ and so is the independence number. But the largest clique subdivision is of order $O(n^{\frac{1}{2}})$, much smaller.

Suppose we have a clique subdivision of order $u = 10n^{\frac{1}{2}}$. Then by the Theorem (??) the edge density in U is at most $\frac{3}{4}$. At least $\frac{1}{4}$ of pairs are nonadjacent. For each pair, we need a path containing at least one vertex. Thus the number of vertices that are used is at least $\frac{1}{4}\binom{u}{2} > n$, a contradiction.

Conjecture 11.5 (Hadwiger)**:** If $\chi(G) \geq k$ then G contains a minor of K_k. (H is a minor of G if we can delete vertices and edges, and contract edges to obtain H. Contracting means replacing adjacent vertices by a single vertex.)

This is known up to $k = 6$.

Lecture 12
Tue. 3/15/11

§1 Martingales and tight concentration

Chernoff bounds give tight concentration for sums of *independent* random variables. We generalize Chernoff's inequality for not necessarily independent random variables.

Definition 12.1: A **martingale** is a sequence of random variables $X_0, X_1, \ldots, X_m$ such that for $0 \leq i < m$,

$$\mathbb{E}(X_{i+1}|X_i, \ldots, X_0) = X_i.$$

The natural example is a gambler in a "fair" casino. Let X_i be the gambler fortune at time i. The gambler starts with X_0 dollars. Given X_i, $\mathbb{E}(X_{i+1}) = X_i$, as needed. The simplest martingale is where the gambler flips a coin, with \$1 stakes each time, normalizing so that $X_0 = 0$. Letting $Y_1, \ldots, Y_m$ be independent random variables with $P(Y_i = 1) = P(Y_i = -1) = \frac{1}{2}$, $X_i = Y_1 + \cdots + Y_i$ has distribution S_i.

Theorem 12.2 (Azuma's Inequality)**:** Let $c = X_0, \ldots, X_m$ be a martingale with $|X_i - X_{i-1}| \leq 1$ for all $0 \leq i < m$. Let $\lambda > 0$. Then

$$P(X_m - c > \lambda\sqrt{m}) < e^{-\lambda^2/2}$$
$$P(|X_m - c| > \lambda\sqrt{m}) < 2e^{-\lambda^2/2}$$

Note Chernoff's bound is a special case.

Proof. By shifting we may assume $X_0 = 0$. Set $\alpha = \frac{\lambda}{\sqrt{m}}$ and $Y_i = X_i - X_{i-1}$, so $|Y_i| \leq 1$ and

$$\mathbb{E}(Y_i | X_{i-1}, \ldots, X_0) = 0.$$

We imitate the proof of Chernoff's bound for the Y_i. By Karamata Majorization and convexity of e^x,

$$\mathbb{E}(e^{\alpha Y_i} | X_{i-1}, \ldots, X_0) \leq \cosh \alpha = \frac{e^\alpha + e^{-\alpha}}{2} \leq e^{\alpha^2/2}.$$

Then

$$\begin{aligned}
\mathbb{E}(e^{\alpha X_m}) &= \mathbb{E}\left(\prod_{i=1}^{m} e^{\alpha Y_i}\right) \\
&= \mathbb{E}\left[\prod_{i=1}^{m-1} e^{\alpha Y_i} \mathbb{E}(e^{\alpha Y_m} | X_0, \ldots, X_{n-1})\right] \\
&\leq e^{(m-1)\alpha^2}/2 \cdot e^{\alpha^2/2} \\
&= e^{(m-1)\alpha^2/2} e^{\alpha^2/2} \\
&= e^{m\alpha/2}
\end{aligned}$$

So

$$\begin{aligned}
P(X_m > \lambda\sqrt{m}) &= P(e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}) \\
&< \mathbb{E}(e^{\alpha X_m}) e^{-\alpha\lambda\sqrt{m}} \\
&\leq e^{-\alpha^2 m/2 - \alpha\lambda\sqrt{m}} = e^{-\lambda^2/2}.
\end{aligned}$$

□

§2 Graph exposure martingales

Let $G(n,p)$ be the underlying probability space and f be a graph theoretic function.

Definition 12.3: Label potential edges $\{i,j\} \subseteq [n]$ by $e_1,\ldots,e_m$ setting $m = \binom{n}{2}$. We define the **edge exposure martingale** $X_0,\ldots,X_m$ by giving values $X_i(H)$, with

$$\begin{aligned} X_0(H) &= \mathbb{E}(f(G)) \\ X_i(H) &= \mathbb{E}(f(G)|e_j \in G \iff e_j \in H, 1 \le j \le i) \\ X_m(H) &= f(H) \end{aligned}$$

First expose i pairs $e_1,\ldots,e_i$ and see if they are in H. The remaining edges are not seen and considered to be random.

For example, $f(H)$ could be the chromatic number of H.

Definition 12.4: Define $X_1,\ldots,X_n$ by

$$X_i(H) = \mathbb{E}(f(G)|\text{for } x,y \le i, \{x,y\} \in G \iff \{x,y\} \in H).$$

(Add a vertex each time and look at the induced subgraph. Note the vertex exposure martingale is a subsequence of the edge exposure martingale, with suitable ordering of vertices.)

Definition 12.5: A graph theoretic function satisfies the **edge Lipschitz condition** whenever if H and H' differ in at most one edge, $|f(H) - f(H')| \le 1$. It satisfies the **vertex Lipschitz condition** if whenever H, H' differ in at most one vertex (in terms of the edges emanating from the vertex).

Theorem 12.6: When f satisfies the edge Lipschitz condition the corresponding edge exposure martingale satisfies $|X_i - X_{i-1}| \le 1$.

When f satisfies the vertex Lipschitz condition the corresponding vertex exposure martingale satisfies $|X_i - X_{i-1}| \le 1$.

For example the chromatic number satisfies both conditions.

Theorem 12.7: Let $G = G(n,p)$. Let $\mu = \mathbb{E}(\chi(G))$. Then

$$P(|\chi(G) - \mu| > \mu\sqrt{n-1}) < 2e^{-\lambda^2/2}.$$

The chromatic number is very concentrated around its mean. (The mean is hard to actually compute, but we can still get concentration results.)

Proof. Consider the vertex exposure martingale $X_1,\ldots,X_n$ on $G(n,p)$ with $f(G) = \chi(G)$. Since f satisfies the vertex Lipschitz condition, the result follows by Azuma's inequality (12.2). (Note we omitted X_0.) □

Lecture 13
Thu. 3/17/11

§1 Chromatic number

Theorem 13.1: Let $p = n^{-\alpha}$ where $\frac{5}{6} < \alpha \leq 1$ is fixed. Let $G = G(n,p)$. Then there exists $v = v(n,p)$ such that, almost everywhere (i.e. the probability goes to 1 as $n \to \infty$), $v \leq \chi(G) \leq v + 3$.

Proof.

Lemma 13.2: Almost always, every $c\sqrt{n}$ vertices of $G = G(n,p)$ is 3-colorable.

Proof. Let T be a minimal set that is not 3-colorable. Then for any x, $T - \{x\}$ is 3-colorable, and x must have internal degree at least 3. So T contains at least $\frac{3|T|}{2}$ edges. Let $t = |T|$. The probability of this occurring for some set T with $t \leq c\sqrt{n}$ vertices is at most

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{\frac{3t}{2}} \leq \sum_{t=4}^{c\sqrt{n}} \left(\frac{ne}{t}\right)^t \left(\frac{te}{3}\right)^{\frac{3t}{2}} p^{\frac{3t}{2}} = o(1). \tag{12}$$

(There are $\binom{n}{t}$ ways to choose t vertices. There are $\binom{t}{2}$ possible edges in T; we need at least $3t/2$ of them.) □

Let $\varepsilon > 0$ be arbitrarily small and $v = v(p,n,\varepsilon)$ be the least integer such that $P(\chi(G) \leq v) > \varepsilon$. Define $Y(G)$ to be the minimal size of a vertex subset S such that $G - S$ may be v-colored. This Y satisfies the vertex Lipschitz condition.

Apply the vertex exposure martingale on $G(n,p)$ to Y. Let $\mu = \mathbb{E}(Y)$. By Azuma's inequality,

$$P(Y \geq \mu + \lambda\sqrt{n-1}) < e^{-\lambda^2/2} = \varepsilon$$
$$P(Y \leq \mu - \lambda\sqrt{n-1}) < e^{-\lambda^2/2} = \varepsilon.$$

Let λ be such that $e^{-\lambda^2/2} = \varepsilon$. But $P(Y = 0) > \varepsilon$, so $\mu < \lambda\sqrt{n-1}$. Thus $P(Y \geq 2\lambda\sqrt{n-1}) < \varepsilon$.

With probability at least $1 - \varepsilon$ there is a v-coloring of all but at most $2\lambda\sqrt{n-1}$ vertices. Then taking $c = 2\lambda$ in the lemma we can 3-color the rest almost always (say, with probability more than $1 - \varepsilon$ for large enough n), so we will need at most 3 more colors. Choice of v says that there is at least probability $1 - \varepsilon$ that v colors will be needed. So

$$P(v \leq \chi(G) \leq v + 3) \geq 1 - 3\varepsilon.$$

□

§2 A general setting

Let $\Omega = A^B$ be the set of all functions $g : B \to A$. Let $p_{ab} = P(g(b) = a)$ with the values of $g(b)$ independent of each other. Fix a gradation

$$\phi = B_0 \subset B_1 \subset \cdots \subset B_m = B.$$

For example, let B be the set of all unordered pairs of vertices on n vertices, and $A = \{0, 1\}$, denoting whether an edges is there or missing. In $G(n, p)$, $p_{1b} = p$ and $p_{0b} = 1 - p$.

Let $L : A^B \to \mathbb{R}$ be a functional (for example the clique number of a graph). Define a martingale $X_0, X_1, \ldots, X_m$ by setting

$$X_i(h) = \mathbb{E}(L(g)|g(b) = h(b) \text{ for all } b \in B_i).$$

Note X_0 is a constant, $\mathbb{E}(L(g))$ and $X_m = L$.

L satisfies the Lipschitz condition relative to the gradation if for all $0 \le i < m$, h, h' differ only on $B_{i+1} - B_i$ implies $|L(h') - L(h)| \le 1$.

Theorem 13.3: Let L satisfy the Lipschitz condition relative to a gradation. Then the corresponding martingale satisfies $|X_{i+1}(h) - X_i(h)| \le 1$ for all $0 \le i < m$ and $h \in A^B$.

Proof. Group things appropriately.

Let H be the family of all h' that agree with h on B_{i+1}. Then

$$X_{i+1}(h) = \sum_{h' \in H} L(h') w_{h'}$$

where $w_{h'}$ is the conditional probability that $g = h'$ given that $g = h$ on B_{i+1}. For every $h' \in H$ let $H[h']$ be the family of h^* that agree with h' on all points except possibly $B_{i+1} - B_i$. The $H(h')$ partition the family of h^* that agree with h on B_i.

$$X_i(h) = \sum_{h' \in H} \sum_{h^* \in H(h')} L(h^*) q_{h^*} w_{h'}$$

where q_{h^*} is the conditional probability that g agrees with h^* on B_{i+1} given that it agrees with h on B_i. Using the Triangle Inequality,

$$\begin{aligned} |X_{i+1}(h) - X_i(h)| &= \left| \sum_{h' \in H} w_{h'} \left[L(h') - \sum_{h^* \in H(h')} L(h^*) q_{h^*} \right] \right| \\ &= \sum_{h' \in H} w_{h'} \sum_{h^* \in H(h')} |q_{h^*}(L(h') - L(h^*))| \\ &\le \sum_{h' \in H} w_{h'} \sum_{h^* \in H(h')} q_{h^*} = 1. \end{aligned}$$

□

Theorem 13.4 (General Azuma's inequality)**:** Let L satisfy the Lipschitz condition relative to a gradation of length m and $\mu = \mathbb{E}(L(g))$. Then for all $\lambda > 0$,

$$P(L(g) \geq \mu + \lambda\sqrt{m}) < e^{-\lambda^2/2}$$
$$P(L(g) \leq \mu - \lambda\sqrt{m}) < e^{-\lambda^2/2}.$$

Example 13.5: Let g be the random function from $\{1, \ldots, n\}$ to itself, all n^n functions equally likely. Let $L(g)$ be the number of values not hit, i.e. the number of y such that $g(x) = y$ has no solution. By linearity of expectation,

$$\mu = \mathbb{E}(L(g)) = n\left(1 - \frac{1}{n}\right)^n \in \left[\frac{n-1}{e}, \frac{n}{e}\right].$$

Set $B_i = \{1, \ldots, i\}$. Note L satisfies the Lipschitz condition. By the general Azuma's inequality (13.4),

$$P\left(\left|L(g) - \frac{n}{e}\right| > \lambda\sqrt{n} + 1\right) < 2e^{-\lambda^2/2}.$$

Lecture 14
Tue. 3/29/11

§1 Talagrand's Inequality

First, a motivating example. Consider a random permutation of $\{1, \ldots, n\}$ how long of an increasing subsequence can we expect? With high probability it is $O(n^{\frac{1}{2}})$. Azuma's inequality gives a concentration of $O(n^{\frac{1}{2}})$, which is bad, especially in the lower part. Talagrand's inequality gives a concentration of $O(n^{\frac{1}{4}})$. (The whole distribution is known now known; the concentration is $O(n^{\frac{1}{6}})$; the distribution for $\frac{2\sqrt{n}-X}{n^{\frac{1}{6}}}$ approaches the Tracy-Widom distribution.)

Talagrand's inequality gives concentration around the *median*. (If the concentration is low, then the mean is close to the median.)

Theorem 14.1 (Talagrand's Inequality)**:** Let $\Omega = \prod_{i=1}^n \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$ and $\vec{x} = (x_1, \ldots, x_n) \in \Omega$.

Define $\rho(A, \vec{x})$ to be the minimum value such that for all $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ with $|\vec{\alpha}| = 1$ there exists $\vec{y} = (y_1, \ldots, y_n) \in A$ with $\sum_{x_i \neq y_i} \alpha_i \leq \rho(A, \vec{x})$. Now matter what α you choose with length at most 1, you can move from $\vec{x}$ to some vector $\vec{y}$ in A. I.e. it measures the minimum cost of moving from $\vec{x}$ to a $\vec{y} \in A$ by changing coordinates when a suitably restricted adversary sets the cost of each change. (Note $\vec{y}$ may depend on $\vec{\alpha}$.)

For any real $t \geq 0$, let $A_t = \{\vec{x} \in \Omega : \rho(A, \vec{x}) \leq t\}$. (Note $A_0 = A$.) Then

$$P(A)(1 - P(A_t)) \leq e^{-\frac{t^2}{4}}.$$

In particular if $P(A) \geq \frac{1}{2}$ and t is large then all but a small proportion of Ω is within distance t of A.

Take $\Omega = \{0,1\}^n$ with the uniform distribution. Let τ be the Hamming distance. Then $\rho(A,\vec{x}) \geq \min_{\vec{y}\in A} \tau(\vec{x},\vec{y}) n^{-\frac{1}{2}}$. (The adversary takes all $\alpha_i = \frac{1}{\sqrt{n}}$.

Suppose to move from $\vec{x}$ to A the values $x_1, \ldots, x_l$ must be changed. Then $\rho(A,\vec{x}) \geq l^{\frac{1}{2}}$. Let $U(A,\vec{x})$ be the set of $\vec{s} = (s_1,\ldots,s_n) \in \{0,1\}^n$ such that if there exists $\vec{y} \in A$ with $x_i \neq y_i$ then $s_i = 1$. Then $\rho(A,\vec{x})$ is the minimum ρ such that with $|\vec{\alpha}| = 1$, there exists $\vec{s} \in U(A,\vec{x})$ with $\vec{\alpha}\cdot\vec{s} \leq \rho$. Let $V(A,\vec{x})$ be the convex hull of $U(A,\vec{x})$.

Theorem 14.2:

$$\rho(A,\vec{x}) = \min_{\vec{v}\in V(A,\vec{x})} |\vec{v}|.$$

Proof. The case $\vec{x} \in A$ is obvious. Assume $\vec{x} \notin A$.

Let $\vec{v}$ achieve the minimum. The hyperplane through $\vec{v}$ perpendicular to the line from the origin to $\vec{v}$ separates the origin from $V(A,\vec{x})$. (Else by convexity there would be a closer point.) For all $\vec{s} \in V(A,\vec{x})$, $\vec{s}\cdot\vec{v} \geq \vec{v}\cdot\vec{v}$.

Set $\vec{\alpha} = \frac{\vec{v}}{|\vec{v}|}$ so $|\vec{\alpha}| = 1$. Then all $\vec{s} \in U(A,\vec{x}) \subseteq V(A,\vec{x})$ satisfy $\vec{s}\cdot\vec{\alpha} \geq \vec{v}\cdot\frac{\vec{v}}{|\vec{v}|}$.

Conversely, pick any $\vec{\alpha}$ with $|\vec{\alpha}| = 1$. Then $\vec{\alpha}\cdot\vec{v} = |\vec{v}|$. As $\vec{v} \in V(A,\vec{x})$, we can write $v = \sum_i \lambda_i \vec{s_i}$ where $\vec{s_i} \in U(A,\vec{x})$ and $\lambda_i \geq 0$, $\sum \lambda_i = 1$.

Then $|\vec{v}| \geq \sum \lambda_i(\vec{\alpha}\cdot\vec{s_i})$ and hence some $\vec{\alpha}\cdot\vec{s_i} \leq |\vec{v}|$. □

Theorem 14.3: Let $\Omega = \{0,1\}^n$, then $\rho(A,\vec{x})$ is the Euclidean distance from $\vec{x}$ to the convex hull of A. Furthermore,

$$\int_\Omega e^{\frac{1}{4}\rho^2(A,\vec{x})}\, d\vec{x} \leq \frac{1}{P(A)}.$$

Proof. (of Talagrand's inequality from Theorem 14.3) Fix A and consider random variables $X = \rho(A,\vec{x})$. Then by Markov's inequality

$$P(\overline{A_t}) = P(X > t) = P(e^{\frac{x^2}{4}} > e^{\frac{t^2}{4}}) \leq \mathbb{E}(e^{\frac{X^2}{4}})e^{-\frac{t^2}{4}}.$$

□

Lecture 15
Thu. 3/31/11

§1 Applications of Talagrand's inequality

Let $\Omega = \prod_{i=1}^n \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $h : \Omega \to \mathbb{R}$. Under certain constraints, Talagrand's inequality will show h is concentrated.

Definition 15.1: A function $h : \Omega \to \mathbb{R}$ is Lipschitz if $|h(x) - h(y)| \leq 1$ whenever x, y differ in at most 1 coordinate.

h is k-Lipschitz

Definition 15.2: A function $f : \mathbb{N} \to \mathbb{N}$ is h-**certifiable** if whenever $h(x) \geq s$ there is $I \subseteq \{1, \ldots, n\}$ with $|I| \leq f(s)$ such that all $y \in \Omega$ that agree with x on the coordinates I have $h(y) > s$.

Example 15.3: For example consider $G(n,p)$ as the product of all $\binom{n}{2}$ coin flips. Let $h(G)$ be the number of triangles in G. Then h is certifiable with $f(s) = 3s$.

If an graph G has at least s triangles, take the edges of those triangles; there are at most $3s$. Any graph which agrees with G on those edges has at least s triangles as well. Warning: h is not Lipschitz; $\frac{h(x)}{n}$ is.

We'd expect variance on the order of n^2: We make $\binom{n}{2}$ coin flips, so there's variance on the order of n^2 on the number of edges; the number of edges correlates linearly with the number of triangles.

Theorem 15.4: Assume h is Lipschitz and f-certifiable. Let $X = h(x)$ where x is a random element of Ω. Then for all b and t,

$$P(X \leq b - t\sqrt{f(b)})P(X \geq b) \leq e^{-\frac{t^2}{4}}.$$

If h is k-Lipschitz, replace with $X \leq b - tk\sqrt{f(b)}$.

Proof. Set $A = \{x : h(x) < b - t\sqrt{f(b)}\}$. Suppose $h(y) \geq b$. We claim $y \notin A_t$. Let I be a set of indices of size at most $f(b)$ that certifies $h(y) \geq b$. Define $\alpha_i = 0$ when $i \in I$ and $\alpha_i = |I|^{-\frac{1}{2}}$ when $i \in I$, so $|\alpha| = 1$. If $y \in A_t$, then there exists a $z \in A$ that differs from y in at most $t|I|^{-\frac{1}{2}} \leq t\sqrt{f(b)}$ coordinates of I (so that the distance is at most $(t|I|^{\frac{1}{2}})|I|^{-\frac{1}{2}}$) and at arbitrarily many coordinates outside I.

Let y' agree with y on I and z outside I. By certification, $h(y') \geq b$. Since y' and z differ in at most $t\sqrt{f(b)}$ coordinates. By Lipschitz, $h(z) \geq h(y') - t\sqrt{f(b)} \geq b - t\sqrt{f(b)}$. Hence $z \notin A$, contradiction.

This shows $P(X \geq b) \leq P(\overline{A_t})$. By Talagrand's inequality,

$$P(X < b - t\sqrt{f(b)})P(X \geq b) \leq P(A)(1 - P(A_t)) \leq e^{-\frac{t^2}{4}}.$$

By continuity we can replace "$<$" with "$\leq$". □

Usually pick b to be the median, or $b - tk\sqrt{f(b)}$ to be the median. Note if $m = b - t\sqrt{f(b)}$ then usually $b \approx m + t\sqrt{f(m)}$, so the probability of being much larger than m is small.

Example 15.5: Let $x = (x_1, \ldots, x_n)$ where the x_i are independent and uniformly chosen from $[0,1]$. There is probability 0 that two of them match, so the ordering of the elements is basically a random permutation.

Let $X = h(x)$ be the length of the longest increasing subsequence of X. Elementary methods give $c_1\sqrt{n} < X < c_2\sqrt{n}$ almost surely. Note X is Lipschitz. Applying Azuma's inequality we get $|X - \mu| < s$ almost surely if $s \gg n$, no good.

We use Talagrand's inequality. Note X is certifiable with $f(s) = s$; any x' which agrees with x on an increasing sequence of length s has length of longest increasign sequence at least s. By Theorem 15.4 with m equal to the median (which is on the order $\sqrt{n}$),

$$P(X \leq m - t\sqrt{m})P(X \geq m) \leq e^{-\frac{t^2}{4}},$$

giving concentration $O(n^{\frac{1}{4}})$.

§2 Correlation inequalities

Let $G = G(n,p)$. Let H be the event that G is Hamiltonian, let P be the event that P is planar. We want to compare $P(H \wedge P)$ and $P(H)P(P)$. H and P are negatively correlated if $P(H \wedge P) \leq P(H)P(P)$, independent if $P(H \wedge P) = P(H)P(P)$ and positively correlated if $P(H \wedge P) \geq P(H)P(P)$. Note H is monotone increasing (if we add edges, Hamiltonian paths become more likely) and P is monotone decreasing (if we add edges, the graph is less likely to be planar).

We expect $P(P|H) \leq P(P)$, so $P(P \wedge H) \leq P(P|H)P(H) \leq P(P)P(H)$. This inequality is a special case of the FKG ineqality of 1971.

Definition 15.6: A is a **monotone decreasing family** of subsets of $\{1,\ldots,n\}$ if whenever $A' \in A$ and $A'' \subseteq A$, we have $A'' \in A$.

Theorem 15.7: If A and B are monotone decreasing families of subsets of $\{1,\ldots,n\}$. Then $|A \cap B|2^n \geq |A||B|$.

Lecture 16
Tue. 4/5/11

§1 Four-function theorem

Theorem 16.1 (Ahlswede-Daykin four function theorem)**:** Suppose $n \geq 1$ and set $N = \{1,\ldots,n\}$. Let $P(N)$ be the power set of N. Let $\mathbb{R}_{\geq 0}$ be the set of nonnegative reals. For a function $\rho : P(N) \to \mathbb{R}_{\geq 0}$ and $\mathcal{A} \subseteq P(N)$, let

$$\rho(\mathcal{A}) = \sum_{A \in \mathcal{A}} \rho(A).$$

For $\mathcal{A}, \mathcal{B} \subseteq P(N)$, define

$$\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}$$
$$\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}.$$

Let $\alpha, \beta, \geq, \delta : P(N) \to \mathbb{R}_{\geq 0}$. If for every $A, B \subseteq N$,

$$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B), \tag{13}$$

then, for every $\mathcal{A}, \mathcal{B} \subseteq P(N)$,

$$\alpha(\mathcal{A})\beta(\mathcal{B}) \leq \gamma(\mathcal{A} \cup \mathcal{B})\delta(\mathcal{A} \cap \mathcal{B}). \tag{14}$$

Proof. We may modify $\alpha, \beta, \gamma, \delta$ by defining $\alpha(A) = 0$ for every $A \notin \mathcal{A}$ and $\beta(B) = 0$ for every $B \notin \mathcal{B}$, $\gamma(C) = 0$ for every $C \notin \mathcal{A} \cup \mathcal{B}$, and $\delta(D) = 0$ for every $D \notin \mathcal{A} \cap \mathcal{B}$. Note (13) still holds. (If $A \in \mathcal{A}, B \in \mathcal{B}$, neither the LHS and RHS changes. Else the LHS is 0.) Thus we may assume that $\mathcal{A} = \mathcal{B} = \mathcal{A} \cup \mathcal{B} = \mathcal{A} \cap \mathcal{B} = P(N)$.

We induct on n. For $n = 1$, $P(N) = \{\phi, N\}$. For each $\rho \in \{\alpha, \beta, \gamma, \delta\}$ let $\rho_0 = \rho(\varphi)$ and $\rho_1 = \rho(N)$. (13) gives

$$\begin{aligned}\alpha_0\beta_0 &\leq \gamma_0\delta_0\\ \alpha_0\beta_1 &\leq \gamma_1\delta_0\\ \alpha_1\beta_0 &\leq \gamma_1\delta_0\\ \alpha_1\beta_1 &\leq \gamma_1\delta_1.\end{aligned}$$

(Note the RHS do not range over all $\gamma_i\delta_j$.) We need $(\alpha_0 + \alpha_1)(\beta_0 + \beta_1) \leq (\gamma_0 + \gamma_1)(\delta_0 + \delta_1)$. If $\gamma_1 = 0$ or δ_0 we're good. Otherwise

$$\begin{aligned}\gamma_0 &\geq \frac{\alpha_0\beta_0}{\delta_0}\\ \delta_1 &\geq \frac{\alpha_1\beta_1}{\gamma_1}\end{aligned}$$

so it suffices to show $\left(\frac{\alpha_0\beta_0}{\delta_0} + \gamma_1\right)\left(\delta_0 + \frac{\alpha_1\beta_1}{\gamma_1}\right) \geq (\alpha_0+\alpha_1)(\beta_0+\beta_1)$. This is true as rearranging gives

$$(\gamma_1\delta_0 - \alpha_0\beta_1)(\gamma_1\delta_0 - \alpha_1\beta_0) \geq 0.$$

Suppose the theorem holds for $n-1$ ($n \geq 2$); we prove it for n. Put $N' = N\backslash\{n\}$ and define for each $\rho \in \{\alpha, \beta, \gamma, \delta\}, A \in P(N')$

$$\rho'(A) = \rho(A) + \rho(A \cup \{n\})$$

since this makes

$$\rho'(P(N')) = \rho(P(N)).$$

Note (13) would follow from applying the induction hypothesis to $\alpha', \beta', \gamma', \delta'$. To use the induction hypothesis we need to check (13) for these new functions: for all $A', B' \subseteq N'$,

$$\alpha'(A')\beta'(B') \leq \gamma'(A' \cup B')\delta'(A' \cap B').$$

Now

$$\begin{aligned}\overline{\alpha}(\phi) &= \alpha(A') & \overline{\alpha}(T) &= \alpha(A' \cup \{n\})\\ \overline{\beta}(\phi) &= \beta(B') & \overline{\beta}(T) &= \beta(B' \cup \{n\})\\ \overline{\gamma}(\phi) &= \gamma(A' \cup B') & \overline{\gamma}(T) &= \gamma(A' \cup B' \cup \{n\}\\ \overline{\delta}(\phi) &= \delta(A' \cap B') & \overline{\delta}(T) &= \delta(A' \cap B' \cup \{n\}.\end{aligned}$$

By (13)

$$\overline{\alpha}(S)\overline{\beta}(R) \leq \overline{\gamma}(S \cup R)\overline{\delta}(S \cap R)$$

for all $S, R \subseteq T$. Hence using the $n = 1$ case,

$$\begin{aligned}\alpha'(A')\beta'(B') &= \overline{\alpha}(P(T))\overline{\beta}(P(T)) \\ &\leq \overline{\gamma}(P(T))\overline{\delta}(P(T)) \\ &= \gamma'(A' \cup B')\delta'(A' \cap B').\end{aligned}$$

□

Definition 16.2: A **lattice** is a poset in which every two elements x, y have a unique minimal upper bound (join) $x \vee y$ and a unique maximal lower bound $c \wedge y$. A lattice is distributive if for all $x, y, z \in L$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

Equivalently, $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$. For $X, Y \subseteq L$, define

$$\begin{aligned}X \vee Y &= \{x \vee y : x \in X, y \in Y\} \\ X \wedge Y &= \{x \vee y : x \in X, y \in Y\}\end{aligned}$$

Any $L \subseteq P(N)$ where $N = \{1, \ldots, n\}$, where posets are ordered by inclusion, is a finite distributive lattice.Conversely, every finite distributive lattice is of this form.

Corollary 16.3 (Four function theorem for distributive lattices)**:** Let L be a finite distributive lattice and $\alpha, \beta, \gamma, \delta : L \to \mathbb{R}_{\geq 0}$. Then the same theorem holds.

The simplest case is $\alpha, \beta, \gamma, \delta = 1$. Then we get the following.

Corollary 16.4: Let L be a finite distributive lattice and $X, Y \subseteq L$. Then

$$|X||Y| \leq |X \vee Y| \cdot |X \wedge Y|.$$

Corollary 16.5: Let $\mathcal{A} \subseteq P(N)$ and $\mathcal{A}\backslash\mathcal{A} = \{A\backslash A' : A, A' \subseteq \mathcal{A}\}$. Then

$$|\mathcal{A}\backslash\mathcal{A}| \geq |\mathcal{A}|.$$

Proof. Let L be a finite distributive lattice on $P(N)$. By Corollary 16.4 with $\mathcal{B} = \{N\backslash F : F \in \mathcal{A}\}$,

$$|\mathcal{A}|^2 = |\mathcal{A}| \cdot |\mathcal{B}| \leq |\mathcal{A} \cup \mathcal{B}| \cdot |\mathcal{A} \cap \mathcal{B}| = |A\backslash\mathcal{A}|^2.$$

□

§2 FKG inequality

Definition 16.6: Let L be a finite distributive lattice. A function $\mu : L \to \mathbb{R}_{\geq 0}$ is called **log-supermodular** if $\mu(x)\mu(y) \leq \mu(x \wedge y)\mu(x \vee y)$, **increasing** if $\mu(x) \leq \mu(y)$ for all $x \leq y$ and **decreasing** if $\mu(x) \geq \mu(y)$ for all $x \leq y$.

Theorem 16.7 (FKG inequality)**:** Let L be a finite distributive lattice and $\mu : L \to \mathbb{R}_{\geq 0}$ be a log-supermodular function. Let $f, g : L \to \mathbb{R}_{\geq 0}$ be increasing functions. Then

$$\left(\sum_{x \in L} \mu(x) f(x)\right)\left(\sum_{x \in L} \mu(x) g(x)\right) \leq \left(\sum_{x \in L} \mu(x) f(x) g(x)\right)\left(\sum_{x \in L} \mu(x)\right).$$

Proof. Define $\alpha, \beta, \gamma, \delta$ by

$$\begin{aligned} \alpha(x) &= \mu(x) f(x) \\ \beta(x) &= \mu(x) g(x) \\ \gamma(x) &= \mu(x) f(x) g(x) \\ \delta(x) &= \mu(x). \end{aligned}$$

We claim these functions satisfy (13); then the conclusion follows from the four function theorem. Indeed,

$$\begin{aligned} \alpha(x)\beta(y) &= \mu(x) f(x) \mu(y) g(y) \\ &\leq \mu(x \wedge y) f(x) \mu(x \vee y) g(y) \\ &\leq \mu(x \wedge y) f(x \vee y) \mu(x \vee y) g(x \vee y) \\ &= \delta(x \wedge y) \gamma(x \vee y). \end{aligned}$$

□

Same holds if both f, g decreasing; just reverse γ, δ. If one is increasing and the other is decreasing, inequality reverses.

Lecture 17
Thu. 4/7/11

§1 Applications of FKG inequality

Above, it is helpful to view μ as a measure on L. We can define for any $f : L \to \mathbb{R}_{\geq 0}$ its expectation

$$\langle f \rangle = \frac{\sum_{x \in L} \mu(x) f(x)}{\sum_{x \in L} \mu(x)}.$$

With this notation we can write the FKG inequality as

$$\langle fg \rangle \geq \langle f \rangle \langle g \rangle.$$

By considering $P(N)$ as a probability space, $P(\mathcal{A}) = \frac{|\mathcal{A}|}{2^n}$.

Lemma 17.1 (Kleitman's lemma)**:** Let $\mathcal{A}, \mathcal{B} \subseteq P(N)$ be monotone increasing, and $\mathcal{C}, \mathcal{D} \subseteq P(N)$ be monotone decreasing. Then

$$P(\mathcal{A} \cap \mathcal{B}) \geq P(\mathcal{A})P(\mathcal{B})$$
$$P(\mathcal{C} \cap \mathcal{D}) \geq P(\mathcal{C})P(\mathcal{D})$$
$$P(\mathcal{A} \cap \mathcal{C}) \geq P(\mathcal{C})P(\mathcal{C}).$$

In other words, $2^n|\mathcal{A} \cap \mathcal{B}| \geq |\mathcal{A}||\mathcal{B}|$, etc.

Proof. Let $f : P(N) \to \mathbb{R}_{\geq 0}$ be the characteristic function of $\mathcal{A}$. Let g be the characteristic function of $\mathcal{B}$ and $\mu \equiv 1$, which is log-supermodular. Applying FKG gives

$$P(\mathcal{A} \cap \mathcal{B}) = \langle fg \rangle \geq \langle f \rangle \langle g \rangle = P(\mathcal{A})P(\mathcal{B}).$$

The others follow similarly. □

For a real vector $(p_1, \ldots, p_n)$ with $0 \leq p_i \leq 1$, consider the probability space where for each $A \subseteq N$,

$$P(A) = \prod_{i \in A} p_i \prod_{i \notin A} (1 - p_i),$$

obtained by picking each $i \in N$ with probability p_i independently of the other elements.

For each $\mathcal{A} \subseteq P(N)$, let $P_p(\mathcal{A})$ denote its probability in this space. Define $\mu = \mu_p$ by $\mu(A) = P_p(A)$. Note μ is log-supermodular; in fact $\mu(A)\mu(B) = \mu(A \cup B)\mu(A \cap B)$. Thus Kleitman's lemma generalized to the following.

Theorem 17.2: For any $p = (p_1, \ldots, p_n)$, $P_p(\mathcal{A} \cap \mathcal{B}) \geq P_p(\mathcal{A})P_p(\mathcal{B})$, and similarly with the other inequalities.

For example, suppose $A_1, \ldots, A_k$ are arbitrary subsets of N and suppose we pick $A \subseteq N$ by choosing each $i \in N$ with probability p independent of the other elements. Then applying the theorem $k - 1$ times, (the family of sets intersecting some A_i is monotone increasing)

$$P(A \text{ intersects each } A_i) \geq \prod_{i=1}^{k} P(A \text{ intersects } A_i).$$

Note this fails if we pick a subset of a random set uniformly at random. By viewing N as the $n = \binom{m}{2}$ edges on $V = \{1, \ldots, m\}$, we can get a correlation inequality for random graphs. Let $G = G(m, p)$. A property of graphs Q is **monotone increasing** if whenever G has Q and whenever H is obtained from G by adding edges, then H has Q as well. For monotone decreasing, replace "adding edges" with "deleting edges."

Theorem 17.3: Let Q_1, Q_2, Q_3, Q_4 be graph properties. Let Q_1, Q_2 be monotone increasing and Q_3, Q_4 be monotone decreasing. Let $G = G(m, p)$. Then

$$P(G \in Q_1 \cap Q_2) \geq P(G \in Q_1)P(G \in Q_2)$$
$$P(G \in Q_3 \cap Q_4) \geq P(G \in Q_3)P(G \in Q_4)$$
$$P(G \in Q_1 \cap Q_3) \leq P(G \in Q_1)P(G \in Q_3)$$

Example 17.4: Let Q be the Hamiltonian property; it is monotone increasing. Let P be the planarity property; it is monotone decreasing. Then

$$P(G \in P \cap H) \leq P(G \in H)P(G \in P).$$

(The number of labeled planar graphs on n vertices is asymptotic to $\alpha n^{\beta}\gamma^{n}$.)

Definition 17.5: A **linear extension** of a poset is a total ordering the preserves inequality relations.

Theorem 17.6 (XYZ Theorem)**:** Let P be a poset with n elements $a_1, \ldots, a_n$. For a uniformly random linear extension,

$$P(a_1 \leq a_2 \wedge a_1 \leq a_3) \geq P(a_1 \leq a_2)P(a_1 \leq a_3).$$

Lecture 18
Tue. 4/12/11

§1 Pseudorandomness

Here are two examples where pseudorandomness comes into play.

Theorem 18.1 (Szemerédi)**:** Every set $A \subseteq \mathbb{N}$ of positive density contains arbitrarily long arithmetic progressions. (The density is $\limsup_{n\to\infty} \frac{|A\cap[1,n]|}{n}$.)

Proof. (Sketch) There are proofs using graph theory (Szemerédi regularity), ergodic theory, Fourier analysis (gives estimates, Gowers), and hypergraph regularity.

If the set acts like a random set, then it has long arithmetic progression Otherwise the density increments a lot somewhere, and we have a subsequence with fewer elements and greater density; apply induction. □

Theorem 18.2 (Green-Tao)**:** The primes contain arbitrarily long arithmetic progressions.

Proof. (Sketch) The primes form a dense subset of a pseudorandom set R, and behaves much like a dense subset of $\mathbb{N}$. □

Probabilistic methods give existence, but we prefer an explicit construction, because it can increase our understanding of the problem, and more importantly, can be used in algorithms. By an explicit construction we mean an algorithm, guaranteed to work, in polynomial time with respect to the parameters of the desired structure.

For example, find an explicit construction for a Ramsey graph, a graph of n vertices that has no clique or independent set of size $2\log_2 n$. Erdős's question, to find (an explicit construction of) a family of graphs G whose largest clique or independent set has order $O(\log|G|)$, is still unanswered.

Explicit constructions are known for several other problems.

§2 Quadratic residue tournament

Definition 18.3: A **tournament** (V, E) is a directed graph with one edge between every pair of vertices, i.e. an orientation of a complete graph. We say x beats y to mean $(x, y) \in E$. Given permutation π of V, $(x, y) \in E$ is consistent if x precedes y in π.

Find a ranking with the most consistent edges.

Let $c(\pi, T)$ be the number of consistent edges of T with respect to π, and let $c(\pi) = \max c(\pi, T)$. Note $c(T) \geq \frac{1}{2}\binom{n}{2}$. (If π' is the reverse of π, then $c(\pi) + c(\pi') = \binom{n}{2}$.) As an exercise, show $c(T) = \frac{1}{2} + \Omega(n^{\frac{3}{2}})$.

The probabilistic method shows there exists T with $c(T) = \left(\frac{1}{2} + o(1)\right)\binom{n}{2}$; a more involved proof shows $c(T) = \frac{1}{2}\binom{n}{2} + O(n^{\frac{3}{2}})$.

Example 18.4: Find T for which $c(T)$ is small.

Solution. Let $p \equiv 3 \pmod 4$ be a prime and $T = T_p$ be a tournament on $\mathbb{F}_p$. Let (i, j) be a directed edge iff $j - i$ is a square modulo p. (-1 is not a perfect square.)

Theorem 18.5:

$$c(T_p) = \frac{1}{2}\binom{p}{2} + O(p^{\frac{3}{2}} \ln p).$$

Proof. For $y \in \mathbb{F}_p$, define

$$\chi(y) = y^{\frac{p-1}{2}} = \begin{cases} 0, & \text{if } y = 0 \\ 1, & \text{if } y \text{ a quadratic residue} \\ -1, & \text{else} \end{cases}$$

Let $D = [d_{ij}]_{i,j=0}^{p-1}$ be an exponential matrix with $d_{ij} = \chi(ij - j)$. For distinct j, l,

$$\begin{aligned} \sum_{i \in \mathbb{F}_p} d_{ij} d_{il} &= \sum_i \chi(i-j)\chi(i-l) \\ &= \sum_{i \neq j,l} \chi(i-j)\chi(i-l) \\ &= \sum_{i \neq j,l} \chi\left(\frac{i-j}{i-l}\right) \\ &= \sum_{i \neq j,l} \chi\left(1 + \frac{l-j}{i-l}\right) \\ &= \sum_{t \neq 0,1} \chi(t) \\ &= 0 - \chi(0) - \chi(1) = -1 \end{aligned}$$

since $\frac{l-j}{i-l}$ runs through everthing except 0,1. For $A, B \subseteq \mathbb{F}_p$, let $e(A, B)$ be the number of edges $(a, b) \in A \times B$. Then

$$\sum_{i \in A} \sum_{j \in B} d_{ij} = e(A, B) - e(B, A).$$

Lemma 18.6:

$$\left|\sum_{i\in A}\sum_{j\in B} d_{ij}\right| \le |A|^{\frac{1}{2}}|B|^{\frac{1}{2}}p^{\frac{1}{2}}.$$

Proof. By Cauchy-Schwarz,

$$\begin{aligned}\left(\sum_{i\in A}\sum_{j\in B} d_{ij}\right)^2 &\le |A|\sum_{i\in A}\left(\sum_{j\in B} d_{ij}\right)^2 \\ &\le |A|\sum_{i\in \mathbb{F}_p}\left(\sum_{j\in B} d_{ij}\right)^2 \\ &\le |A|\sum_{i\in \mathbb{F}_p}\left(|B| + 2\sum_{j<l;j,l\in B} d_{ij}d_{il}\right) \\ &\le |A||B|p + 2\sum_{j<l;j,l\in B}\sum_{i\in\mathbb{F}_p} d_{ij}d_{il} \\ &\le |A||B|p - 2\binom{|B|}{2} \le |A||B|p,\end{aligned}$$

where we used $\sum_{i\in\mathbb{F}_p} d_{ij}d_{il} = -1$. Take square roots. □

Let r be the smallest integer such that $2^r \ge p$. Let π be arbitrary permutation of T_p, $\pi = (\pi_1, \ldots, \pi_p)$ and let $\pi' = (\pi_p, \ldots, \pi_1)$. We need to show $c(\pi, T_p) \le \frac{1}{2}\binom{p}{2} + O(p^{\frac{3}{2}}\ln p)$, or equivalently $c(\pi, T_p) - c(\pi', T_p) = O(p^{\frac{3}{2}}\ln p)$. Take a_1, a_2 such that $p = a_1 + a_2$ and $a_1, a_2 \le 2^{r-1}$. Let A_1 be the set of first a_1 elements, and A_2 be the last a_2 elements. By Lemma (18.6), $e(A_1, A_2) - e(A_2, A_1) \le (a_1a_2p)^{\frac{1}{2}} \le 2^{r-1}p^{\frac{1}{2}}$. Now continue to partition A_1 and A_2 into two sets, and of size at most 2^{r-2}, and so on. Continuing until we are down to singleton sets and adding up, $c(\pi, T_p) - c(\pi', T_p) \le 2^{r-1}p^{\frac{1}{2}}r = O(p^{\frac{3}{2}}\ln p)$. □

The quadratic residue tournament also gives vertices without k-dominating sets.

Lecture 19
Thu. 4/14/11

§1 Eigenvalues and expanders

Definition 19.1: $G = (V, E)$ is called a (n, d, c)-expander if

1. $|V| = n$,
2. it is d-regular, and

3. for every $W \subseteq V$ with $N(W) \leq \frac{n}{2}$, $|N(W)| \geq c|W|$. (Here $N(W)$ is the set of all $v \in V \backslash W$ adjacent to some vertex in W.

Definition 19.2: A family of linear expanders of density (or degree) d and expansion c is a sequence $\{G_i\}_{i=1}^{\infty}$ of (n_i, d, c)-expanders with $n_i \to \infty$ as $i \to \infty$.

Note similar definitions can be made for graphs with maximum degree at most d (but the regular case is nicer). This has many applications in theoretical computer science.

Definition 19.3: Let $G = (V, E)$ be a graph with n vertices. Its **adjacency matrix** $A = A_G$ has

$$a_{ij} = \begin{cases} 1 & \text{if } ij \in E \\ 0 & \text{if } ij \notin E. \end{cases}$$

Note A is symmetric, so has real eigenvalues and a complete set of real orthogonal eigenvectors.

Suppose G is d-regular. Then the largest eigenvalue of A is d: indeed, the all ones vector is associated to the eigenvalue d; d is the largest since the sum of all entries of A^p is the total number of walks of length p (which is nd^p), which is at least the number of closed walks of length p (which is $\mathrm{Tr}(A^p)$). Hence $nd^p \geq \sum_{i=1}^{n} \lambda_i^p$ where λ_i are the eigenvalues. Or use Perron-Frobenius.

Let $\lambda = \lambda(G)$ be the second largest eigenvalue. For subsets $B, C \subseteq V$ let $c(B, C)$ be the number of ordered pairs $(b, c) \in B \times C$ which are edges.

Theorem 19.4: For every partition $V = B \cup C$, $e(B, C) \geq \frac{(d-\lambda)|B||C|}{n}$.

Proof. Let $|V| = n$, $b = |B|$, and $c = |K| = n - b$. Let $D = dI$. Consider

$$\begin{aligned} \langle (D - A)x, x \rangle &= \sum_{u \in V} \left(d(x(u))^2 - \sum_{v \in N(u)} x(u)x(v) \right) \\ &= d \sum_{u \in V} x(v)^2 - \sum_{uv \in E} x(u)x(v) \\ &= \sum_{uv \in E} (x(u) - x(v))^2. \end{aligned}$$

Define x by $x(v) = -c$ for all $v \in B$, $x(v) = b$ for all $v \in C$, and $x(v) = 0$ for all other values. Note $\sum_{v \in V} x(v) = 0$.

We claim that A and $D - A$ have the same eigenvalues. Note if μ is an eigenvalue of A then $d - \mu$ is an eigenvalue of $D - A$. Note x is orthogonal to the (constant) eigenvector corresponding to the smallest eigenvalue 0 of $D - A$. The eigenvectors of $D - A$ are orthogonal and form a basis for $\mathbb{R}^n$. Now x is a linear combination of the other eigenvectors. Since $d - \lambda$ is the second smallest smallest eigenvalue of $D - A$,

$$\langle (D - A)x, x \rangle \geq (d - \lambda)\langle x, x \rangle = (d - \lambda)(bc^2 + cb^2) = (d - \lambda)bcn.$$

But choosing x as mentioned, the LHS is $e(B, C)(b + c)^2$; divide by $(b + c)^2$. □

Corollary 19.5: Keeping the same assumptions, G is a (n, d, c)-expander with

$$c = \frac{d - \lambda}{2d}.$$

Proof. Let $|B| \leq \frac{n}{2}$. Let $C = \overline{B}$. The above shows that $e(B, C) \geq \frac{(d-\lambda)|B||C|}{n} \geq \frac{(d-\lambda)|B|}{2}$. Since G is d-regular,

$$|N(B)| \geq \frac{(d - \lambda)|B|}{2d}.$$

□

Alon improved this to $c = \frac{2(d-\lambda)}{3d-2\lambda}$.

Theorem 19.6: If G is a (n, d, c)-expander then $\lambda \leq d - \frac{d^2}{4+2c^2}$.

How small can λ be?

Theorem 19.7 (Alon Nilli)**:**

$$\lambda \geq 2\sqrt{d-1}\left(1 - O\left(\frac{1}{\operatorname{diam} k}\right)\right).$$

(Alon Nilli is a pseudonym of Noga Alon.) A Ramanujan graph is a graph with $\lambda \leq 2\sqrt{d-1}$.

Theorem 19.8 (Expander mixing lemma)**:** Let G be a d-regular graph. For every $B, C \subseteq V(G)$, let λ be the eigenvalue with second largest absolute value. Suppose $|B| = bn$, $|C| = cn$. Then

$$|e(B, C) - bcdn| \leq \lambda\sqrt{bc}n.$$

Corollary 19.9:

$$\left|e(B) - \frac{1}{2}b^2dn\right| \leq \frac{1}{2}\lambda bn.$$

(Take $B = C$, $e(B, C) = 2e(B)$.)

The number of walks of length p is nd^p. What if we want walks missing a linear-size subset? For an expander, it's exponentially smaller. (If we restrict to a subset of half the size, it is "approximately" a $\frac{d}{2}$-regular graph on half the vertices.) This is useful in algorithms, especially in Monte Carlo for amplification, for example, in primality testing.

Lecture 20
Thu. 4/21/11

§1 Quasi-random graphs

Definition 20.1: For graphs G, H, define $N^*_G(H)$ to be the number of labeled induced copies of H in G. Define $N_G(H)$ to be the number of labeled copies of H in G (not necessarily induced).

Note

$$N_G(H) = \sum_{L \text{ contains copy of } H} N_G^*(L).$$

(The sum is over L obtained by adding edges to H.) Suppose the eigenvalues of the adjacency matrix are $\lambda_1, \ldots, \lambda_n$ with $|\lambda_1| \geq \cdots \geq |\lambda_n|$. For a vertex v of G and $S \subseteq V(G)$, let

1. $N(v)$ be the set of neighbors of v
2. $e(S)$ be the set of edges inside S
3. $e(B, C)$ be the number of pairs in $B \times C$ which are edges, so $e(S, S) = 2e(S)$.

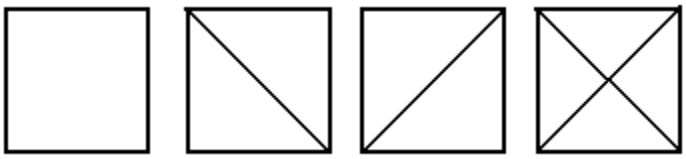
Definition 20.2: Define the following properties of random graphs:

1. $P_1(s)$: For every H on s vertices, $N_G^*(H) = (1 + o(1))n^s 2^{-\binom{s}{2}}$. (Expected number of copies of H)
2. P_2: $N_G(C_4) \leq (1 + o(1)) \left(\frac{n}{2}\right)^4$. (Expected number of 4-cycles)
3. P_3: $|\lambda_2| = o(n)$. (Second eigenvalue small)
4. P_4: For every $S \subseteq V(G)$, $e(S) = \frac{1}{4}|S|^2 + o(n^2)$ (Edges uniformly distributed)
5. P_5: For every $S, T \subseteq V(G)$, $e(S, T) = \frac{1}{2}|S||T| + o(n^2)$.
6. P_6: $\sum_{u,v \in V} \left||N(u) \cap N(v)| - \frac{1}{n}4\right| = o(n^3)$.

Theorem 20.3: All properties are equivalent for d-regular graph on n vertices with $d = \left(\frac{1}{2} + o(1)\right) n$.

Proof. $P_1(4) \implies P_2 \implies P_3 \implies P_4 \implies P_5 \implies P_2$ and $P_5 \implies P_1(s)$ for all s.

$P_1(4) \Rightarrow P_2$: We have the right count for each of the following graphs.



Then

$$N_G(C_4) = \sum_L N_G^*(L) = 4(1 + o(1))n^4 2^{-6}.$$

$P_2 \Rightarrow P_3$: Note $\mathrm{Tr}(A^4) = N_G(C_4) + O(n^3) \leq \left(\frac{n}{2}\right)^2 + o(n^4)$ (it counts the number of 4-cycles plus degenerate 4-cycles). But $\mathrm{Tr}(A^4) = \sum_{i=1}^n \lambda_i^4 \geq \lambda_1^4 + \lambda_2^4$. Since $\lambda_1 = d = \left(\frac{1}{2} + o(1)\right) n$, $\lambda_2^4 = o(n^4)$, giving $|\lambda_2| = o(n)$.

$P_3 \Rightarrow P_4$: Proof omitted.

$P_4 \Rightarrow P_5$: First suppose S and T are disjoint. Then

$$e(S,T) = e(S \cup T) - e(S) - e(T) = \frac{1}{4}(|S|+|T|)^2 - \frac{|S|^2}{4} - \frac{|T|^2}{4} + O(n^2) = \frac{1}{2}|S||T| + o(n^2).$$

If they aren't disjoint, rewrite in terms of the three sets $S \backslash T, T \backslash S, S \cap T$.

$P_5 \Rightarrow P_6$: Since G is d-regular with $d = \left(\frac{1}{2} + o(1)\right) n$. Fix $v \in G$ and let

$$B_1 = \left\{u : |N(u) \cap N(v)| \geq \frac{n}{4}\right\}$$
$$B_2 = \left\{u : |N(u) \cap N(v)| < \frac{n}{4}\right\}.$$

Let $C = N(v)$. Now

$$\begin{aligned}\sum_{u \in B_1} \left||N(u) \cap N(v)| - \frac{n}{4}\right| &= \sum_{u \in B_1} \left(|N(u) \cap N(v)| - \frac{n}{4}\right) \\ &= e(B_1, C) - \frac{n}{4}|B_1| \\ &= \frac{1}{2}|B_1|d + o(n^2) - |B_1|\frac{n}{4} \\ &= o(n^2).\end{aligned}$$

Similarly $\sum_{u \in B_2} \left||N(u) \cap N(v)| - \frac{n}{4}\right| = o(n^2)$. Now sum over $v \in V$.

$P_6 \Rightarrow P_2$: The number of walks of length 4 is $\sum_{u,v} |N(u) \cap N(v)|^2 = n^2 \left(\frac{n}{4}\right)^2 + o(n^4)$. But the LHS is $N_G(C_4) + O(n^3)$.

$P_5 \Rightarrow P_1$: We try to build copies of H, one vertex at a time.

FIXFIX

Suppose H has vertex set $[s]$. Then (v_a, v_b) is an edge iff (a, b) is an edge. After i steps, we have picked a walk $v_1, \ldots, v_i$ and subsets V_j^i with $|V_j^i| = (1 + o(1))2^{-i}n$ and V_j^i being the set of vertices connected to v_i. Now (v_a, u) with $u \in V_j^i$ is an edge iff (a, j) is an edge. Pick $v_{i+1} \in V_{i+1}^i$. In total the number of copies is

$$\prod_{i=1}^{s} 2^{1-i} n(1 + o(1)) = (1 + o(1))n^2 2^{-\binom{s}{2}}.$$

□

Definition 20.4: A quasirandom graph is one with about half the edges and satisfying any of the following equivalent properties.

Lecture 21
Tue. 4/26/11

§1 Dependent Random Choice

Suppose H is sparse or small, G is larger than H and dense, and we want to show that H is a subgraph of G. It is helpful to find a "rich" subset U that is large and such that all, or almost all, small subsets of U have many common neighbors. Then we can embed H. We assume H is bipartite (but the method can be adapted for nonbipartite graphs), and embed H one vertex at a time in U.

We don't want to take vertices independently from one another (think of the case that G is a union of two cliques). Instead, we pick a small random set of vertices T and let $U = N(T)$, where $N(T)$ denote the vertices adjacent to all vertices of T. This favors subsets with large common neighborhood.

Lemma 21.1 (Dependent random choice)**:** Let $a, d, m, n, r \in \mathbb{N}$. Let $G = (V, E)$ be a graph with $|V| = n$ vertices and average degree $d = \frac{2|E(G)|}{n}$. If there is $t \in \mathbb{N}$ such that

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a$$

then G contains a subset U of size at least a such that every r vertices in U have at least m common neighbors.

Proof. We we the alteration method; pick U as noted and then delete a few vertices.

Pick a set Γ of t vertices at random with repetition. Set $A = N(\Gamma)$ and $X = |A|$. Then because the probability that a given vertex is a neighbor of v is $\frac{|N(v)|}{n}$,

$$\mathbb{E}(X) = \sum_{v \in G}\left(\frac{|N(v)|}{n}\right)^t = n^{-t}\sum_{v\in G}|N(v)|^t \geq n^{1-t}\left(\frac{\sum_{v\in G}|N(v)|}{n}\right)^t = \frac{d^t}{n^{t-1}}$$

using Power Mean.

Let Y be the number of r-sets $S \subseteq A$ with $|N(S)| < m$. For a given S,

$$P(S \subseteq A) = \left(\frac{|N(S)|}{n}\right)^t$$

(if $S \subseteq A$, the vertices of T have to be adjacent to all vertices in S) and

$$\mathbb{E}(Y) < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Now $\mathbb{E}(X - Y) > a$. Fix a choice of T such that $X - Y > a$. For each $S \subseteq A$ with $|S| = r$ and $|N(S)| < m$, delete a vertex in S from A. Let U be the remaining subset. □

Definition 21.2: $\mathrm{ex}(n, H)$ is the maximum number of edges of a graph on n vertices with no copy of H.

For example, by Turan's Theorem, $\mathrm{ex}(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor$. For H not bipartite, this gives an asymptotic formula for ex.

Theorem 21.3:

$$\mathrm{ex}(H) = \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right) \binom{n}{2}.$$

Theorem 21.4: If $H = (A \cup B, E)$ is bipartite graph in which all vertices in B have degree at most r, then

$$\mathrm{ex}(n, H) \leq cn^{2-\frac{1}{r}}, \quad c = c(H).$$

If $H = K_{r,s}$ and $s \geq r!$ then

$$\mathrm{ex}(n, H) = \Theta(n^{2-\frac{1}{r}}).$$

Proof. Let $a = |A|$, $b = |B|$, $m = a + b$, $t = r$, $c = \max(a^{\frac{1}{r}}, \frac{3(a+b)}{r})$. Suppose G has n vertices and at least $cn^{2-\frac{1}{r}}$ edges. From the lemma, there exists U with $|U| \geq a$ such that all subsets of size r have at least $m = a + b$ common neighbors. So it suffices to show the following.

Lemma 21.5: Let $H = (A \cup B, E)$ be bipartite, such that $a = |A|, b = |B|$, and all vertices of B have degree at most r. If G has a subset U with $|U| \geq a$ and all subsets $S \subseteq U$ with $|S| = r$ have $|N(S)| \geq m = a + b$, then H is a subgraph of G.

Proof. We find an embedding $f : A \cup B \to V(G)$. Map $f : A \to U$ arbitrarily injectively. Label the vertices of B by $v_1, \ldots, v_b$. We need $f(v_i)$ to be adjacent to all $f(a)$ with $a \in N(v_i)$. The condition on U allows us to greedily embed. □

The second part comes from a construction from algebraic geometry. □

Let Q_r denote the r-dimensional cube on 2^r vertices. Its vertices are in $\{0, 1\}^r$, with two vertices adjacent if they differ in one position. Note Q_r ir r-regular. The Burr-Erdös conjecture is that $r(Q_r) = O(2^r) = O(|V(Q_r)|)$.

Theorem 21.6:

$$r(Q_r) \leq 2^{3r}.$$

Proof. Set $N = 2^{3r}$. The denser color has at least $\frac{1}{2}\binom{N}{2} \geq 2^{-\frac{7}{3}}N^2$ edges. Let G be the graph of this color. The average degree is at least $2^{-\frac{4}{3}N}$. Let $t = \frac{3}{2}r, m = 2^r, a = 2^{r-1}$. Then the inequality in Lemma 21.1 is satified so there is $U \subseteq V(G)$ with $|U| \geq a$, every r vertices of U have at least $m = 2^r$ common neighbors. Then Q_r is a subgraph of this denser color. □

Note we didn't use any properties of Q_r except it's bipartite, the number of its vertices and the maximum degree of a vertex.

We just need "almost all" subsets to have large common neighborhood; this gives tighter bounds.

Theorem 21.7: There exists G with n vertices with edge density $\frac{1}{2}$ such that for any $U \subseteq V(G)$ with $|U| = |\Omega(n)|$, there exist $u, v \in U$ with only $o(n)$ common neighbors.

This shows the limitations of dependent random choice.

Lecture 22
Thu. 4/28/11

§1 Dependent Random Choice

For a number of application, "almost all" small subsets of U having many common neighbors is enough.

Lemma 22.1: Let $G = (A, B, E)$ by a bipartite graph and $|E| = c|A||B|$. For every $0 < \varepsilon < 1$ there exists $U \subseteq A$ such that $|U| \geq \frac{c|A|}{2}$ and at least a $(1-\varepsilon)$ fraction of the ordered pairs in U have at least $\frac{\varepsilon c^2|B|}{2}$ common neighbors.

Proof. Pick $v \in B$ uniformly at random. Let $X = |N(v)|$. We bound $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) \geq \mathbb{E}(X)^2 = (c|A|)^2.$$

Let $T = (a_1, a_2) \in A \times A$ and call T bad if $|N(T)| < \frac{\varepsilon c^2|B|}{2}$. Now choosing $v \in B$ at random,

$$P(T \subseteq N(v)) = \frac{|N(T)|}{|B|}$$

since there are $|B|$ possibilities for B and $N(T)$ of them are adjacent to both elements of T. If T is bad, this probability is less than $\frac{\varepsilon c^2}{2}$.

Let Z be the number of bad pairs in $N(v)$. We have

$$\mathbb{E}(Z) \leq \frac{\varepsilon c^2}{2} A^2$$

(since the probability of a bad T being in $N(v)$ is at most $\frac{\varepsilon c^2}{2}$ and there are $|A|^2$ pairs) and

$$\mathbb{E}\left(X^2 - \frac{Z}{\varepsilon}\right) = \mathbb{E}(X^2) - \frac{1}{\varepsilon}\mathbb{E}(Z) \geq \frac{c^2|A|^2}{2}.$$

Thus there exists v so that $X^2 - \frac{Z}{\varepsilon} \geq \frac{c^2|A|^2}{2}$. Then $|X| \geq \frac{c|A|}{2}$ and $|Z| \leq \varepsilon X^2$. Set $U = N(v)$. □

This can be generalized to n-tuples instead of pairs.

For A, B sets of integers, define the sumset and partial sumsets

$$A + B = \{a + b \mid a \in A, n, b \in B\}$$
$$A \overset{G}{+} B = \{\, a + b |\, (a, b) \in E\}\,.$$

For A an arithmetic proression of length $\frac{n}{2}$ plus $\frac{n}{2}$ "random vertices," $|A \overset{G}{+} A| = O(n)$ and $|A + A| = \Omega(n^2)$.

Theorem 22.2 (Balog-Szemerédi)**:** If $|A| = |B| = n$, G has at least cn^2 edges and $|A \overset{G}{+} B| \leq Cn^2$, then there exists $A' \subseteq A$ and $B' \subseteq B$ such that $|A'|, |B'| \geq c'n$, and $|A' + B'| \leq C'n$, where c' and C' depend on c and C alone.

Lemma 22.3: Let $G = (A, B, E)$ with $|A| = |B| = n$ and $|E| = cn^2$. Then there exist $A' \subseteq A, B' \subseteq B$, each of size at least $\frac{cn}{8}$ and there are at least $2^{-12}c^5n^2$ paths of length 3 between any $a' \in A'$ and $b' \in B'$.

Proof. (of 22.2 given 22.3) Let A', B' be as in the lemma above. Given $y \in A' + B'$, we can find $a \in A'$ and $b \in B'$ such that $y = a + b$. To each such y there correspond at least $2^{-12}c^5n^2$ pairs (a', b') such that a, b', a', b is a path. Writing

$$y = (a + b) = \underbrace{(a + b')}_{x} - \underbrace{(b' + a')}_{x'} + \underbrace{(a' + b)}_{x''},$$

we have that each $y \in A' + B'$ corresponds to $2^{-12}c^5n^2$ triplets in $(A \overset{G}{+} B)^3 =: X^3$. Moreover the triplets corresponding to different y are distinct.

Since $|X| \leq Cn$ there are at most C^3n^3 such triples. Then

$$|A' + B'| = \frac{|X|^3}{2^{-12}c^5n^2} \leq 2^{12}C^3c^{-5}n.$$

□

Proof. (of Lemma 22.3) Let $A_1 \subseteq A$ consist of vertices of degree at least $\frac{cn}{2}$. Let $c_1 = \frac{e(A_1,B)}{|A_1||B|}$. Note $e(A_1, B) \geq \frac{cn^2}{2}$. Also $c_1 \geq c$, $c_1 \geq \frac{cn^2/2}{|A_1||B|} = \frac{cn/2}{|A_1|}$.

Apply Lemma 22.1 to A_1, B (c replaced by c_1, $\varepsilon = \frac{c}{16}$) to get $U \subseteq A_1$ with $|U| \geq \frac{c_1|A_1|}{2} \geq \frac{cn}{4}$ and at most $\frac{c|U|^2}{16}$ ordered pairs in U are "bad," i.e. have less than $\frac{\varepsilon c_1^2 n}{2} \geq \frac{c^3 n}{32}$ common neighbors.

Let $A' \subseteq U$ be those vertices a in at most $\frac{c|U|}{8}$ bad pairs (a, a'). The number of bad pairs is at least $|U \backslash A'| \cdot \frac{c|U|}{8}$, giving $|U \backslash A'| \leq \frac{|U|}{2}$ and $|A'| \geq \frac{|U|}{2} \geq \frac{cn}{8}$.

Let B' be the set of vertices in B with at least $\frac{c|U|}{4}$ neighbors in U. By counting and choice of U, $|B'| \geq \frac{cn}{4}$. Pick $a \in A'$ and $b \in B'$. By choice of A', a has common neighbors with all but a small fraction of U, and b has many neigbhors in U. This gives paths of length 3. The theorem follows after some calculation.

□

Lecture 23
Tue. 5/3/11

§1 Crossing number, incidences, and sum-product estimates

Definition 23.1: For a graph $G = (V, E)$ let the **crossing number** $\mathrm{cr}(G)$ be the minimum number of crossings in any drawing of G.

Note the following facts:

1. (Farey's theorem) If a planar graph can be represented using curves, then it can be represented using line segments. (Disk representation theorem) Any planar graph can be represented as nonoverlapping disks, with adjacency if they touch.
2. $\mathrm{cr}(G) = 0$ iff G is planar.
3. $\mathrm{cr}(G) \leq \binom{|E|}{2}$.
4. If $n \geq 3$ and G is planar, $m \leq 3n - 6$. (Consider a triangulation.) For any planar G, $m \leq 3n$.
5. $\mathrm{cr}(G) \geq m - 3n$. (Keep pulling out edges one by one.)

We use the probabilistic method to amplify this simple bound and prove the following.

Theorem 23.2 (Crossing lemma)**:** If G has $m \geq 4n$ edges then

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

Proof. Let $t = \mathrm{cr}(H)$

Pick each vertex with probability p independent of the other vertices. Let H be the induced subgraph with these picked vertices. Then

$$\begin{aligned} \mathbb{E}(|V(H)|) &= pn \\ \mathbb{E}(|E(H)|) &= p^2 m \\ \mathbb{E}(\mathrm{cr}(H)) &\leq p^4 t \end{aligned}$$

The inequality comes from the fact that for a crossing to appear in H, the four vertices involved must be in H. Using $\mathrm{cr}(H) \geq |E(H)| - 3|V|$, we get

$$p^4 t \geq \mathbb{E}(\mathrm{cr}(H)) \geq \mathbb{E}(|E(H)|) - 3\mathbb{E}(|V(H)|) = p^2 m - pn.$$

or

$$t \geq p^{-2} m - 3p^{-3} n.$$

To maximize the right-hand side take $p = \frac{4n}{m}$; this gives the desired result. □

Theorem 23.3 (Szemerédi-Trotter incidence theorem)**:** Let P be a set of points in $\mathbb{R}^2$ and L be a set of lines in $\mathbb{R}^2$. Let $l(P, L)$ be the set of pairs $(p, \ell) \in P \times L$ which are incident. Then there is a constant c such that

$$l(P, L) \leq 4(m^{\frac{2}{3}} n^{\frac{2}{3}} + m + n)$$

where $m = |L|$ and $n = |P|$.

Note that the dominant term depends on the relative sizes of m and n.

Proof. We can assume all lines have at least one point and all points are on at least one line. Consider the graph (V, E) where V is the set of points, $E = (p, p')$ is an edge if p and p' are the closest points on a line $\ell \in L$. Then $|V| = n$ and $|I| = |E| + |L|$. The graph is embedded in the plane, and the number of crossings is at most the number of pairs of lines. If $|E| < 4n$, then $|I| \leq m_4 n$; else by the crossing lemma,

$$\frac{(|I| - m)^3}{64n^2} \leq \operatorname{cr}(G) \leq \binom{m}{2} \leq \frac{m^2}{2}$$

and the bound follows:

$$\begin{aligned}(|I| - m)^3 &\leq 32m^2n^2 \\ |I| &\leq 32^{\frac{1}{3}} m^{\frac{2}{3}} n^{\frac{2}{3}} + m.\end{aligned}$$

The extra n comes from removing the initial assumption. □

Example 23.4 (Unit distance problem)**:** Given n points in $\mathbb{R}^2$, how many unit distances can you have?

Solution. (Sketch) Look at unit circles around the points, and say two vertices are adjacent if they are adjacent in a circle. The maximum number of edges between any two vertices is at most 2 (we're counting them once for each circle they're in); use a variant of the crossing lemma for nonsimple graphs.

For $A, B \subseteq \mathbb{R}$,

$$\begin{aligned}A + B &= \{a + b \mid a \in A, b \in B\} \\ A \cdot B &= \{a \cdot b \mid a \in A, b \in B\}.\end{aligned}$$

Let $|A| = n$. Then $|A + A| \geq 2n - 1$ with equality iff A is an arithmetic progression. To prove this note that if $a_1 < \ldots < a_n$ then

$$a_1 + a_1 < a_1 + a_2 < \cdots < a_1 + a_n < a_2 + a_n < \cdots < a_n + a_n.$$

The following is a generalization.

Theorem 23.5 (Freiman's Theorem)**:** If $|A + A| \leq Cn$ then A is a dense subset of a generalized arithmetic progression.

By taking exponentials, we get a similar result for geometric progressions and $A \cdot A$. The following says that we can't have a set with $A + A$ and $A \cdot A$ both small, i.e. looking both like an arithmetic and geometric progression.

Theorem 23.6 (Elekes)**:**

$$\max(|A + A|, |A \cdot A|) \geq c|A|^{\frac{5}{4}}.$$

Proof. Let $\ell_{a,b}$ be the line $y = a(x - b)$. Note $\ell_{a,b}$ contains the point $(c + b, c \cdot a) \in P$ for all $c \in A$. Let $L = \{\ell_{a,b} \mid a, b \in A\}$; note $|L| = |A|^2$. Let $P = (A + A) \times (A \cdot A)$; then $|P| = |A + A||A \cdot A|$. We obtain a bound for $|P|$.

By the incidence theorem,

$$c(|L|^{\frac{2}{3}}|P|^{\frac{2}{3}} + |L| + |P|) \geq I(P, L) \geq |L||A|.$$

Some calculation finishes the problem. □

Lecture 24
Thu. 5/5/11

§1 Independence number of triangle-free graphs

Suppose G is a graph G with n vertices and maximum degree d. Then by greedy picking,

$$\alpha(G) \geq \frac{n}{d+1}.$$

Theorem 24.1 (Ajtai, Komlós, Szemerédi, Sheover)**:** If G is triangle-free, then

$$\alpha(G) \geq \frac{n \log d}{8d}.$$

Proof. We may assume $d \geq 16$. Let W be an independent set of G, chosen randomly among all independent sets. For each $v \in V(G)$, let

$$X_v = d|\{v\} \cap W| + |N(v) \cap W|$$

where $N(v)$ is the set of all neighbors of v.

Claim 24.2:

$$\mathbb{E}(X_v) \geq \frac{1}{4} \log d.$$

Proof. Let H be the induced subgraph of G with vertex set $V(G) - (\{v\} \cup N(v))$. Fix an independent set S in H. Let $X \subseteq N(v)$ be the nonneighbors of S; these are precisely the elements of $N(v)$ that can be added to S and still give an independent graph. Note none of the lements of X are connected to each other, because they are all adjacent to v and the graph is triangle free.

It suffices to show

$$\mathbb{E}(X_v | W \cap V(H) = S) \geq \frac{\log d}{4}$$

for all choices of S. Conditioning on $W \cap V(H) = S$, there are $2^x + 1$ choices for W: either W contains v, or W does not contain v and contains a subset of W. Then since v contributes d to X_v, and the average size of a subset of X is $\frac{x}{2}$ ($|X| = x$),

$$\mathbb{E}(X_v | W \cap V(H) = S) = \frac{d + \frac{x}{2}2^x}{2^x} \geq \frac{1}{4} \log d$$

by computation. □

By the claim, $\mathbb{E}(\sum_v X_v) \geq \frac{n}{4} \ln d$. But $\sum_v X_v \leq 2d|W|$ since for $v \in W$, X_v gets contribution d from v and for each $u \in N(v)$, X_v gets a contribution 1 from u. Thus, taking expected values, $\mathbb{E}(|W|) \geq \frac{n \log d}{8d}$. Thus there exists W so that $|W| \geq \frac{n \log d}{8d}$. □

Consider $R(3,k)$. It is the minimum number of vertices in a complete graph such that in any coloring with red of blue, there is a red triangle of blue K_k. Alternatively, it is the minimum n such that any triangle-free graph on n vertices has an independent set of size n. By Pigeonhole, $R(3,k) \geq k^2$. By Lovász Local Lemma, $R(3,k) \geq \frac{ck^2}{(\ln k)^2}$.

Theorem 24.3:

$$R(3,k) = \Theta\left(\frac{k^2}{\ln k}\right).$$

Proof. For the lower bound, randomly order all $\binom{N}{2}$ $e_1, \ldots, e_{\binom{N}{2}}$. Add the e_i in order unless e_i makes a triangle with edges already inside. With high probability this gives a triangle-free graph with no independent set of size n. A lot of computation here. (30 pages)

Let G have at least $n = \frac{8k^2}{\log k}$ vertices and be triangle-free. If the maximum degree is at least k we are done. Otherwise by the previous theorem, $\alpha(G) \geq \frac{n \log k}{8k} = k$. □

The cast $R(4,k)$ is unsolved; we just know $\left(\frac{k}{\log k}\right)^{\frac{5}{2}} \leq R(4,k) \leq \frac{ck^3}{(\log k)^2}$.

§2 Local Coloring

If G has n vertices, with n large, and every subset of size $10^{-20}n$ vertices is 3-colorable, is G 100-colorable? No.

Theorem 24.4: For all positive integers k, there exists $\varepsilon > 0$ such that for all n sufficiently large, there exists G on n vertices with $\chi(G) \geq k$ and $\chi(G[S]) \leq 3$ for every S of size εn. I.e. we can't detect global colorability by local colorability.

Proof. Let $c > 2kH\left(\frac{1}{k}\right) \ln 2$ where H is the entropy function $H(x) = -x \ln x - (1-x)\ln(1-x)$. Let $\varepsilon = \frac{1}{4} e^{-s} 3^3 c^{-3}$. (In relation to the question above, note for $n = 100$, $\varepsilon > 10^{-20}$.) Set $p = \frac{c}{n}$ and consider $G(n,p)$. If $\chi(G) \leq k$ then $\alpha(G) \geq \frac{n}{k}$. The expected number of independent sets is (for convenience, assume $k|n$)

$$\begin{aligned}\mathbb{E}\left(\text{number of independent sets of size } \frac{n}{k}\right) &= \binom{n}{n/k}(1-p)^{\binom{n/k}{2}} \\ &< 2^{n\left(H\left(\frac{1}{k}\right)+o(1)\right)} \\ &= e^{-\frac{cn}{2k^2}(1+o(1))} = o(1)\end{aligned}$$

where we used Stirling's formula to get $\binom{n}{an} = 2^{n(H(a)+o(1))}$.

If there exists S, $|S| \leq \varepsilon n$ with $\chi(G[S]) \geq 4$, then there exists a minimum such S and every vertex in $G[S]$ would have degree at least 3. (If there is a vertex v of degree smaller than 3, and if $S \backslash \{v\}$ is colored with at most 3 colors, then v can be colored with one of those

3 colors.) Then $\mathbb{E}(G[S]) \geq \frac{3t}{2}$. We show it is unlikely for any subgraph to have so many edges.

Now

$$P\left(\exists S : |S| \leq \varepsilon n, \, S \text{ has at least } \frac{3|S|}{2} \text{ edges}\right) \leq \sum_{t \leq \varepsilon n} \binom{n}{t} \binom{\binom{t}{2}}{\frac{3t}{2}} \left(\frac{c}{n}\right)^{\frac{3t}{2}} = o(1)$$

by using the estimate that the summand is at most $\left[\left(\frac{en}{t}\right)\left(\frac{te}{n}\right)\left(\frac{c}{n}\right)^{\frac{3}{2}}\right]^t$ (cf. (12)). □

Lecture 25
Tue. 5/10/11

§1 Weierstrass Approximation Theorem

The theorem tells us we can approximate continuous real functions by polynomials.

Theorem 25.1: The set of real polynomials over $[0,1]$ is dense in the set of continuous functions $C[0,1]$ (the topology being determined by the norm $|f| = \max |f|$).

In other words, for all continuous real functions $f : [0,1] \to \mathbb{R}$ and $\varepsilon > 0$, there exists a polynomial $p(x)$ such that $|p(x) - f(x)| \leq \varepsilon$ for all $x \in [0,1]$.

Proof. (Bernstein) We will use the fact that the binomial distribution is tightly distributed around its mean.

Since f is uniformly continuous, there exists $\delta > 0$ such that if $x, x' \in [0,1]$ and $|x - x'| \leq \delta$ then $|f(x) - f(x')| \leq \frac{\varepsilon}{2}$. In addition f is bounded; take M such that $|f(x)| \leq M$ for all $x \in [0,1]$.

Let $B(n,x)$ be a binomial random variable with n independent trials and probability of success x for each of them. Note

$$\begin{aligned} P(B(n,x) = j) &= \binom{n}{j} x^j (1-x)^{n-j} \\ \mu = \mathbb{E}(B(n,x)) &= xn \\ \sigma &= \sqrt{nx(1-x)} < \sqrt{n}. \end{aligned}$$

By Chebyshev's inequality,

$$P(|B(n,x) - nx|) > n^{\frac{2}{3}} \leq \left(\frac{\sqrt{n}}{n^{\frac{2}{3}}}\right) = \frac{1}{n^{\frac{1}{3}}}.$$

Thus there exists n such that $P(|B(n,x) - nx| > n^{\frac{2}{3}}) < \frac{\varepsilon}{4M}$ and $\frac{1}{n^{\frac{1}{3}}} < \delta$. Define

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right).$$

We claim that $|P_n(x) - f(x)| \le \varepsilon$. For every $x \in [0,1]$,

$$\begin{aligned}
|P_n(x) - f(x)| &\le \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} \left| f\left(\frac{i}{n}\right) - f(x) \right| \\
&\le \sum_{i, |i-nx| \le n^{\frac{2}{3}}} \binom{n}{i} x^i (1-x)^{n-i} \left| f\left(\frac{i}{n}\right) - f(x) \right| \\
&\quad + \sum_{i, |i-nx| > n^{\frac{2}{3}}} \binom{n}{i} x^i (1-x)^{n-i} \left(\left| f\left(\frac{i}{n}\right) \right| + |f(x)| \right) \\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{4M} \cdot 2M = \varepsilon.
\end{aligned}$$

(To see that the first sum at most $\frac{\varepsilon}{2}$, note that $\left|\frac{i}{n} - x\right| \le n^{-\frac{1}{3}}$ there and $\frac{1}{n^{\frac{1}{3}}} < \delta$.) □

§2 Antichains

Definition 25.2: A family F of subsets of $[n]$ is an **antichain** if no set in F is contained in another.

Theorem 25.3: Let F be an antichain. Then

$$\sum_{A \in F} \frac{1}{\binom{n}{|A|}} \le 1.$$

Proof. Pick a random permutation of $1, \ldots, n$. Consider the family of sets $C_\sigma = \{\{\sigma(1)\}, \{\sigma(1), \sigma(2)\}, \ldots, \{$
Let $X = F \cap C_\sigma$ because for any pair of sets in C_σ, one is inside the other.

Note $X = \sum_A X_A$ where A is the indicator random variable for the event $A \in C_\sigma$. Since $P(A \in C_\sigma) = \frac{1}{\binom{n}{|A|}}$,

$$1 \ge \mathbb{E}(X) = \sum_{A \in \mathcal{F}} \mathbb{E}(X_A) = \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}}.$$

□

Corollary 25.4 (Sperner's Theorem)**:** Let $\mathcal{F}$ be an antichain. Then $|\mathcal{F}| \le \binom{n}{\lfloor \frac{n}{2} \rfloor}$.

Proof. Note $\binom{n}{x}$ is maximal at $x = \lfloor \frac{n}{2} \rfloor$ so the theorem gives

$$1 \ge \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \ge \frac{|\mathcal{F}|}{\binom{n}{\lfloor \frac{n}{2} \rfloor}}.$$

□

§3 Discrepancy

Let Ω be a set, and let $A \subseteq 2^\Omega$ be a collection of subsets of Ω. We want to color the elements of Ω with red or blue such that all $S \in A$ have roughly the same number of red and blue elements.

Label red and blue by -1 and 1.

Definition 25.5: For a coloring $\chi : \Omega \to \{-1, 1\}$, let $\chi(S) = \sum_{a \in S} \chi(a)$. Define

$$\operatorname{disc}(A, x) = \max_{S \in A} |\chi(S)|$$

and define the **discrepancy** of A to be

$$\operatorname{disc}(A) := \min_{\chi : \Omega \to \{-1,1\}} \operatorname{disc}(A, \chi).$$

We want to show that under certain constraints, the discrepancy is small.

Theorem 25.6: Let A be a family of n subsets of an m-element set Ω. Then

$$\operatorname{disc}(A) \leq \sqrt{2m} \ln 2n.$$

Proof. Let $\chi : \Omega \to \{-1, 1\}$ be random. For $S \subseteq \Omega$ let X_S be the indicator random variable for the event $|\chi(S)| > \alpha$. If $|S| = a$, then by Chernoff estimate,

$$\mathbb{E}(X_S) = P(|\chi(S)| > \alpha) < 2e^{-\frac{\alpha^2}{2a}} \leq 2^{-\frac{\alpha^2}{2m}} = \frac{1}{n}.$$

Let $X = \sum_{S \in A} X_S$. Then

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{S \in A} X_S\right) = \sum_{S \in A} \mathbb{E}(X_S) < n \cdot \frac{1}{n} = 1.$$

Hence there exists a coloring χ such that $X = 0$, i.e. all sets in A have discrepancy at most α, i.e. $\operatorname{disc}(A) \leq \alpha$. □

Lecture 26
Thu. 5/12/11

§1 Discrepancy

We improve the bound from last time.

Theorem 26.1 (Spencer)**:** Suppose $|\Omega| = n$ and $A \subseteq 2^\Omega$, $|A| = n$. Then

$$\operatorname{disc}(A) \leq 11\sqrt{n}.$$

The assumption that $|\Omega| = n$ is not necessary can be dropped, i.e. $|\Omega| \geq n$ is okay.

Proof. A random coloring will not work. Instead, we consider a partial coloring $\chi : \Omega \to \{-1, 0, 1\}$ when $\chi(a) = 0$ means a is uncolored.

Lemma 26.2: There exists a partial coloring χ with at most $10^{-9}n$ uncolored points such that $|\chi(S)| \leq 10\sqrt{n}$ for every $S \in A$.

(We will then partially color the remaining points, and so on, to get a geometric series.)

Proof. Let $A = \{S_1, \ldots, S_n\}$ and $\chi : \Omega \to \{0, 1\}$ be random. For $1 \leq i \leq n$ define b_i to be the closest integer to $\frac{\chi(S_i)}{20\sqrt{n}}$. In particular $b_i = 0$ when

$$-10\sqrt{n} < \chi(S_i) < 10\sqrt{n}.$$

Let $p_j = P(b_i = j)$. Chernoff's estimate gives

$$\begin{aligned} p_0 &> 1 - 2e^{-50} \\ p_{-1} = p_1 &< e^{-50} \\ &\vdots \\ p_{-s} = p_s &< e^{50(2s-1)^2}. \end{aligned}$$

We bound the entropy $H(b_i)$:

$$H(b_i) = \sum_{j \in \mathbb{Z}} -p_j \log_2 p_j \leq \varepsilon := 3 \cdot 10^{-20}.$$

Consider the n-tuple $(b_1, \ldots, b_n)$. Then

$$H(b_1, \ldots, b_n) \leq \sum_{i=1}^{n} H(b_i) \leq \varepsilon n.$$

If a random variable Z assumes no value with probability 2^{-t}, then the entropy is large, $H(z) \geq t$. By the contrapositive there exists a n-tuple $(b_1, \ldots, b_n)$ such that $P((b_1, \ldots, b_n) = (s_1, \ldots, s_n)) \geq 2^{-\varepsilon n}$. All 2^n colorings are equally likely so there exists a set of at least $2^{(1-\varepsilon)n}$ colorings $\chi : \Omega \to \{-1, 1\}$ all having the same value $(b_1, \ldots, b_n)$.

Think of the class C^1 of all 2^n colorings $\chi : \Omega \to \{-1, 1\}$ as the Hamming cube $\{-1, 1\}^n$. We use the following:

Lemma 26.3 (Kleitman)**:** If $D \subseteq C$ and $|D| \geq \sum_{i \leq r} \binom{n}{i}$ with $r \leq \frac{n}{2}$, then D has diameter at least $2r$.

We may take $r = \alpha n$ as long as $\alpha < \frac{1}{2}$ and $2^{H(\alpha)} \leq 2^{1-\varepsilon}$. (Note $\binom{n}{\alpha n} = 2^{n(H(\alpha)+o(1))}$.)) Calculation then gives we can take

$$\alpha = \frac{1}{2}(1 - 10^{-9}).$$

(Use the Taylor expansion $H\left(\frac{1}{2}-x\right) \sim 1-\frac{2}{\ln 2}x^2$ for x small.) The diameter of C^1 is at least $n(1-10^{-9})$. Let $x_1, x_2 \in C^1$ be of maximal distance. Set $\chi = \frac{x_1-x_2}{2}$; this leaves all but $10^{-9}n$ uncolored because χ_1, χ_2 have the same b-vector. Then

$$|\chi(S_i)| \leq 10\sqrt{n}$$

for all $S_i \in A$, as needed. □

Lemma 26.4: Let $|A| = n$ and $|\Omega| = r$ with $r \leq 10^{-9}n$. Then there exists a partial coloring χ of Ω with at most 10^{-40} points uncolored, such that

$$|\chi(S)| < 10\sqrt{r}\sqrt{\ln\left(\frac{n}{r}\right)}$$

for all $S \in A$.

The proof is similar to the first lemma.

Let χ^1 be a coloring of all but $10^{-9}n$ of the elements, χ^2 be a coloring of all remaining elements but $10^{-49}n$, and χ^3 be a coloring of all remaining elements but $10^{-89}n$, and so on. Let $\chi = \chi^1 + \chi^2 + \cdots$. Then

$$|\chi(S)| < |\chi^1(S)| + |\chi^2(S)| + \cdots < 10\sqrt{n} + 10\sqrt{10^{-9}n}\sqrt{\ln 10^9} + \cdots < 11\sqrt{n}.$$

□