Discrete Math, 12th Problem Set (August 12)

REU 2003

Instructor: Laszlo Babai
Scribe: Ben Wieland

1 Min Deg

Definition 1.1. The *minimum degree* $\min\deg(G)$ of a permutation group G is the minimum of the degrees of the nonidentity elements of G.

For example, $\min\deg(S_n) = 2$, $\min\deg(A_n) = 3$, $\min\deg(D_n)$ is $n-2$ if n is even and $n-1$ if n is odd. $\mathrm{AGL}(d,q)$ acts on ${\mathbb{F}_q}^d$, which has $n = q^d$ elements, has $\min\deg = q^d - q^{d-1} = n(1 - \frac{1}{q}) \geq \frac{n}{2}$.

Exercise 1.2. (A. Bochert c. 1895) If $G < S_n$ is doubly transitive and not a giant (not A_n or S_n), then $\min\deg(G) \geq \frac{n}{4} - 2$.

Theorem 1.3. *(Jordan) If $G < S_n$ is primitive, then $\min\deg(G) \to \infty$ as a function of n.*

For example, $S_k < S_n$ with $n = \binom{k}{2} \sim \frac{k^2}{2}$ acting on two-element subsets, is primitive. The degree of a transposition in this induced action is $2(k-2)$ and $\min\deg = 2(k-2) \sim \sqrt{8n}$. The classification of finite simple groups can be used to show that this is minimal among primitive permutation groups of degree n, but elementary arguments already give the right order of magnitude; they show that the minimum degree of a primitive permutation group must be at least $c\sqrt{n}$.

2 Diameter of S_n

If $A \subset \Omega = \{1, \ldots, n\}$, then set $\mu(A) = \frac{|A|}{n}$, the normalized size of A. If we have a permutation $\sigma_k \in S_n$ with support A and we wish smaller support, then we can use $\sigma_{k+1} = [\sigma_k, \sigma_k^\tau]$, which has support at most $3|A \cap A^\tau|$. If $\mu(A) = p$ then $E(\mu(A \cap A^\tau)) = p^2$, with τ random in a transitive group. Thus replacing σ by $[\sigma, \sigma^\tau]$ changes our proportion from p to $3p^2$. Iterating, we get about $\operatorname{supp}\sigma_k = (3p)^{2^k}$, which quickly tends to 0 if $3p < 1$. If $3p$ is bounded away from 1, then in about $\log\log n$ steps this will hit $\frac{1}{n}$.

If $\sigma_{k+1} = [\sigma_k, \sigma_k^\tau]$ and b_k is the word length of σ_k, then $b_{k+1} = 4b_k + n^c$, as the random τ has word length n^c. This gives $b_k \sim n^c 5^k = n^c \log^c n$, which is "quasi-polynomial" and certainly bounded by $n^{c+\varepsilon}$.

But there is a danger that A and A^τ will commute (especialy when $|A|$ is small; then A and A^τ are likely to be disjoint).

Exercise 2.1. With $A = \operatorname{supp} \sigma$, suppose $x^\sigma = y$, $x' = x^{\tau^{-1}} \in A$, and $y' = y^{\tau^{-1}} \notin A$. Then $[\sigma, \sigma^\tau] \neq 1$.

Thus our goal is to obtain τ such that τ^{-1} keeps x in the support, sends $y = x^\sigma$ out of the support, and makes $A \cap A^\tau$ small. This requires triple transitivity. We do a random walk on $\Omega^{(3)}$, the space of ordered triples with no repetition, with edges labeled by generators of our group. This graph is strongly connected by triple transitivity. After N^c (now $N = n(n-1)(n-2)$) steps, the distribution of vertices will be nearly random: a random word of length N^c sends a given vertex to any other vertex with probability $\frac{1\pm\varepsilon}{N}$. With probability $\frac{1\pm\varepsilon}{n(n-1)}$ it sends a given x' (in A) to x and a given y' (not in A) to y. Conditioned on this, it still sends any other z' to z with probability $\frac{1\pm\varepsilon}{n-2}$, so by linearity of expectation, we still have the expected value of $|A \cap A^\tau|$, conditioned on $x' \mapsto x$ and $y' \mapsto y$ as $\frac{1}{n} + \frac{(|A|-2)^2}{n-2}(1 \pm \varepsilon)$, which means that the proportion of Ω in the overlap is approximately the square of the proportion of A.

3 Markov Chains

Read the handout "Finite Markov Chains."

Definition 3.1. A *stochastic matrix* is a matrix with nonnegative entries and row sums of 1. A *doubly stochastic matrix* is a matrix A such that A and A^{tr} (transpose) are both stochastic. A stochastic matrix is *ergodic* if it (precisely, the digraph of its nonzero entries) is strongly connected and aperiodic.

Exercise 3.2. An eigenvalue of a stochastic matrix has norm at most 1.

Exercise 3.3. If a stochastic matrix is strongly connected, show

1. The geometric multiplicity of 1 is 1.

2. Let its period be r and $\omega \in \mathbb{C}$ with $\omega^r = 1$. If λ is an eigenvalue, then so is $\omega\lambda$.

Exercise$^+$ 3.4. If a stochastic matrix is ergodic, then the algebraic multiplicity of the eigenvalue 1 is 1 and all other eigenvalues have norm strictly less than 1. (This is almost equivalent to the following theorem.)

Theorem 3.5 (Perron–Frobenius). *If T is an ergodic matrix, then $T^\infty = \lim_{k\to\infty} T^k$ exists.*

Observation 3.6. Since $TT^\infty = T^\infty$, we must have $Tx = x$ for x a column of T, but since the geometric multiplicity is 1, this eigenvector is a multiple of the all 1s vector. Thus all rows are equal; they are the unique stationary distribution π.

Exercise 3.7. If a Markov chain is ergodic, then for any initial distribution q_0, $\lim_{k\to\infty} q_0 T^k = \pi$, the stationary distribution.

Exercise 3.8. If T is ergodic and doubly stochastic, then the stationary distribution is uniform.

One way to guarantee that T is doubly stochastic is to ask $T = T^{tr}$. A walk on a Cayley graph is always doubly stochastic. Having the generating set S be closed under inverses makes $T = T^{tr}$, but we do that only to invoke the Laundau-Odlyzko theorem.

Theorem 3.9 (Landau–Odlyzko). *A random walk on a regular undirected graph of degree Δ, diameter d and n vertices has an eigenvalue gap*

$$\gamma = 1 - \max_{\lambda \neq 1} |\lambda| \geq \frac{c}{n\Delta d}$$

The eigenvalue gap tells us about the rate of convergence to the stationary distribution. This is always true, but easier if T is symmetric. Then the spectral theorem gives us an orthonormal basis $e_1, \dots, e_n$, with $e_i T = \lambda_i e_i$. for T. We can choose $e_1 = (1, \dots, 1)/\sqrt{n}$. Then for C the rotation that sends those eigenvectors to the standard basis, $C^{-1}TC$ is diagonal, with entries its eigenvalues. This "separation of coordinates" makes raising the matrix to powers easy: $C^{-1}T^kC = (C^{-1}TC)^k$, a diagonal matrix with entries λ_i^k. This exponentially decays to the diagonal matrix with entries $(1, 0, \dots, 0)$. Conjugating by the inverse of C then gives T^∞.

To measure convergence of T^k to T^∞, we need a measure of the size of a matrix and apply it to $T^k - T^\infty$, which conjugated by C gives a diagonal matrix with entries $(0, \lambda_2, \dots, \lambda_n)$.

Definition 3.10. The *operator norm* or *matrix norm* $\|A\|$ of a matrix A is $\max\limits_{x\neq 0} \frac{\|xA\|}{\|x\|}$, where the norm $\|x\|$ of vectors is the ℓ^2-norm: $\sqrt{\sum x_i^2}$. The *Frobenius norm* $\|A\|_F$ is the ℓ^2 norm on the entries: $\sqrt{\sum a_{ij}^2}$.

Exercise 3.11. $\|A\| \leq \|A\|_F \leq \sqrt{n}\|A\|$ (or is it $n\|A\|$?). The all 1s matrix and the identity matrix show that these are sharp.

Exercise 3.12. If $A = A^{tr}$ then $\|A\| = \max_\lambda |\lambda|$. *Hint.* Use the spectral theorem.

Exercise 3.13. If C is an orthogonal matrix, $\|AC\| = \|CA\| = \|A\|$.

Thus $\|T^k - T^\infty\| = \|C^{-1}(T^k - T^\infty)\| = \lambda_2^k = (1-\gamma)^k < e^{-\gamma k}$. So for $\|T^k - T^\infty\| < \varepsilon$ it suffices to have $k \geq \frac{1}{\gamma} \ln \frac{1}{\varepsilon}$.

Now we need to relate this bound on the operator norm to what we care about: the deviation of the distribution from uniform, $\sum_{j=1}^{n} \left| p_{ij}^{(k)} - \frac{1}{n} \right|$.

Set $\varepsilon = \frac{\delta}{n}$ and $A = T^k - T^\infty$. If $\|A\| < \varepsilon$, then $\|A\|_F^2 < n^2\varepsilon^2 < \delta$, so for all i and j, $|A_{ij}| < n\varepsilon = \delta$. So to be within δ of a uniform distribution, we need $\|A\| < \varepsilon = \frac{\delta}{n}$, which we can achieve with $k > \frac{\ln(\frac{1}{\varepsilon})}{\gamma} \sim \frac{2\ln n}{\gamma}$ (since δ is a constant) and by Landau–Odlyzko, $\gamma > \frac{c}{N\Delta d} > cn^{-7}$ ($N = n^3$, $\Delta \leq n$, $d \leq n^3$), so we can let k be $O(n^7)$, which leads to a similar diamater and we have proved

Theorem 3.14. *If $S_n = \langle S \rangle$ and one of the generators has degree less than $.3n$ then the diameter of the Cayley graph is $O(n^7 \log^c n)$.*

Theorem 3.15. *Every chain of subgroups in S_n has length at less than $2n$.*

Exercise 3.16. Use the above theorem to show that a minimal set of generators of S_n has fewer than $2n$ elements and thus $\Delta < 2n$.

4 Ramsey Theory

Definition 4.1. A subset $S \subset G$ is *product-free* if it contains no solutions to $xy = z$. It is *triangle-free* if it contains no solution to $xyz = 1$, with x, y, z not all the same. Let $\alpha(G)$ be the size of the largest triangle-free subset of G. Let $\widetilde{\alpha}(G) = \frac{\alpha(G)}{|G|}$.

We showed that for abelian groups we could find a product-free subset of size $\frac{2}{7}|G|$. We cannot achieve such a constant fraction in a triangle-free subset. That is, there exists a sequence of groups such that $\widetilde{\alpha}(G) \to 0$, but the instructor can only show that it goes to 0 very slowly.

The sequence of groups $G_k = \mathbb{Z}_3^k$ model higher-dimensional versions of the Set game. $\alpha(G_k)$ is the number of cards that can contain no Set.

Exercise 4.2. $L = \lim \sqrt[k]{\alpha(G_k)}$ exists. $2 \leq L \leq 3$.

Exercise 4.3. $L \geq \sqrt[4]{20}$. *Hint.* Find 20 Set cards without a Set.

Conjecture 4.4. $L = 3$.

Theorem 4.5 (van der Waerden). *For all k and r, if we color the natural numbers with r colors, we can find a monochromatic arithmetic progression of k terms.*

Exercise 4.6. This is equivalent to the claim that for all k and r, there exists N such that we can color the first N natural numbers with r colors such that there exists a k-term arithmetic progression.

Theorem 4.7 (Szemerédi, 1974). *For all k, ε, there exists N such that for any $S \subset [N]$ with $|S| \geq \varepsilon N$, S contains a k-term arithmetic progression.*

This theorem, conjectured by Erdős-Turán and sometimes called the "density version of van der Waerden's theorem" was proved by Szemerédi using a Ramsey-type theorem for graphs (The Szemerédi Lemma). A noteworthy later proof due to Furstenberg uses a fixed-point theorem and ergodic theory.

Definition 4.8. An ordered collection of t elements $x_1, \dots, x_t$ of $[t]^N$ is a *combinatorial line* if for each of the N coordinates, the t elements all have the same value $x_{1i} = x_{2i} = \dots = x_{ti}$ or $x_{ji} = j$ for all j.

Theorem 4.9 (Hales–Jewett). *For all t and r, there exists N sucht if we color $[t]^N$ by r colors, there exists a monochromatic combinatorial line.*

This is called the "combinatorial essence" of van der Waerden's theorem. Also, there are two more parameters: it should be that we color the a-dimensional spaces and fine a b-dimensional space, all of whose a-dimensional subspaces are the same color.

Exercise 4.10. Use Hales–Jewett to prove van der Waerden

Exercise 4.11. State the density version of Hales–Jewett, proved by Furstenberg and Katznelson.

Exercise 4.12. Use Furstenberg–Katznelson to prove that $\widetilde{\alpha}(\mathbb{Z}_3^k) \to 0$.