# A THEORIST'S TOOLKIT

# E0 206

*Arindam Khan,*

*Anand Louis*

# ADMIN INFO

- Grading (tentative):
  - Homeworks: 40%
  - Project/independent study: 25%
  - Final exam: 30%
  - Scribing lectures: 5%

- Intended audience: Students considering further study in theoretical computer science.

# ACADEMIC HONESTY

- You may do your homeworks in groups of size at most 3. You must write the names of all your collaborators in the submissions.

- At the discretion of the instructors, you may be asked further questions regarding your submissions, including but not limited to, solving additional problems in the presence of the instructors and within a stipulated time.

- In case you are found to have used unfair means in any aspect of this course, the instructors, at their discretion, may impose any penalty ranging from reducing 1 grade to awarding an F in this course.

# RESEARCH IN THEORETICAL COMPUTER SCIENCE

- Designing algorithms/mechanisms/cryptosystems
  - E0 225
  - Approximation algorithms
  - Randomized algorithms
  - Streaming algorithms
  - Learning algorithms
  - ……..
- Proving computational lower bounds

# RANDOMIZED ALGORITHMS

- Quicksort, hashing, etc.

- Given an array of n numbers promised to be one of the following two cases
  - All elements are 0
  - $n/2$ elements are 0 and $n/2$ elements are 1

- Any deterministic algorithm will take time $\Omega(n)$.

# RANDOMIZED ALGORITHMS

- Pick a random location on the array and check its value.

$\Pr[Alg\ outputs\ correct\ answer] = \frac{1}{2}$. Running time $= O(1)$.

- Repeat $k$ times

$\Pr[Alg\ outputs\ correct\ answer] = 1 - \frac{1}{2^k}$. Running time $= O(k)$.

# APPROXIMATION ALGORITHMS

- Min s-t cut

Given a graph $G = (V, E)$, compute a set $S \subset V$ such that number of edges that have one vertex in S and one vertex in $V \setminus S$ is minimized.

This problem has a polynomial time algorithm.

- Flow based algorithms
- LP based algorithms

- Max-cut

Given a graph $G = (V, E)$, compute the set $S \subset V$ such that number of edges that have one vertex in S and one vertex in $V \setminus S$ is ~~minimize~~ maximized.

- This problem is NP-hard!

- Random cut cuts half the edges!

Add each vertex to $S$ independently with probability ½ and $V \setminus$ S with probability ½.

- Fix an edge $\{u, v\}$

$$\Pr[\{u, v\}\ is\ cut] = \Pr[u \in \mathrm{S}] * \Pr[v \in V \setminus S] + \Pr[u \in V \setminus S] * \Pr[v \in S]$$
$$= \frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{2} = \frac{1}{2}$$

# MAX-CUT

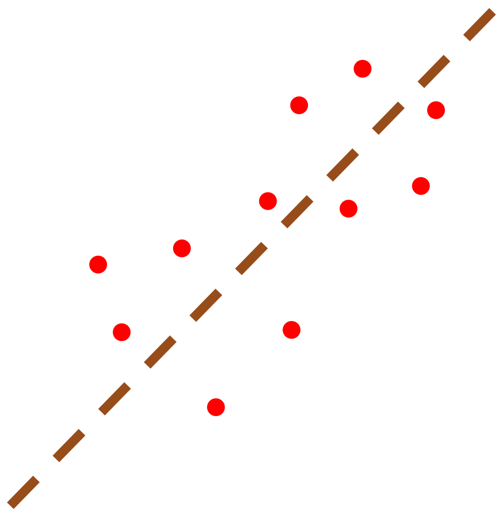
- $\mathrm{E}[number\ of\ edges\ cut] = \frac{1}{2}|E| \geq \frac{1}{2}$ max-cut
- Convex optimization based approaches give 0.878 approximation!
  - Uses geometric and probabilistic tools.
- Best possible (assuming a certain complexity theoretic conjecture)!

# LINEAR ALGEBRAIC TOOLS

- Top singular vector gives the best fit line!
- Top few singular vectors give best fit subspace.

- Low rank approximation of matrices.

- Many other applications!



Noisy data points

# EIGENVALUES OF GRAPHS

- For a $d$-regular graph, the largest eigenvalue of its adjacency matrix is $d$.
- Graph is disconnected if and only if second largest eigenvalue is $= d$.
- If second largest eigenvalue is $\ll d$, then graph is "well connected".
- If graph has $k$ eigenvalues "close" to $d$, then it can be partitioned into k parts that are "sparsely" connected with each other.

# BOOLEAN FUNCTIONS AND DISCRETE FOURIER ANALYSIS

$$f:\{-1,1\}^n \to \{-1,1\}\ \mathbb{R}$$

- Boolean functions can be uniquely represented as multilinear polynomials.

- Property testing, probabilistically checkable proofs, etc.

- Can be extended to non-Boolean functions as well.

# LINEARITY TESTING

$$f(x+y) = f(x) + f(y)\ \forall x, y, z \in \{-1,1\}^n$$

- Sample $x, y, z \sim \{-1,1\}^n$ independently and uniformly at random, and check if $f(x+y) = f(x) + f(y)$
- If $f$ is "far" from being a linear function, then the probability of test accepting is "small".

# (TOOLS commonly USED IN) RESEARCH IN THEORETICAL COMPUTER SCIENCE

- **Probabilistic methods**: Moments, Lovasz local lemma, martingales, random graphs,  applications, etc.
- **Information theory**: Shearer's Lemma, entropy and compression, Pinsker's lemma, KL-divergence, applications, etc.
- **Linear algebraic tools**: Courant-Fischer theorem, SVD, Cheeger's inequality, expanders, etc.
- **Discrete Fourier analysis:** and applications, etc.
- **Multiplicative weights update method:** and applications, online convex optimization, etc.

# QUESTIONS?