Lecture #21

Last Time: Grothendieck's Inequality and Lovasz theta Fctn

(new vectors from Taylor expansion of sine; $\alpha \le \theta \le \overline{\chi}$)

Today: Planted Clique and Random Matrix Theory

Erdos-Renyi $G(n, \frac{1}{2})$

$\uparrow$ # vertices $\quad \nwarrow$ probability ~~of~~ each edge, independent

How large is the largest clique (whp)?

$$\mathbb{E}[\#\text{ k-cliques}] = \binom{n}{k} 2^{-\binom{k}{2}} \le n^k 2^{-\binom{k}{2}} = 2^{k(\log n - (\frac{k-1}{2}))}$$

If $k = (2+\delta)\log n$ then

$$\mathbb{E}[\#\text{ k-cliques}] \lessapprox 2^{-\delta \log^2 n} = n^{-\delta \log n}$$

Fact 1: whp, $\omega(G) = (2 \pm o(1))\log n$

$\uparrow$ largest clique

Planted clique $G(n, \frac{1}{2}, k)$

- choose $H \leftarrow G(n, \frac{1}{2})$
- choose k vertices u.a.r, plant clique on them to get G

Can we recover the planted clique?

Fact: there is a $n^{O(\log n)}$-time algorithm to solve planted clique, whenever $k \ge (2+\delta)\log n$

Idea: • Brute-force search for a $(2+\delta)\log n$-sized clique

• Find all common neighbors

What about polynomial time algorithms?

Fact 2: There is a polynomial time algorithm, that succeeds (whp) if $k \geq c\sqrt{n \log n}$

Proof: A vertex u has degree

$$\deg(u) = \begin{cases} k-1 + \mathrm{Bin}(n-k, \frac{1}{2}) & u \text{ in planted clique} \\ \mathrm{Bin}(n, \frac{1}{2}) & \text{else} \end{cases}$$

whp (e.g. by Chernoff bound) the following holds

(1) $\forall u \notin$ planted clique, $\deg(u) < \frac{n}{2} + \frac{c}{4}\sqrt{n \log n}$

(2) $\forall u \in$ planted clique, $\deg(u) \geq \frac{n-k}{2} - \frac{c}{4}\sqrt{n \log n} + k$

Thus the k highest degree nodes are exactly those in planted clique. ☒

Planted clique is one of most fundamental average-case problems, with seeming (?) gap between inefficent and efficient algorithms

Applications: Cryptography, Nash equilibrium, motifs in biological networks, community detection

Is there a polynomial time algorithm for detecting smaller planted cliques?

Theorem 1: [Alon–Krivelevich–Sudakov] there is a polynomial time algorithm for planted clique that succeeds (whp) if $k \geq C\sqrt{n}$

the main tool will be random matrix theory:

Let $A \in \mathbb{R}^{n \times n}$ be an symmetric matrix with random entries

$$A_{ij} = \begin{cases} \text{random } \pm 1 & \text{if } i<j, \text{ or } i=j \\ A_{ji} & \text{else} \end{cases}$$

Theorem 2: With high probability, $\|A\|_{op} \leq (2+o(1))\sqrt{n}$

i.e. $\|A\|_{op} = O(\sqrt{n})$

We will prove a weaker bound, but using elementary tools

Step #1: For any fixed unit vector x,

$$\underbrace{|x^T A x|}_{\text{random variable with variance } 2\sum_{i<j}(x_i x_j)^2 + \sum_i x_i^4 = \left(\sum_i x_i^2\right)^2} \leq C\sqrt{n} \quad \text{with probability } 1-e^{-10n}$$

for some large enough $C>0$

Step #2: Let $\Sigma \subseteq S^{n-1}$ be a maximal $\frac{1}{4}$-net — i.e.

(1) $\forall x, y \in \Sigma, \ \|x-y\|_2 \geq \frac{1}{4}$

(2) $\forall z \in S^{n-1}, \ \exists x \in \Sigma$ with $\|x-z\|_2 < \frac{1}{4}$ (maximal)

Now let z be a unit vector, and maximize $|z^T A z| = \|A\|_{op}$

then let x be as in (2). Then by triangle inequality:

$$|x^T A x| \geq |z^T A z| - \underbrace{|(z-x)^T A x|}_{\leq \frac{\|A\|_{op}}{4}} - \underbrace{|x^T A (z-x)|}_{\leq \frac{\|A\|_{op}}{4}} - \underbrace{|(z-x)^T A (z-x)|}_{\frac{\|A\|_{op}}{16}}$$

Thus $|x^T A x| \geq \frac{\|A\|_{op}}{4}$. this implies:

$$\Pr[\|A\|_{op} > c\sqrt{n}] \leq \Pr[\exists x \in \Sigma, |x^T A x| > \frac{c}{4}\sqrt{n}]$$
$$\leq |\Sigma| e^{-10n} \quad (*)$$

Now we need to bound $|\Sigma|$ to finish the proof.

Observation: $\{B(x, \frac{1}{10})\}_{x \in \Sigma}$ are all disjoint

Moreover their union is contained in $B(0, \frac{11}{10})$. Hence

$$|\Sigma| \leq \frac{vol(B(0, \frac{11}{10}))}{vol(B(0, \frac{1}{10}))} = 11^n$$

which with (*) finishes the proof $\boxtimes$

The same proof works verbatim for asymmetric A's, but where we consider $y^T A x$, $\forall x, y \in \Sigma$ instead

---

Spectral Algorithm for Finding Planted Cliques

Given G, construct A with $A_{ij} = \begin{cases} +1 & \text{if } i=j \\ +1 & \text{if } i \neq j, (i,j) \in E \\ -1 & \text{else} \end{cases}$

Let u' be the top eigenvector of A

Set $T =$ top k coordinates of u' then set

$$C = \{u \mid u \text{ has atleast } \frac{4}{5}k \text{ neighbors in } T\}$$

Lemma [Main]: If $k \geq C\sqrt{n}$ for sufficiently large C, then whp T contains at least $\frac{4}{5}k$ vertices of the planted clique

Fact 3: whp, no vertex u outside of the planted clique is adjacent to more than $\frac{4}{7}k$ vertices in the planted clique $\quad (\frac{4}{7}k + \frac{1}{5}k < \frac{4}{5}k)$

Together, these imply Theorem 1 since the main lemma guarantees every node in p.c. will be added; Fact 3 guarantees no other node will be.

Proof [Sketch] We will decompose A as

$$A = \begin{array}{c c} & \begin{array}{cc} k & n-k \end{array} \\ \begin{array}{c} k \\ n-k \end{array} & \left[\begin{array}{c|c} J_S & 0 \\ \hline 0 & 0 \end{array}\right] \end{array} + \left[\begin{array}{c|c} 0 & \pm 1 \\ \hline \pm 1 & \pm 1 \end{array}\right]$$

assuming planted clique are first k coordinates.

It follows from Theorem 2 (+ asymmetric extension) that $\|E\|_{op} \leq 3C\sqrt{n}$

Moreover $\|M\|_{op} = k$, and its top eigenvector u is:

$$u = [\underbrace{\tfrac{1}{\sqrt{k}}, \tfrac{1}{\sqrt{k}} \ldots \tfrac{1}{\sqrt{k}}}_{k}, \underbrace{0, 0 \ldots 0}_{n-k}]$$

Intuition: If $k \geq 800C\sqrt{n}$, then $\|A\|_{op} \approx \|M\|_{op}$, and their top eigenvectors will be close.

In particular, from standard bounds (Wedin's theorem):

$$\sin \Theta(u,u') \le \frac{2\|E\|_{op}}{k}$$

difference btwn largest and second largest eigenvalue

Roughly, this means that for $k = 800C\sqrt{n}$ we have

$$\langle u, u' \rangle \ge 1 - \frac{1}{100}$$

Hence we have that the following conditions hold:

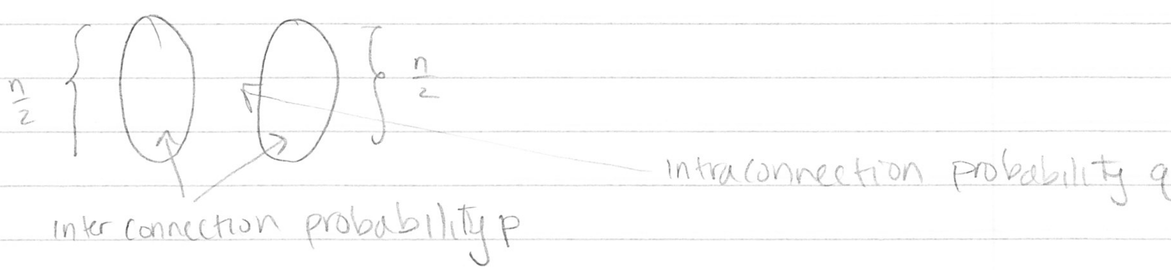
(1) at least $\frac{4}{5}k$ of the first k coordinates of u' have value at least $\frac{1}{2\sqrt{k}}$

(2) at most $\frac{1}{5}k$ of the last $n-k$ coordinates of u' have value at least $\frac{1}{2\sqrt{k}}$

Thus among the k largest coordinates, at least $\frac{4}{5}k$ must come from the planted clique, as desired $\boxtimes$

Many other average-case problems are amenable to spectral techniques

Planted Bisection $G(n,p,q)$



Can we recover the hidden bisection?

Theorem [McSherry]: There is a spectral algorithm that succeeds when

$$\frac{p-q}{p} \geq \frac{c \log n}{pn}$$

Intuition: Decompose A (adjacency matrix) as:

$$A = \left[\begin{array}{c|c} p & q \\ \hline q & p \end{array}\right] + E$$

Then if $\|E\|_{op}$ is sufficiently small, the top two eigenvectors of A can be used to recover the hidden bisection, extends to more general stochastic block model

semirandom model

Thought experiment: What if an "adversary" can remove edges in random part of graph, in planted clique? Do the algorithms still work?

No, but [Feige-Krauthgamer] show Lovasz Theta Fctn does!