PRINCETON UNIVERSITY FALL '24 COS 521:ADVANCED ALGORITHMS

Homework 1

Out: *Nov 18* Due: *Dec 4*

Instructions:

- Upload your solutions (to the non-extra-credit) as *a single* PDF file (one PDF total) to Gradescope. Please anonymize your submission (do not list your name in the PDF title or in the document itself). If you forget, it's OK.
- If you choose to do extra credit, upload your solution to the extra credits as a single PDF file to Gradescope. Please again anonymize your submission.
- You may discuss ideas for solutions with any classmates, textbooks, the Internet, etc. Please attach a brief "collaboration statement" listing any collaborators at the end of your PDF. **You must write up your solutions individually.**
- For each problem, you should aim to keep your writeup below one page. For some problems, this may be infeasible, and for some problems you may write significantly less than a page. This is not a hard constraint, but part of the assignment is figuring out how to easily convince the grader of correctness, and to do so concisely. "One page" is just a guideline: if your solution is longer because you chose to use figures (or large margins, display math, etc.) that's fine.
- Each problem is worth ten points (even those with multiple subparts).

Problems:

§1 (10 points, On the Courant Fisher Theorem)

(a) (7 points) Let A, B be symmetric, real matrices with eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \lambda_n(A)$ (and similarly for B). Prove that for every k, $\lambda_k(A) + \lambda_n(B) \leq \lambda_k(A+B) \leq \lambda_k(A)+\lambda_1(B)$. Use this claim to establish that $|\lambda_k(A+B)-\lambda_k(A)| \leq \max\{\lambda_1(B), |\lambda_n(B)|\}$.

(b) (3 points) Let A be the adjacency matrix of a not necessarily regular graph G with m edges and n vertices with eigenvalues $\lambda_1 \geq \lambda_2 \ldots \lambda_n$. Prove that $\lambda_1 \geq 2m/n$.

§2 (10 points, spectral norm of a random matrix via union bound) Let R be a random symmetric matrix with uniformly random ± 1 entries. In this problem, you will establish that $\|R\|_2 \leq C\sqrt{nlogn}$ via a different method. You can assume the following Hoeffding's inequality that we discussed in the course early on. Let $X_1, X_2, \ldots, X_n$ be independent random variables such that each X_i takes values in $[a_i, b_i]$. Then, for any $t > 0$, $\Pr[|\sum_i X_i - \mathbb{E}[\sum_i X_i]| \geq t] \leq 2\exp(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2})$.

You should also recall that $\|R\|_2 = \max_{x\neq 0} \|Rx\|_2 / \|x\|_2 = \max_{x\neq 0} |x^\top Rx| / \|x\|_2^2$. In this problem, for any $1 > \epsilon > 0$, let S_ϵ be a finite set of N_ϵ unit vectors in $\mathbb{R}^n$ such that for every unit vector $u \in \mathbb{R}^n$, there is a vector $u' \in S$ such that $\|u - u'\|_2 \leq \epsilon$.

(a) (3 points) Prove that for every unit vector u and $t \geq 0$, $\Pr[|u^\top Ru| \geq t] \leq 2e^{-t^2/2}$. (Hint: note that for a unit vector u, $1 = \|u\|_2^2 \cdot \|u\|_2^2 = \sum_{i=1}^n u_i^4 + 2\sum_{i<j} u_i^2 u_j^2$.)

(b) (1 point) Prove that $\Pr[\exists u \in S_\epsilon, |u^\top Ru| \geq t] \leq 2N_\epsilon e^{-t^2/2}$.

(c) (2 point) Prove that for every unit vector u, and any ± 1-entry matrix B, $|u^\top Bu| \leq n^C$ for some $C > 0$. What's the smallest C for which you can establish this claim?

(d) (4 points) In this part, you can assume without proof that for every $\epsilon > 0$, there is an S_ϵ of size $N_\epsilon \leq (c/\epsilon)^n$. Using this and the results of the previous parts, argue that $\Pr[\|R\|_2 \geq O(\sqrt{n \log n})] \leq 1/n$ (Hint: write every unit vector $u = v + e$ such that $v \in S_\epsilon$ and e has length at most ϵ. What value of ϵ should you choose?)

(e) (Extra Credit) Prove the assumption in part (d). That is, prove that there is an S_ϵ as described in part (2) of size $(c/\epsilon)^n$ for some $c > 0$.

§3 (10 points, combinatorial algorithm for recovering planted communities) In the class, we saw a spectral algorithm for recovering a planted communities when the edge densities within the communities p and across the communities q satisfy $p - q \gg 1/\sqrt{n}$. Here, we will see a simple combinatorial algorithm that succeeds when $p - q \geq \Omega(1)$.

Let G be a graph on n vertices (n is even) chosen as follows: 1) Pick an arbitrary S of size $n/2$, 2) For each pair i, j of vertices such that $i, j \in S$ or $i, j \notin S$, include $\{i, j\}$ in G with probability p, 3) For each pair i, j such that $i \in S, j \notin S$ or $i \notin S, j \in S$, include $\{i, j\}$ in G with probability q. Suppose that $p - q > c$ for some fixed constant $c > 0$.

Consider the following algorithm: 1) pick a vertex v, 2) Output $\hat{S}$ obtained by including in $\hat{S}$ the $n/2$ vertices that have the fewest common neighbors with v.

Prove that for large enough n, with probability at least 0.99 over the draw of G, $\hat{S}$ either equals S or $V \setminus S$. (Hint: suppose WLOG that $v \in S$. Compute the expected number of common neighbors between v and any vertex in S and similarly between v and any vertex in $V \setminus S$. Now use Chernoff Bound.)

§4 (10 points, self-improving planted clique algorithm) In the class, we saw that we can distinguish between a graph $G \sim G(n, 1/2)$ and $G \sim G(n, 1/2, k)$ (i.e., $G \sim G(n, 1/2)$ with an added k clique) in polynomial time if $k \geq c\sqrt{n}$ for some $c > 0$.

Find an algorithm that for any $t \in \mathbb{N}$, runs in time $n^{O(t)}$ and succeeds in the same goal for $k \geq \sqrt{n/2^t}$. (Hint: suppose you were given, in addition, a set S of t vertices in the planted clique if there was one. Can you now reduce the problem to graphs on a smaller number of vertices?)