CS264: Homework #5

Due by midnight on Wednesday, October 29, 2014

Instructions:

(1) Form a group of 1-3 students. You should turn in only one write-up for your entire group.

(2) Turn in your solutions at `http://rishig.scripts.mit.edu/cs264-bwca/submit-paper.html`. You can contact the TA (Rishi) at `bwca-staff@lists.stanford.edu`. Please type your solutions if possible and feel free to use the LaTeX template provided on the course home page.

(3) Students taking the course for a letter grade should complete all exercises and problems. Students taking the course pass-fail should complete the Exercises, and can skip the Problems.

(4) Write convincingly but not excessively. Exercise solutions rarely need to be more than 1-2 paragraphs. Problem solutions rarely need to be more than a half-page (per part), and can often be shorter.

(5) You may refer to your course notes, and to the textbooks and research papers listed on the course Web page *only*. You cannot refer to textbooks, handouts, or research papers that are not listed on the course home page. Cite any sources that you use, and make sure that all your words are your own.

(6) If you discuss solution approaches with anyone outside of your team, you must list their names on the front page of your write-up.

(7) Exercises are worth 5 points each. Problem parts are labeled with point values.

(8) No late assignments will be accepted.

Lecture 9 Exercises

None! (Though see the problems, below.)

Lecture 10 Exercises

Exercise 37

Consider a random graph G drawn from the distribution $\mathcal{G}(n, \frac{1}{2})$ described in lecture. Prove that, for every $\epsilon > 0$, the number of edges crossing every bisection of G is between $(1-\epsilon)\frac{n^2}{8}$ and $(1+\epsilon)\frac{n^2}{8}$ with probability $1 - o(1)$, where the $o(1)$ term tends to 0 as $n \to \infty$.

[Hint: what is the expected number of edges crossing each bisection? How many different bisections are there? To prove that everything is close to its expectation, look up "Chernoff-Hoeffding bounds" and the "Union Bound."]

Exercise 38

Recall the heuristic for finding a planted k-clique that simply returns the k vertices with the highest degrees. Consider a semirandom adversary A, which given a graph G with a planted k-clique C and non-clique edge probabilities $\frac{1}{2}$, returns a graph $A(G)$ with the same clique C and possibly fewer non-clique edges.

Prove that there is a constant $c > 0$ and a semirandom adversary such that, with high probability over the choice of G, the "top k degrees" algorithm fails to return a clique of $A(G)$ when $k = cn$.

Exercise 39

Recall that $h(G)$ denotes the optimal value of the semidefinite program (SDP) described in lecture. Prove that if $\hat{G}$ is the graph G plus one additional edge, then

$$h(G) \leq h(\hat{G}) \leq h(G) + 1.$$

[Hint: Show that the constraints of (SDP) imply that $|x_{ij}| \leq 1$ for every i, j.]

Exercise 40

Suppose you are given, as a black box, an algorithm that computes the *value* of the minimum bisection of a graph G. Show how to use this black box to reconstruct the actual minimum bisection $(S, \bar{S})$.

[You can assume that the minimum bisection of G is unique.]

Problems

Problem 17

(10 points) The point of this problem is prove a weaker version of the key technical lemma from Lecture #9 (the fact that the kernel of a random matrix A is an almost Euclidean subspace).

Consider an $m \times n$ matrix A, with $m = \frac{n}{2}$, chosen uniformly at random among all such matrices with $\{0, 1\}$ entries. Prove that there is a constant $\delta > 0$ (independent of n) such that, with high probability (approaching 1 as $n \to \infty$) over the choice of A, every non-zero vector $x \in \{0, 1\}^n$ that satisfies $Ax = 0$ (over $\mathbb{Z}_2$) has a "1" in at least δn components.

Problem 18

The goal of this problem is to extend the compressive sensing result from Lecture #9 to the case where the ground truth vector z is only *approximately* k-sparse. Let I denote the k coordinates of z that maximize the ℓ_1 norm $\sum_{i \in I} |z_i|$, and let r denote the residual ℓ_1 norm $\sum_{i \notin I} |z_i|$ on the coordinates outside of I.

Assume that A is $\frac{1}{4}\sqrt{\frac{n}{k}}$-good, as in lecture. Let w denote the optimal solution to the ℓ_1-minimizing linear program from lecture.

(a) (7 points) Prove that

$$\|z - w\|_1 \leq 2\|(z - w)_I\|_1 + 2r.$$

[This follows from similar maneuvers to the derivation in lecture.]

(b) (3 points) Conclude that the computed vector w is almost the same as the ground truth vector z, in the sense that $\|z - w\|_1 \leq 4r$.

[Use a result from lecture.]

Problem 19

(10 points) Consider the following heuristic for finding a clique in a graph with vertices $V = \{1, 2, \ldots, n\}$:

1. $S = \emptyset$.

2. While there is a vertex not in S connected to every vertex in S, add the lowest-indexed such vertex to S.

3. Return S.

Note that this heuristic always returns a clique. Prove that in a random graph G drawn from $\mathcal{G}(n, \frac{1}{2})$, the expected size of the clique returned is at least $c \log n$, where $c > 0$ is some constant independent of n.

[Hint: two potentially useful tools are the "principle of deferred decisions" and the expected value of a geometric random variable.]