PRINCETON UNIVERSITY FALL '16 COS 521:ADVANCED ALGORITHMS

Homework 3

Out: *Nov 7* Due: *Nov 18*

1. Compute the mixing time (both upper and lower bounds) of a graph on $2n$ nodes that consists of two complete graphs on n nodes joined by a single edge. (Hint: Use elementary probability calculations and reasoning about "probability fluid"; no need for eigenvalues.)

2. Let M be the Markov chain of a 5-regular undirected graph that is connected. Each node has self-loops with probability 1/2. We saw in class that 1 is an eigenvalue with eigenvector $\vec{1}$. Show that every other eigenvalue has magnitude at most $1 - 1/10n^2$. (Hint: First show that a connected graph cannot have 2 eigenvalues that are 1.) What does this imply about the mixing time for a random walk on this graph from an arbitrary starting point?

3. This question will study how mixing can be much slower on directed graphs. Describe an n-node directed graph (with max indegree and outdegree at most 5) that is fully connected but where the random walk takes $\exp(\Omega(n))$ time to mix (and the walk ultimately does mix). Argue carefully.

4. Describe an example (i.e., an appropriate set of n points in $\Re^n$) that shows that the Johnson-Lindenstrauss dimension reduction method — the transformation described in Lecture, with an appropriate scaling— does *not* preserve ℓ_1 distances within even factor 2. (Extra credit: Show that no *linear transformation* suffices, let alone J-L.)

5. (Dimension reduction for SVM's with margin) Suppose we are given two sets P, N of unit vectors in $\Re^n$ with the guarantee that there exists a hyperplane $a \cdot x = 0$ such that every point in P is on one side and every point in N is on the other. Furthermore, the ℓ_2 distance of each point in P and N to this hyperplane is at least ϵ. Then show using the Johnson Lindenstrauss lemma (hint: you can use it as a black box) that a random linear mapping to $O(\log n/\epsilon^2)$ dimensions and such that the points are still separable by a hyperplane with margin $\epsilon/2$.

6. Recall that $G(n, 1/2)$ is the random graph on n nodes in which each edge is present with probability exactly 1/2. In the planted clique problem, you are given a graph $G \sim G(n, 1/2)$ with a clique "planted" on some k special vertices. In a previous homework, you showed that with high probability, $G \sim G(n, 1/2)$ contains no clique of size more than $2 \log(n)$ thus, if $k \gg 2 \log(n)$, the added clique is the unique maximum clique in G. In this question, we explore a spectral algorithm for the planted clique problem.

   a) Show that the second largest eigenvalue of the adjacency matrix of $G \sim G(n, 1/2)$ is at most $O(\sqrt{n})$ with high probability. (Hint: Use the method from the class that bounds the largest eigenvalue of a random matrix.)

b) Show that the second largest eigenvalue of the adjacency matrix of $G \sim G(n, 1/2)+$ k-clique is at least $k/2$ whenever $k > 4\sqrt{n}$.

c) Use a) and b) to give an algorithm that with high probability correctly detects whether a k-clique has been added to a random graph for $k > 4\sqrt{n}$. Use matlab, scipy or any other package to compute the eigenvalues of $G(n, 1/2)$ and $G(n, 1/2) +$ k-clique for $n = 400, 800, 1200$ and $k \in [\sqrt{n}/4, 4\sqrt{n}]$. Include a table with the top 3 eigenvalues. (Do 3 repetitions with newly sampled graphs for each n to see if the eigenvalue distribution is pretty stable over the samples.) Report your results. Do they agree with the calculations made in a) and b) above?

d) (Extra Credit) Can you recover the vertices of the added clique from the second eigenvector of $G \sim G(n, 1/2) + k$-clique for k as above?

7. Implement the portfolio management appearing in the notes for Lecture 13 ("Going with the Slope: Offline, Online and Randomly") in any programming environment and check its performance on S& P stock data (download from http://ocobook.cs.princeton.edu/links.htm). Include your code as well as the final performance (i.e., the percentage gain achieved by your strategy).

8. (Extra credit) Calculate the eigenvectors and eigenvalues of the n-dimensional boolean hypercube, which is the graph with vertex set $\{-1, 1\}^n$ and x, y are connected by an edge iff they differ in exactly one of the n locations. (Hint: Use symmetry extensively.)