PRINCETON UNIVERSITY FALL '13 COS 521: ADVANCED ALGORITHMS

Homework 3

Out: *Nov 5* Due: *Nov 14*

1. Compute the mixing time (both upper and lowerbounds) 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; no eigenvalues.)

2. Let M be the Markov chain of a d-regular 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: Recall how we showed in lecture that a connected graph cannot have 2 eigenvalues that are 1. Then use Courant-Fisher.) What does this imply about the mixing time for a random walk on this graph (with 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 indegre 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 that shows that the Johnson-Lindenstrauss dimension reduction method does *not* preserve ℓ_1 distances.

5. Recall that $G(n, 1/2)$ is the random graph on n nodes in which each edge is present with probability exactly $1/2$. (a) Show that with high probability it does not contain any clique of size more than $3 \log n$. (b) Use matlab, scipy or any other package to compute the eigenvalues of $G(n, 1/2)$ for $n = 400, 800, 1200$. Include a table with the top 5 eigenvalues. (Do 3 repetitions with newly sampled graphs for each n to see if the eigenvalue distribution is pretty stable over the samples.) What's your best guess for the value of the second largest eigenvalue as a function of n? (c) In the *planted clique* problem somebody takes such a graph and plants a clique of size K for some K. Clearly, if $K \gg 3 \log n$ the presence of this clique is identifiable *in principle*. Let's see when this presence shows up in the graph's spectrum. Try $K \in [\sqrt{n}/4, 4\sqrt{n}]$ and see for which value of K does the planted clique change the spectrum. (Use as high a value of n as your system can handle.) Report your results.

6. The course webpage links to a grayscale photo. Interpret it as an $n \times m$ matrix and run SVD on it. What is the value of k such that a rank k approximation gives a reasonable approximation (visually) to the image? What value of k gives an approximation that looks high quality to your eyes? Attach both pictures and your code. (In matlab you need mat2gray function.) Extra credit: Try to explain from first principles why SVD works for image compression at all.

7. (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.)

8. (Extra Credit) Prove the following extension of Johnson-Lindenstrauss dimension-reduction: Given n points in $\Re^m$ there is a way to reduce number of dimensions to $O(\log n/\epsilon^2)$ that preserves the *angle* between all points up to ϵ.