CS 225 - Pseudorandomness Prof. Salil Vadhan

Problem Set 3

Harvard SEAS - Spring 2015 *Due: Fri. Mar. 13, 2015*

Your problem set solutions must be typed (in e.g. LaTeX) and submitted electronically to `cs225-hw@seas.harvard.edu`. You are allowed 12 late days for the semester, of which at most 5 can be used on any individual problem set. (1 late day = 24 hours exactly). Please name your file `PS3-lastname.*`.

The problem sets may require a lot of thought, so be sure to start them early. You are encouraged to discuss the course material and the homework problems with each other in small groups (2-3 people). Identify your collaborators on your submission. Discussion of homework problems may include brainstorming and verbally walking through possible solutions, but should not include one person telling the others how to solve the problem. In addition, each person must write up their solutions independently, and these write-ups should not be checked against each other or passed around.

Strive for clarity and conciseness in your solutions, emphasizing the main ideas over low-level details. Do not despair if you cannot solve all the problems! Difficult problems are included to stimulate your thinking and for your enjoyment, not to overwork you. *ed problems are extra credit.

Problem 4.2. (More Combinatorial Consequences of Spectral Expansion) Let G be a graph on N vertices with spectral expansion $\gamma = 1 - \lambda$. Prove that:

1. The *independence number* $\alpha(G)$ is at most $(\lambda/(1+\lambda))N$, where $\alpha(G)$ is defined to be the size of the largest independent set, i.e. subset S of vertices s.t. there are no edges with both endpoints in S.

2. The *chromatic number* $\chi(G)$ is at least $(1+\lambda)/\lambda$, where $\chi(G)$ is defined to be the smallest number of colors for which the vertices of G can be colored s.t. all pairs of adjacent vertices have different colors.

3. The *diameter* of G is $O(\log_{1/\lambda} N)$.

Recall that computing $\alpha(G)$ and $\chi(G)$ exactly are **NP**-complete problems. However, the above shows that for expanders, nontrivial bounds on these quantities can be computed in polynomial time.

Problem 4.5. (Near-Optimal Sampling) Describe an algorithm for Sampling that tosses $O(m+\log(1/\varepsilon)+\log(1/\delta))$ coins, makes $O((1/\varepsilon^2)\cdot\log(1/\delta))$ queries to a function $f : \{0,1\}^m \to [0,1]$, and estimates $\mu(f)$ to within $\pm\varepsilon$ with probability at least $1-\delta$. (Hint: use expander walks to generate several sequences of coin tosses for the pairwise-independent averaging sampler, and compute the answer via a "median of averages".) It turns out that these bounds on the randomness and query/sample complexities are each optimal up to constant factors (for most parameter settings of interest).

Problem 4.8. (The Replacement Product) Given a D_1-regular graph G_1 on N_1 vertices and a D_2-regular graph G_2 on D_1 vertices, consider the following graph $G_1 \circledr G_2$ on vertex set $[N_1] \times [D_1]$: vertex (u, i) is connected to (v, j) iff (a) $u = v$ and (i, j) is an edge in G_2, or (b) v is the i'th neighbor of u in G_1 and u is the j'th neighbor of v. That is, we "replace" each vertex v in G_1 with a copy of G_2, associating each edge incident to v with one vertex of G_2.

1. Prove that there is a function g such that if G_1 has spectral expansion $\gamma_1 > 0$ and G_2 has spectral expansion $\gamma_2 > 0$ (and both graphs are undirected), then $G_1 \circledr G_2$ has spectral expansion $g(\gamma_1, \gamma_2, D_2) > 0$. (Hint: Note that $(G_1 \circledr G_2)^3$ has $G_1 \circledz G_2$ as a subgraph.)

2. Show how to convert an explicit construction of constant-degree (spectral) expanders into an explicit construction of degree 3 (spectral) expanders.

3*. Without using Theorem 4.14, prove an analogue of Part 1 for edge expansion. That is, there is a function h such that if G_1 is an $(N_1/2, \varepsilon_1)$ edge expander and G_2 is a $(D_1/2, \varepsilon_2)$ edge expander, then $G_1 \circledr G_2$ is a $(N_1 D_1/2, h(\varepsilon_1, \varepsilon_2, D_2))$ edge expander, where $h(\varepsilon_1, \varepsilon_2, D_2) > 0$ if $\varepsilon_1, \varepsilon_2 > 0$. (Hint: given any set S of vertices of $G_1 \circledr G_2$, partition S into the clouds that are more than "half-full" and those that are not.)

4. Prove that the functions $g(\gamma_1, \gamma_2, D_2)$ and $h(\varepsilon_1, \varepsilon_2, D_2)$ must depend on D_2, by showing that $G_1 \circledr G_2$ cannot be a $(N_1 D_1/2, \varepsilon)$ edge expander if $\varepsilon > 1/(D_2 + 1)$ and $N_1 \geq 2$.

Problem 4.10. (Unbalanced Vertex Expanders and Data Structures) Consider a $(K, (1-\varepsilon)D)$ bipartite vertex expander G with N left vertices, M right vertices, and left degree D.

1. For a set S of left vertices, a $y \in N(S)$ is called a *unique neighbor* of S if y is incident to exactly one edge from S. Prove that every left-set S of size at most K has at least $(1-2\varepsilon)D|S|$ unique neighbors.

2. For a set S of size at most $K/2$, prove that at most $|S|/2$ vertices outside S have at least δD neighbors in $N(S)$, for $\delta = O(\varepsilon)$.

Now we'll see a beautiful application of such expanders to data structures. Suppose we want to store a small subset S of a large universe $[N]$ such that we can test membership in S by probing just 1 bit of our data structure. A trivial way to achieve this is to store the characteristic vector of S, but this requires N bits of storage. The hashing-based data structures mentioned in Section 3.5.3 only require storing $O(|S|)$ words, each of $O(\log N)$ bits, but testing membership requires reading an entire word (rather than just one bit.)

Our data structure will consist of M bits, which we think of as a $\{0, 1\}$-assignment to the right vertices of our expander. This assignment will have the following property.

Property Π: For all left vertices x, all but a $\delta = O(\varepsilon)$ fraction of the neighbors of x are assigned the value $\chi_S(x)$ (where $\chi_S(x) = 1$ iff $x \in S$).

3. Show that if we store an assignment satisfying Property Π, then we can probabilistically test membership in S with error probability δ by reading just one bit of the data structure.

4. Show that an assignment satisfying Property Π exists provided $|S| \leq K/2$. (Hint: first assign 1 to all of S's neighbors and 0 to all its nonneighbors, then try to correct the errors.)

It turns out that the needed expanders exist with $M = O(K \log N)$ (for any constant ε), so the size of this data structure matches the hashing-based scheme while admitting (randomized) 1-bit probes. However, note that such bipartite vertex expanders do *not* follow from explicit spectral expanders as given in Theorem 4.39, because the latter do not provide vertex expansion beyond $D/2$ nor do they yield highly imbalanced expanders (with $M \ll N$) as needed here. But in Chapter 5, we will see how to explicitly construct expanders that are quite good for this application (specifically, with $M = K^{1.01} \cdot \text{polylog} N$).