1 Overview

We will continue discussing finding sparsest cuts. In particular, we will finish up from last time and then show you a “spectral” or eigenvalue approach.

2 Quick Review.

Last time, we presented a “distance” linear program that produced the minimum distance function that ensured that the average distance between pairs of nodes was at least $1/n^2$. (It was the linear programming dual of fractionally embedding the complete graph into the graph.)

We showed that the sparsest cut measure, i.e.,

$$\mathcal{S} = \min_{S} \frac{\#edges(S, \overline{S})}{|S| \cot |\overline{S}}, $$

was an upper bound on the optimal value. (We did this by showing that the solution for a cut, $(S, \overline{S})$, of assigning $d(e) = 1/|S||overlineS$ is feasible and has cost at most the sparseness of this cut.)

We then wished to show how, given a solution to the distance LP of value V, we could produce a cut of size $O(V \log n)$.

We began by noting that for a graph with a pair of nodes with distance greater than $1/2$, that one could produce a cut of size $O(V_r n^2 \log n)$, where V_r is the volume that is cut off from the graph. Repeatedly doing so, if possible, would eventually produce a relatively balanced cut, with sparseness $O(V \log n)$ as required.

We, however, required the condition that the remaining graph was always high diameter.

If this is not the case, we will use the following lemma.

Lemma 1

There is a polynomial time algorithm, that given a solution to the distance linear program of value W, if there is a node r, that is within $1/4n^2$ of at least $n/2$ other nodes, S, finds a cut of sparsity $O(W)$.

Proof:

Again, the cut will be a shortest path level. That is we assign distance 0 to all the nodes in S and perform Dijkstra's (or grow a ball around it uniformly.) One of the cuts encountered in this process will have the desired sparseness.

We argue this as follows. First, we note that

$$\sum_i d(i,S) \geq 1/4n,$$

where $d(i,S)$ is the distance from i to the closest node in S (0 for those in S.) This can be proven using condition on the LP that

$$\sum_{ij} d(i,j) \geq 1$$

as follows.

We note that

$$d(i,j) \leq d(i,r) + d(r,j) \leq d(i,S) + 2/4n^2 + d(j,S).$$

Now, we sum the inequality over all pairs i,j and note that the term $d(i,S)$ appears $2n$ times. In particular, we get

$$1 \leq \sum_{ij} d(i,j) \leq 2n \sum_i d(i,S) + 1/2.$$

This implies that

$$\sum_i d(i,S) \geq 1/4$$

as desired.

Now, we consider growing the ball from S in units of ϵ (some tiny amount.) We consider a cut at level i such place (i.e., we have grown out by $i\epsilon$). We define C_i is the cutsize and n_i as the number of nodes left *outside* of the ball.

We note that

$$W \geq \sum_i \epsilon C_i.$$

We also note that

$$\sum_i d(i,S) \leq \sum_{i>1} n_i \epsilon.$$

This requires a bit of thought, but follows thinking about $d(i,S)$ as the number of intervals of length ϵ to get to S and noting that a node at i intervals away is counted at least i times in the sum on the right hand side.

Note that the sparsity of the ith cut is

$$C_i/n_i(n).$$

Furthermore, we have

$$\frac{\sum_i \epsilon C_i}{\sum_i \epsilon n_i} \leq 4Wn.$$

Thus, at least for one i we have the ratio is of C_i/n_i is at most $4Wn$ and the sparsity is $O(W)$.

□

Thus, given an distance linear programming solution, we can find a cut of sparsity $O(W \log n)$ using either repeated shortest paths or the lemma above.

3 A spectral approach.

An alternative approach that is popular in practice is to use eigenvalues. In particular, let us consider mapping each node to a point on the real number line so that the nodes are reasonably well spread but the edges are not too stretched out.

3.1 Another relaxation.

For some intuition, let's restrict attention so that each node is mapped to one of two places on the number line.

In this case, for any cut S, we map the nodes in S to $-x$ and those in $\overline{S}$ to $y = x|S|/|\overline{S}|$. The squared length of an edge is $(x+y)^2$, the total squared length between all pairs is $|S||\overline{S}|(x+y)^2$. The ratio is the sparseness of the cut. That is, the following value is a lower bound on the sparsity of any cut

$$\min_x \frac{\sum_{e=u,v} ||x_u - x_v||^2}{\sum_{ij} ||x_i - x_j||^2},$$

where x_i is an assignment for node i to a real number. That is, x is a vector with an element for each node in the graph. This has been called the Raleigh quotient.

3.2 Random Walks

This problem happens to be solvable efficiently. There are many interpretations of the problem.

For example, one can view this a physical problem of having a bunch of nodes with repulsive forces between nodes and springs for edges, and finding the minimum energy configuration on a line.

One can also note that the relaxation corresponds to a certain random walk on the graph. That is, consider a particle at a node chooses with probability $1/(d+1)$ to stay and probability $1/(d+1)$ to go to a randomly chosen neighbor. It is to see that if this walk is allowed to continue infinitely, that the probability that one is at any node is proportional to its degree as follow. Intuitively, for graphs that are sparse this will happen slowly whereas for graphs that are not this will happen quickly.

For example, for the line graph the time to get even close to uniform is $\Omega(n^2)$ whereas for the complete graph it occurs in 1 timestep.

One can think of the current probability of being at a node as a vector on the nodes. For now, we assume that all the degrees are the same. One can think of the transition function as a matrix with $1/(d+1)$ on the diagonals, and $1/(d+1)$ on the off-diagonals.

Now, the vector consisting of all ones, is an eigenvector of the matrix with eigenvalue 1. Any other eigenvector has no larger eigenvalue.

The notion of eigenvectors have some associated facts. A central one is that there is a set of eigenvectors that forms an orthonormal basis of your vectors. Thus, any vector can be written as

$$x = a_1 v_1 + a_2 v_2 + \cdot a_n v_n,$$

where $v_1, \ldots v_n$ are eigenvectors. Furthermore, multiplying the transition matrix by x, t times yields a vector

$$x_t = a_1 \lambda_1^t v_1 + a_2 \lambda_2^t v_2 + \cdot a_n \lambda_n^t v_n,$$

where λ_i is the i eigenvector. As you might see, one continuously gets closer and closer to a vector that looks like the vector with the highest eigenvector.

Furthermore, if the vector x was initially orthogonal to the first eigenvector, the vector will converge to the second eigenvector. Moreover, the second eigenvector tells you the speed at which you go to the uniform distribution. Thus, the second eigenvalue perhaps has something to do with the sparseness.

Indeed, what is the value. Well, the second eigenvalue is orthogonal to the first, and it is the one whose norm reduces by the least. That is, it optimizes the following function

$$\min_{x:\sum_i x_i = 0, |x|=1} \sum_{e=(u,v)} ||x_u - x_v||^2,$$

since for some edge the averaging operation decreases the norm by

With a bit of thought, one can show this is equivalent to minimizing the raleigh quotient.