Spectral Graph Theory and its Applications Lecture 17

Sparsifying Graphs, I

Lecturer: Daniel A. Spielman November 2, 2004

17.1 Announcements

- There will be no class on Thursday, Nov 4th.
- Problem 2c is false. It is hereby removed from the problem set.

17.2 Sparsifying Graphs

Many graph algorithms run faster on sparse graphs then dense graphs. I think that the first researchers to find a way to take advantage of this were Benczur and Karger, who introduced the idea of sparsification. Sparsifying a graph is a process for producing a sparser graph that is roughly equivalent to the original for the purpose under consideration. One can use sparsifiers to speed up graph algorithms by first sparsifying their input, and then running the algorithm on the sparse graph. This idea has since been exploited in many places, which I should list. In this class, we will see how to use sparsifiers to speed up linear system solvers.

Today, we will take the first step in this direction. Given a graph with adjacency matrix A, we will show how to produce a sparse weighted graph with adjacency matrix $\tilde{A}$ such that A is close to $\tilde{A}$ in a spectral norm. In particular, if D is the diagonal matrix of degrees of A, then we will guarantee that all eigenvalues of $D^{-1}(\tilde{A} - A)$ are small. Later in the semester we will exploit this fact algorithmically. It is possible to obtain similar results if the original graph is weighted, but we ignore this case for simplicity of exposition.

To build $\tilde{A}$, we will flip a coin for each edge of A and, based on the outcome, decide whether or not to keep that edge. The coins will be biased, and if we decide to keep an edge, we will assign it a weight inversely proportional to our probability of keeping it. We apply these weights so that the expected weighted degree of each node remains approximately the same. To make sure that no node looses all its edges, we must be careful to make sure that each node will retain some edges. On the other hand, to reduce the total number of edges we must make sure to get rid of most edges from high degree nodes. To meet both of these objectives, we choose a parameter δ that will govern the number of edges in $\tilde{A}$, and decide to keep an edge between nodes i and j with probability $p_{i,j}$, where

$$p_{i,j} = \begin{cases} \frac{\delta}{\min(d_i, d_j)} & \text{if } \delta < \min(d_i, d_j) \text{, and} \\ 1 & \text{otherwise.} \end{cases}$$

We now prove

Lemma 17.2.1. *The expected number of edges in $\tilde{A}$ is at most δn.*

Proof. Let $X_{i,j}$ be a random variable that is 1 if we keep the edge between nodes i and j, and 0 otherwise. Then, the expected number of edges is

$$\mathbf{E}\left[\sum_{(i,j)\in E} X_{i,j}\right] = \sum_{(i,j)\in E} \frac{\delta}{\min(d_i, d_j)} \leq \sum_{(i,j)\in E} \left(\frac{\delta}{d_i} + \frac{\delta}{d_j}\right) = \delta n.$$

□

One can also use a modified Hoeffding inequality to prove that the number of edges is highly concentrated around its expectation.

17.3 Relation to the trace

Each edge (i, j) that we do put in $\tilde{A}$ will get weight $1/p_{i,j}$. So, if we define

$$\Delta = D^{-1}(\tilde{A} - A),$$

then

$$\Delta_{i,j} = \begin{cases} \frac{1}{d_i}(\frac{1}{p_{i,j}} - 1) & \text{with probability } p_{i,j}, \text{ and} \\ -\frac{1}{d_i} & \text{with probability } 1 - p_{i,j}. \end{cases}$$

Note that

$$\mathbf{E}\left[\Delta_{i,j}\right] = 0.$$

Let's also observe that $\Delta_{i,j}$ is never too big:

Claim 17.3.1.

$$|\Delta_{i,j}| \leq 1/\delta.$$

Proof. In order for $\Delta_{i,j}$ to not be fixed to zero, it must be the case that $\delta < \min(d_i, d_j)$. So, $1/d_i < 1/\delta$, which takes care of one case. On the other hand,

$$\frac{1}{d_i}\frac{1}{p_{i,j}} = \frac{\min(d_i, d_j)}{d_i \delta} \leq \frac{1}{\delta}.$$

Now, subtracting $1/d_i$ from this term cannot make its magnitude exceed $1/\delta$ because $1/d_i < 1/\delta$. □

As in the last class, we will bound $\lambda_{max}(\Delta)$ by first proving a bound on $\mathbf{E}\left[\mathrm{Tr}\left(\Delta^k\right)\right]$. We will prove

Lemma 17.3.2. *For even k,*

$$\mathbf{E}\left[\mathrm{Tr}\left(\Delta^k\right)\right] \leq \frac{n(2k)^k}{\delta^{k/2}}.$$

Applying Markov's inequality, we then derive

Theorem 17.3.3. *For every $\alpha \geq 1$ and even integer k,*

$$P\left[\lambda_{max}(\Delta) \geq \alpha \frac{2kn^{1/k}}{\sqrt{\delta}}\right] < \alpha^{-k}.$$

This theorem is particularly helpful when we choose k to be approximately $\log_2 n$, in which case $n^{1/k}$ becomes 2. If δ is then larger than $\log^2 n$, the denominator becomes roughly $O(\log n/\sqrt{\delta})$. Had we worked directly from the theorem of Furedi and Komlos, as proved by Van Vu, we would not be able to prove such a theorem unless $\delta = \Omega(n)$.

17.4 Analysis of the Trace

As in the previous lecture, we will bound the trace by observing that

$$\left(\Delta^k\right)_{v_0,v_k} = \sum_{v_1,\ldots,v_{k-1}} \prod_{i=1}^{k} \Delta_{v_{i-1},v_i},$$

and so

$$\mathbf{E}\left[\left(\Delta^k\right)_{v_0,v_k}\right] = \sum_{v_1,\ldots,v_{k-1}} \mathbf{E}\left[\prod_{i=1}^{k} \Delta_{v_{i-1},v_i}\right].$$

As before, this expectation will become zero if there is any edge that is used just once. What is different from the previous lecture is that the matrix is not symmetric. However, $\Delta_{i,j}$ does determine $\Delta_{j,i}$. So that we can write the product as a product of expectations of terms involving the same edge, we will prove:

Lemma 17.4.1. *For all edges (t,r) and integers $k \geq 1$ and $l \geq 0$ such that $k + l \geq 2$,*

$$\mathbf{E}\left[\Delta_{r,t}^k \Delta_{t,r}^l\right] \leq \frac{1}{c^{k+l-1}} \frac{1}{d_r}.$$

Proof. To make life easier, we'll just prove this in the case $k + l = 2$. The other cases are similar. We note that $\Delta_{r,t} = \tilde{A}_{r,t}/d_r$ and $\Delta_{t,r} = \tilde{A}_{r,t}/d_t$, so we can compute

$$\begin{aligned}
\mathbf{E}\left[\Delta_{r,t}^k \Delta_{t,r}^l\right] &= \frac{1}{d_r^k d_t^l}\left(p_{r,t}\left(\frac{1-p_{r,t}}{p_{r,t}}\right)^2 + (1-p_{r,t})\right) \\
&= \frac{1}{d_r^k d_t^l}\left(\frac{1-p_{r,t}}{p_{r,t}}\right) \\
&\leq \frac{1}{d_r^k d_t^l}\left(\frac{1}{p_{r,t}}\right) \\
&= \frac{1}{d_r^k d_t^l}\left(\frac{\min(d_r, d_t)}{\delta}\right).
\end{aligned}$$

To finish the proof, we observe

$$\frac{\min(d_r, d_t)}{d_r d_t} = \frac{1}{\max(d_r, d_t)} \leq \frac{1}{d_r},$$

and

$$\frac{\min(d_r, d_t)}{d_r^2} \leq \frac{1}{d_r}.$$

□