Lecture 13: Spectral sparsfication via sampling CSE 525, Spring 2019
Instructor: James R. Lee

1 Sparsification via effective resistances

Consider the *graph sparsification problem:* Given a graph $G = (V, E)$, we want to approximate G (in a sense to be defined shortly) by a sparse graph $H = (V, E')$. Generally we would like that $E' \subseteq E$ and moreover $|E'|$ is as small as possible—say $O(n)$ or $O(n \log n)$ where $n = |V|$.

1.1 Laplacians of graphs

In everything that follows, we will consider n-vertex graphs with vertex set $V = \{1, 2, \ldots, n\}$. For an edge $e \in E$ with $e = \{i, j\}$ and $i < j$, we define the vector $x_e = e_i - e_j$ where $\{e_i\}$ are the standard basis vectors in $\mathbb{R}^n$. We also define the $n \times n$ matrix $L_e = x_e x_e^T$. Notice that this matrix has $(L_e)_{ii} = (L_e)_{jj} = 1$ and $(L_e)_{ij} = (L_e)_{ji} = -1$; the rest of the entries are zero. This matrix has rank one and is positive semidefinite: For every $v \in \mathbb{R}^n$, we have

$$v^T L_e v = (v_i - v_j)^2 .$$

Now for a graph $G = (V, E)$, we define the *(combinatorial) Laplacian of G* by

$$L_G = \sum_{e \in E} L_e .$$

It should be clear that L_G is also PSD (since it is a nonnegative sum of PSD matrices) and

$$v^T L_G v = \sum_{\{i,j\} \in E} (v_i - v_j)^2 . \tag{1.1}$$

If G is equipped with nonnegative edge weights $\{w_e \geqslant 0 : e \in E\}$, we define the corresponding weighted Laplacian by $L_G = \sum_{e \in E} w_e L_e$.

1.1.1 Spectral sparsification

Spielman and Teng introduced the following notion of *spectral* graph approximation. Consider (possibly weighted) graphs H and G. We say that H *ε-spectrally approximates* G for some $\varepsilon > 0$ if

$$(1 - \varepsilon) L_G \preceq L_H \preceq (1 + \varepsilon) L_G . \tag{1.2}$$

Recall that this equivalent to requiring that for every $v \in \mathbb{R}^n$, we have

$$(1 - \varepsilon) v^T L_G v \leqslant v^T L_H v \leqslant (1 + \varepsilon) v^T L_G v .$$

From this expression, we see that spectral approximation is stronger than cut approximation. Indeed, consider any subset of vertices $S \subseteq V$ and the corresponding characteristic vector $v = \mathbf{1}_S$. Then $v^T L_G v = w_G(E(S, \bar{S}))$ and $v_T L_H v = w_H(E(S, \bar{S}))$, where these two expressions are meant to represent the weight of the edges in G that cross the cut $(S, \bar{S})$ and the weight of the edges in H that cross the cut $(S, \bar{S})$, respectively. In particular, (1.2) entails that the weight of every cut in H should be within $1 \pm \varepsilon$ of the weight of the corresponding cut in G.

1.2 Random sampling

We will prove the following theorem.

Theorem 1.1 (Spielman-Srivastava 2008)**.** *For every $\varepsilon > 0$, the following holds. For every unweighted, connected graph $G = (V, E)$, there exists a weighted graph $H = (V, E')$ such that $E' \subseteq E$ and $|E| \leqslant O(\frac{n \log n}{\varepsilon^2})$ and H ε-spectrally approximates G.*

To see that weights are necessary in Theorem 1.1, consider the case when G is an n-clique. In that case, we will need to put large weights on any sparse graph H that approximates G (recalling that, in particular, the weight of every cut in H should be close to the size of the corresponding cut in G).

A sampling algorithm. Suppose we have some probabilities $p_e \geqslant 0$ such that $\sum_{e \in E} p_e = 1$. Then we can consider the following algorithm: Set all edge weights $w_e := 0$ for every $e \in E$. For $i = 1, 2, \ldots, k$, sample an edge e (independent from previous choices) with probability p_e and update

$$w_e := w_e + \frac{1}{k p_e} .$$

Let H be the corresponding (random) weighed graph.

It is easy to see that $L_H = \sum_{i=1}^{k} Z_i$ where $Z_i = \frac{1}{k p_{e^{(i)}}} L_{e^{(i)}}$ and $e^{(i)}$ is the edge that we sampled in the ith iteration. Moreover, we can calculate for any $i \in \{1, \ldots, k\}$:

$$\mathbb{E}[Z_i] = \sum_{e \in E} p_e \frac{1}{k p_e} L_e = \frac{1}{k} L_G .$$

Therefore $\mathbb{E}[L_H] = L_G$. The expectation of our "approximator" is equal to L_G. Moreover, our approximator L_H is a sum of i.i.d. random matrices. In order to achieve (1.2) using a small value of k, we need concentration for this sum. And in order to achieve concentration using the approach of Lecture 14, we need a bound on the eigenvalues of individual summands.

Claim 1.2. *For any $\kappa \geqslant 1$, if $Z_i \preceq \kappa \, \mathbb{E}[Z_i]$ (with probability one), then we can choose $k = O(\frac{\kappa}{\varepsilon^2} \log n)$ and achieve (1.2) with high probability.*

Proof. Theorem 1.3 of Lecture 14 states that under our assumptions, we have

$$\mathbb{P}\left[(1 - \varepsilon) L_G \preceq \sum_{i=1}^{k} Z_i \preceq (1 + \varepsilon) L_G \right] \geqslant 1 - 2n e^{-\varepsilon^2 k} / 4\kappa .$$

Plugging in some $k = O(\frac{\kappa}{\varepsilon^2} \log n)$, we can get this expression to be at least $1 - 1/n$. □

So we are left to choose sampling probabilities p_e that can guarantee $Z_i \preceq \kappa \, \mathbb{E}[Z_i]$ for some reasonable value of κ. A natural choice would be *uniform* sampling: $p_e = 1/|E|$. But this can fail to give non-trivial bounds: Suppose that G is the union of two disjoint $n/2$-cliques connected by a single edge a. In order for H to spectrally approximate G, we had better include the edge a (otherwise the corresponding cut in H will have weight zero, while it has non-zero weight in G). But this would mean we need to sample $\Theta(n^2)$ edges if we do uniform random sampling (since G contains $\Theta(n^2)$ edges, only one of which is a). That doesn't yield a particularly sparse graph.

Remark 1.3. It is not too difficult to see that no matter what choice we make for $\{p_e\}$ we will need at least $\Omega(n \log n)$ edges to be sampled. Consider G to be the n-path. In that case, we will need to sample every edge at least once to approximate all $n - 1$ cuts in G. The standard coupon collector bound dictates that we will need at least $\Omega(n \log n)$ samples.

1.2.1 Effective resistances

Instead, we need to choose the probabilities $\{p_e\}$ so that important edges (like the single edge a in the preceding example) have a very good chance of being sampled. To do this, we will set p_e to be proportional to the *effective resistance* of the edge e. We'll see more about effective resistances (and their relationship to random walks) in the next lecture.

For now, we can simply give the definition: For every edge e, we set

$$\mathsf{R}_e := \mathrm{Tr}(L_e L_G^+)\,,$$

where we recall that L_G^+ is the *pseudo-inverse of* L_G.

Since G is connected, L_G has rank $n-1$. That is easy to see from the formula (1.1). If v is a multiple of the all-ones vector $(1,1,\ldots,1)$, then $v^T L_G v = 0$. On the other hand, if $v^T L_G v = \sum_{\{i,j\}\in E}(v_i - v_j)^2 = 0$ and G is connected, it must be that $v_1 = v_2 = \cdots = v_n$. In other words, $L_G v = 0$ implies that v is a multiple of $(1,1,\ldots,1)$. Thus L_G has rank $n-1$. That means we can write $L_G = \sum_{i=2}^n \lambda_i w_i w_i^T$ for some orthonormal family of vectors $\{w_i\} \subseteq \mathbb{R}^n$ and $\lambda_i > 0$. One then defines $L_G^+ = \sum_{i=1}^n \frac{1}{\lambda_i} w_i w_i^T$. Now we have $L_G L_G^+ = I_{\mathrm{im}(L_G)}$. We will also need the positive square root: $L_G^{+/2} = \sum_{i=1}^n \frac{1}{\sqrt{\lambda_i}} w_i w_i^T$.

Finally we define our sampling probabilities $p_e = \frac{\mathsf{R}_e}{n-1}$. First, let's verify that these are indeed probabilities. Recalling that $L_e = x_e x_e^T$, we have $\mathsf{R}_e = \mathrm{Tr}(L_e L_G^+) = x_e^T L_G^+ x_e \geqslant 0$ since L_G^+ is PSD (because L_G is PSD). Moreover, using linearity of the trace and the definition of L_G:

$$\sum_{e\in E} \mathsf{R}_e = \mathrm{Tr}\left(\sum_{e\in E} L_e L_G^+\right) = \mathrm{Tr}(L_G L_G^+) = \mathrm{Tr}(I_{\mathrm{im}(L_G)}) = n-1\,,$$

where the last line follows because, as we discussed, L_G has rank $n-1$.

Note that with these sampling probabilities, we have

$$Z_i \preceq \kappa \cdot \mathbb{E}[Z_i] \iff \frac{n-1}{k\mathsf{R}_e} L_e \preceq \kappa \frac{1}{k} L_G \iff L_e \preceq \kappa \frac{\mathsf{R}_e}{n-1} L_G$$

By Claim 1.2 with $\kappa = n-1$, we are done with the proof of Theorem 1.1 after we verify the following.

Lemma 1.4. *For every edge $e \in E$, we have $L_e \preceq \mathsf{R}_e L_G$.*

Proof. We need to prove that $v^T L_e v \leqslant \mathsf{R}_e v^T L_G v$ for every $v \in \mathbb{R}^n$. But we need only prove this for $v \perp (1,\ldots,1)$ since $(1,\ldots,1) \in \ker(L_G) \cap \ker(L_e)$. Since $\mathrm{im}(L_G^{+/2})$ contains every vector orthogonal to $(1,\ldots,1)$, it suffices to prove the inequality for $v = L_G^{+/2} w$ for any $w \in \mathbb{R}^n$ with $w \perp (1,\ldots,1)$:

$$(L_G^{+/2} w)^T L_e (L_G^{+/2} w) \leqslant \mathsf{R}_e (L_G^{+/2} w)^T L_G (L_G^{+/2} w)$$

Using the fact that L_G^+ is symmetric, and $w \perp \ker(L_G)$, the RHS is $\mathsf{R}_e w^T w$. On the other hand, the LHS is $w^T L_G^{+/2} L_e L_G^{+/2} w \leqslant \|L_G^{+/2} L_e L_G^{+/2}\| w^T w \leqslant \mathrm{Tr}(L_G^{+/2} L_e L_G^{+/2}) w^T w = \mathsf{R}_e w^T w$, completing the proof. □

Remark 1.5. The proof is a bit easier to understand in the following way: We want to prove $L_e \preceq \mathsf{R}_e L_G$. It would be nice to simply multiply on the left and right by $L_G^{+/2}$ yielding

$$L_G^{+/2} L_e L_G^{+/2} \preceq \mathsf{R}_e I_{\mathrm{im(G)}}\,.$$

This latter statement is true because the maximum eigenvalue of $L_G^{+/2} L_e L_G^{+/2}$ is certainly at most its trace which is equal to R_e.

If A and B are symmetric matrices and C is invertible, then $CAC^T \preceq CBC^T \iff A \preceq B$ (this is an easy exercise). On the other hand, if C is singular (like $L_G^{+/2}$), then we need to be careful about what happens on $\ker(C)$.