Introduction

The goal of graph sparsification is to take a dense graph and find some sparse (weighted) subgraph which preserves the spectrum of the graph. Sparsification provides a powerful tool for converting large graphs into a much more succinct representation where a lot of the global and local properties of the graph are preserved. Formally, given a dense graph G, we want to find a sparse (weighted) graph H such that the following property holds.

$$(1-\epsilon)L_G \preceq L_H \preceq (1+\epsilon)L_G$$

That is, we require the following bounds on the quadratic form for all $x \in \mathbb{R}^n$.

$$(1-\epsilon)x^T L_G x \leq x^T L_H x \leq (1+\epsilon)x^T L_G x$$

In particular, if x is the characteristic vector of a cut $(S, \overline{S})$ the we obtain a tight relationship between the cut in G and the weighted cut in H.

$$(1-\epsilon)E_G(S,\overline{S}) \leq E_H(S,\overline{S}) \leq (1+\epsilon)E_G(S,\overline{S})$$

Note that this is a very strong requirement. Recall that in our lectures on thin trees we obtained results of the form where $1 \leq E_T(S,\overline{S}) \leq 2E_G(S,\overline{S})$. Note that this condition additionally implies a strong preservation of the spectrum of L_G in L_H. That is, for all $i \in [n]$ the ith eigenvalue is preserved.

$$(1-\epsilon)\lambda_i(G) \leq \lambda_i(H) \leq (1+\epsilon)\lambda_i(G)$$

This type of approximation will hold even for the pseudoinverse, so the effective resistances are preserved. We will show that, while the graph sparsification requirement is strong, we can obtain such a sparsification with high probability through a straight-forward sampling algorithm. The applications of such a sparsification are abundant; in particular, any question of the form $L_G x = b$ can be reframed as $L_H x \approx b$.

Sparsification through Sampling

Perhaps, the simplest idea to build a sparse version of G is to sample edges uniformly at random until the spectrum of H, the sampled graph is close enough to G. The trouble with this algorithm is that it may need to sample $O(n^2)$ edges just to obtain a connected graph. For example, consider two cliques connected by a single cut edge. We will need to sample a quadratic number of edges in expectation to have a resonable chance of having sampled the cut edge. At the same time, in order to have a *multiplicative* bound relating the eigenvalues of the two graphs, we need H to be connected.

Here, we present a sparsification algorithm of Spielman and Srivastava, which fixes this particular problem.

Sparsify(G, q):

- Sample an edge $e \in E$ with probability $p_e = \frac{R_{\text{eff}}(e)}{n-1}$
- add e to H with weight $\frac{1}{p_e q}$
- Repeat q times, sampling independently wiht replacement

Theorem 1. *For some constant c and $q = cn\log n/\epsilon^2$, with probability $p > 1/2$, $H \leftarrow$ Sparsify(G,q) satisfies $(1-\epsilon)L_G \preceq L_H \preceq (1+\epsilon)L_G$.*

This means that after sampling $\Theta(n \log n)$ edges, we can construct a sparsified version of G that preserves the cut structure at least half the time. We can repeat this procedure $\Theta(\log n)$ times (or simply sample more edges) to boost the probability of success to $\Theta(1/\text{poly } n)$. In what follows, we will argue why the algorithm produces a sparsifier as desired.

Proof. First, recall that the sum of effective resistances over all edges is $n-1$, so p_e induces a legal probability distribution over the edges. Then, consider the diagonal $m \times m$ matrix S defined as $S[e,e] =$ weight of e in H. We can see that S is equivalently defined as follows.

$$S[e,e] = (\text{number of times } e \text{ is selected}) \cdot \frac{1}{p_e q}$$

Note that the Laplacian of H can be computed as $L_H = B^T SB$, where S selects and re-weights the edges of $L_G = B^T B$. Moreover, we can see that $\mathbb{E}[L_H] = L_G$. This is most easily seen because $\mathbb{E}[S] = I$, due to the fact that q edges are sampled with probability p_e, and upon sampling e, $\frac{1}{p_e q}$ is added to the weight. An interesting aspect of this algorithm is that it relies on the fact that $\mathbb{E}[S] = I$ and the fact that S is far from I – remember that S is sparse, and only $O(n \log n)$ of its $O(n^2)$ entries will be non-zero.

Now we will argue that L_H and L_G are spectrally similar with high probability. Recall we defined $Y = BL^\dagger B^T$, which has the property that $Y^2 = BL^\dagger B^T BL^\dagger B = BL^\dagger LL^\dagger B^T = Y$. We also note that $\mathrm{Im}(Y) = \mathrm{Im}(B)$. Clearly, $\mathrm{Im}(Y) \subseteq \mathrm{Im}(B)$. If $z \in \mathrm{Im}(B)$ then $z = Bx$ for some x, so $Yz = BL^\dagger B^T Bx = Bx = z$, and $z \in \mathrm{Im}(Y)$. With this in mind, we show the following lemma.

Lemma 2. *If $\|YSY - Y\|_2 \le \epsilon$, then $(1-\epsilon)L_G \preceq L_H \preceq (1+\epsilon)L_G$.*

To see this lemma, consider some z in the image of B, and therefore the image of Y.

$$z^T YSYz = xB^T BL^\dagger B^T SBL^\dagger B^T Bx = x^T B^T SBx = x^T L_H x \tag{1}$$

$$z^T Yz = x^T B^T YBx = x^T L_G x \tag{2}$$

$$z^T z = x^T B^T Bx = x^T L_G x \tag{3}$$

Note that it should not be surprising that $z^T z = z^T Yz$, as Y acts as the identity for $z \in \mathrm{Im}(Y)$. We know that the matrix norm maximizes the quadratic forms in (1) and (2), so if $\|YSY - Y\|_2 \le \epsilon$ then $|x^T L_H x - x^T L_G x| \le \epsilon$. Thus, we get the following inequalities

$$-\epsilon \le \frac{x^T L_H x - x^T L_G x}{x^T L_G x} \le \epsilon$$

which rearrange to the desired conditions.

Finally, we use a Chernoff Bound to argue that the $\|YSY - Y\|_2 \le \epsilon$ with probability $p > 1/2$.

Lemma 3. *Let D be some probability distribution over $\Omega \subseteq \mathbb{R}^d$ where for all $z \in \Omega$, $\|\mathbb{E}[zz^T]\|_2 \le 1$ and $\|z\|_2 \le M$. Suppose we take q independent samples from D, $z_1, z_2, \ldots, z_q$. Then the expected difference between the empirical mean and the expected value of the outer products of $z \in \Omega$ is bounded as follows.*

$$\mathbb{E}\left[\left\|\frac{1}{q}\sum_i z_i z_i^T - \mathbb{E}[zz^T]\right\|_2\right] \le \min\left(cM\sqrt{\frac{\log q}{q}}, 1\right)$$

That is, if we have a distribution over a subset of $\mathbb{R}^d$ and the expected norm of the outer product of a sample is bounded by 1 and every supported z has bounded norm, then we get concentration about the expected value of $\mathbb{E}[zz^T]$. So consider the following distribution: $z = \frac{Y(\cdot,e)}{\sqrt{p_e}}$ with probability p_e.

Then, for every z, the norm is bounded as follows.

$$\|z\|_2 = \sqrt{\frac{\langle Y(\cdot,e), Y(e,\cdot)\rangle}{p_e}} = \sqrt{\frac{Y(e,e)}{p_e}} = \sqrt{\frac{R_{\text{eff}}}{\frac{R_{\text{eff}}(e)}{n-1}}} = \sqrt{n-1}$$

Additionally, we can see expected value of zz^T is Y.

$$\mathbb{E}\left[zz^T\right] = \sum_e p_e \frac{Y(\cdot, e)Y(\cdot, e)^T}{p_e} = Y$$

Thus, if we allow $q = \Theta(\frac{c^2 n \log n}{\epsilon^2})$, we can see that the conditions will be met with probability $p \geq 1/2$.

$\square$

As a conclusion, we note that sparsifying a complete graph can be seen as one way of sampling an Erdös-Rényi graph $H \sim G(n, q/\binom{n}{2})$. Also, if we know that our graph is complete, a random d-regular graph actually gives better sparsification. In fact, we know that for a d-regular graph, if we define $\sigma_2 = \max(\lambda_2(A), -\lambda_n(A))$ then $\sigma_2 \leq 2\sqrt{d-1} + o(1)$.