Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications.

For two symmetric matrix $A, B \in \mathbb{R}^{n \times n}$ we write

$$A \preceq B$$

iff $B - A \succeq 0$, i.e., $B - A$ is a PSD matrix. In other words, $A \preceq B$ iff for any vector $x \in \mathbb{R}^n$,

$$x^T A x \leq x^T B x$$

Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $A$ and $\tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_n$ be the eigenvalues of $B$. It follows that if $A \preceq B$, then for all $i$, $\lambda_i \leq \tilde{\lambda}_i$.

**Definition 14.1.** Given a graph $G = (V, E)$ and $\epsilon > 0$, we say a (weighted) graph $H = (V, E')$ is a $1 \pm \epsilon$-spectral sparsifier of $G$ if

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

Ideally, we want $H$ to be a subgraph of $G$ which has much fewer edges than $G$. An immediate consequence of the above definition is that all eigenvalues of $H$ approximate eigenvalues of $H$ up to multiplicative $1 \pm \epsilon$ error.

It is also not hard to see that if $H$ is a $1 \pm \epsilon$-spectral sparsifier of $G$ then it preserves the size of all cuts of $G$. In particular, for a set $S \subseteq V$, recall $1^S$ is the indicator vector of the set $S$. It follows that for a graph $G$,

$$1^S L_G 1^S = \sum_{i \sim j} (1_i^S - 1_j^S)^2 = \sum_{i \sim j} |\{i, j\} \cap S| = 2|E(S, \overline{S})|$$

So, if $H$ is a $1 \pm \epsilon$-spectral sparsifier of $G$ we have

$$(1 - \epsilon) 1^S L_G 1^S \leq 1^S L_H 1^S \leq (1 + \epsilon) 1^S L_G 1^S,$$

so the (weighted) size of every cut in $H$ is within $1 \pm \epsilon$ multiplicative factor of the same cut in $G$.

**Theorem 14.2** (Speilman-Srivastava). For every graph $G = (V, E)$ and $\epsilon > 0$, there is a weighted graph $H$ that is a subgraph of $G$ such that $H$ is a $1 \pm \epsilon$-spectral sparsifier of $G$ and that $H$ has at most $O(n \log n/\epsilon^2)$ many edges.

The first idea that come to mind is to construct an unbiased estimator: Let $X$ be a random matrix defined as follows: For every edge $e \in E$, $X = L_e/p_e$ with probability $p_e$. Then, observe that

$$E[X] = \sum_e p_e L_e = \sum_e L_e = L_G.$$ 

So, $X$ is an unbiased estimator. And, the main question is how to choose the probabilities such that concentration bounds can kick in and imply $X \approx E[X]$.

Let us start with a simple case of a complete graph. If $G$ is a complete graph, we can simply let $p_e = 1/(\binom{n}{2})$ for all edges. It then follows that $O(n \log n/\epsilon^2)$ many samples are enough to approximate the complete
Lecture 14: Spectral Sparsification of Graphs

Figure 14.1: Barbell Graph

graph. However, it turns out that a uniform distribution does not necessarily work out in a general graph.
For example, if $G$ is a Barbell graph, i.e., union of two $K_n$ connected by an edge (see Figure 14.1), then, if
we want to down-size $G$ to $O(n \log n)$ edges we need to let $p_e = O(\log n)/n$ for all edges, but then the single
edge connecting the two complete graphs won’t be chosen with high probability. So, $H$ is disconnected with
high probability and it cannot be a spectral sparsifier of $G$ for any $\epsilon < 1$. In the rest of this section we will
see how to choose the edge probabilities $p_e$.

14.0.1 Reduction to Isotropic Case

First, it turns out that we can reduce the graph sparsification problem to a linear algebraic problem. First, let
us recall the generalized eigenvalue problem. In the generalized eigenvalue problem we are given a symmetric
matrix $A$ and a PSD matrix $B$ and we want to find

$$\max_x x^T Ax$$

In the special case that $B$ is the identity matrix, the solution of the above problem is exactly the largest
eigenvector of $A$. We can solve the above problem by reducing it to an eigenvalue problem.

$$\max_x x^T B^{1/2} AB^{-1/2} B^{1/2} x$$

So, to find the solution to the generalized eigenvalue problem it is enough to find the largest eigenvector
$y$ of the matrix $B^{-1/2} AB^{-1/2}$ and then let $x = B^{-1/2} y$. Note that, here we are using the fact that $B$ is PSD;
otherwise $B^{-1/2}$ is not well defined.

Now, let us go back to the spectral sparsifier problem. Suppose $H$ is a $1 \pm \epsilon$-spectral sparsifier of $G$. It
follows that for all $x \in \mathbb{R}^n$.

$$1 - \epsilon \leq \frac{x^T L_H x}{x^T L_G x} \leq 1 + \epsilon$$

By a similar analogy, it follows that for all $y$,

$$1 - \epsilon \leq \frac{y^T L_G^{-1/2} L_H L_G^{-1/2} y}{y^T y} \leq 1 + \epsilon$$

So, the above inequality implies that the matrix $L_G^{-1/2} L_H L_G^{-1/2}$ is approximately equal to the identity matrix.

Remark 14.3. There is a technical problem here: since $L_G$ has a zero eigenvalue the inverse of $L_G$ is
not well-defined. In the above calculation, we take the inverse with respect to positive eigenvalues of $G$;
in particular if $L_G = \sum \lambda_i v_i v_i^T$, we let $L_G^{-1/2} = \sum_{\lambda_i > 0} \frac{1}{\sqrt{\lambda_i}} v_i v_i^T$. We ignore this fact in the rest of our
calculations for the simplicity of the argument.
Now, we reformulate the spectral sparsification problem as follows:

**Theorem 14.4.** Given \( n \times n \) PSD matrices, \( E_1, \ldots, E_m \) such that
\[
\sum_{i=1}^{m} E_i = I,
\]

For any \( \epsilon > 0 \), there is a subset \( S \) of them of size \( O(n \log n/\epsilon^2) \) and a set of weights \( w_i \) for each \( i \in S \) such that
\[
(1 - \epsilon)I \preceq \sum_{i \in S} w_i E_i \preceq (1 + \epsilon)I
\]

Let us discuss how we can reduce the sparsification problem to the above theorem. Say our graph \( G \) has \( m \) edges. For edge \( e_i \) define
\[
E_i = L_G^{-1/2} L_e L_G^{-1/2}.
\]

First, observe that each \( E_i \) is a PSD matrix, and furthermore,
\[
\sum_{i=1}^{m} E_i = \sum_{i=1}^{m} L_G^{-1/2} L_e L_G^{-1/2} = L_G^{-1/2} \left( \sum_{i=1}^{m} L_e \right) L_G^{-1/2} = L_G^{-1/2} L_G L_G^{-1/2} = I.
\]

So, roughly speaking by multiplying the Laplacians of the edges of \( G \) by \( L_G^{-1/2} \) on both sides we are normalizing the space such that every direction look the same. We are reducing the graph spectral sparsification problem to a linear algebraic problem of finding a sparsifier of the sum of PSD matrices that add up to the identity matrix.

### 14.0.2 Finding the Spectral Sparsifier

Now, as before, let
\[
X = \frac{E_i}{p_i}
\]

with probability \( p_i \). Similar to before, \( \mathbb{E}[X] = I \); also \( X \) is a distribution over PSD matrices. To prove the concentration we used the matrix Chernoff bound we proved in the previous lecture:

**Theorem 14.5.** Let \( X \) be a random \( n \times n \) PSD matrix. Suppose that \( \|X\| \leq \alpha \) with probability 1 and \( \mathbb{E}[S] = I \). Let \( X_1, \ldots, X_k \) be independent copies of \( X \), then for any \( \epsilon > 0 \),
\[
\mathbb{P}\left[ (1 - \epsilon)I \preceq \frac{1}{k} (X_1 + \cdots + X_k) \preceq (1 + \epsilon)I \right] \geq 1 - 2ne^{-\epsilon^2 k/4\alpha}.
\]

So, this says that to prove Theorem 14.4 it is enough to choose \( k = O(\alpha \log n/\epsilon^2) \) many copies of \( X \). To finish the proof all we need to choose are the probabilities \( p_i \). Here comes the important choice; we need to choose \( p_i \)'s such that \( \alpha \leq O(n) \).

First, suppose we let \( p_i \) be uniform, i.e., \( p_i = 1/m \) for all \( i \). Then, we need to choose \( \alpha \) such that for all \( i \),
\[
\frac{E_i}{1/m} \leq \alpha I.
\]

But it turns out that in the worst case we have to let \( \alpha = m \).
The idea is to let $p_i \propto \text{Tr}(E_i)$. Let us first find the normalizing constant: Suppose $p_i = \beta \text{Tr}(E_i)$. Then,

$$\sum_i p_i = \beta \sum_i \text{Tr}(E_i) = \beta \text{Tr} \left( \sum_i E_i \right) = \beta \text{Tr}(I) = \beta n$$

So, we should let $\beta = 1/n$. It follows that $p_i = \beta \text{Tr}(E_i) = \text{Tr}(E_i)/n$.

Now, we claim that for all $i$,

$$\frac{E_i}{\text{Tr}(E_i)/n} \preceq \alpha I$$

for $\alpha = n$. This will complete the proof of Theorem 14.4. To show the above it is enough to show with probability 1,

$$\frac{E_i}{\text{Tr}(E_i)} \preceq I \iff \lambda_{\text{max}}(\frac{E_i}{\text{Tr}(E_i)}) \leq 1.$$

But this is true for any PSD matrix simply because $\text{Tr}(E_i) \geq \lambda_{\text{max}}(E_i)$.

### 14.1 Back to Spectral Sparsification

In the previous section we saw that we should choose each $E_i$ with probability $\text{Tr}(E_i)/n$. Translating this back to the setting of graph sparsification; recall that for edge $e_i$, $E_i = L_G^{-1/2} L_e L_G^{-1/2}$. So, we should sample every edge $e$ of $G$ with probability

$$p_e = \frac{\text{Tr}(L_G^{-1/2} L_e L_G^{-1/2})}{n}$$

The quantity

$$\text{Tr}(L_G^{-1/2} L_e L_G^{-1/2}) = b_e^T L_G^{-1} b_e$$

is exactly the effective resistance of the edge $e$.

The following simple algorithm can be used to construct a $1 \pm \epsilon$-spectral sparsifier of $G$:

1. For $i = 1$ to $O(n \log n/\epsilon^2)$

2. Sample each edge $e$ of $G$ with probability $p_e = \text{Tr}(L_G^{-1/2} L_e L_G^{-1/2})/n$. If the edge $e$ is sampled weight it by $1/p_e$. 