COS 598I Lecture 1: Graph sparsification

Lecturer: Pravesh Kothari   Scribe: Ştefan Tudose

Spring 2025

1 Introduction

Let $G = (V, E)$ be a finite graph and let $w : E \to \mathbb{R}_+$ be a weight function. Given a set $S \subset V$ of vertices, we define $E(S)$ as all the edges in E between a vertex in S and one in $V - S$. We define the *cut* of S in G as

$$\mathrm{cut}_G(S) = \sum_{e \in E(S)} w_e$$

If no weights are specified, then it is understood that all the weights are identically one, in which case $\mathrm{cut}_G(S)$ is just the number of edges between S and $V - S$.

Definition: Given a graph G, a subgraph G' is said to be an *ε-cut-sparsifier* of G if

$$\mathrm{cut}_{G'}(S) \in (1 - \varepsilon, 1 + \varepsilon)\mathrm{cut}_G(S)$$

for any subset S of V.

The main result that we are going to prove is that given a graph on n vertices , there exists an ε-cut-sparsifier of G with $O(n \log n/\varepsilon^2)$ edges. Let us now mention a result which will play a key role in the proof:

Theorem (Matrix Khintchine inequality)**.** Let $A_1, A_2, \ldots, A_k$ be symmetric $n \times n$ real matrices and let $s_1, s_2, \ldots, s_k$ be i.i.d. symmetric Bernoulli random variables, i.e. $\mathbb{P}(s_i = 1) = \mathbb{P}(s_i = -1) = \frac{1}{2}$. Then

$$\mathbb{E}\|s_1 A_1 + \ldots + s_k A_k\| \leq c\|A_1^2 + \ldots + A_k^2\|^{1/2}\sqrt{\log n}$$

for some universal constant $c > 0$.

In section 2 we introduce the graph Laplacian, which is a crucial tool in the proof, and we explore some of its properties. In section 3 we state and prove the main result.

2 The graph Laplacian

2.1 Definition

Given a graph $G = (V, E)$ on n vertices (with no weights, or equivalently all weights equal to one), we define the graph Laplacian as

$$L_G = \begin{pmatrix} \deg v_1 & & \\ & \ddots & \\ & & \deg v_n \end{pmatrix} - A$$

where A is the adjacency matrix given by $A_{ij} = 1_{(v_i,v_j)\in E}$. Note that L_G is a symmetric matrix.

We can define for each edge $e = (i, j)$ the matrix L_e which has $(L_e)_{ii} = (L_e)_{jj} = 1,\ (L_e)_{ij} = (L_e)_{ji} = -1$ and all the other entries zero. We then have

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

Note that if $x \in \mathbb{R}^n$ then

$$x^T L_e x = (x_i - x_j)^2$$

so

$$x^T L_G x = \sum_{(i,j)\in E} (x_i - x_j)^2$$

This shows in particular that L_G is a positive semidefinite matrix.

If the graph G has a weight function $w : E \to \mathbb{R}_+$ one can similarly define the graph Laplacian as

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

in which case L_G is again symmetric and since

$$x^T L_G x = \sum_{(i,j)\in E} w_e(x_i - x_j)^2$$

it is also positive semidefinite.

2.2 The pseudoinverse

Consider the (unweighted) graph $G = (V, E)$ on n vertices. The graph Laplacian is not invertible since $(1, 1, \ldots, 1)$ is in the kernel of L_G. Nonetheless, L_G is a symmetric matrix, so if we let $v_1, \ldots, v_n$ be an orthonormal basis (of $\mathbb{R}^n$) of eigenvectors with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$L_G = \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

Since L_G is positive semidefinite, we have $\lambda_i \geq 0$, so we can define its *pseudoinverse*

$$L^+ = \sum_{\lambda_i>0} \frac{1}{\lambda_i} v_i v_i^T$$

and in particular we can also write

$$(L^+)^{1/2} = \sum_{\lambda_i>0} \frac{1}{\sqrt{\lambda_i}} v_i v_i^T$$

Define the following normalized version of L_e:

$$\tilde{L}_e = (L^+)^{1/2} L_e (L^+)^{1/2}$$

Note that

$$\sum_{e\in E} \tilde{L}_e = (L^+)^{1/2} L_G (L^+)^{1/2} = I_{(\mathrm{Ker} L_G)^\perp}$$

Indeed, a simple computation using $L_G = \sum_{i=1}^{n} \lambda_i v_i v_i^T,\ (L^+)^{1/2} = \sum_{\lambda_i > 0} \frac{1}{\sqrt{\lambda_i}} v_i v_i^T$, shows that

$$(L^+)^{1/2} L_G (L^+)^{1/2} = \sum_{\lambda_i > 0} v_i v_i^T$$

which is the projection on the eigenspaces corresponding to strictly positive eigenvalues. We will need the following in the proof of the main theorem:

Lemma: We have that

$$\sum_{e \in E} \|\tilde{L}_e\| < n$$

Proof. Note that $L_e = (e_i - e_j)(e_i - e_j)^T$ where e_i is the vector having 1 on position i and 0 everywhere else. Then $\tilde{L}_e = \left((L^+)^{1/2}(e_i - e_j)\right)\left((L^+)^{1/2}(e_i - e_j)\right)^T$ so $\tilde{L}_e = xx^T$ for some vector x. This implies that $\|\tilde{L}_e\| = \mathrm{Tr}\ \tilde{L}_e$ and therefore

$$\sum_{e \in E} \|\tilde{L}_e\| = \mathrm{Tr} \sum_{e \in E} \tilde{L}_e = \sum_{j=1}^{n} \langle \sum_{e \in E} \tilde{L}_e v_j, v_j \rangle = \sum_{j=1}^{n} \langle \sum_{\lambda_i > 0} v_i v_i^T v_j, v_j \rangle = \sum_{\lambda_i > 0} 1 < n$$

□

As an observation which will be useful later, $w_e \tilde{L}_e$ can also be written as xx^T by choosing $x = \sqrt{w_e}(L^+)^{1/2}(e_i - e_j)$. This implies in particular that $\|w_e \tilde{L}_e\| = \mathrm{Tr}(w_e \tilde{L}_e)$.

2.3 Connection to cuts and sparsification

Let $G = (V, E)$ be a graph with weight function $w : E \to \mathbb{R}_+$. Given a subset $S \subset V$ of vertices, we can choose the vector y as follows: $y_i = \frac{1}{2}$ if vertex i is in S and $y_i = -\frac{1}{2}$ otherwise. Then

$$y^T L_e y = 1_{e \in E(S)}$$

and so

$$y^T L_G y = \sum_{e \in E} w_e y^T L_e y = \sum_{e \in E} w_e 1_{e \in E(S)} = \sum_{e \in E(S)} w_e = \mathrm{cut}_G(S)$$

Consider now the unweighted graph $G = (V, E)$ and consider a weighted subgraph $G' = (V, E')$ with weight function $w : E' \to \mathbb{R}_+$. The condition that G' is an ε-cut-sparsifier of G is

$$|\mathrm{cut}_{G'}(S) - \mathrm{cut}_G(S)| \le \varepsilon \mathrm{cut}_G(S) \text{ for any } S \subset V$$

which by the above observation can be equivalently written as

$$|y^T L_{G'} y - y^T L_G y| \le \varepsilon y^T L_G y \text{ for any } y \in \{\pm 1/2\}^n$$

If the stronger condition

$$|y^T L_{G'} y - y^T L_G y| \le \varepsilon y^T L_G y \text{ for any } y \in \mathbb{R}^n$$

holds, then in particular G' is an ε-cut-sparsifier of G. Since $(L^+)^{1/2}$ is invertible, we can let $y = (L^+)^{1/2}x$ and the stronger condition is equivalent to

$$|x^T \tilde{L}_{G'} x - x^T I_{(KerL_G)^\perp} x| \leq \varepsilon x^T I_{(KerL_G)^\perp} x \text{ for any } x \in \mathbb{R}^n$$

or equivalently[1]

$$|x^T \left(\sum_{e \in E'} w_e \tilde{L}_e \right) x - x^T x| \leq \varepsilon x^T x \text{ for any } x \in \mathbb{R}^n$$

which in turn is equivalent to

$$\left\| \sum_{e \in E'} w_e \tilde{L}_e - I_n \right\| \leq \varepsilon$$

To summarize, if

$$\left\| \sum_{e \in E'} w_e \tilde{L}_e - I_n \right\| \leq \varepsilon$$

then $G' = (V, E')$ with $w : E' \to \mathbb{R}_+$ is an ε-cut-sparsification of $G = (V, E)$.

3 Graph sparsification

We now state the main result:

Theorem (Spielman-Srivastava '06)**.** Given a graph $G = (V, E)$ on n vertices, there exists an ε-cut-sparsifier $G' = (V, E')$ with

$$|E'| = O(n \log n / \varepsilon^2)$$

Proof. We will construct by induction the subgraphs with edges $E = E_0 \supset E_1 \supset E_2 \supset \ldots$ with the following three properties for $i \geq 1$:

$$\frac{1}{2}|E_{i-1}| \leq |E_i| \leq \frac{3}{4}|E_{i-1}| \tag{1}$$

$$\sum_{e \in E_i} \|w_e^i \tilde{L}_e\| < 2n \tag{2}$$

$$\left\| \sum_{e \in E_i} w_e^i \tilde{L}_e - \sum_{e \in E_{i-1}} w_e^{i-1} \tilde{L}_e \right\| \leq c_0 \sqrt{\frac{n \log n}{|E_{i-1}|}} \tag{3}$$

where $w^i : E_i \to \mathbb{R}_+$ is the weight of $G_i = (V, E_i)$ (which we will construct inductively as well) and $c_0 > 0$ is a universal constant .

The result follows quickly once we have constructed the sets E_i as above. Indeed, by (3)

$$\left\| \sum_{e \in E_k} w_e^k \tilde{L}_e - I_n \right\| \leq \sum_{i=1}^{k} \left\| \sum_{e \in E_i} w_e^i \tilde{L}_e - \sum_{e \in E_{i-1}} w_e^{i-1} \tilde{L}_e \right\| \leq c_0 \sqrt{n \log n} \sum_{i=1}^{k} \frac{1}{\sqrt{|E_{i-1}|}} \leq C \frac{\sqrt{n \log n}}{\sqrt{|E_k|}}$$

[1]The only thing to note is that if $x \in \mathrm{Ker} L_G$ then $(L^+)^{1/2}x = 0$.

where we have used that $|E_i| \geq (4/3)^{k-i}|E_k|$. Thus, if we want an ε-cut-sparsifier, it suffices to have

$$\sqrt{\frac{n \log n}{|E_k|}} \sim \varepsilon$$

so $|E_k| \sim \dfrac{n \log n}{\varepsilon^2}$ and therefore $G_k = (V, E_k)$ is an ε-cut-sparsifier with $O(n \log n/\varepsilon^2)$ edges.

For the induction step, assume that we are given the subgraph $G_i = (V, E_i)$ with the weight function $w^i : E_i \to \mathbb{R}_+$ and we want to construct $G_{i+1} = (V, E_{i+1})$ and its weight function $w^{i+1} : E_{i+1} \to \mathbb{R}_+$. As a preliminary remark, by the induction hypothesis we can show like above that

$$\left\| \sum_{e \in E_i} w_e^i \tilde{L}_e - I_n \right\| \leq C \frac{\sqrt{n \log n}}{\sqrt{|E_i|}}$$

which implies by the triangle inequality that

$$\left\| \sum_{e \in E_i} w_e^i \tilde{L}_e \right\| \leq 1 + C \frac{\sqrt{n \log n}}{\sqrt{|E_i|}} < 2$$

since we have $|E_i| \geq |E_k|$ and $\sqrt{\dfrac{n \log n}{|E_k|}} \sim \varepsilon$.

The plan is now as follows: the edges with a "large" value of $\|w_e^i \tilde{L}_e\|$ will be kept in E_{i+1} (with the same weights), and we will choose a a subset of the edges with a "small" value of $\|w_e^i \tilde{L}_e\|$ to keep in E_{i+1} and their weights will double.

We know by (2) that

$$\sum_{e \in E_i} \|w_e^i \tilde{L}_e\| < 2n$$

This implies that half of the terms in the sum above are less than $4n/|E_i|$. To see that, let $|E_i| = 2m$ and label the edges in E_i as $e_1, e_2, \ldots, e_{2m}$ so that

$$\|w_{e_1}^i \tilde{L}_{e_1}\| \leq \|w_{e_2}^i \tilde{L}_{e_2}\| \leq \ldots \leq \|w_{e_{2m}}^i \tilde{L}_{e_{2m}}\|$$

Then

$$m \cdot \|w_{e_m}^i \tilde{L}_{e_m}\| \leq \sum_{j=m+1}^{2m} \|w_{e_j}^i \tilde{L}_{e_j}\| < 2n$$

so $\|w_{e_j}^i \tilde{L}_{e_j}\| \leq \|w_{e_m}^i \tilde{L}_{e_{2m}}\| < 2n/m = 4n/|E_i|$ for any $j \leq m$.

We will leave edges $e_{m+1}, \ldots, e_{2m}$ in E_{i+1} with their weights untouched. For the rest of the edges, we use the following

Lemma: There exists a choice of signs $s_1, \ldots, s_m \in \{-1, 1\}$ such that

$$\left\| \sum_{j=1}^{m} s_j w_{e_j}^i \tilde{L}_{e_j} \right\| \leq c_0 \sqrt{\frac{n \log n}{|E_i|}}$$

for some universal constant $c_0 > 0$.

Proof. By the matrix Khintchine inequality we have

$$\mathbb{E}\left\|\sum_{j=1}^{m} s_j w^i_{e_j}\tilde{L}_{e_j}\right\| \leq c\left\|\sum_{j=1}^{m}(w^i_{e_j}\tilde{L}_{e_j})^2\right\|^{1/2}\sqrt{\log n}$$

where $s_1,\ldots,s_n$ are i.i.d. symmetric Bernoulli random variables. Since $w_{e_j}\tilde{L}_{e_j}$ is a matrix of the form xx^T for some $x\in\mathbb{R}^n$, it follows that $(w^i_{e_j}\tilde{L}_{e_j})^2=(x^Tx)(xx^T)=\|w^i_{e_j}\tilde{L}_{e_j}\|w^i_{e_j}\tilde{L}_{e_j}$. We therefore have

$$\sum_{j=1}^{m}(w^i_{e_j}\tilde{L}_{e_j})^2=\sum_{j=1}^{m}\|w^i_{e_j}\tilde{L}_{e_j}\|w^i_{e_j}\tilde{L}_{e_j}\leq\frac{4n}{|E_i|}\sum_{j=1}^{m}w^i_{e_j}\tilde{L}_{e_j}\leq\frac{4n}{|E_i|}\sum_{e\in E_i}w^i_e\tilde{L}_e$$

where the inequality is in the sense of positive semidefinite matrices, i.e. $A\leq B$ iff $v^TAv\leq v^TBv$ for all $v\in\mathbb{R}^n$. This implies that

$$\left\|\sum_{j=1}^{m}(w^i_{e_j}\tilde{L}_{e_j})^2\right\|\leq\frac{4n}{|E_i|}\left\|\sum_{e\in E_i}w^i_e\tilde{L}_{e_j}\right\|\leq\frac{8n}{|E_i|}$$

□

Out of the two signs in the previous lemma, one sign corresponds to at most half of the edges, assume without the loss of the generality that it's the plus sign. We then keep all the edges with $s_j=1$ in E_{i+1} and double their weights. We don't keep any of the edges with $s_j=-1$ in E_{i+1}.

To summarize what we've done so far, we started with the edges $e_1,\ldots,e_{2m}$ of E_i and we decided that that we construct E_{i+1} as follows

$$\begin{cases} e_j\in E_{i+1},\ w^{i+1}_{e_j}=w^i_{e_j} & \text{if } m+1\leq j\leq 2m\\ e_j\in E_{i+1},\ w^{i+1}_{e_j}=2w^i_{e_j} & \text{if } j\leq m \text{ and } s_j=1\\ e_j\notin E_{i+1} & \text{if } j\leq m \text{ and } s_j=-1\end{cases}$$

We will now check that E_{i+1} and $w^{i+1}:E_{i+1}\to\mathbb{R}_+$ satisfy the three assumptions from our induction hypothesis. Note that because we chose the sign with at most half of the first m edges, we have

$$\frac{1}{2}|E_i|\leq|E_{i+1}|\leq\frac{3}{4}|E_i|$$

which gives assumption (1) for $i+1$.

We also have by construction

$$\left\|\sum_{e\in E_{i+1}}w^{i+1}_e\tilde{L}_e-\sum_{e\in E_i}w^i_e\tilde{L}_e\right\|=\left\|\sum_{j=1}^{m}s_jw^i_{e_j}\tilde{L}_{e_j}\right\|\leq c_0\sqrt{\frac{n\log n}{|E_i|}}$$

which is precisely assumption (2).

Finally, since $|\mathrm{Tr}\ A|\leq n\|A\|$ for any $n\times n$ matrix, note that

$$\left|\sum_{e\in E_{i+1}}\|w^{i+1}_e\tilde{L}_e\|-\sum_{e\in E_i}\|w^i_e\tilde{L}_e\|\right|=\left|\mathrm{Tr}\left(\sum_{e\in E_{i+1}}w^{i+1}_e\tilde{L}_e-\sum_{e\in E_i}w^i_e\tilde{L}_e\right)\right|\leq n\left\|\sum_{e\in E_{i+1}}w^{i+1}_e\tilde{L}_e-\sum_{e\in E_i}w^i_e\tilde{L}_e\right\|$$

which by the above implies

$$\left| \sum_{e \in E_{i+1}} \|w_e^{i+1} \tilde{L}_e\| - \sum_{e \in E_i} \|w_e^i \tilde{L}_e\| \right| \leq nc_0 \sqrt{\frac{n \log n}{|E_i|}}$$

Writing the analogous relation for all the smaller indices, summing them up and using

$$\sum_{e \in E} \|\tilde{L}_e\| < n$$

we obtain

$$\sum_{e \in E_{i+1}} \|w_e^{i+1} \tilde{L}_e\| \leq n \left(1 + c_0 \sqrt{n \log n} \sum_{j=0}^{i-1} \frac{1}{\sqrt{|E_j|}} \right) \leq n \left(1 + C \sqrt{\frac{n \log n}{|E_i|}} \right) < 2n$$

which is assumption (3). □

It turns out that using the matrix Khintchine inequality is lossy and one can do much better by a deterministic choice of sign. It follows from the work of Batson-Spielman-Srivastava that the following is true

Theorem. Given $A_1, \ldots, A_k$ symmetric $n \times n$ matrices of rank 1 such that $\|A_i\| \leq \rho$ for all $1 \leq i \leq k$, then there exists a choice of signs $s_1, \ldots, s_k \in \{-1, 1\}$ such that

$$\|s_1 A_1 + \ldots + s_k A_k\| \leq c\rho^{1/2}$$

for a universal constant $c > 0$.

Note that the lemma we have used in the proof above gave[2] the bound $c\rho^{1/2}\sqrt{\log n}$, so this theorem is a significant improvement as it gets rid of the $\sqrt{\log n}$ term. It's immediate to see that if we use this theorem instead of our lemma, the error will now be

$$\left\| \sum_{e \in E_k} w_e^k \tilde{L}_e - I_n \right\| \leq \sum_{i=1}^{k} \left\| \sum_{e \in E_i} w_e^i \tilde{L}_e - \sum_{e \in E_{i-1}} w_e^{i-1} \tilde{L}_e \right\| \leq c_0 \sqrt{n} \sum_{i=1}^{k} \frac{1}{\sqrt{|E_{i-1}|}} \leq C \frac{\sqrt{n}}{\sqrt{|E_k|}} \sim \varepsilon$$

so we can obtain an ε-cut-sparsifier with $O(n/\varepsilon^2)$ edges.

[2] In the case of our proof we had $A_j = w_{e_j}^i \tilde{L}_{e_j}$ and $\rho = \dfrac{4n}{|E_i|}$.