CS 388R: Randomized Algorithms, Fall 2021 October 26th, 2021

Lecture 17: Spectral Sparsification of Graphs

Prof. Eric Price *Scribe: Raymond Hong, Youguang Chen*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In the last lecture, we learned the matrix concentration inequalities: the Bernstein concentration inequality and the Rudelson-Vershynin (RV) Lemma. At the end of the lecture, we introduced the graph sparsification problem.

In this lecture, we will continue the discussion of spectral sparsification of graphs. In particular, we will cover the following contents:

- introduction to the spectral graph sparsification problem (Section 2),
- effective resistance of edge when we treat the graph as an electrical network (Section 3),
- the graph sparsification algorithm given by Spielman and Srivastava [SS08] (Section 4.1),
- analysis of the algorithm on complete graphs (Section 4.2) and on general connected graphs (Section 4.3).

2 Setup

2.1 Graph representation

Let $G = (V, E)$ be an unweighted and undirected graph with n vertices and m edges. The graph Laplacian of G is defined by a $n \times n$ matrix $L_G = D - A$, where $A \in \mathbb{R}^{n\times n}$ is the graph adjacency matrix such that $A(u, v) = 1$ for $(u, v) \in E$ and 0 otherwise, and $D \in \mathbb{R}^{n\times n}$ is the diagonal matrix of vertex degree, i.e. $D(u, u) = \sum_{v\in V} A(u, v)$.

For the ease of our discussion, we need to express Laplacian in another way. If we orient the edges of G arbitrarily, we can define its signed edge-vertex incidence matrix $U \in \mathbb{R}^{m\times n}$, given by

$$U(e,u) = \begin{cases} 1 & \text{if vertex } u \text{ is edge } e\text{'s head} \\ -1 & \text{if vertex } u \text{ is edge } e\text{'s tail} \\ 0 & \text{otherwise.} \end{cases}$$

We denote u_e^T as the row of U associated with edge $e = (u, v)$, then $u_e = (\chi_u - \chi_v)$, where $\chi_u \in \mathbb{R}^n$ is the elementary unit vector with a coordinate 1 at position u. Now, we can write the Laplacian as

$$L_G = U^T U = \sum_{e\in E} u_e u_e^T. \tag{1}$$

It is easy to see that L_G is a positive semi-definite matrix. For a weighted graph $H = (V, \tilde{E}, w)$, we can write its Laplacian as

$$L_H = U^T W U = \sum_{e \in \tilde{E}} w_e u_e u_e^T, \tag{2}$$

where $W \in \mathbb{R}^{|\tilde{E}| \times |\tilde{E}|}$ is a diagonal matrix with edge weights.

2.2 Graph Sparsification

Our goal of sparsification is to approximate G by a sparse graph H with the same vertices. The question is what would be the appropriate metric to measure how H is close to G? Here we discuss two types of graph sparsifiers: cut-sparsifier and spectral sparsifier.

Definition 1 (Cut-sparsifier)**.** *H is a cut-sparsifier for G with parameter ϵ if*

$$\forall S \subseteq V, \qquad (1-\epsilon) C_G(S) \le C_H(S) \le (1+\epsilon) C_G(S),$$

where $C_G(S)$ and $C_H(S)$ are graph cuts to G and H respectively.

Definition 2 (Spectral sparsifier)**.** *H is a spectral sparsifier for G with parameter ϵ if*

$$\forall x \in \mathbb{R}^n, \qquad (1-\epsilon) x^T L_G x \le x^T L_H x \le (1+\epsilon) x^T L_G x, \tag{3}$$

which is equivalent to

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

The following Proposition states the relation between these two sparsifiers.

Proposition 3. *If $H = (V, \tilde{E}, w)$ is a spectral sparsifier for graph $G = (V, E)$ with parameter ϵ, then H is also a cut-sparsifier for G with parameter ϵ.*

Proof. Let $\mathbf{1}_S \in \{0,1\}^n$ be the indication vector for vertices in S, i.e. $\mathbf{1}_S(v) = 1$ if $v \in S$. By the definition of graph cut, we have

$$\begin{aligned} \forall S \subseteq V, \qquad C_G(S) &= \sum_{e=(u,v) \in E} \mathbb{1}_{\{u \in S, v \notin S\}} \\ &= \sum_{e \in E} |u_e^T \mathbf{1}_S| \\ &= \sum_{e \in E} \mathbf{1}_S^T u_e u_e^T \mathbf{1}_S \\ &= \mathbf{1}_S^T L_G \mathbf{1}_S, \end{aligned}$$

and

$$\begin{aligned} \forall S \subseteq V, \qquad C_H(S) &= \sum_{e=(u,v) \in \tilde{E}} w_e \mathbb{1}_{\{u \in S, v \notin S\}} \\ &= \mathbf{1}_S^T \Big(\sum_{e \in \tilde{E}} w_e u_e u_e^T \Big) \mathbf{1}_S \\ &= \mathbf{1}_S^T L_H \mathbf{1}_S. \end{aligned}$$

Since H is a spectral sparsifier with parameter ϵ, applying (3) with $x = \mathbf{1}_S$, we can get

$$\forall S \subseteq V, \qquad (1-\epsilon)C_G(S) \le C_H(S) \le (1+\epsilon)C_G(S).$$

Thus H is also a cut-sparsifier for graph G with parameter ϵ. $\square$

Since spectral sparsifier implies cut-sparsifier, in this lecture we will focus on the spectral sparsification of graphs.

3 Intuition from Electrical Flows

We identify graph $G = (V, E)$ with an electrical network on n nodes and each edge e corresponds to a resistor of unit resistance (i.e., conductance is 1). We will use the following notations to describe the electrical flows:

- $v \in \mathbb{R}^n$: potentials induced at the vertices.
- $x \in \mathbb{R}^n$: currents injected at the vertices.
- $y \in \mathbb{R}^m$: currents induced in the edges.

By Kirchhoff's current law, due to the charge conservation, for any vertex u, the injected current is equal to the sum of currents in edges around v, i.e. $x(u) = \sum_{e \in E} u_e(u)y(e)$. Thus

$$x = U^T y. \tag{4}$$

By Ohm's law, the current flow in an edge is equal to the potential difference across its ends times its conductance, i.e.

$$y = Uv \tag{5}$$

Combining (4) and (5), we can get $x = U^T y = U^T U v = L_G v$. Denote the Moore-Penrose Pseudoinverse of L_G as L_G^+, we can write the potential as

$$v = L_G^+ x. \tag{6}$$

Now we can define the effective resistance of an edge.

Definition 4 (Effective Resistance)**.** *The effect resistance of and edge is the potential difference induced between them when a unit current is injected at one and extracted at the other.*

To get the expression of the effective resistance r_e of an edge $e = (i, j)$, let the injected current be $x = \chi_i - \chi_j = u_e$, then by (6) we have $v = L_G^+ u_e$. The potential difference between vertex i and j is $v(i) - v(j) = (\chi_i - \chi_j)^T v = u_e^T v = u_e^T L_G^+ u_e$. According to Definition 4, the effective resistance of edge e is

$$r_e = u_e^T L_G^+ u_e. \tag{7}$$

4 Spectral Sparsification of graphs

4.1 Algorithm

Our main idea is to sample edges from G and include them in the sparse graph H. Assume that we sample M edges from G with replacement with probability p_e, and for each round we include the sampled edge into H with weights $\frac{1}{Mp_e}$. Define $z_i = \sqrt{\frac{1}{p_e}}u_e$ with probability p_e for $i = 1, \cdots, M$. We can observe that

$$\mathbb{E}[z_i z_i^T] = \sum_{e \in E} p_e \frac{1}{p_e} u_e u_e^T = U^T U = L_G. \tag{8}$$

By (2) the Laplacain of graph H is

$$\begin{aligned} L_H &= \sum_{e \in E} \frac{(\#\text{ of times } e \text{ sampled })}{Mp_e} u_e u_e^T = \frac{1}{M} \sum_{e \in E} (\#\text{ of times } e \text{ sampled }) \cdot \frac{u_e}{\sqrt{p_e}} \frac{u_e}{\sqrt{p_e}}^T \\ &= \frac{1}{M} \sum_{i=1}^{M} z_i z_i^T, \end{aligned} \tag{9}$$

and the expectation of L_H is

$$\mathbb{E}[L_H] = \mathbb{E}\left[\frac{1}{M} \sum_{i=1}^{M} z_i z_i^T\right] = L_G, \tag{10}$$

i.e. the expectation of L_H is exactly L_G.

Our question is: how many samples M and what probability distribution p_e can make L_H close to L_G in the desired way?

By intuition, the probability p_e should reflects the importance of an edge. For instance, consider an unweighted complete graph, we might expect p_e to be the same for each edge since each edge has the same importance. While for a barbell graph, we would like the edge that connects the two components is more likely to be sampled than other edges. We will set the probability proportional to the edge effective resistance defined in the last section, i.e. $p_e \propto r_e$ (we will show later that $p_e = r_e/(n-1)$ for a connected unweighted graph). The intuition is that the effective resistance of an edge is known to be equal to the probability that the edge appears in a random spanning tree of G, and is proportional to the commute time between the endpoints of the edge.

We summarize the graph sparsification algorithm as the following. In the rest of the lecture, we analyze this algorithm for complete graphs and general connected graphs.

Algorithm 1 $H = \texttt{Sparsify}(G, M)$

Take M samples independently with replacement by the following way:
- choose an edge e of G with probability p_e proportional to edge effective resistance r_e,
- add e to H with weight $\frac{1}{Mp_e}$, summing weights if e is chosen more than once.

4.2 Warmup: Complete Graphs

Consider an unweighted complete graph G, then its Laplacian is:

$$L_G = \begin{bmatrix} n-1 & -1 & \dots & -1 \\ -1 & n-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & \dots & -1 & n-1 \end{bmatrix} = nI_n - \mathbf{1}\mathbf{1}^T. \tag{11}$$

Note that

$$\|L_G\| = \sup_{\|x\|_2=1} x^T L_G x = \sup_{\|x\|_2=1} nx^T x - x^T \mathbf{1}\mathbf{1}^T x = \sup_{\|x\|_2=1} n - \big(\sum_i x_i\big)^2 \leq n. \tag{12}$$

We first show that for a complete graph, controlling the closeness between the quadratic forms $x^T L_H x$ and $x^T L_G x$ is equivalent to bounding the spectral norm of $L_H - L_G$.

Lemma 5. *For an unweighted complete graph G, H is a spectral sparsifier with parameter ϵ is equivalent to $\|L_H - L_G\| \leq n\epsilon$.*

Proof. By the definitions of spectral sparsifier and spectral norm, we need to show that $\forall x \in \mathbb{R}^n$ such that $x \neq 0$,

$$(1-\epsilon)x^T L_G x \leq x^T L_H x \leq (1+\epsilon)x^T L_G x \quad \Longleftrightarrow \quad \frac{|x^T(L_H - L_G)x|}{x^T x} \leq n\epsilon. \tag{13}$$

- If $x \in$span$(\mathbf{1})$, i.e. $x = c\mathbf{1}$ for some c, then $Ux = 0$. Thus $L_G x = U^T U x = 0$ and $L_H x = U^T W U x = 0$. Then (13) holds trivially.
- If $x \perp$span$(\mathbf{1})$, i.e. $\mathbf{1}^T x = 0$, then $L_G x = nx - \mathbf{1}\mathbf{1}^T x = nx$, and $x^T L_G x = nx^T x$. The left hand side of (13) becomes to

  $$-n\epsilon x^T x \leq x^T(L_G - L_H)x \leq n\epsilon x^T x,$$

  which is equivalent to the right hand side of (13).

□

Our goal now is to show that $\mathbb{E}[\|L_H - L_G\|] \leq n\epsilon$. First recall the Rudelson-Vershynin (RV) Lemma [RV05] that we learned in the last lecture.

Lemma 6. *[Rudelson-Vershynin (RV) Lemma] Let $x_1, \cdots, x_m \in \mathbb{R}^n$ be independent random vectors such that*

$$\|x_i\|_2 \leq K (K \geq 1), \qquad \|\mathbb{E}[x_i x_i^T]\| \leq 1.$$

Then

$$\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^m x_i x_i^T - \frac{1}{m}\sum_{i=1}^m \mathbb{E}[x_i x_i^T]\right\|\right] \lesssim K\sqrt{\frac{\log n}{m}},$$

for $K\sqrt{\frac{\log n}{m}} < 1$.

Theorem 7. *Let G is an undirected, unweighted and complete graph and $H =$*`Sparsify`*(G, M) be the sparse graph generated by Algorithm 1. Suppose G and H have Laplacians L_G and L_H respectively. If $M \geq \frac{O(1)}{\epsilon^2} n \log n$,*

$$\mathbb{E}[\|L_H - L_G\|] \leq n\epsilon. \tag{14}$$

Proof. First we determine what is p_e. Note that G is unweighted and complete graph and every edge should have the same effective resistance, thus p_e should be the same for each edge. Since there are $\binom{n}{2}$ edges in the graph, the sampling probability is

$$p_e = \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}. \tag{15}$$

Recall that we previously defined $z_i = \sqrt{\frac{1}{p_e}} u_e$. let $y_i = \frac{z_i}{\sqrt{n}}$, then

$$y_i = \frac{z_i}{\sqrt{n}} = \sqrt{\frac{n-1}{2}} u_e, \tag{16}$$

and $\|y_i\|_2 = \sqrt{n-1}$.

We also have:

$$\text{(by (9))} \qquad \frac{L_H}{n} = \frac{1}{nM} \sum_{i=1}^{M} z_i z_i^T = \frac{1}{M} \sum_{i=1}^{M} y_i y_i^T, \tag{17}$$

$$\text{(by (8))} \qquad \mathbb{E}[y_i y_i^T] = \frac{1}{n} \mathbb{E}[z_i z_i^T] = \frac{L_G}{n}, \tag{18}$$

$$\text{(by (12))} \qquad \| \mathbb{E}[y_i y_i^T] \| = \frac{\|L_G\|}{n} \leq 1. \tag{19}$$

Now we can apply the RV Lemma to upper bound $\mathbb{E}[\|L_H - L_G\|]$:

$$\begin{aligned}
\mathbb{E}[\|L_H - L_G\|] &= n\,\mathbb{E}\left[\left\|\frac{L_H}{n} - \frac{L_G}{n}\right\|\right] \\
&= n\,\mathbb{E}\left[\left\|\frac{1}{M}\sum_{i=1}^{M} y_i y_i^T - \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[y_i y_i^T]\right\|\right] && \text{(by (17) and (18))} \\
&\lesssim n\sqrt{\frac{n \log n}{M}},
\end{aligned}$$

where the last inequality follows by RV Lemma with $K = \sqrt{n-1}$. Thus if $M \geq \frac{O(1)}{\epsilon^2} n \log n$, we have $\|L_H - L_G\| \leq \epsilon n$. □

4.3 General Graphs

4.3.1 Result

We now consider spectral sparsification for a general *undirected, unweighted and connected* graph G. Let H be the sparse graph generated by Algorithm 1. The main result is that if M is sufficiently large, then H is a good sparsifier, i.e. the quadratic forms $x^T L_G x$ and $x^T L_H x$ are close.

Theorem 8. *Let G be an undirected, unweighted and connected graph, and $H =$* `Sparsify`*(G, M) be the sparse graph generated by Algorithm 1. Let L_G and L_H be the Laplacian of G and H respectively. If $1/\sqrt{n} < \epsilon \le 1$ and $M = Cn \log n/\epsilon^2$, then there exits some constant C s.t. with probability at least $1/2$,*

$$\forall x \in \mathbb{R}^n \quad (1-\epsilon)x^T L_G x \le x^T L_H x \le (1+\epsilon)x^T L_G x. \tag{20}$$

For spectral sparsification of *weighted* graphs, please refer to [SS08]. Basically, we can use the similar idea of sampling in Algorithm 1. But the probability used for sampling edge should be proportional to $w_e r_e$, and the weight added to H becomes to $w_e/(Mp_e)$. Besides, the analysis is very similar to what we show in this section.

We will prove Theorem 8 in the rest of this section.

4.3.2 Analysis

We first look at some properties of the Laplacian $L_G = U^T U$ and the matrix $R = U L_G^+ U^T$ in Lemma 9 and Lemma 10 respectively.

Lemma 9. *The Laplacian matrix L_G has the following properties:*

(i) $\ker(L_G) = \ker(U) = \mathrm{span}(\mathbf{1})$, where $\ker(L_G)$ is the kernel of matrix L_G.

(ii) $L_G^+ L_G = L_G L_G^+$ is the projection onto the span of the non-zero eigenvectors of L_G.

Proof.

(i) Since $L_G = U^T U$,

$$\begin{aligned} x \in \ker(L_G) \iff \forall y \in \mathbb{R}^n, \quad y^T U^T U x = 0 &\iff Ux = 0 \iff \|Ux\|_2^2 = 0 \\ &\iff \sum_{(u,v)\in E} (x(u) - x(v))^2 = 0 \\ &\iff x(u) = x(v), \forall (u,v) \in E \\ &\iff x \in \mathrm{span}(\mathbf{1}), \text{ since } G \text{ is connected graph.} \end{aligned}$$

(ii) From (i), the rank of L_G is $n-1$, since L_G is symmetric, we can diagonalize it as

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

where $\lambda_1, \cdots, \lambda_{n-1}$ are the $(n-1)$ nonzero eigenvalues of L_G and $v_1, \cdots, v_{n-1}$ are the corresponding eigenvectors. Then the Moore-Penrose Pseudoinverse of L_G is

$$L_G^+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} v_i v_i^T.$$

Note that

$$L_G^+ L_G = L_G L_G^+ = \sum_{i=1}^{n-1} v_i v_i^T,$$

then $L_G^+ L_G = L_G L_G^+$ is the projection onto the span of the non-zero eigenvectors of L_G. Thus, $L_G L_G^+ = L_G^+ L_G$ is an identity on the range space of L_G, which is denoted by $\text{im}(L_G)$.

□

Lemma 10. *Matrix R has the following properties:*

(i) R is a projection matrix.

(ii) $im(R) = im(U)$.

(iii) The eigenvalues of R are 1 with multiplicity of $n-1$ and 0 with multiplicity of $m-n+1$.

(iv) $r_e = R(e,e) = \|R(\cdot,e)\|^2$, where $R(\cdot,e)$ is the column corresponding to edge e in R.

(v) (Foster's Resistance Theorem) $\sum_{e\in E} r_e = n-1$.

Proof.

(i) Since $R = UL_G^+U^T$,

$$\begin{aligned} RR = UL_G^+U^TUL_G^+U^T &= UL_G^+L_GL_G^+U^T \\ &= UL_G^+U^T \qquad (L_G^+L_G \text{ is identity on } \text{im}(L_G) \text{ by Lemma 9.(ii)}) \\ &= R. \end{aligned}$$

(ii) $\text{im}(R) = \text{im}(UL_G^+U^T) \subseteq \text{im}(U)$. To see the other inclusion, pick arbitrary $y \in\text{im}(U)$, we can choose $x \perp\text{ker}(U)\text{=span}(\mathbf{1})$, s.t. $y = Ux$. Then

$$\begin{aligned} Ry = UL_G^+U^TUx &= UL_G^+L_Gx \\ &= Ux \qquad (L_G^+L_Gx = x \text{ for } x \perp\text{ker}(U)) \\ &= y. \end{aligned}$$

(iii) Let λ and v be an eigenvalue and corresponding eigenvector of R, then $\lambda^2 v = \lambda Rv = RRv = Rv = \lambda v$ since $RR = R$ by (i). Thus $\lambda^2 = \lambda$, i.e. the eigenvalues of R are all 0 or 1.

From (ii) and Lemma 9.(i), we can conclude that $\dim(\text{im}(R)) = \dim(\text{im}(U)) = \dim(\text{im}(L_G))$ $= n-1$. Therefore R has $n-1$ eigenvalues that are 1, and $m-n-1$ eigenvalues that are 0.

(iv) Recall the definition of the effective resistance, we have $r_e = u_e^T L_G^+ u_e = R(e,e)$. The second equality follows from $R(e,e) = R^2(e,e) = R(\cdot,e)^T R(\cdot,e)$, since R is a projection matrix and is symmetric.

(v) From (iii) and (iv),

$$\sum_{e\in E} r_e = \sum_{e\in E} R(e,e) = Tr(R) = n-1.$$

□

We describe the outcome of H by a random diagonal matrix $S \in \mathbb{R}^{m\times m}$ with diagonals defined by

$$S(e,e) = \frac{\#\text{ of times edge } e \text{ is sampled}}{Mp_e}. \tag{21}$$

Now we can use S to rewrite the Laplacian of H, i.e L_H as

$$L_H = \sum_{e\,\mathbb{E}} \frac{\#\text{ of times edge } e \text{ is sampled}}{Mp_e} u_e u_e^T = \sum_{e\in E} S(e,e) u_e u_e^T = U^T S U. \tag{22}$$

Our goal is to show that H is a good sparsifier of G, i.e. we need to show that $x^T L_H x$ and $x^T L_G x$ are close. We start by reducing the proble of preserving $x^T L_G x$ to that of preserving $y^T R y$. In particular, we can prove the following lemma, which says that if S does not distort $y^T R y$ too much then $x^T L_H x$ and $x^T L_G x$ are close.

Lemma 11. *If S is a nonnegative diagonal matrix such that*

$$\|RSR - RR\| \le \epsilon, \tag{23}$$

then $\forall x \in \mathbb{R}^n$,

$$(1-\epsilon)x^T L_G x \le x^T L_H x \le (1+\epsilon) x^T L_G x. \tag{24}$$

Proof. By the definition of the spectral norm, the assumption of (23) is equivalent to

$$\sup_{y\in\mathbb{R}^m, y\neq 0} \frac{|y^T R(S-I)Ry|}{y^T y} \le \epsilon.$$

We restrict our attention to vectors in im(U). Let y ∈im(U), by Lemma 10.(ii), $\exists z \in \mathbb{R}^m$, s.t. $y = Rz$, and thus $Ry = RRz = Rz = y$. Then we have

$$\begin{aligned}
\sup_{y\in \mathrm{im}(U), y\neq 0} \frac{|y^T R(S-I)Ry}{y^T y} &= \sup_{y\in \mathrm{im}(U), y\neq 0} \frac{|y^T (S-I)y|}{y^T y} \\
&= \sup_{x\in\mathbb{R}^n, Ux\neq 0} \frac{|x^T U^T (S-I) U x|}{x^T U^T U x} \qquad (\text{rewrite } y = Ux \text{ for some } x \in \mathbb{R}^n) \\
&= \sup_{x\in\mathbb{R}^n, Ux\neq 0} \frac{|x^T (L_H - L_G) x|}{x^T L_G x} \le \epsilon. \qquad (L_G = U^T U,\ L_H = U^T S U)
\end{aligned}$$

Rearranging yields the results in (24) for all $x \notin$ker(U). When $Ux = 0$, $x^T L_H x = x^T U^T S U x = 0$, (24) holds trivially.

□

We can now finish the proof of Theorem 8 by showing that $\|RSR - RR\|$ is likely to be small, which is the assumption of the previous Lemma.

Proof of Theorem 8. First we use the RV Lemma to bound $\mathbb{E}[\|RSR - RR\|]$. Sampling M edges from G is corresponding to sampling M columns from R, so we can write

$$\begin{aligned} RSR = \sum_{e \in E} S(e,e) R(\cdot,e) R(\cdot,e)^T &= \sum_{e \in E} \frac{\#\text{ of times edge } e \text{ is sampled}}{M p_e} R(\cdot,e) R(\cdot,e)^T \\ &= \frac{1}{M} \sum_{e \in E} (\#\text{ of times edge } e \text{ is sampled}) \cdot \frac{R(\cdot,e)}{\sqrt{p_e}} \frac{R(\cdot,e)^T}{\sqrt{p_e}} \\ &= \frac{1}{M} \sum_{i=1}^{M} y_i y_i^T, \end{aligned}$$

where vectors $y_1, \cdots, y_M$ are sampled independently with replacement from the distribution

$$y = \frac{1}{\sqrt{p_e}} R(\cdot,e) \quad \text{with probability } p_e. \tag{25}$$

The expectation of $y_i y_i^T$ for each i is

$$\mathbb{E}[y_i y_i^T] = \sum_{e \in E} p_e \frac{1}{p_e} R(\cdot,e) R(\cdot,e) = R^T R = RR = R. \tag{26}$$

The spectral norm of R is 1 by Lemma 10.(iii), so $\|\,\mathbb{E}[y_i y_i^T]\| = \|R\| = 1$.

By Fosters' Resistance Theorem (i.e. Lemma 10.(v)), sum of effective resistance is $\sum_{e \in E} r_e = n-1$ and thus the sampling probability is $p_e = \frac{r_e}{n-1}$. Then we have

$$\|y_i\|_2 = \|\frac{1}{\sqrt{p_e}} R(\cdot,e)\| = \sqrt{\frac{n-1}{r_e}} \|R(\cdot,e)\| = \sqrt{n-1}, \tag{27}$$

since $\|R(\cdot,e)\| = \sqrt{r_e}$ by Lemma 10(iv).

Now we can apply the RV Lemma: by taking $M = Cn \log n / \epsilon^2$ for some constant C, we have

$$\mathbb{E}[\|RSR - RR\|] = \mathbb{E}\left[\left\|\frac{1}{M}\sum_{i=1}^{M} y_i y_i^T - \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}[y_i y_i^T]\right\|\right] \le \sqrt{n-1}\sqrt{\frac{\log n}{M}} \le \frac{\epsilon}{2}. \tag{28}$$

By Markov's inequality,

$$\mathbb{P}\left[\|RSR - RR\| \ge \epsilon\right] \le \frac{1}{2}. \tag{29}$$

Then the theorem follows by combining this with the results from Lemma 11.

□

References

[RV05] M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM*, 54(4):21, 2007.

[SS08] D.A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6), pp.1913-1926. 2011.