CS 388R: Randomized Algorithms Fall 2017

Lecture 19 — November 9, 2017

Prof. Eric Price *Scribe: Niels Kornerup, Naren Manoj, Ryan Rice*

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

1 Overview

Let's say we want to come up with a smaller representation for a given graph G which preserves most of the properties of G. Previously, we discussed a *cut sparsifier*, which provides an approximation for all cuts in the graph. Today, we'll generalize this to determine a spectral sparsifier for our graph with $O(\epsilon^{-2} n \log n)$ edges and the same amount of time.

1.1 Notation

Unless otherwise stated, the following notation holds consistent throughout these notes.

If we have a scalar s, then $\|s\|$ represents the absolute value of s.

If we have a vector v, then $\|v\|$ represents the l^2 norm of v.

If we have a matrix M, then $\|M\|$ represents the spectral norm of M. The spectral norm is defined below, and should not be confused with the Frobenius norm.

Given a vector x and a positive definite matrix M, we define $\|x\|_M$ to be $\sqrt{x^T M x}$.

Definition 1 (Spectral Norm)**.** *The spectral norm of a matrix M is the value of its largest singular value. This can be written as:*

$$\|M\| = \sup_{\|v\| \neq 0} \frac{v^T M v}{v^T v}$$

or, equivalently:

$$\|M\| = \sup_{\|v\| = 1} v^T M v$$

Generally, we use capital letter variables to denote matrices and lowercase letter variables to denote vectors and scalars.

If we are discussing a graph and we use the variables n and m without prior definition, then take n to be the number of vertices in the graph and m to be the number of edges in the graph.

2 Intuition and Problem Definition

First, recall the definition of the Laplacian.

Definition 2 (Graph Laplacian)**.** *Suppose we have a graph $G = (V, E)$. Further, let $n = |V|$ and $m = |E|$.*

Define the matrix $D \in \mathbb{R}^{n \times n}$ below:

$$D_{i,j} = \begin{cases} \deg(i) & i = j \\ 0 & \textit{otherwise} \end{cases}$$

and the adjacency matrix $A \in \mathbb{R}^{n \times n}$:

$$A_{i,j} = \begin{cases} 1 & (i,j) \in E \\ 0 & \textit{otherwise} \end{cases}$$

The Laplacian Matrix is then given by:

$$L_G = D - A$$

Now, suppose that we view our graph as an electrical circuit. Let's say that the edge weights are resistors with resistance $1/w_e$. If set electric potentials at each node, how much power is spent?

To answer this question, recall the below fundamental equations (V represents voltages, I represents current, P represents power, and R represents resistance):

$$V = IR$$
$$P = IV$$

And note that we can eliminate the current term by substituting the second equation into the first:

$$P = \frac{V^2}{R}$$

For convenience, replace resistance with conductance, denoted as C. Note that $C = 1/R$, which allows us to write:

$$P = V^2 C$$

Now, define the vector x such that x_i represents the electric potential at a given node in the graph. Let w_e be the weight of edge e. Substituting these gives us:

$$\begin{aligned} P(x) &= \sum_{e=(u,v) \in E} w_e \cdot (x_u - x_v)^2 \\ &= x^T L_G x \end{aligned}$$

We'd like to preserve this function while sparsifying the graph Laplacian. We now have a more concrete way to express the objective we'd like to accomplish. If L_H is our sparsified Laplacian, then we want:

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

Note that this is equivalent to stating that the matrices

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

and

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

are positive semidefinite (recall that a matrix is positive semidefinite if it has no negative eigenvalues; equivalently, for any real vector x and a matrix $M \in \mathbb{S}_+^n$, we must have $x^T M x \geq 0$). We can thus express the same objective in a more simpler way:

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

which is equivalent to

$$|x^T(L_H - L_G)x| \leq \epsilon x^T L_G x \tag{1}$$

and

$$P_H(x) \in (1 \pm \epsilon)P_H(x) \tag{2}$$

Before we begin, we point out a connection between a topic we covered earlier and the topic we are covering now.

Lemma 3. *Every spectral sparsifier is also a cut sparsifier.*

Proof. Let H be a spectral sparsifier of G. Let S be any cut. If we define 1_S to be the indicator vector for S ($1_{S_i} = 1 \Leftrightarrow i \in S$), then:

$$1_S^T L_G 1_S = P_G(1_S) = \sum_{e \in E} w_e (1_{S_u} - 1_{S_v})^2$$

where $(1_{S_u} - 1_{S_v})^2$ is 1 if (u, v) is an edge being cut from S to $\overline{S}$. Hence, $P_H(1_S)$ is the approximated cut size of S, and $P_G(1_S)$ is the true cut size. Since H is a spectral sparsifier by 2,

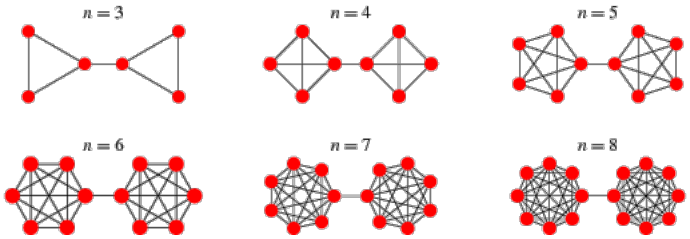
$$cut_H(S, \overline{S}) = P_H(1_S) \in (1 \pm \epsilon)P_G(1_S) = (1 \pm \epsilon)cut_G(S, \overline{S})$$

implying that H is a cut sparsifier as well. □

3 Spectral Sparsification Algorithm

When we are sparsifying our graph, we should ask ourselves the question, "what properties of the graph would we like to preserve?" In the rest of this lecture, we'll focus on sampling edges from the graph with a certain probability. The hope is that our newly sparsified graph representation doesn't end up storing *all* the edge connectivity information while still retaining other key information about the graph.

Intuitively, there are certain edges we care more about than others. For instance, consider a barbell graph, examples of which are shown below.



Observe that if our sparsifier somehow drops the edge connecting the two large components on either side, then we lose *a lot* more connectivity information than if we dropped some edge within the two large components. Hence, we should try to weight some important edges more highly in our distribution.

Our problem is now twofold. First, we have to construct a reasonable distribution over edges from which we can sample to build our representation. Second, we have to demonstrate that such a scheme actually attains the objective we stated earlier.

Solving the second problem is more straightforward than the first, so we do this below.

Define u_e for edge $e = (a, b)$ to be $(e_a - e_b)$ where e_x is the elementary vector (all zeroes except for a 1 at the xth position). With some quick algebra, we can show that the Laplacian is equal to the below expression:

$$\sum_{e \in E} w_e u_e u_e^T$$

In our algorithm, we will sample m edges with replacement from the distribution that we compute. In particular, let the probability that we pick edge e be p_e. We then add this edge to our sparsified representation with the modified weight:

$$\frac{w_e}{mp_e}$$

To analyze this algorithm, let $z_e = \sqrt{\frac{w_e}{mp_e}} u_e$ with probability p_e and the zero vector otherwise.

Therefore the expectation of the Laplacian of this new graph after one round of sampling is:

$$\begin{aligned}
\mathbb{E}[L_H] &= \mathbb{E}\left[\sum_{e\in E} z_e z_e^T\right] \\
&= \sum_{e\in E} \mathbb{E}\left[z_e z_e^T\right] \\
&= \sum_{e=(a,b)\in E} p_e \left(\sqrt{\frac{w_e}{mp_e}}(e_a - e_b)\right)\left(\sqrt{\frac{w_e}{mp_e}}(e_a - e_b)\right)^T \\
&= \frac{1}{m}\sum_{e=(a,b)\in E} \left(\sqrt{w_e}(e_a - e_b)\right)\left(\sqrt{w_e}(e_a - e_b)\right)^T \\
&= \frac{1}{m}\sum_{e\in E} u_e u_e^T \\
&= \frac{L_G}{m}
\end{aligned}$$

Over all m rounds of sampling, we have:

$$L_H = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\sum_{e\in E} z_e z_e^T\right] = \frac{1}{m}\sum_{i=1}^{m} L_G = L_G$$

Note that this result tells us that we can approximate the Laplacian of the original graph using such a scheme. However, we also want to approximate the spectral norm between the differences of the Laplacians with high probability. Demonstrating an instantiation for the distribution over edges and demonstrating that we can attain the desired result is the subject of the remainder of this lecture.

3.1 Complete Graphs

To gain some intuition for solving the generalized problem, let's solve the problem for a special case - complete graphs.

Consider a complete graph G with weight 1 on all edges. Then:

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

Notice that

$$\begin{aligned}
x^T L_G x &= x^T(nI_n - \mathbf{1}\mathbf{1}^T)x \\
&= x^T(nx - \mathbf{1}\mathbf{1}^T x) \\
&= n\|x\|^2 - x^T\mathbf{1}\mathbf{1}^T x \\
&\leq n\|x\|^2 \text{ because } \mathbf{1}\mathbf{1}^T \text{ is positive semidefinite}
\end{aligned}$$

As a result, we can claim:

$$\frac{x^T(L_H - L_G)x}{\|x\|^2} \leq \frac{x^T(L_H - L_G)x}{x^T L_G x} \leq \epsilon n$$

Since the spectral norm is defined as $\sup_{\|x\|\neq 0} \frac{x^T(L_H - L_G)x}{x^T x} \leq \epsilon n$, we have that by 1 the condition:

$$\|L_H - L_G\| \leq \epsilon n.$$

is sufficient for L_H to be a spectral sparsifier.

For complete graphs, observe that no edge is more important than another. Therefore, we should create a uniform distribution over edges ($p_i = p_j$ for all $i, j \in E$). Specifically, this yields:

$$p_e = \frac{1}{\binom{n}{2}} = \Theta\left(\frac{1}{n^2}\right)$$

Now, suppose we perform $q = O(\frac{n \log n}{\epsilon^2})$ rounds of sampling for some suitable parameter ϵ. Notice that

$$\|z_e\| \leq \sqrt{\frac{w_e}{p_e}}\sqrt{u_e u_e^T} = \Theta\left(\sqrt{\frac{1}{p_e}}\right) = \Theta(n)$$

To finish our proof, we need one tool introduced in the previous lecture - the Rudelson Vershynin (RV) Lemma. The proof of this lemma was covered in the previous lecture.

Lemma 4 (Rudelson Vershynin (RV) Lemma)**.** *If $x_1, \ldots, x_m$ are independent and identically distributed (i.i.d) random vectors in $\mathbb{R}^n$, such that they are uniformly bounded, so $\|x\| \leq k$, and $\left\|\mathbb{E}[xx^T]\right\| \leq 1$, then*

$$\left\|\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} x_i x_i^T - \mathbb{E}[xx^T]\right]\right\| \leq k\sqrt{\frac{\log n}{m}} \tag{3}$$

where x is any x_i (since they are identically distributed).

Armed with this fact, we can now complete our proof. First, note that z_e is a bounded random variable. Now, set:

$$y_e = \frac{z_e}{\sqrt{n}}$$

It thus follows that:

$$\left\|\mathbb{E}[y_e y_e^T]\right\|_2 = \left\|\frac{z_e z_e^T}{\sqrt{n}^2}\right\|_2 \leq 1$$

and:

$$\|y_e\| \leq O(\sqrt{n})$$

which gives us a setting to use the RV Lemma 3. This analysis now yields:

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^{m} y_i y_i^T - \mathbb{E}[yy^T]\right\|\right] &= \mathbb{E}\left[\left\|\frac{1}{n}\left(\frac{1}{m}\sum_{i=1}^{m} z_i z_i^T - L_G\right)\right\|\right] \\
&= \frac{1}{n}\,\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^{m} z_i z_i^T - L_G\right\|\right] \\
&\leq \sqrt{n}\sqrt{\frac{\log n}{(n\log n)/\epsilon^2}} \\
&= \epsilon
\end{aligned}$$

As a result, we obtain:

$$\mathbb{E}\left[\left\|\frac{1}{m}\sum_{i=1}^{m} z_i z_i^T - L_G\right\|\right] = \mathbb{E}[\|L_H - L_G\|] \leq \epsilon n$$

as desired. Therefore, we have that L_H is a spectral sparsifier to L_G.

3.2 Non-complete Graphs

The tricky part is to now generalize our previous analysis to general graphs. A similar procedure will work, but in order for our previous analysis to apply, we have to perform a few slight modifications. First, we introduce some terminology and tools.

Definition 5 (Spherical Laplacian)**.** *A Laplacian for a graph G is spherical if all of the Laplacian's eigenvalues are within a constant factor of one another.*

Intuitively, bounding the differences between spherical Laplacians will be pretty easy. In particular, if we can send all the eigenvalues to some constant range, then we can easily bound the difference between the spectral norms. Hence, our objective is to compute some reconditioning of the Laplacian which lends itself nicely to such a property.

Define $U \in \mathbb{R}^{m\times n}$ to be the matrix below:

$$U = \begin{bmatrix} — & u_1^T & — \\ — & u_2^T & — \\ & \vdots & \\ — & u_m^T & — \end{bmatrix}$$

Consider the following claims.

Claim 6. *For an unweighted graph G,*

$$L_G = U^T U$$

Proof.

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

□

Claim 7. *If L_G is the Laplacian Matrix for a graph G, then $L_G \in \mathbb{S}_+^n$.*

Proof. $x^T L_G x = x^T U^T U x = \|Ux\| \geq 0$ for any real vector x. □

And note the following facts taken from physics:

Fact 8 (Kirchoff's Current Law)**.** *The current entering a junction must equal the current exiting the same junction.*

Fact 9 (Ohm's Law)**.** *Recall that $V = IR$. Replacing R with $1/C$ yields:*

$$I = VC$$

Now, consider injecting a current vector $x \in \mathbb{R}^n$ where x_i represents the amount of current injected at node i. Suppose that as a result, we obtain some induced potentials $v \in \mathbb{R}^n$ at the vertices and some induced currents $y \in \mathbb{R}^m$ across the edges. Note that the potential *differences* across the edges are now given by the product Uv. Now, using these facts, we can write:

$$y = Uv$$

and:

$$U^T y = x$$

Substituting the first equation into the second equation yields:

$$\begin{aligned} x &= U^T y \\ &= U^T U v \\ &= L_G v \end{aligned}$$

where the last line follows from 6. Since every matrix with real entries must have a pseudoinverse, we can write:

$$v = L_G^+ x$$

Additionally, recall the definition of effective resistance.

Definition 10 (Effective Resistance)**.** *The effective resistance between two junctions is the potential difference induced when a unit current is injected at the first junction and extracted from the second.*

To inject a unit current across an edge e, we can set $x = u_e$. Now, multiplying both sides of our previous equation by u_e^T will give us the effective resistance across an edge:

$$r_e = u_e^T v = u_e^T L_G^+ u_e$$

To generalize, note that:

$$R = UL_G^+U^T$$

has the values of the effective resistances down the major diagonal. We are now ready to make our central claim.

Theorem 11 (Suitable probabilities for a spectral sparsifier)**.** *If we set $p_e \propto r_e$, then we can obtain a Laplacian L_H which approximates the original Laplacian L_G with high probability. In our case, we can use $p_e = r_e/(n-1)$.*

Proof. Our proof will roughly proceed as follows. We will first show that our sampling method doesn't distort the matrix of effective resistances by too much. More precisely, we'll show that the sampling method results in an effective resistance matrix which is reasonably close to the original effective resistance matrix, where we measure "reasonably close" in terms of the spectral norm of the difference between the two matrices. We can then use this to show that the sparsifier preserves the norm with respect to its Laplacian. This final result is exactly what we are aiming for.

To prove this theorem, we need a few more tools. First, let's explore some of the helpful properties of R.

Lemma 12. *R is a projection matrix.*

Proof. Set M below: $M = UL_G^+U^T$. Then

$$\begin{aligned} M^2 &= UL_G^+U^TUL_G^+U^T \\ &= UL_G^+L_GL_G^+U^T \\ &= UL_G^+U^T \end{aligned}$$

The second line follows from the fact that $U^TU = L_G$ and the last line follows from the definition of the pseudoinverse. □

Lemma 13. *The eigenvalues of R consist of only 0s and 1s with appropriate multiplicity.*

Proof. Use the above. If λ is an eigenvalue, then we have:

$$\begin{aligned} Rv &= \lambda v \\ R^2v &= R\lambda v \\ &= \lambda Rv \\ &= \lambda^2 v \end{aligned}$$

Thus, λ^2 must be the same eigenvalue as λ, and we have:

$$\lambda^2 = \lambda$$

This has restricted our space of possible eigenvalues to 0 and 1. Now, we can prove that there exists at least one positive eigenvalue via contradiction. Suppose that R has only 0 eigenvalues. It thus follows that R is nilpotent. Then, there exists some n for which:

$$R^n = 0$$

Furthermore, note that as a result of R being a projection matrix, we have:

$$R^2 = R$$

and therefore:

$$R^n = R = 0$$

However, it is clearly untrue that $R = 0$. We may therefore conclude that R has at least one positive eigenvalue. In particular, R has at least one eigenvalue of 1 and the remaining eigenvalues must be 0.

Let's determine the multiplicity of these eigenvalues. Observe that if an eigenvalue has value 0, then its associated eigenvector is in the nullspace of R. Furthermore, note that every vector in the nullspace is an eigenvector associated with eigenvalue 0:

$$Rv = \lambda v = 0v = 0$$

Furthermore, if an eigenvalue has value 1, then its associated eigenvector is in the span of R:

$$Rv = \lambda v = v$$

Now, recall the rank-nullity theorem. Apply it to our case:

$$\text{rk}(R) + \text{nul}(R) = m$$

Observe that R is a projection matrix to a space of size $n-1$. From this, it follows that the size of the nullspace of R is $m-n+1$. Hence, we have that the multiplicity of the 0 eigenvalue is $m-n+1$ and the multiplicity of the 1 eigenvalue is $n-1$.

□

We also need the following fact:

Lemma 14 (Foster's Theorem)**.** *The sum of the effective resistances on a connected graph is exactly* $n-1$.

Proof.

$$\begin{aligned}\sum_{e\in E} R_{eff}(e) &= \sum_{e\in E} u_e^T L_G^+ u_e \\ &= \sum_{e\in E} (UL_G^+U^T)_{e,e} \\ &= \text{Tr}(UL_G^+U^T)\end{aligned}$$

Observe that our problem is now equivalent to showing $\text{Tr}(R) = n - 1$. However, previously, we showed that there are only $n - 1$ eigenvalues with value 1 in R, and the rest are equal to 0. The desired result immediately follows (recall the well-known fact that states that the trace of a matrix is equal to the sum of its eigenvalues).

□

Observe that Foster's Theorem immediately validates our probability distribution. To see this, consider adding the selection probabilities of each edge. Note that the numerator evaluates to $n - 1$.

Now, let $S \in \mathbb{R}^{m \times m}$ be the matrix defined below:

$$S_{ij} = \begin{cases} (\text{number of times edge } i \text{ is sampled if } i = j)\,/(mp_e) \\ 0 \text{ otherwise} \end{cases}$$

Our sampling method yields an effective resistance matrix RSR. We'd like to show that this isn't too far from the original effective resistance matrix R. More precisely:

Lemma 15 (Sparsification Process Preserves Spectral Norm of Effective Resistance Matrices)**.**

$$\|RSR - R\| \leq \epsilon$$

for a suitable ϵ*.*

Proof. Rewrite RSR:

$$\begin{aligned} RSR &= \sum_e R_e S_{ee} R_e^T \text{ where } R_e \text{ is the } e\text{th column.} \\ &= \sum_e S_{ee} R_e R_e^T \end{aligned}$$

Let s_e be the number of times edge e was sampled. We can rewrite the above equations now:

$$\begin{aligned} RSR &= \sum_e S_{ee} R_e R_e^T \\ &= \sum_e \frac{s_e}{mp_e} R_e R_e^T \\ &= \sum_e \frac{s_e}{m} \cdot \frac{R_e}{\sqrt{p_e}} \cdot \frac{R_e^T}{\sqrt{p_e}} \\ &= \frac{1}{m} \sum_{i=1}^{m} y_i y_i^T \end{aligned}$$

where y_i is a vector drawn independently with replacement from the distribution:

$$y = R_e / \sqrt{p_e}$$

which implies:

$$\begin{aligned}\mathbb{E}[yy^T] &= \sum_e \frac{1}{p_e} R_e R_e^T p_e \\ &= \sum_e R_e R_e^T \\ &= RR \\ &= R\end{aligned}$$

where the last line follows from R being a projection matrix. Since $\|R\| = 1$ (a result arising directly from our analysis of the eigenvalues of R), we must have $\mathbb{E}[yy^T] = 1$. Additionally, we can prove an upper bound on $\|y\|$:

$$\begin{aligned}\|y\| &\leq \left\| \frac{1}{\sqrt{p_e}} R_e \right\| \\ &= \frac{1}{\sqrt{p_e}} \|R_e\| \\ &= \frac{1}{\sqrt{p_e}} \sqrt{r_e} \\ &= \sqrt{\frac{r_e}{p_e}} \\ &= \sqrt{n-1}\end{aligned}$$

Note that we've satisfied the conditions to apply the RV Lemma. We now have:

$$\begin{aligned}\mathbb{E}\left[\|RSR - R\|\right] &= \mathbb{E}\left[\left\| \frac{1}{m} \sum_{i=1}^{m} y_i y_i^T - \mathbb{E}[yy^T] \right\|\right] \\ &\leq \sqrt{n-1} \cdot \sqrt{\frac{\log m}{m}} \\ &= \frac{\epsilon}{2}\end{aligned}$$

By Markov's Inequality, we have that

$$\mathbb{P}\left[\|RSR - R\| \geq \epsilon\right] \leq \frac{\mathbb{E}\left[\|RSR - R\|\right]}{\epsilon} \leq \frac{1}{2}$$

which implies that we're pretty close no less than half the time. To characterize this "pretty close", we can set ϵ below:

$$\epsilon = 2\sqrt{\frac{(n-1)\log m}{m}}$$

which concludes the proof of this lemma.

□

Observe that the below will now allow us to complete the proof of our main theorem entirely.

Lemma 16 (Small Spectral Norm Deviation Between Effective Resistance Matrices Implies Sparsifier)**.** *Suppose that:*

$$\|RSR - R\| \leq \epsilon$$

Then:

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

Proof. First, note that $L_H = U^T S U$, which is a minor extension of 6.

Observe:

$$\begin{aligned}
\|RSR - R\| &= \|RSR - RR\| \\
&= \|R(S-I)R\| \\
&= \sup_{\|y\| \neq 0} \frac{\left\|y^T R(S-I)Ry\right\|}{y^T y} \\
&\leq \epsilon
\end{aligned}$$

First, restrict the above to only consider vectors for which $Ry = y$ (clearly, such vectors exist since we demonstrated earlier that R has eigenvalues of 1). Then, further restrict our attention to those vectors y that can be expressed as Ux for some vector x:

$$\begin{aligned}
\|RSR - R\| &= \sup_{\|y\| \neq 0} \frac{\left\|y^T R(S-I)Ry\right\|}{y^T y} \\
&\geq \frac{\left\|y^T (S-I)y\right\|}{y^T y} \\
&\geq \frac{\left\|x^T U^T (S-I)Ux\right\|}{x^T U^T U x} \\
&= \frac{\left\|x^T U^T SUx - x^T U^T Ux\right\|}{x^T U^T U x} \\
&= \frac{\left\|x^T L_H x - x^T L_G x\right\|}{x^T LG_x}
\end{aligned}$$

Since we are given $\|RSR - R\| \leq \epsilon$, we have:

$$\epsilon \geq \frac{\left\|x^T L_H x - x^T L_G x\right\|}{x^T LG_x}$$

which immediately yields the desired inequality.

□

Notice that the proof of the above lemma concludes the proof of our central theorem, so we're finally done!

□

References

[1] Rajeev Motwani, Prabhakar Raghavan. Randomized Algorithms. *Cambridge University Press*, 0-521-47465-5, 1995.

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