

A Survey of Spectral Sparsification

Robi Bhattacharjee - robibhat@mit.edu

Victor Pontis - vpontis@mit.edu

December 9, 2013

Abstract

A spectral sparsification algorithm takes a graph as input and returns a new graph with few edges that preserves up to an approximation a term called the Laplacian of the graph. The Laplacian of the graph relates to the connectivity of the graph. The Laplacian is a strong property so graphs with similar Laplacians also have similar cut values as well as other important properties.

In this paper we will give a background to spectral graph theory to a reader with an undergraduate course in computer science. We will give the necessary background work, including material in linear algebra and effective resistance, needed for the analysis of the algorithms presented. We will then describe two sampling algorithms: one that considers vertex degree and one that considers effective resistance giving intuition behind both algorithms.

1 Background

1.1 Introduction to Spectral Sparsification

A sparse graph is a graph where the number of edges is on the order of the number of vertices (up to a logarithmic factor). A sparsification algorithm takes a general graph and returns a sparse graph that is an approximation of the input graph. A good sparsification algorithm may produce a graph with $\tilde{O}(n)$ edges when a complete graph can have up to $n^2/2$ edges. Benczur and Karger introduced the notion of graph sparsifiers in their 1996 Paper “Approximating $s - t$ Minimum Cuts in $\tilde{O}(n^2)$ time” [?] with the notion of cut sparsifiers. Their paper defines a class of sparsification algorithms called cut sparsifiers which return a sparse graph that approximately preserves the cut between any two sets of vertices. They used their sparsification algorithm to improve runtimes on mincut and maxflow.

In this paper we will be studying spectral sparsification algorithms. Spectral sparsification can be thought of as a generalization of cut sparsifiers that preserves a term called the Laplacian, L_G , of a graph G as defined here

$$L_G(u, v) = \begin{cases} -w(u, v) & \text{if } u \neq v \\ \sum_z w(u, z) & \text{if } u = v. \end{cases} \quad (1)$$

we can also write the Laplacian in its quadratic form

$$x^T L_G x = \sum w_{(u,v)} (x_u - x_v)^2 \quad (2)$$

We can see from the quadratic form of the Laplacian that it is a generalization of cut sparsification. Consider any set $S \subset V$ and let x_S be the vector defined as $x(u) = 1$ if $u \in S$ and $x(u) = 0$ otherwise. Then it is clear that $x_S^T L_G x$ is just equal to the sum of the weights of the edges that span S to $V - S$, which we will denote in the future as $E(S, V - S)$. This shows that spectral sparsifiers are at least as strong as cut sparsifiers (they are in fact strictly stronger, but this requires more analysis).

Next, let \tilde{G} be a graph with the same vertices as G . \tilde{G} is said to be a σ -spectral-sparsifier if for all x

$$\frac{1}{\sigma} x^T L_{\tilde{G}} x \leq x^T L_G x \leq \sigma x^T L_{\tilde{G}} x \quad (3)$$

Let \tilde{E} be the set of edges (with weights) of \tilde{G} . Then \tilde{G} is said to be a σ -cut-sparsifier if for every $S \subset V$

$$\frac{1}{\sigma} \sum_{e \in \tilde{E}(S, V-S)} w_e \leq \sum_{e \in E(S, V-S)} w_e \leq \sigma \sum_{e \in \tilde{E}(S, V-S)} w_e \quad (4)$$

It is worth noticing that all spectral sparsifiers are also cut sparsifiers. This can easily be seen by plugging in x_S for any $S \subset V$ into (??)

For the rest of the paper, we will only be considering the sparsification of unweighted graphs. Sparsification of weighted graphs is a bit more involved but heavily draws on the concepts and algorithms used in unweighted graphs. In some cases we can simply scale unweighted graph algorithms to work on weighted graphs.

1.2 Linear Algebra Background

Here we go over some linear algebra that will be useful for the analysis of graphs and sparsification algorithms.

The spectral theorem for Hermitian Operators in linear algebra states that for any symmetric matrix, all of its eigenvalues are real and that its eigenvectors can be chosen to form an orthonormal basis of its domain. This theorem is central to our analysis of spectral sparsification.

In addition to the spectral theorem, we also use a list of basic linear algebra facts and notation listed below.

1. If a symmetric matrix has orthonormal eigenvectors x_1, x_2, \dots, x_n with eigenvalues $0 = \lambda_1 < \lambda_2 < \dots < \lambda_n$ then for all $x \neq 0$, $\lambda_1 \leq \frac{x^T A x}{x^T x} \leq \lambda_n$.
2. For a given matrix M we use $\lambda_i(M)$ to denote the i th smallest eigenvalue. This definition will only be applied to matrices with real eigenvalues.
3. The 2-norm of a symmetric matrix $\|A\|$ is equal to the absolute value of its largest eigenvalue.

1.3 Graph Properties

A closely related topic to sparsification is conductance. The conductance of a cut in a graph is a measure of the connectivity among vertices on both sides of the cut and the

number of edges crossing the cut. A set of vertices will have low conductance if there are a lot of edges on both sides of the cut but not many edges in the cut. We define the volume of a set S of vertices so that $\text{Vol}(S) = \sum_{i \in S} d_i$. Then we have a definition for the conductance of a set of vertices:

$$\Phi_G(S) = \frac{|E(S, V - S)|}{\min(\text{Vol}(S), \text{Vol}(V - S))} \quad (5)$$

We then get the conductance of the full graph by minimizing over all S such that S is a nonempty set of vertices

$$\Phi_G = \min_{\emptyset \neq S \subset V} \Phi_G(S) \quad (6)$$

Conductance is closely related to the Laplacian of a graph as we show in Cheeger's Inequality below. Before describing this, we first define the normalized Laplacian. The normalized Laplacian is the matrix

$$\mathcal{L}_G = D^{-1/2} L_G D^{-1/2} \quad (7)$$

Where D is the diagonal matrix defined by $D(u, v) = 0$ if $u \neq v$ and $D(v, v) = d_v$ (d_v denotes the degree of vertex v).

The normalized laplacian of a graph satisfies a similar equation to (??):

$$x^T \mathcal{L}_G x = \sum_{(u,v) \in E} \left(\frac{x_u}{\sqrt{d_u}} - \frac{x_v}{\sqrt{d_v}} \right)^2 \quad (8)$$

2 Sampling by vertex degree

2.1 Cheeger's Inequality

Cheeger's Inequality connects the conductance Φ_G with $\lambda_2(\mathcal{L})$. It states that

Theorem 1. $2\Phi_G \geq \lambda_2(\mathcal{L}) \geq \Phi_G^2/2$

Proof. First, consider the set S such that $\Phi_S = \Phi_G$ and such that $\text{Vol}(S) \leq \text{Vol}(V - S)$. Let $e, f > 0$ be positive real numbers that will be specified later. Then let x denote the vector defined by

$$x(v) = \begin{cases} e\sqrt{d_v} & \text{if } v \in S \\ -f\sqrt{d_v} & \text{if } v \notin S \end{cases}$$

Then it follows from equation (??) that

$$\frac{x^T \mathcal{L} x}{x^T x} = \frac{(e + f)^2 |E(S, V - S)|}{e^2 \text{Vol}(S) + f^2 \text{Vol}(V - S)}$$

We now select e, f such that $\sum_{v \in S} e d_v = \sum_{v \in V - S} f d_v$. This implies that x is orthogonal to the vector y defined by $y(v) = \sqrt{d_v}$. Therefore, we must have that $\frac{x^T \mathcal{L} x}{x^T x} \geq \lambda_2(\mathcal{L})$ since y can easily be verified to be the vector with eigenvalue 0 with respect to \mathcal{L} . However, we have that

$$\lambda_2(\mathcal{L}) \leq \frac{x^T \mathcal{L} x}{x^T x} = \frac{(e+f)^2 |E(S, V-S)|}{e^2 \text{Vol}(S) + f^2 \text{Vol}(V-S)} = \frac{(e+f)^2 |E(S, V-S)|}{(e+f)e \text{Vol}(S)} = \frac{(e+f)\Phi_G}{e} \leq 2\Phi_G$$

By substituting that $\sum_{v \in S} e d_v = \sum_{v \in V-S} f d_v$ and by using the fact that since $\text{Vol}(S) \leq \text{Vol}(V-S)$, $e \geq f$. This gives us one side of the inequality.

The other side of the inequality has a more complicated proof, so we will only provide a bit of intuition as to how the proof goes. The basic idea is to start with the eigenvector x such that $\mathcal{L}x = \lambda_2 x$. Then, sort the vertices v based on how large $x(v)$ is to get $x(v_1) \leq x(v_2) \leq \dots \leq x(v_n)$. The idea then becomes to show that we can find i such that $\lambda_2 \geq \Phi_{S_i}^2/2$ where $S_i = \{v_1, v_2, \dots, v_i\}$. The intuition behind this is somewhat analogous to the proof of the other direction in the inequality. We want to group large values of $x(v)$ together and smaller values of $x(v)$ together so that the terms in the laplacian that account for λ_2 , there has to be a decent number of edges spanning from S_i to $V - S_i$.

□

2.2 Algorithm

In this section we present an algorithm for generating a spectral sparsifier \tilde{G} of an unweighted graph G with high conductivity. The basic idea of the algorithm is to independently select edge e with probability p_e . If an edge is selected, it is added into \tilde{G} and given a weight of $1/p_e$. The main result is that for an accurate selection of p_e (which will be detailed later), the graph \tilde{G} will be a good spectral sparsifier of G with high probability.

Rigorizing the above, define $\text{Sample}(G, \epsilon, p, \lambda)$ where G is a graph with $\lambda_2(L_G) = \lambda$, and p, ϵ are real numbers between 0 and 1/2 to be the following routine.

1. Create \tilde{G} to be a graph with the same vertices as G and initially no edges
2. Set $k = \max(\log(3/p), \log n)$
3. Set $\Upsilon = (\frac{12k}{\epsilon\lambda})^2$
4. For every $(i, j) \in E$ set $p_{i,j} = \min(1, \frac{\Upsilon}{\min(d_i, d_j)})$
5. For every $(i, j) \in E$ place an edge of weight $1/p_{i,j}$ between vertices i and j in \tilde{G} with probability $p_{i,j}$
6. Output \tilde{G}

This routine basically selects edges with probabilities inversely proportional to the degrees to their vertices. The constant Υ is necessary to ensure that \tilde{G} does not deviate too far from G . We will now present the main result about this algorithm.

2.3 Analysis

Theorem 2. $\tilde{G} = \text{Sample}(G, \epsilon, p, \lambda)$ satisfies with probability at least $1 - p$

1. \tilde{G} is a $(1 + \epsilon)$ -approximation of G
2. \tilde{G} has at most $\frac{288nk^2}{\epsilon^2\lambda^2}$ edges

The proof of this theorem is fairly involved, but draws on many simple and intuitive concepts. Instead of providing a fully rigorized proof (which one can easily find in [?]), we will provide a proof sketch that gives intuition instead of rigor.

At a very high level, the basic idea of the proof is to prove that

1. \tilde{G} is NOT a $(1 + \epsilon)$ approximation of G with probability at most $2p/3$
2. \tilde{G} has strictly more than $\frac{288nk^2}{\epsilon^2\lambda^2}$ edges with probability at most $p/3$.

Since $P(A \cup B) \leq P(A) + P(B)$, the probability that \tilde{G} doesn't satisfy the desired properties would be at most p meaning that it works with probability $1 - p$. We now show how to obtain the two claims above.

Lemma 3. *Let \tilde{L} denote the Laplacian of \tilde{G} . Then if $\|D^{-1/2}(L - \tilde{L})D^{-1/2}\| \leq \delta$, \tilde{G} is a σ -approximation of G where $\sigma = \frac{\lambda}{\lambda - \delta}$*

Proof. We desire to show that $x^T Lx$ is close to $x^T \tilde{L}x$ for all x . A key observation is that for both L and \tilde{L} , the vector $t = (1, 1, 1, \dots, 1)$ satisfies $t^T L t = t^T \tilde{L} t = 0$. Consequently, let z be the unique vector such that $x = z + ct$ for some scalar c and such that $z^T t = 0$. Then it follows that $x^T Lx = z^T Lz$ and $x^T \tilde{L}x = z^T \tilde{L}z$. Using this, we have the following:

$$x^T \tilde{L}x = z^T \tilde{L}z = z^T Lz + z^T (\tilde{L} - L)z$$

We now make the substitution $y = D^{1/2}z$ yielding that

$$x^T \tilde{L}x = y^T D^{-1/2} L D^{-1/2} y + y^T D^{-1/2} (\tilde{L} - L) D^{-1/2} y$$

However, it is easy to see that $x^T Lx = y^T D^{-1/2} L D^{-1/2} y$. Also, the absolute value of $y^T D^{-1/2} L D^{-1/2} y$ is at least $\lambda|y|^2$ by the fact that y is orthogonal $D^{1/2}t$ which is the eigenvector of $D^{-1/2} L D^{-1/2}$ that corresponds to the eigenvalue 0 and by the fact that λ is the smallest non-zero eigenvalue of $D^{-1/2} L D^{-1/2}$. Also, the absolute value of $y^T D^{-1/2} (\tilde{L} - L) D^{-1/2} y$ is at most $\delta|y|^2$. This means the ratio of $y^T D^{-1/2} (\tilde{L} - L) D^{-1/2} y$ to $y^T D^{-1/2} L D^{-1/2} y$ is at most δ/λ . From this, it immediately follows that

$$x^T Lx(1 - \delta/\lambda) \leq x^T \tilde{L}x \leq x^T Lx(1 + \delta/\lambda)$$

which immediately gives us our desired result.

This lemma suggests that in order to prove the original theorem, since we already know that $\lambda_2(\mathcal{L}) = \lambda$, it suffices to analyze the distribution of $\|D^{-1/2}(\tilde{L} - L)D^{-1/2}\|$.

To do this, we notice that $\|D^{-1/2}(\tilde{L} - L)D^{-1/2}\| \leq \|D^{-1/2}(\tilde{A} - A)D^{-1/2}\| + \|D^{-1/2}(\tilde{D} - D)D^{-1/2}\|$ by the triangle inequality and by the fact that $A + L = D$ and $\tilde{A} + \tilde{L} = \tilde{D}$ where A denotes the adjacency matrix and D denotes the diagonal matrix (with diagonal entries vertex degrees). We analyze each of those terms separately. The motivation for doing this will become clear when we analyze the term $\|D^{-1/2}(\tilde{A} - A)D^{-1/2}\|$

Finding the distribution of $\|D^{-1/2}(\tilde{D} - D)D^{-1/2}\|$ is relatively easy. Suppose d_1, d_2, \dots, d_n are the diagonal entries of D and $\tilde{d}_1, \tilde{d}_2, \tilde{d}_3, \dots, \tilde{d}_n$ be the diagonal entries of \tilde{D} . It follows quite easily that the largest eigenvalue of $D^{-1/2}(\tilde{D} - D)D^{-1/2}$ will just be the maximum value of $|(\tilde{d}_i - d_i)/d_i| = |1 - \frac{\tilde{d}_i}{d_i}|$. Now each \tilde{d}_i can be expressed as the sum of a bunch of random variables (namely, the weights of the edges of vertex i),

and as a result of this we can get a bound on $\Pr[||D^{-1/2}(\tilde{D} - D)D^{-1/2}|| \geq \epsilon\lambda/3] \leq p/3$ without too much difficulty.

The distribution of $||D^{-1/2}(\tilde{A} - A)D^{-1/2}||$ is a little more tricky to analyze. First of all, for any matrix B , conjugating by an invertible matrix C to get CBC^{-1} preserves the eigenvalues of B . Consequently, it is equivalent to analyze the distribution of $||\Delta|| = ||D^{-1}(\tilde{A} - A)||$. Let the eigenvalues of Δ be f_1, f_2, \dots, f_n and let $m \geq 2$ be an even integer. Then the eigenvalues of Δ^m are $f_1^m, f_2^m, \dots, f_n^m$. Now suppose f_1 is the eigenvalue with the largest absolute value. Then it follows that $f_1^m \leq \sum_i f_i^m$ which means that $|f_1| \leq (\sum_i f_i^m)^{1/m}$. The key observation is that $\sum_i f_i^m = \text{Tr}(\Delta^m)$. Substituting this, we get that

$$||D^{-1}(\tilde{A} - A)|| \leq (\text{Tr}(\Delta^m))^{1/m} \quad (9)$$

This equation is very useful because for any r ,

$$\Pr[||D^{-1}(\tilde{A} - A)|| > r] \leq \Pr[(\text{Tr}(\Delta^m))^{1/m} > r] = \Pr[(\text{Tr}(\Delta^m)) > r^m] \quad (10)$$

□

The nice thing about this is that we no longer have to mess around with eigenvalues and only care about the trace of the matrix Δ^m . In the final paper, we will analyze the term $\text{Tr}[\Delta^m]$ further by relating Δ^m to m iterations of a stochastic process in $\tilde{G} - G$.

3 Sampling by Effective Resistance

3.1 Effective Resistance Background

The effective resistance in graphs derived from electrical networks shows great promise as a new, powerful measure of distance and gives important contributions to the analysis of spectral graph theory. In the overview, we will define effective resistance by modelling the graph as an electrical network, connect effective resistance to spectral graph theory, and look at the relationship between commute time and effective resistance. Then we examine an algorithm that samples edges with probabilities proportional to their effective resistance and give an analysis.

3.1.1 Calculating Effective Resistance

In order to understand effective resistance, we need to remember two laws from electromagnetism, Ohm's and Kirchoff's. Ohm's Law states that $V = ir$ if V is voltage, i current, and r resistance. Kirchoff's Law requires the conservation of current, stating the current into any vertex is equal to the current leaving the vertex. The effective resistance between two vertices u and v is then equal to the voltage differential between the two when a unit of current is from u to v .¹

Calculating the effective resistance on a small circuit is easy. When we have resistors in parallel we simply add them and when we have resistors in series we combine their inverses in a formula you should remember from electromagnetism. We can then

¹This definition of effective resistance may seem a little strange at first but it relates to the analysis in [?] which gives a nice overview of effective resistance.

apply these two methods to combine all of the resistors into one, yielding the effective resistance between two vertices. This gets much more complicated as the circuit gets larger and we will give another expression of effective resistance in terms of eigenvalues of the Laplacian.

3.1.2 Connection to Laplacian

Intuitively we can think of effective resistance as a property that would be preserved with the Laplacian because they are both local properties that look at the entire graph. In the Laplacian you can weigh \mathbf{x} so that it picks out vertices and edges and makes them important. You can do this with every edge so it locally makes every part of the graph important. Now consider the effective resistance between u and v , any edge on the graph that we add will decrease the effective resistance. The effect of the decrease in resistance will be larger the closer the new edge is to u and v . Effective resistance at each considers the whole graph but places importance on edges in the nearby subgraph.

We can relate the idea of effective resistance to the Laplacian and spectral sparsification. We can define a matrix B such that $B_{m \times n}$ is the *signed edge-vertex incidence matrix* [?] where

$$B(e, v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

and $W_{m \times m}$ is defined as

$$W(e, e) = w_e \quad (12)$$

which is the diagonal matrix with the weight of an edge at each diagonal. We can then define our Laplacian L as $B^T W B$.

Kirchoff's law then tells us $B^T \mathbf{i} = \mathbf{i}_{\text{ext}}$ where \mathbf{i}_{ext} is a vector of the currents injected at the vertices. And then Ohm's law tells us that the current of an edge is equal to the potential difference across it divided by the resistance or, in another expression, multiplied by the conductance, giving us $\mathbf{i} = W B \mathbf{v}$. Combining these two laws gives us a relation between the voltage difference and the current injected at the vertices, $\mathbf{i}_{\text{ext}} = B^T (W B \mathbf{v}) = L \mathbf{v}$. This gives us an intuitive notion of the relationship between the effective resistance the Laplacian of a graph.

Now we define the pseudoinverse of the Laplacian and find the effective resistance in terms of the pseudoinverse. The Laplacian has an eigenvalue of 0 so its discriminant is zero making it non-invertible. But we are able to take the pseudoinverse so that LL^+ is a projection into the space spanned by the Laplacian. Recall from the spectral theorem that the Laplacian can be decomposed to eigenvectors in the following manner

$$L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^T \quad (13)$$

where the λ_i 's are eigenvalues and the u_i 's are their corresponding orthonormal eigenvectors. We can then define the *Moore-Penrose Pseudoinverse* of L as

$$L^+ = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} u_i u_i^T \quad (14)$$

and we also get the following properties from the pseudoinverse

$$L^+ \mathbf{w} = \mathbf{0}$$

and for all $w \perp \mathbf{1}$

$$L^+ \mathbf{w} = \mathbf{v} \text{ such that } L\mathbf{v} = \mathbf{w} \text{ and } \mathbf{v} \perp \mathbf{1}$$

We are then able to express the effective resistance between vertices a and b , R_{ab} as vectors in the pseudoinverse of the Laplacian as follows

$$R_{ab} = (\mathbf{e}_a - \mathbf{e}_b)^T L^+ (\mathbf{e}_a - \mathbf{e}_b) \quad (15)$$

$$= L_{aa}^+ - 2L_{ab}^+ + L_{bb}^+ \quad (16)$$

Through another formulation we can see that

$$R_e = BL^+ B^T(e, e) \quad (17)$$

where R_e is the effective resistance of edge e .

3.1.3 Connection to commute time

We now define the commute time on an graph between two vertices and relate it to effective resistance. In a random walk on a graph we start at one vertex and at each step we randomly choose an outgoing edge to follow. The hit time between u and v is the expected time a random walk will take to traverse from u to v . The commute time from u to v is the sum of the hit time from u to v and the hit time from v to u . u and v have a small commute time if there are a lot of short paths between the two vertices.

We are able to relate commute time to effective resistance with the following lemma

Lemma 4. *If we look at two vertices u and v and call the commute time C_{uv} and the effective resistance R_{uv} we can get a relation between the two in the form $C_{uv} = 2mR_{uv}$.*

We will not prove this Lemma but it is proven in [?] and gives us a nice connection between commute time and effective resistance.

In the final version of the paper we will show how commute time gives a good notion of graph connectivity so that if we preserve commute time a lot of other properties are also preserved.

3.2 Algorithm

Spielman and Srivastava introduced an alternative sampling algorithm for generating a spectral sparsifier of $O(n \log n / \epsilon^2)$ edges in nearly linear time based on effective resistance [?]. Their paper defines a sparsify algorithm that runs as follows

$$H = \text{Sparsify}(G, q)$$

Choose a random edge e with probability p_e proportional to $w_e R_e$ and add e to H with weight w_e/qp_e . Take q samples independently with replacement, summing weights if an edge is chosen more than once.

The paper also proves the following theorem

Theorem 5. *Suppose we have G and $H = \text{Sparsify}(G, Q)$ and $1/\sqrt{n} < \epsilon \leq 1$. If $q = O(n \log n/\epsilon^2)$ and n is sufficiently large then with probability at least half H is a $(1 + \epsilon)$ -spectral approximation of G .*

It makes sense to sample with probability inversely related to effective resistance because this corresponds to the effective conductance of an edge which is proportional to how strong the edge connects the two vertices. If the effective conductance is high we know that the edge very important in the connectivity of a graph and we must include it.

3.3 Analysis

The proof of correctness for the sampling algorithm involves a lot of linear algebra and here we will go over but not prove a few of the key lemmas.

Lemma 6 (Projection Matrix). *(i) Π is a projection matrix. (ii) $\text{im}(\Pi) = \text{im}(W^{1/2}B) = W^{1/2}\mathbb{B}$. (iii) The eigenvalues of Π are 1 with multiplicity $n - 1$ and 0 with multiplicity $m - n + 1$. (iv) $\Pi(e, e) = \|\Pi(\cdot, e)\|^2$. W is defined in (??) and B is defined in (??).*

This projection matrix allows us to get a bound on the Laplacian through the following lemma

Lemma 7. *Suppose S is a nonnegative diagonal matrix such that*

$$\|\Pi S \Pi - \Pi \Pi\|_2 \leq \epsilon$$

Then

$$\forall x \in \mathbb{R}^n. (1 - \epsilon)x^T Lx \leq \tilde{L} \leq (1 + \epsilon)x^T Lx,$$

where $L = B^T W B$ and $\tilde{L} = B^T W^{1/2} S W^{1/2} B$.

4 Comparing the two algorithms

4.1 The related intuition between vertex degree and effective resistance

Here we will relate the two algorithms described above and give some shared intuition.

4.2 Combinatorics vs linear algebra

Verifying cut sparsifiers is NP-hard while verifying spectral sparsifiers is not. In spectral sparsification we have a harder problem that reduces the size of our solution set.

Why we need to partition when looking at just degree and not when looking at effective resistance. Look at the case of two complete graphs connected by one edge and how the different sampling algorithms handle this.

Here we will discuss if linear algebra is needed and contemplate replacing certain linear algebra techniques with a combinatorial analysis.

5 Conclusion

Here we give some conclusions about our project. Relating the paper to our original goals and talk about potential work in the future.