Twice-Ramanujan Sparsifiers *

Joshua Batson † Daniel A. Spielman ‡ Nikhil Srivastava §

October 31, 2018

Abstract

We prove that for every $d > 1$ and every undirected, weighted graph $G = (V, E)$, there exists a weighted graph H with at most $\lceil d\,|V| \rceil$ edges such that for every $x \in \mathbb{R}^V$,

$$1 \leq \frac{x^T L_H x}{x^T L_G x} \leq \frac{d + 1 + 2\sqrt{d}}{d + 1 - 2\sqrt{d}},$$

where L_G and L_H are the Laplacian matrices of G and H, respectively.

1 Introduction

A sparsifier of a graph $G = (V, E)$ is a sparse graph H that is 'similar' to G. We consider the spectral notion of 'similar' introduced by Spielman and Teng [ST04]: we say that H is a κ-approximation of G if for all $x \in \mathbb{R}^V$,

$$1 \leq \frac{x^T L_H x}{x^T L_G x} \leq \kappa. \tag{1}$$

In the case where G is the complete graph, optimal sparsifiers are supplied by *Ramanujan Graphs* [LPS88, Mar88]. These are d-regular graphs H that κ-approximate the complete graph, for

$$\kappa = \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}}.$$

In this paper, we prove that every graph can be approximated at least this well with only twice as many edges (as a d-regular graph has $dn/2$ edges).

*This material is based upon work supported by the National Science Foundation under Grant CCF-0634957. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

†Yale College.

‡Program in Applied Mathematics and Department of Computer Science, Yale University.

§Department of Computer Science, Yale University.

Theorem 1.1. *For every $d > 1$, every undirected weighted graph G with n vertices contains a weighted subgraph H with $\lceil dn \rceil$ edges (i.e., average degree $2d$) that satisfies:*

$$1 \leq \frac{x^T L_H x}{x^T L_G x} \leq \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}.$$

We remark that while the edges of H are subset of the edges of G, the weights of edges in H and G will typically be different.

At the end of the paper, we observe that our proof provides a deterministic algorithm for computing the graph H in time $O(dn^3 m)$.

2 Prior Work

Spielman and Teng [ST04] introduced the notion of sparsification that we consider, and proved that $(1+\epsilon)$-approximations with $\widetilde{O}(n/\epsilon^2)$ edges could be constructed in $\widetilde{O}(m)$ time. They were inspired by the notion of sparsification introduced by Benczur and Karger [BK96] for cut problems, which only required inequality (1) to hold for all $x \in \{0,1\}^V$; Benczur and Karger showed how to construct graphs H meeting this guarantee with $O(n \log n/\epsilon^2)$ edges in $O(m \log^3 n)$ time.

Spielman and Srivastava [SS08] recently proved the existence of spectral sparsifiers with $O(n \log n/\epsilon^2)$ edges, and showed how to construct them in $\widetilde{O}(m)$ time. They conjectured that it should be possible to find such sparsifiers with only $O(n/\epsilon^2)$ edges. We resolve this conjecture.

Very recently, partial progress was made towards the conjecture by Goyal, Rademacher and Vempala [GRV08], who show how to find graphs H with only $2n$ edges that $O(\log n)$-approximate bounded degree graphs G under the cut notion of Benczur and Karger.

We remark that all of these constructions were randomized. Ours is the first deterministic algorithm to achieve the guarantees of any of these papers.

3 Preliminaries

3.1 The Incidence Matrix and the Laplacian

Let $G = (V, E, w)$ be a connected weighted undirected graph with n vertices and m edges and edge weights $w_e > 0$. If we orient the edges of G arbitrarily, we can write its Laplacian as $L = B^T W B$, where $B_{m \times n}$ is the *signed edge-vertex incidence matrix*, given by

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

and $W_{m\times m}$ is the diagonal matrix with $W(e,e) = w_e$. Denote the row vectors of B by $\{b_e\}_{e\in E}$ and note that $b_{(u,v)}^T = (\chi_v - \chi_u)$. It is well known that that $\mathrm{im}(B) \subseteq \mathbb{R}^m$ is the *cut space* of G [GR01].

It is immediate that L is positive semidefinite since:

$$x^T L x = x^T B^T W B x = \|W^{1/2} B x\|_2^2 \geq 0 \quad \text{for every } x \in \mathbb{R}^n.$$

We also have $\ker(L) = \ker(W^{1/2}B) = \mathrm{span}(\mathbf{1})$, since

$$\begin{aligned} x^T L x = 0 &\iff \|W^{1/2} B x\|_2^2 = 0 \\ &\iff \sum_{uv\in E} w_{uv}(x(u) - x(v))^2 = 0 \\ &\iff x(u) - x(v) = 0 \quad \text{for all edges } (u,v) \\ &\iff \ x \text{ is constant, since } G \text{ is connected.} \end{aligned}$$

3.2 The Pseudoinverse

Since L is symmetric we can diagonalize it and write

$$L = \sum_{i=1}^{n-1} \lambda_i u_i u_i^T$$

where $\lambda_1, \ldots, \lambda_{n-1}$ are the nonzero eigenvalues of L and $u_1, \ldots, u_{n-1}$ are a corresponding set of orthonormal eigenvectors. The *Moore-Penrose Pseudoinverse* of L is then defined as

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

Notice that $\ker(L) = \ker(L^+)$ and that

$$LL^+ = L^+L = \sum_{i=1}^{n-1} u_i u_i^T,$$

which is simply the projection onto the span of the nonzero eigenvectors of L (which are also the eigenvectors of L^+). Thus, $LL^+ = L^+L$ is the identity on $\mathrm{im}(L) = \ker(L)^\perp = \mathbb{R}^n \setminus \mathrm{span}(\mathbf{1})$.

3.3 The Sherman-Morrison Formula

We use the following well-known theorem from Linear Algebra, which describes the behavior of the inverse of a matrix under rank-one updates (see [GV96, Section 2.1.3]).

Lemma 3.1. *If A is a nonsingular $n \times n$ matrix and $\boldsymbol{\pi}$ is a vector, then*

$$(A + \boldsymbol{\pi}\boldsymbol{\pi}^T)^{-1} = A^{-1} - \frac{A^{-1}\boldsymbol{\pi}\boldsymbol{\pi}^T A^{-1}}{1 + \boldsymbol{\pi}^T A^{-1}\boldsymbol{\pi}}.$$

4 The Main Result

As in [SS08], we begin by considering the $m \times m$ matrix $\Pi = W^{1/2} B L^{+} B^{T} W^{1/2}$.

We will exploit the properties of Π stated in the following lemma. See [SS08] for a proof.

Lemma 4.1 (Projection Matrix)**.** *(i) Π is a projection matrix. (ii) $\mathrm{im}(\Pi) = \mathrm{im}(W^{1/2}B) \subseteq \mathbb{R}^m$. (iii) The eigenvalues of Π are 1 with multiplicity $n-1$ and 0 with multiplicity $m-n+1$.*

Each column (and row) of Π is associated with an edge of G. Let $\boldsymbol{\pi}_e$ denote the restriction of the column associated with edge e to the image of Π, under an arbitrary choice of orthonormal basis for the image of Π. By restricting to this space, we have

$$\sum_e \boldsymbol{\pi}_e \boldsymbol{\pi}_e^T = I.$$

When $H = (V, F, \tilde{w})$ is a subgraph of $G = (V, E, w)$, there is a simple relationship between L_H and L_G via the $m \times m$ diagonal matrix

$$S_H(e,e) = \frac{\tilde{w}_e}{w_e};$$

specifically, we have

$$L_H = B^T W^{1/2} S_H W^{1/2} B,$$

and

$$\Pi S_H \Pi = \sum_e S_H(e,e) \boldsymbol{\pi}_e \boldsymbol{\pi}_e^T.$$

This immediately gives the following identity, which reduces the task of sparsification to choosing a small, well-conditioned subset of the columns of Π, up to rescaling.

Lemma 4.2 (Condition Number)**.** *If S_H is defined as above, then*

$$\lambda_{\min}(\Pi S_H \Pi) \leq \frac{x^T L_H x}{x^T L_G x} \leq \lambda_{\max}(\Pi S_H \Pi) \qquad \forall x \perp \ker(L_G) = \mathrm{span}(\mathbf{1}),$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the smallest and largest eigenvalues of the restriction of $\Pi S_H \Pi$ to $\mathrm{im}(\Pi)$.

Proof. Notice that

$$\begin{aligned}
\frac{x^T L_H x}{x^T L_G x} &= \frac{x^T B^T W^{1/2} S_H W^{1/2} B x}{x^T B^T W B x} \\
&= \frac{y^T \Pi^T S_H \Pi y}{y^T \Pi \Pi y} \quad \text{for some } y \in \mathrm{im}(\Pi), \text{ since } \mathrm{im}(W^{1/2}B) = \mathrm{im}(\Pi) \\
&= \frac{y^T \Pi S_H \Pi y}{y^T y} \quad \text{since } \Pi = \Pi^T \text{ and } \Pi y = y \text{ for } y \in \mathrm{im}(\Pi).
\end{aligned}$$

But this is just the Rayleigh quotient of $\Pi S_H \Pi$, and therefore bounded between $\lambda_{\min}(\Pi S_H \Pi)$ and $\lambda_{\max}(\Pi S_H \Pi)$ for $y \in \mathrm{im}(\Pi)$. $\square$

So what we have to do is choose a small set of coefficients s_e so that $\sum_e s_e \boldsymbol{\pi}_e \boldsymbol{\pi}_e^T$ is well-conditioned. To this end we define two 'barrier' potential functions which measure the quality of the eigenvalues of a matrix.

Definition 4.3. *For $u, l \in \mathbb{R}$ and A a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2, \ldots, \lambda_{n-1}$, define:*

$$\Phi^u(A) \triangleq \mathrm{Tr}(uI - A)^{-1} = \sum_i \frac{1}{u - \lambda_i} \quad \textit{Upper potential function.}$$

$$\Phi_l(A) \triangleq \mathrm{Tr}(A - lI)^{-1} = \sum_i \frac{1}{\lambda_i - l} \quad \textit{Lower potential function.}$$

As long as $A \prec uI$ and $A \succ lI$ (i.e., $\lambda_{\max}(A) < u$ and $\lambda_{\min}(A) > l$), these potential functions measure how far the eigenvalues of A are from the barriers u and l. In particular, they blow up as any eigenvalue approaches a barrier, since then $uI - A$ (or $A - lI$) approaches a singular matrix.

To prove the theorem, we will construct a sequence of matrices

$$0 = A^{(0)}, A^{(1)}, \ldots, A^{(s)}, \ldots A^{(S)}$$

along with positive constants $u_0, l_0, \delta_u, \delta_l, \epsilon_u$ and ϵ_l which satisfy the following conditions:

1. Initially, the potentials are

$$\Phi^{u_0}(A^{(0)}) = \epsilon_u \quad \text{and} \quad \Phi_{l_0}(A^{(0)}) = \epsilon_l.$$

2. Each matrix is obtained by a rank-one update of the previous one — specifically by adding some multiple of a column of Π.

$$A^{(s+1)} = A^{(s)} + t\boldsymbol{\pi}\boldsymbol{\pi}^T \quad \text{for some } \boldsymbol{\pi} \in \{\boldsymbol{\pi}_e\}_{e \in E} \text{ and } t \geq 0.$$

3. If we shift the barriers u and l by δ_u and δ_l respectively at each step, then the upper and lower potentials do not increase. For every $s = 0, 1, \ldots S$,

$$\Phi^{u+\delta_u}(A^{(s+1)}) \leq \Phi^u(A^{(s)}) \leq \epsilon_u \quad \text{for } u = u_0 + s\delta_u.$$

$$\Phi_{l+\delta_l}(A^{(s+1)}) \leq \Phi_l(A^{(s)}) \leq \epsilon_l \quad \text{for } l = l_0 + s\delta_l.$$

4. No eigenvalue ever jumps accross a barrier. For every $s = 0, 1, \ldots S$,

$$\lambda_{\max}(A^{(s)}) < u_0 + s\delta_u$$

$$\lambda_{\min}(A^{(s)}) > l_0 + s\delta_l.$$

To complete the proof we will choose $u_0, l_0, \delta_u, \delta_l, \epsilon_u$ and ϵ_l so that after $S = dn$ steps, the condition number of $A^{(S)}$ is bounded by

$$\frac{\lambda_{\max}(A^{(S)})}{\lambda_{\min}(A^{(S)})} \leq \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}.$$

The main technical challenge is to show that there is always a choice of $\boldsymbol{\pi}\boldsymbol{\pi}^T$ to add to the current matrix which allows us to shift *both* barriers up by a constant without increasing either potential. We achieve this in the following 3 lemmas.

Lemma 4.4 (Upper Barrier Shift). *Suppose $\lambda_{\max}(A) < u$ and $\boldsymbol{\pi}$ is any vector. If*

$$\frac{1}{t} \geq \frac{\boldsymbol{\pi}^T((u+\delta_u)I - A)^{-2}\boldsymbol{\pi}}{\Phi^u(A) - \Phi^{u+\delta_u}(A)} + \boldsymbol{\pi}^T((u+\delta_u)I - A)^{-1}\boldsymbol{\pi} \triangleq U_A(\boldsymbol{\pi})$$

then

$$\Phi^{u+\delta_u}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T) \leq \Phi^u(A) \quad \text{and } \lambda_{\max}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T) < u + \delta_u.$$

That is, if we add t times $\boldsymbol{\pi}\boldsymbol{\pi}^T$ to A and shift the upper barrier by δ_u, then we do not increase the upper potential.

Proof. Let $u' = u + \delta_u$. By the Sherman-Morrison formula, we can write the updated potential as:

$$\begin{aligned}
\Phi^{u+\delta_u}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T) &= \mathrm{Tr}(u'I - A - t\boldsymbol{\pi}\boldsymbol{\pi}^T)^{-1} \\
&= \mathrm{Tr}\left((u'I - A)^{-1} + \frac{t(u'I-A)^{-1}\boldsymbol{\pi}\boldsymbol{\pi}^T(u'I-A)^{-1}}{1 - t\boldsymbol{\pi}^T(u'I-A)^{-1}\boldsymbol{\pi}}\right) \\
&= \mathrm{Tr}(u'I - A)^{-1} + \frac{t\mathrm{Tr}(\boldsymbol{\pi}^T(u'I-A)^{-1}(u'I-A)^{-1}\boldsymbol{\pi})}{1 - t\boldsymbol{\pi}^T(u'I-A)^{-1}\boldsymbol{\pi}} \\
&\qquad \text{since Tr is linear and } \mathrm{Tr}(XY) = \mathrm{Tr}(YX) \\
&= \Phi^{u+\delta_u}(A) + \frac{t\boldsymbol{\pi}^T(u'I-A)^{-2}\boldsymbol{\pi}}{1 - t\boldsymbol{\pi}^T(u'I-A)^{-1}\boldsymbol{\pi}} \\
&= \Phi^u(A) - (\Phi^u(A) - \Phi^{u+\delta_u}(A)) + \frac{\boldsymbol{\pi}^T(u'I-A)^{-2}\boldsymbol{\pi}}{1/t - \boldsymbol{\pi}^T(u'I-A)^{-1}\boldsymbol{\pi}}
\end{aligned}$$

Substituting $1/t \geq U_A(\boldsymbol{\pi})$ gives $\Phi^{u+\delta_u}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T) \leq \Phi^u(A) < \infty$ for all $t \in [0, 1/U_A(\boldsymbol{\pi})] \setminus \Gamma$, where $\Gamma = \{t : \{\lambda_i(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T)\} \ni (u + \delta_u)\}$ is the set where $h(t) = \Phi^{u+\delta_u}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T)$ is not defined.

Assume for contradiction that $\lambda_{\max}(A + t_0\boldsymbol{\pi}\boldsymbol{\pi}^T) \geq (u + \delta_u)$ for some $t_0 \in [0, 1/U_A(\boldsymbol{\pi})]$. Since $\lambda_{\max}(A + t\boldsymbol{\pi}\boldsymbol{\pi}^T)$ is continuous and nondecreasing in t and $\lambda_{\max}(A) < u$, there must be some least $0 < t_1 \leq t_0$ for which $\lambda_{\max}(A + t_1\boldsymbol{\pi}\boldsymbol{\pi}^T) = u + \delta_u$. Now h is defined on $[0, t_1)$ and $h(t) \uparrow \infty$ as $t \uparrow t_1$. But this is impossible since $h(t) \leq \Phi^u(A)$ on $[0, 1/U_A(\boldsymbol{\pi})] \setminus \Gamma$. $\square$

Lemma 4.5 (Lower Barrier Shift)**.** *Suppose* $\lambda_{\min}(A) > l + \delta_l$ *and* $\boldsymbol{\pi}$ *is any vector. If*

$$\frac{1}{t} \leq \frac{\boldsymbol{\pi}^T(A-(l+\delta_l)I)^{-2}\boldsymbol{\pi}}{\Phi_{l+\delta_l}(A)-\Phi_l(A)} - \boldsymbol{\pi}^T(A-(l+\delta_l)I)^{-1}\boldsymbol{\pi} \triangleq L_A(\boldsymbol{\pi})$$

then

$$\Phi_{l+\delta_l}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) \leq \Phi_l(A) \quad \text{and } \lambda_{\min}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) > l+\delta_l.$$

That is, if we add t *times* $\boldsymbol{\pi}\boldsymbol{\pi}^T$ *to* A *and shift the lower barrier by* δ_l*, then we do not increase the lower potential.*

Proof. We proceed as in the proof for the upper potential. Let $l' = l + \delta_l$. By Sherman-Morrison, we have:

$$\begin{aligned}
\Phi_{l+\delta_l}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) &= \mathrm{Tr}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T - l'I)^{-1} \\
&= \mathrm{Tr}\left((A-l'I)^{-1} - \frac{t(A-l'I)^{-1}\boldsymbol{\pi}\boldsymbol{\pi}^T(A-l'I)^{-1}}{1+t\boldsymbol{\pi}^T(A-l')^{-1}\boldsymbol{\pi}}\right) \\
&= \mathrm{Tr}(A-l'I)^{-1} - \frac{t\mathrm{Tr}(\boldsymbol{\pi}^T(A-l'I)^{-1}(A-l'I)^{-1}\boldsymbol{\pi})}{1+t\boldsymbol{\pi}^T(A-l'I)^{-1}\boldsymbol{\pi}} \\
&= \Phi_{l+\delta_l}(A) - \frac{t\boldsymbol{\pi}^T(A-l'I)^{-2}\boldsymbol{\pi}}{1+t\boldsymbol{\pi}^T(A-l'I)^{-1}\boldsymbol{\pi}} \\
&= \Phi_l(A) + (\Phi_{l+\delta_l}(A) - \Phi_l(A)) - \frac{\boldsymbol{\pi}^T(A-l'I)^{-2}\boldsymbol{\pi}}{1/t+\boldsymbol{\pi}^T(A-l'I)^{-1}\boldsymbol{\pi}}
\end{aligned}$$

Rearranging shows that $\Phi_{l+\delta_l}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) \leq \Phi_l(A)$ when $1/t \leq L_A(\boldsymbol{\pi})$. It is immediate that $\lambda_{\min}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) > l+\delta_l$ since $\lambda_{\min}(A+t\boldsymbol{\pi}\boldsymbol{\pi}^T) \geq \lambda_{\min}(A)$. □

Lemma 4.6 (Both Barriers)**.** *If* $\Phi^u(A) \leq \epsilon_u, \Phi_l(A) \leq \epsilon_l$*, and* $\epsilon_u, \epsilon_l, \delta_u$ *and* δ_l *satisfy*

$$0 \leq \frac{1}{\delta_u} + \epsilon_u \leq \frac{1}{\delta_l} - \epsilon_l \tag{2}$$

then there exists an e *for which*

$$L_A(\boldsymbol{\pi}_e) \geq U_A(\boldsymbol{\pi}_e).$$

Proof. We will show that

$$\sum_e L_A(\boldsymbol{\pi}_e) \geq \sum_e U_A(\boldsymbol{\pi}_e),$$

from which the claim will follow. We begin by bounding

$$
\begin{aligned}
\sum_e U_A(\boldsymbol{\pi}_e) &= \frac{\sum_e \boldsymbol{\pi}_e^T((u+\delta_u)I-A)^{-2}\boldsymbol{\pi}_e}{\Phi^u(A)-\Phi^{u+\delta_u}(A)} + \sum_e \boldsymbol{\pi}_e^T((u+\delta_u)I-A)^{-1}\boldsymbol{\pi}_e \\
&= \frac{((u+\delta_u)I-A)^{-2}\bullet(\sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T)}{\Phi^u(A)-\Phi^{u+\delta_u}(A)} + ((u+\delta_u)I-A)^{-1}\bullet(\sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T) \\
&= \frac{\mathrm{Tr}((u+\delta_u)I-A)^{-2}}{\Phi^u(A)-\Phi^{u+\delta_u}(A)} + \mathrm{Tr}((u+\delta_u)I-A)^{-1} \\
&\qquad \text{since } \sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T = \Pi = I \text{ and } X\bullet I = \mathrm{Tr}(X) \\
&= \frac{\sum_i \frac{1}{(u+\delta_u-\lambda_i)^2}}{\sum_i \frac{1}{u-\lambda_i} - \sum_i \frac{1}{u+\delta_u-\lambda_i}} + \Phi^{u+\delta_u}(A) \\
&= \frac{\sum_i \frac{1}{(u+\delta_u-\lambda_i)^2}}{\delta_u \sum_i \frac{1}{(u-\lambda_i)(u+\delta_u-\lambda_i)}} + \Phi^{u+\delta_u}(A) \\
&\le \frac{1}{\delta_u} + \Phi^{u+\delta_u}(A) \quad \text{as } \sum_i \frac{1}{(u-\lambda_i)(u+\delta_u-\lambda_i)} \ge \sum_i \frac{1}{(u+\delta_u-\lambda_i)^2} \\
&\le \frac{1}{\delta_u} + \Phi^u(A) \le \frac{1}{\delta_u} + \epsilon_u.
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\sum_e L_A(\boldsymbol{\pi}) &= \frac{\sum_e \boldsymbol{\pi}_e^T((A-(l+\delta_l))^{-2}\boldsymbol{\pi}_e}{\Phi_{l+\delta_l}(A)-\Phi_l(A)} - \sum_e \boldsymbol{\pi}_e^T(A-(l+\delta_l)I)^{-1}\boldsymbol{\pi}_e \\
&= \frac{(A-(l+\delta_l)I)^{-2}\bullet(\sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T)}{\Phi_{l+\delta_l}(A)-\Phi_l(A)} - (A-(l+\delta_l)I)^{-1}\bullet(\sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T) \\
&= \frac{\mathrm{Tr}(A-(l+\delta_l)I)^{-2}}{\Phi_{l+\delta_l}(A)-\Phi_l(A)} - \mathrm{Tr}(A-(l+\delta_l)I)^{-1} \\
&\qquad \text{since } \sum_e \boldsymbol{\pi}_e\boldsymbol{\pi}_e^T = \Pi = I \text{ and } X\bullet I = \mathrm{Tr}(X) \\
&= \frac{\sum_i \frac{1}{(\lambda_i-l-\delta_l)^2}}{\sum_i \frac{1}{\lambda_i-l-\delta_l} - \sum_i \frac{1}{\lambda_i-l}} - \sum_i \frac{1}{\lambda_i-l-\delta_l}. \\
&\ge \frac{1}{\delta_l} - \sum_i \frac{1}{\lambda_i - l},
\end{aligned}
$$

by Claim 4.7.

Putting everything together, we find that

$$
\sum_e U_A(\boldsymbol{\pi}_e) \le \frac{1}{\delta_u} + \epsilon_u \le \frac{1}{\delta_l} - \epsilon_l \le \sum_e L_A(\boldsymbol{\pi}_e),
$$

as desired. □

Claim 4.7. *If $\lambda_i > l$ for all i, $0 \leq \sum(\lambda_i - l)^{-1} \leq \epsilon_l$, and $1/\delta_l - \epsilon_l \geq 0$, then*

$$\frac{\sum_i \frac{1}{(\lambda_i - l - \delta_l)^2}}{\sum_i \frac{1}{\lambda_i - l - \delta_l} - \sum_i \frac{1}{\lambda_i - l}} - \sum_i \frac{1}{\lambda_i - l - \delta_l} \geq \frac{1}{\delta_l} - \sum_i \frac{1}{\lambda_i - l}. \tag{3}$$

Proof. We have

$$\delta_l \leq 1/\epsilon_l \leq \lambda_i - l,$$

for every i. So, the denominator of the left-most term on the left-hand side is positive, and the claimed inequality is equivalent to

$$\begin{aligned}
&\sum_i \frac{1}{(\lambda_i - l - \delta_l)^2} \\
&\geq \left(\sum_i \frac{1}{\lambda_i - l - \delta_l} - \sum_i \frac{1}{\lambda_i - l}\right)\left(\frac{1}{\delta_l} + \sum_i \frac{1}{\lambda_i - l - \delta_l} - \sum_i \frac{1}{\lambda_i - l}\right) \\
&= \left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)}\right)\left(\frac{1}{\delta_l} + \delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)}\right) \\
&= \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)} + \left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)}\right)^2,
\end{aligned}$$

which, by moving the first term on the RHS to the LHS, is just

$$\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)^2(\lambda_i - l)} \geq \left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)}\right)^2.$$

But by Chauchy-Schwartz:

$$\begin{aligned}
\left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)(\lambda_i - l)}\right)^2 &\leq \left(\delta_l \sum_i \frac{1}{\lambda_i - l}\right)\left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)^2(\lambda_i - l)}\right) \\
&\leq (\delta_l \epsilon_l)\left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)^2(\lambda_i - l)}\right) \\
&\quad \text{since } \Phi_l(A) \leq \epsilon_l \\
&\leq 1\left(\delta_l \sum_i \frac{1}{(\lambda_i - l - \delta_l)^2(\lambda_i - l)}\right) \\
&\quad \text{since } \frac{1}{\delta_l} - \epsilon_l \geq 0,
\end{aligned}$$

and so (3) is established. ∎

Proof of Theorem 1.1. All we need to do now is set $\epsilon_u, \epsilon_l, \delta_u$, and δ_l in a manner that satisfies Lemma 4.6 and gives a good bound on the condition number.

Then, we can take $A^{(0)} = 0$ and construct $A^{(s+1)}$ from $A^{(s)}$ by choosing any vector $\boldsymbol{\pi}_e$ with

$$L_{A^{(s)}}(\boldsymbol{\pi}_e) \geq U_{A^{(s)}}(\boldsymbol{\pi}_e)$$

(such a vector is guaranteed to exist by Lemma 4.6) and setting $A^{(s+1)} = A^{(s)} + t\boldsymbol{\pi}_e\boldsymbol{\pi}_e^T$ for any $t \geq 0$ satisfying:

$$L_{A^{(s)}}(\boldsymbol{\pi}_e) \geq \frac{1}{t} \geq U_{A^{(s)}}(\boldsymbol{\pi}_e).$$

It is sufficient to take

$$\begin{aligned} \delta_l &= 1 & \epsilon_l &= \frac{1}{\sqrt{d}} & l_0 &= -n/\epsilon_l \\ \delta_u &= \frac{\sqrt{d}+1}{\sqrt{d}-1} & \epsilon_u &= \frac{\sqrt{d}-1}{d+\sqrt{d}} & u_0 &= n/\epsilon_u. \end{aligned}$$

We can check that:

$$\begin{aligned} \frac{1}{\delta_u} + \epsilon_u &= \frac{\sqrt{d}-1}{\sqrt{d}+1} + \frac{\sqrt{d}-1}{\sqrt{d}(\sqrt{d}+1)} \\ &= 1 - \frac{1}{\sqrt{d}} \\ &= \frac{1}{\delta_l} - \epsilon_l \end{aligned}$$

so that (2) satisfied.

The initial potentials are $\Phi^{\frac{n}{\epsilon_u}}(0) = \epsilon_u$ and $\Phi_{\frac{n}{\epsilon_l}}(0) = \epsilon_l$. After dn steps, we have

$$\begin{aligned} \frac{\lambda_{\max}(A^{(dn)})}{\lambda_{\min}(A^{(dn)})} &\leq \frac{n/\epsilon_u + dn\delta_u}{-n/\epsilon_l + dn\delta_l} \\ &= \frac{\frac{d+\sqrt{d}}{\sqrt{d}-1} + d\frac{\sqrt{d}+1}{\sqrt{d}-1}}{d-\sqrt{d}} \\ &= \frac{d+2\sqrt{d}+1}{d-2\sqrt{d}+1}, \end{aligned}$$

as desired. $\square$

To turn this proof into an algorithm, one must first compute the vectors $\boldsymbol{\pi}_e$, which can be done in time $O(n^2m)$. For each iteration of the algorithm, we must compute $((u+\delta_u)I - A)^{-1}$, $((u+\delta_u)I - A)^{-2}$, and the same matrices for the lower potential function. This computation can be performed in time $O(n^3)$. Finally, we can decide which edge to add in each iteration by computing $U_A(\boldsymbol{\pi}_e)$ and $L_A(\boldsymbol{\pi}_e)$ for each edge, which can be done in time $O(n^2m)$. As we run for dn iterations, the total time of the algorithm is $O(dn^3m)$.

References

[BK96] András A. Benczúr and David R. Karger. Approximating s-t minimum cuts in $O(n^2)$ time. In *Proceedings of The Twenty-Eighth Annual ACM Symposium On The Theory Of Computing (STOC '96)*, pages 47–55, New York, USA, May 1996. ACM Press.

[GR01] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer, 2001.

[GRV08] Navin Goyal, Luis Rademacher, and Santosh Vempala. Expanders via random spanning trees. `http://arXiv.org/abs/0807.1496`, July 2008.

[GV96] G. H. Golub and C. F. Van Loan. *Matrix Computations, 3rd. Edition*. The Johns Hopkins University Press, Baltimore, MD, 1996.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Mar88] G. A. Margulis. Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, July 1988.

[SS08] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In Richard E. Ladner and Cynthia Dwork, editors, *STOC*, pages 563–568. ACM, 2008. Full version available at `http://arXiv.org/abs/0803.0929`.

[ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing (STOC-04)*, pages 81–90, New York, June 13–15 2004. ACM Press. Full version available at `http://arxiv.org/abs/cs.DS/0310051`.