Graph Sparsification by Effective Resistances*

Daniel A. Spielman
Program in Applied Mathematics and
Department of Computer Science
Yale University
New Haven, CT, USA
spielman@cs.yale.edu

Nikhil Srivastava
Department of Computer Science
Yale University
New Haven, CT, USA
nikhil.srivastava@yale.edu

ABSTRACT

We present a nearly-linear time algorithm that produces high-quality sparsifiers of weighted graphs. Given as input a weighted graph $G = (V, E, w)$ and a parameter $\epsilon > 0$, we produce a weighted subgraph $H = (V, \tilde{E}, \tilde{w})$ of G such that $|\tilde{E}| = O(n \log n/\epsilon^2)$ and for all vectors $x \in \mathbb{R}^V$

$$(1-\epsilon)\sum_{uv \in E}(x(u)-x(v))^2 w_{uv} \leq \sum_{uv \in \tilde{E}}(x(u)-x(v))^2 \tilde{w}_{uv} \leq (1+\epsilon)\sum_{uv \in E}(x(u)-x(v))^2 w_{uv}. \quad (1)$$

This improves upon the sparsifiers constructed by Spielman and Teng, which had $O(n \log^c n)$ edges for some large constant c, and upon those of Benczúr and Karger, which only satisfied (1) for $x \in \{0,1\}^V$. We conjecture the existence of sparsifiers with $O(n)$ edges, noting that these would generalize the notion of expander graphs, which are constant-degree sparsifiers for the complete graph.

A key ingredient in our algorithm is a subroutine of independent interest: a nearly-linear time algorithm that builds a data structure from which we can query the approximate effective resistance between any two vertices in a graph in $O(\log n)$ time.

Categories and Subject Descriptors

G.2.2 [**Discrete Mathematics**]: Graph Theory

General Terms

Algorithms, Theory

Keywords

Spectral graph theory, electrical flows, random sampling

1. INTRODUCTION

The goal of sparsification is to approximate a given graph G by a sparse graph H on the same set of vertices. If H is close to G in some appropriate metric, then H can be used as a proxy for G in computations without introducing too much error. At the same time, since H has very few edges, computation with and storage of H should be cheaper.

The notion of graph sparsification was introduced in [4] by Benczúr and Karger to accelerate cut algorithms whose running time depends on the number of edges. They gave a nearly-linear time procedure which takes a graph G on n vertices and $\epsilon > 0$ and outputs a weighted subgraph H with $O(n \log n/\epsilon^2)$ edges such that the weight of every cut in H is within a factor of $(1 \pm \epsilon)$ of its weight in G. This was used to turn Goldberg and Tarjan's $\tilde{O}(mn)$ max-flow algorithm [15] into an $\tilde{O}(n^2)$ algorithm for approximate *st*-mincut, and appeared more recently as the first step of an $\tilde{O}(n^{3/2}+m)$-time $O(\log^2 n)$ approximation algorithm for sparsest cut [19].

The cut-preserving guarantee of [4] is equivalent to satisfying (1) for all $x \in \{0,1\}^n$, which are the characteristic vectors of cuts. Spielman and Teng [23, 24] devised stronger sparsifiers which extend (1) to all $x \in \mathbb{R}^n$, but have $O(n \log^c n)$ edges for some large constant c. They used these sparsifiers to construct preconditioners for symmetric diagonally-dominant matrices, which led to the first nearly-linear time solvers for such systems of equations.

In this work, we construct sparsifiers that achieve same guarantee as Spielman and Teng's but with $O(n \log n/\epsilon^2)$ edges, thus improving on both [4] and [23]. Our sparsifiers are subgraphs of the original graph and can be computed in $\tilde{O}(m)$ time by random sampling, where the sampling probabilities are given by the effective resistances of the edges. While this is conceptually much simpler than the recursive partitioning approach of [23], we need to solve $O(\log n)$ linear systems to compute the effective resistances quickly, and we do this by Spielman and Teng's solver.

1.1 Our Results

Our main idea is to include each edge of G in the sparsifier H with probability proportional to its effective resistance. 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 (see, e.g., [8] or [5]), and was proven in [6] to be proportional to the commute time between the endpoints of the

*This material is based upon work supported by the National Science Foundation under Grants No. CCF-0707522 and 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
STOC'08, May 17–20, 2008, Victoria, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-047-0/08/05 ...$5.00.

edge. We show how to approximate the effective resistances of edges in G quickly and prove that sampling according to these approximate values yields a good sparsifier.

To define effective resistance, identify $G = (V, E, w)$ with an electrical network on n nodes in which each edge e corresponds to a link of conductance w_e (i.e., a resistor of resistance $1/w_e$). Then the effective resistance R_e across an edge e is the potential difference induced across it when a unit current is injected at one end and extracted at the other. Our algorithm can now be stated as follows:

$H = \textbf{Sparsify}(G, q)$
Choose a random edge e of G 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.

Recall that the *Laplacian* of a weighted graph is given by $L = D - A$ where A is the weighted adjacency matrix $(a_{ij}) = w_{ij}$ and D is the diagonal matrix of weighted degrees. Notice that the quadratic form associated with L is just $x^T Lx = \sum_{uv \in E}(x(u) - x(v))^2 w_{uv}$. Let L be the Laplacian of G and let $\tilde{L}$ be the Laplacian of H. Our main theorem is that if q is sufficiently large, then the quadratic forms of L and $\tilde{L}$ are close:

Theorem 1. *Suppose G and $H = \textbf{Sparsify}(G, q)$ have Laplacians L and $\tilde{L}$ respectively, and $0 < \epsilon \le 1$. If $q = 4C^2 n \log n/\epsilon^2$, where C is the constant in Lemma 5, then with probability at least $1/2$:*

$$\forall x \in \mathbb{R}^n \quad (1 - \epsilon)x^T Lx \le x^T \tilde{L}x \le (1 + \epsilon)x^T Lx. \qquad (2)$$

Sparsifiers that satisfy this condition preserve many properties of the graph. The Courant-Fischer Theorem tells us that

$$\lambda_i = \max_{S:\dim(S)=k} \min_{x \in S} \frac{x^T Lx}{x^T x}.$$

Thus, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of L and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ are the eigenvalues of $\tilde{L}$, then we have

$$(1 - \epsilon)\lambda_i \le \tilde{\lambda}_i \le (1 + \epsilon)\lambda_i,$$

and the eigenspaces spanned by corresponding eigenvalues are related. As the eigenvalues of the normalized Laplacian are given by

$$\lambda_i = \max_{S:\dim(S)=k} \min_{x \in S} \frac{x^T D^{-1/2} L D^{-1/2} x}{x^T x},$$

and are the same as the eigenvalues of the walk matrix $D^{-1}L$, we obtain the same relationship between the eigenvalues of the walk matrix of the original graph and its sparsifier. Many properties of graphs and random walks are known to be revealed by their spectra (see any text on spectral graph theory, e.g. [7] or [18] for details). The existence of sparse subgraphs which retain these properties is interesting its own right; indeed, expander graphs can be viewed as constant degree sparsifiers for the complete graph.

We remark that the condition (2) also implies

$$\forall x \in \mathbb{R}^n \quad \frac{1}{1+\epsilon} x^T L^+ x \le x^T \tilde{L}^+ x \le \frac{1}{1-\epsilon} x^T L^+ x,$$

where L^+ is the pseudoinverse of L. Thus sparsifiers also approximately preserve the effective resistances between vertices, since for vertices u and v, the effective resistance between them is given by the formula $(\chi_u - \chi_v)^T L^+ (\chi_u - \chi_v)$, where χ_u is the elementary unit vector with a coordinate 1 in position u.

We prove Theorem 1 in Section 3. In Section 4, we show how to compute approximate effective resistances in nearly-linear time, which is essentially optimal. The tools we use to do this are Spielman and Teng's nearly-linear time solver [23, 24] and the Johnson-Lindenstrauss Lemma [17, 1]. Specifically, we prove the following theorem, in which $R_{(u,v)}$ denotes the effective resistance between vertices u and v.

Theorem 2. *There is an $\tilde{O}(m/\epsilon^2)$ time algorithm which on input $G = (V, E, w)$ and $\epsilon > 0$ computes a $(24 \log n/\epsilon^2) \times n$ matrix Z such that with probability at least $1 - 1/n$:*

$$(1 - \epsilon)R_{(u,v)} \le \|Z(\chi_u - \chi_v)\|^2 \le (1 + \epsilon)R_{(u,v)}$$

for every pair of vertices $u, v \in V$.

Since Z can be applied to any $(\chi_u - \chi_v)$ simply by subtracting two of its columns, we can query the approximate effective resistance between any pair of vertices (u, v) in time $O(\log n)$, and for *all* the edges in time $O(m \log n)$.

In Section 5, we show that using the approximate resistances for sampling only changes the bound in Theorem 1 by a constant, yielding an $\tilde{O}(m)$ time sparsification algorithm (Corollary 7). In Section 6, we show that H is close to G in some additional ways that make it more useful for preconditioning systems of linear equations. We conclude with the conjecture that sparsifiers with $O(n)$ edges exist.

1.2 Prior Work

In addition to the graph sparsifiers of [4] and [23], there is a large body of work on sparse [3, 2] and low-rank [13, 2, 22, 9, 10] approximations for general matrices. The algorithms in this literature provide guarantees of the form $\|A - \tilde{A}\|_2 \le \epsilon$, where A is the original matrix and $\tilde{A}$ is obtained by entrywise or columnwise sampling of A. This is analogous to merely satisfying (1) for vectors x in the span of the dominant eigenvectors of A; thus, if we were to use these sparsifiers on graphs, they would only preserve the large cuts. Interestingly, our proof uses some of the same machinery as the low-rank approximation result of Rudelson and Vershynin [22] — the sampling of edges in our algorithm corresponds to picking $q = O(n \log n)$ columns at random from a certain rank $(n - 1)$ matrix of dimension $m \times m$.

The use of effective resistance as a distance in graphs has recently gained attention as it is often more useful than the ordinary geodesic distance in a graph. In small-world graphs, all vertices will be close to one another, but those with a smaller effective resistance distance are connected by more short paths. See, for example [12, 11], which use effective resistance/commute time as a distance measure in social network graphs.

2. PRELIMINARIES

2.1 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 $w_e \ge 0$. 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 the span of its columns by $\mathbb{B} = \text{im}(B) \subseteq \mathbb{R}^m$ (also called the *cut space* of G [14]). Note that $b_{(u,v)}^T = (\chi_v - \chi_u)$.

It is immediate that L is positive semidefinite since:

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

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

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

2.2 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 $\text{im}(L) = \ker(L)^\perp = \mathbb{R}^n \setminus \text{span}(\mathbf{1})$. We will rely on this fact heavily in the proof of Theorem 1.

2.3 Electrical Flows

We use the same notation as [16] to describe electrical flows on graphs: for a vector $\mathbf{i}_{\text{ext}}(u)$ of currents injected at the vertices, let $\mathbf{i}(e)$ be the currents induced in the edges (in the direction of orientation) and $\mathbf{v}(u)$ the potentials induced at the vertices. By Kirchoff's current law, the sum of the currents entering a vertex is equal to the amount injected:

$$B^T\mathbf{i} = \mathbf{i}_{\text{ext}}.$$

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

$$\mathbf{i} = WB\mathbf{v},$$

Combining these two facts, we obtain

$$\mathbf{i}_{\text{ext}} = B^T(WB\mathbf{v}) = L\mathbf{v}.$$

If $\mathbf{i}_{\text{ext}} \perp \mathbf{1} = \ker(L)$, then we can write

$$\mathbf{v} = L^+\mathbf{i}_{\text{ext}}$$

by the definition of L^+ in section 2.2.

Recall that the *effective resistance* between two vertices u and v is defined as the potential difference induced between them when a unit current is injected at one and extracted at the other. We will derive an algebraic expression for the effective resistance in terms of L^+. To inject and extract a unit current across the endpoints of an edge $e = (u,v)$, we set $\mathbf{i}_{\text{ext}} = b_e^T = (\chi_v - \chi_u)$, which is clearly orthogonal to $\mathbf{1}$. The potentials induced by this at the vertices are given by $\mathbf{v} = L^+b_e^T$; to measure the potential difference across $e = (u,v)$, we simply multiply by b_e on the left:

$$\mathbf{v}(v) - \mathbf{v}(u) = (\chi_v - \chi_u)^T\mathbf{v} = b_e L^+ b_e^T.$$

It follows that the effective resistance across e is given by $b_e L^+ b_e^T$ and that the matrix BL^+B^T has as its diagonal entries $BL^+B^T(e,e) = R_e$.

3. THE MAIN RESULT

We will prove Theorem 1. Consider the matrix $\Pi = W^{1/2}BL^+B^TW^{1/2}$. Since we know $BL^+B^T(e,e) = R_e$, the diagonal entries of Π are $\Pi(e,e) = \sqrt{W(e,e)}R_e\sqrt{W(e,e)} = w_eR_e$. Π has some notable properties.

LEMMA 3 (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)* $\Pi(e,e) = \|\Pi(\cdot,e)\|^2$.

PROOF. To see (i), observe that

$$\begin{aligned} \Pi^2 &= (W^{1/2}BL^+B^TW^{1/2})(W^{1/2}BL^+B^TW^{1/2}) \\ &= W^{1/2}BL^+(B^TWB)L^+B^TW^{1/2} \\ &= W^{1/2}BL^+LL^+B^TW^{1/2} \quad \text{since } L = B^TWB \\ &= W^{1/2}BL^+B^TW^{1/2} \\ &\qquad \text{since } L^+L \text{ is the identity on } \text{im}(L^+) \\ &= \Pi. \end{aligned}$$

For (ii), we have

$$\text{im}(\Pi) = \text{im}(W^{1/2}BL^+B^TW^{1/2}) \subseteq \text{im}(W^{1/2}B).$$

To see the other inclusion, assume $y \in \text{im}(W^{1/2}B)$. Then we can choose $x \perp \ker(W^{1/2}B) = \ker(L)$ such that $W^{1/2}Bx = y$. But now

$$\begin{aligned} \Pi y &= W^{1/2}BL^+B^TW^{1/2}W^{1/2}Bx \\ &= W^{1/2}BL^+Lx \quad \text{since } B^TWB = L \\ &= W^{1/2}Bx \quad \text{since } L^+Lx = x \text{ for } x \perp \ker(L) \\ &= y. \end{aligned}$$

Thus $y \in \text{im}(\Pi)$, as desired.

For (iii), recall from section 2.1 that $\dim(\ker(W^{1/2}B)) = 1$. Consequently, $\dim(\text{im}(\Pi)) = \dim(\text{im}(W^{1/2}B)) = n-1$. But since $\Pi^2 = \Pi$, the eigenvalues of Π are all 0 or 1, and as it projects onto a space of dimension $n-1$, there must be exactly $n-1$ nonzero eigenvalues.

(iv) follows from $\Pi^2(e,e) = \Pi(\cdot,e)^T\Pi(\cdot,e)$, since Π is symmetric. □

To show that $H = (V, \tilde{E}, \tilde{w})$ is a good sparsifier for G, we need to show that the quadratic forms x^TLx and $x^T\tilde{L}x$

are close. We start by reducing the problem of preserving $x^T Lx$ to that of preserving $y^T \Pi y$. This will be much nicer since the eigenvalues of Π are all 0 or 1, so that any matrix $\tilde{\Pi}$ which approximates Π in the spectral norm (i.e., makes $\|\tilde{\Pi} - \Pi\|_2$ small) also preserves its quadratic form.

Write the outcome of $H = \textbf{Sparsify}(G, q)$ as the following random matrix:

$$S(e,e) = \frac{\tilde{w_e}}{w_e} = \frac{(\#\text{ of times } e \text{ is sampled})}{qp_e}. \qquad (3)$$

$S_{m\times m}$ is a nonnegative diagonal matrix and the random entry $S(e,e)$ specifies the 'amount' of edge e included in H by **Sparsify**. For example $S(e,e) = 1/qp_e$ if e is sampled once, $2/qp_e$ if it is sampled twice, and zero if it is not sampled at all. The weight of e in H is now given by $\tilde{w_e} = S(e,e)w_e$, and we can write the Laplacian of H as:

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

since $\tilde{W} = WS = W^{1/2}SW^{1/2}$. The scaling of weights by $1/qp_e$ in **Sparsify** implies that $\mathbb{E}\tilde{w_e} = w_e$ (since q independent samples are taken, each with probability p_e), and thus $\mathbb{E}S = I$ and $\mathbb{E}\tilde{L} = L$.

We can now prove the following lemma, which says that if S does not distort $y^T \Pi y$ too much then $x^T Lx$ and $x^T \tilde{L} x$ are close.

LEMMA 4. *Suppose S is a nonnegative diagonal matrix such that*

$$\|\Pi S \Pi - \Pi\Pi\|_2 \le \epsilon.$$

Then

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

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

PROOF. The assumption is equivalent to

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

since $\|A\|_2 = \sup_{y\neq 0} |y^T Ay|/y^T y$ for symmetric A. Restricting our attention to vectors in $\mathrm{im}(W^{1/2}B)$, we have

$$\sup_{y\in \mathrm{im}(W^{1/2}B), y\neq 0} \frac{|y^T \Pi(S-I)\Pi y|}{y^T y} \le \epsilon.$$

But by Lemma 3.(ii), Π is the identity on $\mathrm{im}(W^{1/2}B)$ so $\Pi y = y$ for all $y \in \mathrm{im}(W^{1/2}B)$. Also, every such y can be written as $y = W^{1/2}Bx$ for $x \in \mathbb{R}^n$. Substituting this into the above expression we obtain:

$$\begin{aligned}
&\sup_{y\in \mathrm{im}(W^{1/2}B), y\neq 0} \frac{|y^T \Pi(S-I)\Pi y|}{y^T y} \\
&= \sup_{y\in \mathrm{im}(W^{1/2}B), y\neq 0} \frac{|y^T (S-I) y|}{y^T y} \\
&= \sup_{x\in \mathbb{R}^n, W^{1/2}Bx\neq 0} \frac{|x^T B^T W^{1/2} S W^{1/2} Bx - x^T B^T W B x|}{x^T B^T W B x} \\
&= \sup_{x\in \mathbb{R}^n, W^{1/2}Bx\neq 0} \frac{|x^T \tilde{L} x - x^T L x|}{x^T L x} \le \epsilon.
\end{aligned}$$

Rearranging yields the desired conclusion for all $x \notin \ker(W^{1/2}B)$. When $x \in \ker(W^{1/2}B)$ then $x^T Lx = x^T \tilde{L} x = 0$ and the claim holds trivially. □

To show that $\|\Pi S \Pi - \Pi\Pi\|_2$ is likely to be small we use the following concentration result, which is a sort of law of large numbers for symmetric rank 1 matrices. It was first proven by Rudelson in [21], but the version we state here appears in the more recent paper [22] by Rudelson and Vershynin.

LEMMA 5 (RUDELSON & VERSHYNIN, [22] THM. 3.1). *Let $\mathbf{p}$ be a probability distribution over $\Omega \subseteq \mathbb{R}^d$ such that $\sup_{y\in\Omega} \|y\|_2 \le M$ and $\|\mathbb{E}_{\mathbf{p}} yy^T\|_2 \le 1$. Let $y_1 \ldots y_q$ be independent samples drawn from $\mathbf{p}$. Then*

$$\mathbb{E}\left\| \frac{1}{q}\sum_{i=1}^q y_i y_i^T - \mathbb{E}yy^T \right\|_2 \le \min\left(CM\sqrt{\frac{\log q}{q}}, 1\right)$$

where C is an absolute constant.

We can now finish the proof of Theorem 1.

PROOF OF THEOREM 1. **Sparsify** samples edges from G independently with replacement, with probabilities p_e proportional to $w_e R_e$. Since $\sum_e w_e R_e = \mathrm{Tr}(\Pi) = n-1$ by Lemma 3.(iii), the actual probability distribution over E is given by $p_e = \frac{w_e R_e}{n-1}$. Sampling q edges from G corresponds to sampling q columns from Π, so we can write

$$\begin{aligned}
\Pi S \Pi &= \sum_e S(e,e)\Pi(\cdot,e)\Pi(\cdot,e)^T \\
&= \sum_e \frac{(\#\text{ of times } e \text{ is sampled})}{qp_e}\Pi(\cdot,e)\Pi(\cdot,e)^T \quad \text{by (3)} \\
&= \frac{1}{q}\sum_e (\#\text{ of times } e \text{ is sampled}))\frac{\Pi(\cdot,e)}{\sqrt{p_e}}\frac{\Pi(\cdot,e)^T}{\sqrt{p_e}} \\
&= \frac{1}{q}\sum_{i=1}^q y_i y_i^T
\end{aligned}$$

for vectors $y_1, \ldots, y_q$ drawn independently with replacement from the distribution

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

We can now apply Lemma 5. The expectation of yy^T is given by

$$\mathbb{E}yy^T = \sum_e p_e \frac{1}{p_e}\Pi(\cdot,e)\Pi(\cdot,e)^T = \Pi\Pi = \Pi,$$

so $\|\mathbb{E}yy^T\|_2 = \|\Pi\|_2 = 1$. We also have a bound on the norm of y:

$$\frac{1}{\sqrt{p_e}}\|\Pi(\cdot,e)\|_2 = \frac{1}{\sqrt{p_e}}\sqrt{\Pi(e,e)} = \sqrt{\frac{n-1}{R_e w_e}}\sqrt{R_e w_e} = \sqrt{n-1}.$$

Taking $q = 4C^2 n\log n/\epsilon^2$ gives:

$$\begin{aligned}
\mathbb{E}\,\|\Pi S\Pi - \Pi\Pi\|_2 &= \mathbb{E}\left\|\frac{1}{q}\sum_{i=1}^q y_i y_i^T - \mathbb{E}yy^T\right\|_2 \\
&\le C\sqrt{\epsilon^2 \frac{\log(4C^2 n\log n/\epsilon^2)(n-1)}{4C^2 n\log n}} \le \epsilon/2. \qquad (4)
\end{aligned}$$

By Markov's inequality, we have

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

with probability at least 1/2. By Lemma 4, this completes the proof of the theorem. □

4. COMPUTING APPROXIMATE RESISTANCES QUICKLY

It is not clear how to compute the effective resistances $\{R_e\}$ exactly and efficiently. In this section, we show that one can compute constant factor approximations to *all* the R_e in time $\tilde{O}(m)$. In fact, we do something stronger: we build a $O(\log n) \times n$ matrix Z from which the effective resistance between any two vertices (including vertices not connected by an edge) can be computed in $O(\log n)$ time.

Proof of Theorem 2. If u and v are vertices in G, then the effective resistance between u and v can be written as:

$$\begin{aligned} R_{(u,v)} &= (\chi_u - \chi_v)^T L^+ (\chi_u - \chi_v) \\ &= (\chi_u - \chi_v)^T L^+ L L^+ (\chi_u - \chi_v) \\ &= ((\chi_u - \chi_v)^T L^+ B^T W^{1/2})(W^{1/2} B L^+ (\chi_u - \chi_v)) \\ &= \|W^{1/2} B L^+ (\chi_u - \chi_v)^2\|_2^2. \end{aligned}$$

Thus effective resistances are just pairwise distances between vectors in $\{W^{1/2} B L^+ \chi_v\}_{v \in V}$. By the Johnson-Lindenstrauss Lemma, these distances are preserved if we project the vectors onto a subspace spanned by $O(\log n)$ random vectors. We recall the following version of the lemma, due to Achlioptas [1], which allows us to use random ± 1 vectors instead of the more expensive Gaussian ones.

Lemma 6 (Achlioptas, [1] Thm. 1.1). *Given fixed vectors $v_1 \dots v_n \in \mathbb{R}^d$ and $\epsilon > 0$, let $Q_{k \times d}$ be a random $\pm 1/\sqrt{k}$ matrix (i.e., independent Bernoulli entries) with $k \geq 24 \log n / \epsilon^2$. Then with probability at least $1 - 1/n$, we have:*

$$(1-\epsilon)\|v_i - v_j\|_2^2 \leq \|Qv_i - Qv_j\|_2^2 \leq (1+\epsilon)\|v_i - v_j\|_2^2$$

for all pairs $i, j \leq n$.

Our goal is now to compute the projections $\{Q W^{1/2} B L^+ \chi_v\}$ for some appropriate random Q. We construct the matrix $Z = Q W^{1/2} B L^+$ in steps, using both the JL lemma and the linear system solver of Spielman and Teng [23, 24]:

1. Let Q be a random $\pm 1/\sqrt{k}$ matrix of dimension $k \times n$ where $k = 24 \log n / \epsilon^2$.
2. Compute $Y = Q W^{1/2} B$. Note that this takes $2m \times 24 \log n / \epsilon^2 + m = \tilde{O}(m/\epsilon^2)$ time since B has $2m$ entries and $W^{1/2}$ is diagonal.
3. Find the rows of $Z = Y L^+$ by $k = O(\log n)$ calls to the Spielman-Teng solver – these are just solutions z to the systems $zL = y_i$ where y_i, for $1 \leq i \leq k$, are the rows of Y. Each call takes $\tilde{O}(m)$ time.

Thus the construction of Z takes $\tilde{O}(m/\epsilon^2)$ time. We can then find the approximate resistance $\|Z(\chi_u - \chi_v)\|^2 \approx R_{(u,v)}$ for any $u, v \in V$ in $O(\log n)$ time simply by subtracting two columns of Z and computing the norm of their difference. ☐

Using the above procedure, we can compute arbitrarily good approximations to the effective resistances $\{R_e\}$ which we need for sampling in nearly-linear time. It turns out that any constant factor approximation yields a sparsifier, so we are done.

Corollary 7. *Suppose Z_e are numbers satisfying $Z_e \geq R_e/\alpha$ and $\sum_e w_e Z_e \leq \alpha \sum_e w_e R_e$ for some $\alpha \geq 1$. If we sample as in* **Sparsify** *but take each edge with probability $p'_e = \frac{w_e Z_e}{\sum_e w_e Z_e}$ instead of $p_e = \frac{w_e R_e}{\sum_e w_e R_e}$, then H satisfies:*

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

with probability at least $1/2$.

Proof. We note that

$$p'_e = \frac{w_e S_e}{\sum_e w_e S_e} \geq \frac{w_e (R_e/\alpha)}{\alpha \sum_e w_e R_e} = \frac{p_e}{\alpha^2}$$

and proceed as in the proof of Theorem 1. The norm of the random vector y is now bounded by:

$$\frac{1}{\sqrt{p'_e}} \|\Pi(e,\cdot)\|_2 \leq \frac{\alpha}{\sqrt{p_e}} \sqrt{\Pi(e,e)} = \alpha\sqrt{n-1}$$

which introduces a factor of α into the final bound on the expectation, but changes nothing else. ☐

5. AN ADDITIONAL PROPERTY

Corollary 7 suggests that **Sparsify** is quite robust with respect to changes in the sampling probabilities p_e, and that we may be able to prove additional guarantees on H by tweaking them. In this section, we prove one such claim.

The following property is desirable for using H to solve linear systems (specifically, for the construction of *ultrasparsifiers* [23, 24], which we will not define here):

$$\text{For every vertex } v \in V, \quad \sum_{e \ni v} \frac{\tilde{w}_e}{w_e} \leq 2\deg(v). \tag{5}$$

This says, roughly, that not too many of the edges incident to any given vertex get blown up too much by sampling and rescaling. We show how to incorporate this property into our sparsifiers.

Lemma 8. *Suppose we sample the edges of G as in* **Sparsify** *with probabilities that satisfy*

$$p_{(u,v)} \geq \frac{\beta}{n \min(\deg(u), \deg(v))}$$

for some constant $0 < \beta < 1$. Then with probability at least $1 - 1/n$,

$$\sum_{e \ni v} \frac{\tilde{w}_e}{w_e} \leq 2\deg(v) \quad \textit{for all } v \in V.$$

Proof. Fix a vertex v and define i.i.d. random variables $X_1, \dots, X_q$ by:

$$X_i = \begin{cases} \frac{1}{p_e} & \text{with probability } p_e \text{, for each } e \ni v \\ 0 & \text{otherwise (i.e., with probability } 1 - \sum_{e \ni v} p_e) \end{cases}$$

so that X_i indicates which edge incident to v (if any) is chosen in the i^{th} sample of **Sparsify**. Let

$$D_v = \sum_{e \ni v} \frac{\tilde{w}_e}{w_e} = \sum_{e \ni v} \frac{(\#\text{ of times } e \text{ is sampled})}{q p_e} = \frac{1}{q} \sum_{i=1}^{q} X_i.$$

We want to show that with high probability, $D_v \leq 2\deg(v)$ for *all* vertices v. We begin by bounding the expectation and variance of each X_i:

$$\mathbb{E}X_i = \sum_{e \ni v} p_e \frac{1}{p_e} = \deg(v)$$

$$\begin{aligned}\mathbf{Var}(X_i) &= \sum_{e \ni v} p_e \left(\frac{1}{p_e^2} - \frac{1}{p_e} \right) \\ &\leq \sum_{e \ni v} \frac{1}{p_e} \\ &\leq \sum_{(u,v) \ni v} \frac{n \min(\deg(u), \deg(v))}{\beta} \quad \text{by assumption} \\ &\leq \sum_{(u,v) \ni v} \frac{n \deg(v)}{\beta} \\ &= \frac{n \deg(v)^2}{\beta}\end{aligned}$$

Since the X_i are independent, the variance of D_v is just

$$\mathbf{Var}(D_v) = \frac{1}{q^2} \sum_{i=1}^{q} \mathbf{Var}(X_i) \leq \frac{n \deg(v)^2}{\beta q}.$$

We now apply Bennett's inequality for sums of i.i.d. variables (see, e.g., [20]), which says

$$\mathbb{P}[|D_v - \mathbb{E}D_v| > \mathbb{E}D_v] \leq \exp\left(\frac{-(\mathbb{E}D_v)^2}{\mathbf{Var}(D_v)(1 + \frac{\mathbb{E}Z_v}{q})} \right)$$

We know that $\mathbb{E}D_v = \mathbb{E}X_i = \deg(v)$. Substituting our estimate for $\mathbf{Var}(D_v)$ and setting $q \geq 4n \log n / \beta$ gives:

$$\begin{aligned}\mathbb{P}[D_v > 2\deg(v)] &\leq \exp\left(\frac{-\deg(v)^2}{\frac{n \deg(v)^2}{\beta q}(1 + \frac{\deg(v)}{q})} \right) \\ &\leq \exp\left(\frac{-\beta q}{2n} \right) \quad \text{since } 1 + \frac{\deg(v)}{q} \leq 2 \\ &\leq \exp(-2 \log n) = 1/n^2.\end{aligned}$$

Taking a union bound over all v gives the desired result. □

Sampling with probabilities

$$p'_e = p'_{(u,v)} = \frac{1}{2} \left(\frac{\|Zb_e^T\|^2 w_e}{\sum_e \|Zb_e^T\|^2 w_e} + \frac{1}{n \min(\deg(u), \deg(v))} \right)$$

satisfies the requirements of both Corollary 7 (with $\alpha = 2$) and Lemma 8 (with $\beta = 1/2$) and yields a sparsifier with the desired property.

THEOREM 9. *There is an $\tilde{O}(m/\epsilon^2)$ time algorithm which on input $G = (V, E, w), \epsilon > 0$ produces a weighted subgraph $H = (V, \tilde{E}, \tilde{w})$ of G with $O(n \log n / \epsilon^2)$ edges which, with probability at least $1/2$, satisfies both (2) and (5).*

6. CONCLUSION

We conclude by making the conjecture that sparsifiers with $O(n)$ edges exist for general graphs. These would be a generalization of expander graphs, which are sparsifiers for the complete graph. It seems unlikely that such sparsifiers could be found by random sampling or in nearly-linear time, but even a polynomial-time algorithm would be of interest, and might, among other things, lead to a new construction of expanders.

7. REFERENCES

[1] D. Achlioptas. Database-friendly random projections. In *PODS '01*, pages 274–281, 2001.

[2] D. Achlioptas and F. McSherry. Fast computation of low rank matrix approximations. In *STOC '01*, pages 611–618, 2001.

[3] S. Arora, E. Hazan, and S. Kale. A fast random sampling algorithm for sparsifying matrices. In *APPROX-RANDOM '06*, volume 4110 of *Lecture Notes in Computer Science*, pages 272–279. Springer, 2006.

[4] A. A. Benczúr and D. R. Karger. Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time. In *STOC '96*, pages 47–55, 1996.

[5] B. Bollobas. *Modern Graph Theory.* Springer, July 1998.

[6] A. K. Chandra, P. Raghavan, W. L. Ruzzo, and R. Smolensky. The electrical resistance of a graph captures its commute and cover times. In *STOC '89*, pages 574–586, 1989.

[7] F. R. K. Chung. *Spectral Graph Theory.* CBMS Regional Conference Series in Mathematics. American Mathematical Society, 1997.

[8] P. Doyle and J. Snell. *Random walks and electric networks.* Math. Assoc. America., Washington, 1984.

[9] P. Drineas and R. Kannan. Fast monte-carlo algorithms for approximate matrix multiplication. In *FOCS '01*, pages 452–459, 2001.

[10] P. Drineas and R. Kannan. Pass efficient algorithms for approximating large matrices. In *SODA '03*, pages 223–232, 2003.

[11] A. Firat, S. Chatterjee, and M. Yilmaz. Genetic clustering of social networks using random walks. *Computational Statistics & Data Analysis*, 51(12):6285–6294, August 2007.

[12] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *Knowledge and Data Engineering, IEEE Transactions on*, 19(3):355–369, 2007.

[13] A. Frieze, R. Kannan, and S. Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM*, 51(6):1025–1041, 2004.

[14] C. Godsil and G. Royle. *Algebraic Graph Theory.* Graduate Texts in Mathematics. Springer-Verlag, 2001.

[15] A. V. Goldberg and R. E. Tarjan. A new approach to the maximum flow problem. In *STOC '86*, pages 136–146, 1986.

[16] S. Guattery and G. L. Miller. Graph embeddings and laplacian eigenvalues. *SIAM J. Matrix Anal. Appl.*, 21(3):703–723, 2000.

[17] W. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemp. Math.*, 26:189–206, 1984.

[18] J. Kelner. Lecture notes for 18.409, an algorithmist's toolkit. 2007.

[19] R. Khandekar, S. Rao, and U. Vazirani. Graph partitioning using single commodity flows. In *STOC '06*, pages 385–390, 2006.

[20] G. Lugosi. Concentration-of-measure inequalities, 2003. Available at http://www.econ.upf.edu/ lugosi/anu.ps.

[21] M. Rudelson. Random vectors in the isotropic position. *J. of Functional Analysis*, 163(1):60–72, 1999.

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

[23] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC '04*, pages 81–90, 2004. Full version available at http://arxiv.org/abs/cs.DS/0310051.

[24] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. Available at http://www.arxiv.org/abs/cs.NA/0607105, 2006.