Ranking and sparsifying a connection graph

Fan Chung Wenbo Zhao Mark Kempton
University of California, San Diego

July 16, 2014

Abstract

Many problems arising in dealing with high-dimensional data sets involve connection graphs in which each edge is associated with both an edge weight and a d-dimensional linear transformation. We consider vectorized versions of the PageRank and effective resistance which can be used as basic tools for organizing and analyzing complex data sets. For example, the generalized PageRank and effective resistance can be utilized to derive and modify diffusion distances for vector diffusion maps in data and image processing. Furthermore, the edge ranking of the connection graphs determined by the vectorized PageRank and effective resistance are an essential part of sparsification algorithms which simplify and preserve the global structure of connection graphs. In addition, we examine consistencies in a connection graph, particularly in the applications of recovering low-dimensional data sets and the reduction of noises. In these applications, we analyze the effect of deleting edges with high edge rank.

1 Introduction

In this paper, we consider a generalization of graphs, called connection graphs, in which each edge of the graph is associated with a weight and also a "rotation" (which is a linear orthogonal transformation acting on a d-dimensional vector space for some positive integer d). The adjacency matrix and the discrete Laplace operator are linear operators acting on the space of vector-valued functions (instead of the usual real-valued functions) and therefore can be represented by matrices of size $dn \times dn$ where n is the number of vertices in the graph.

Connection graphs arise in numerous applications, in particular for data and image processing involving high-dimensional data sets. To quantify the affinities between two data points, it is often not enough to use only a scalar edge weight. For example, if the high-dimensional data set can be represented or approximated by a low-dimensional manifold, the patterns associated with nearby data points are likely to related by certain rotations [33]. There are many recent developments of related research in cryo-electron microscopy [18, 32], angular synchronization of eigenvectors [14, 31] and vector diffusion maps [33]. In many areas of machine learning, high-dimensional data points in general can be treated by various methods, such as the Principle Component Analysis [22], to reduce vectors into some low-dimensional space and then use the connection graph with rotations on edges to provide the additional information for proximity. In computer vision, there has been a great deal of recent work dealing with trillions of photos that are now available on the web [2]. The feature matching techniques [28] can be used to derive vectors associated with the images. Then the information networks of photos can be built which are exactly connection graphs with rotations corresponding to the angles and positions of the cameras in use. The use of connection graphs can be further traced to earlier work in graph gauge theory for computing the vibrational spectra of molecules and examining the spins associated with vibrations [12].

Many information networks arising from massive data sets exhibit the small world phenomenon. Consequently the usual graph distance is no longer very useful. It is crucial to have the appropriate metric for expressing the proximity between two vertices. Previously, various notions of diffusion distances have been defined [33] and used for manifold learning and dimension reduction. Here we consider two basic notions, the *connection PageRank* and the *connection resistance*, (which are generalizations of the usual PageRank and effective resistance). Both the connection PageRank and connection resistance can then be used to measure relationships between vertices in the connection graph. To illustrate the usage of both metrics, we derive edge ranking using the connection PageRank and the connection resistance. In the applications to cryo-electron microscopy, the edge ranking can help eliminate the superfluous or erroneous edges that appear because of various "noises". We here will use the connection PageRank and the connection resistance as tools for the basis of algorithms that can be used to construct a *sparsifier* which has fewer edges but preserves the global structure of the connection network.

The notion of PageRank was first introduced by Brin and Page [9] in 1998 for Google's Web search algorithms. Although the PageRank was originally designed for the Web graph, the concepts work well for any graph for quantifying the relationships between pairs of vertices (or pairs of subsets) in any given graph. There are very efficient and robust algorithms for computing and approximating PageRank [3, 7, 21, 8]. In this paper, we further generalize the PageRank for connection graphs and give efficient and sharp approximation algorithms for computing the connection PageRank, similar to the algorithm presented in [8].

The effective resistance plays a major role in electrical network theory and can be traced back to the classical work of Kirchhoff [26]. Here we consider a generalized version of effective resistance for the connection graphs. To illustrate the usage of connection resistance, we examine a basic problem on graph sparsification. Graph sparsification was first introduced by Benczúr and Karger [6, 23, 24, 25] for approximately solving various network design problems. The heart of the graph sparsification algorithms is the sampling technique for randomly selecting edges. The goal is to approximate a given graph G on n vertices by a sparse graph $\tilde{G}$, called a sparsifier, with fewer edges on the same set of vertices such that every cut in the sparsifier $\tilde{G}$ has its size within a factor $(1 \pm \epsilon)$ of the size of the corresponding cut in G for some constant ϵ. Spielman and Teng [34] constructed a spectral sparsifier with $O(n \log^c n)$ edges for some large constant c. In [37], Spielman and Srivastava gave a different sampling scheme using the effective resistances to construct an improved spectral sparsifier with only $O(n \log n)$ edges. In this paper, we will construct the connection sparsifier using the weighted connection resistance. Our algorithm is similar to the one found in [37].

In recent work of Bandeira, Singer, and Spielman in [5], they study the $\mathsf{O}(d)$ synchronization problem in which each vertex of a connection graph is assigned a rotation in the orthogonal group $\mathsf{O}(d)$. Our work differs from theirs in that here we examine the problem of assigning a vector in $\mathbf{R}^d$ to each vertex, rather than an orthogonal matrix in $\mathsf{O}(d)$, (see the remark following the proof of Theorem 1). In other words, our connection Laplacian is an operator acting on the space of vector-valued functions. However, their work is closely related to our work in this paper. In particular, they define the connection Laplacian, and use its spectrum to give a measure of how close a connection graph is to being consistent.

A Summary of the Results

Our results can be summarized as follows:

- We review definitions for the connection graph and the connection Laplacian in Section 2. The connection Laplacian is also studied in [33, 5]. In particular, we discuss the notion of "consistency" in a connection graph (which is considered to be the ideal situation for various applications). We give a characterization for a consistent connection graph by using the eigenvalues of the connection Laplacian.
- We introduce the connection PageRank in Section 3. We follow the method of [8] to develop a sublinear time algorithm for computing an approximate connection PageRank vector.
- We define the connection resistance in Section 4 and then examine various properties of the connection resistance.
- We use the connection resistance to give an edge ranking algorithm and a sparsification algorithm for connection graphs in Section 5.
- In Section 6 we propose a method for reducing noise in data by deleting, with high probability, edges having large edge rank. Using probabilistic and spectral techniques as in [10], we prove that for a connection graph, the eigenvalue related to consistency can be substantially reduced by deleting edges with high rank. Consequently, the resulting graph is an improved approximation for recovering a consistent connection graph.

2 Preliminaries

For positive integers m, n and d, we consider a family of matrices, denoted by $\mathcal{F}(m, n, d; \mathbf{R})$ consisting of all $md \times nd$ matrices with real-valued entries. A matrix in $\mathcal{F}(m, n, d; \mathbf{R})$ can also be viewed as a $m \times n$ matrix whose entries are represented by $d \times d$ blocks. A *rotation* is a matrix that is used to perform a rotation in Euclidean space. Namely, a rotation O is a square matrix, with real entries, satisfying $O^T = O^{-1}$ and $\det(O) = 1$. The set of $d \times d$ rotation matrices form the special orthogonal group $\mathsf{SO}(d)$. It is easy to check that all eigenvalues of a rotation O are of norm 1. Furthermore, a rotation $O \in \mathsf{SO}(d)$ with d odd has an eigenvalue 1 (see [17]).

2.1 The Connection Laplacian

Suppose $G = (V, E, w)$ is an undirected graph with vertex set V, edge set E and edge weights $w_{uv} = w_{vu} > 0$ for edges (u, v) in E. Suppose each oriented edge (u, v) is associated with a rotation matrix $O_{uv} \in \mathsf{SO}(d)$ satisfying $O_{uv}O_{vu} = I_{d\times d}$. Let O denote the set of rotations associated with all oriented edges in G. The *connection graph*, denoted by $\mathbb{G} = (V, E, O, w)$, has G as the *underlying graph*. The *connection matrix* $\mathbb{A}$ of $\mathbb{G}$ is defined by:

$$\mathbb{A}(u, v) = \begin{cases} w_{uv}O_{uv} & \text{if } (u, v) \in E, \\ 0_{d\times d} & \text{if } (u, v) \notin E \end{cases}$$

where $0_{d\times d}$ is the zero matrix of size $d \times d$. In other words, for $|V| = n$, we view $\mathbb{A} \in \mathcal{F}(n, n, d; \mathbf{R})$ as a block matrix where each block is either a $d \times d$ rotation matrix O_{uv} multiplied by a scalar weight w_{uv}, or a $d \times d$ zero matrix. The matrix $\mathbb{A}$ is symmetric as $O_{uv}^T = O_{vu}$ and $w_{uv} = w_{vu}$. The diagonal matrix $\mathbb{D} \in \mathcal{F}(n, n, d; \mathbf{R})$ is defined by the diagonal blocks $\mathbb{D}(u, u) = d_u I_{d\times d}$ for $u \in V$. Here d_u is the weighted degree of u in G, i.e., $d_u = \sum_{(u,v)\in E} w_{uv}$.

The *connection Laplacian* $\mathbb{L} \in \mathcal{F}(n, n, d; \mathbf{R})$ of a graph $\mathbb{G}$ is the block matrix $\mathbb{L} = \mathbb{D} - \mathbb{A}$. Recall that for any orientation of edges of the underlying graph G on n vertices and m edges, the combinatorial Laplacian L can be written as $L = B^T W B$ where W is a $m \times m$ diagonal matrix with $W_{e,e} = w_e$, and B is the edge-vertex incident matrix of size $m \times n$ such that $B(e, v) = 1$ if v is e's head; $B(e, v) = -1$ if v is e's tail; and $B(e, v) = 0$ otherwise. A useful observation for the connection Laplacian is the fact that it can be written in a similar form. Let $\mathbb{B} \in \mathcal{F}(m, n, d; \mathbf{R})$ be the block matrix given by

$$\mathbb{B}(e, v) = \begin{cases} O_{uv} & v \text{ is } e\text{'s head}, \\ -I_{d\times d} & v \text{ is } e\text{'s tail}, \\ 0_{d\times d} & \text{otherwise}. \end{cases}$$

Also, let the block matrix $\mathbb{W} \in \mathcal{F}(m, m, d; \mathbf{R})$ denote a diagonal block matrix given by $\mathbb{W}(e, e) = w_e I_{d\times d}$. We remark that, given an orientation of the edges, the connection Laplacian also can alternatively be defined as

$$\mathbb{L} = \mathbb{B}^T \mathbb{W} \mathbb{B}.$$

This can be verified by direct computation.

We have the following useful lemma regarding the Dirichlet sum of the connection Laplacian as an operator on the space of vector-valued functions on the vertex set of a connection graph.

Lemma 1. *For any function $f : V \to \mathbf{R}^d$, we have*

$$f\mathbb{L}f^T = \sum_{(u,v)\in E} w_{uv} \|f(u)O_{uv} - f(v)\|_2^2 \tag{1}$$

where $f(v)$ here is regarded as a row vector of dimension d. Furthermore, an eigenpair $(\lambda_i, \phi_i$ has $\lambda_i = 0$ if and only if $\phi_i(u)O_{uv} = \phi_i(v)$ for all $(u, v) \in E$.

Proof. For equation 1, observe that for a fixed edge $e = (u, v)$,

$$f\mathbb{B}^T(e) = f(u)O_{uv} - f(v).$$

Thus,

$$\begin{aligned} f\mathbb{L}f^T &= (f\mathbb{B}^T)\mathbb{W}(\mathbb{B}f^T) \\ &= (f\mathbb{B}^T)\mathbb{W}(f\mathbb{B}^T)^T \\ &= \sum_{(u,v)\in E} w(u,v)\left\|f(u)O_{uv} - f(v)\right\|_2^2. \end{aligned}$$

Also, $\mathbb{L}$ is symmetric and therefore has real eigenfunctions and real eigenvalues. The spectral decompositions of $\mathbb{L}$ is given by

$$\mathbb{L}_{\mathbb{G}}(u,v) = \sum_{i=1}^{nd} \lambda_i \phi_i(u)^T \phi_i(v).$$

By Equation (1), $\lambda_1 \geq 0$ and $\lambda_i = 0$ if and only if $\phi_i(u)O_{uv} = \phi_i(v)$ for all $\{u,v\} \in E$ and the lemma follows. □

2.2 The Consistency of A Connection Graph

For a connection graph $\mathbb{G} = (V,E,O,w)$, we say $\mathbb{G}$ is *consistent* if for any cycle $c = (v_k, v_1, v_2, \ldots, v_k)$ the product of rotations along the cycle is the identity matrix, i.e. $O_{v_k v_1}\prod_{i=1}^{k-1} O_{v_i v_{i+1}} = I_{d\times d}$. In other words, for any two vertices u and v, the products of rotations along different paths from u to v are the same. In the following theorem, we give a characterization for a consistent connection graph by using the eigenvalues of the connection Laplacian.

Theorem 1. *Let $\mathbb{G}$ be a connected connection graph on n vertices having conncection Laplacian $\mathbb{L}$ of dimension nd, and let L be the Laplacian of the underlying graph G. The following statements are equivalent.*

- *(i) $\mathbb{G}$ is consistent.*
- *(ii) The connection Laplacian $\mathbb{L}$ of $\mathbb{G}$ has d eigenvalues of value 0.*
- *(iii) The eigenvalues of $\mathbb{L}$ are the n eigenvalues of L, each of multiplicity d.*
- *(iv) For each vertex u in G, we can find $O_u \in SO(d)$ such that for any edge (u,v) with rotation O_{uv}, we have $O_{uv} = O_u^{-1}O_v$.*

Proof. *(i)* $\Longrightarrow$ *(ii)*. For a fixed vertex $u \in V$ and an arbitrary d-dimensional vector $\widehat{x}$, we can define a function $\widehat{f} : V \to \mathbf{R}^d$, by defining $\widehat{f}(u) = \widehat{x}$ initially. Then we assign $\widehat{f}(v) = \widehat{f}(u)O_{uv}$ for all the neighbors v of u. Since G is connected and $\mathbb{G}$ is consistent, we can continue the assigning process to all neighboring vertices without any conflict until all vertices are assigned. The resulting function $\widehat{f} : V \to \mathbf{R}^d$ satisfies $\widehat{f}\mathbb{L}\widehat{f}^T = \sum_{(u,v)\in E} w_{uv}\left\|\widehat{f}(u)O_{uv} - \widehat{f}(v)\right\|_2^2 = 0$. Therefore 0 is an eigenvalue of $\mathbb{L}$ with eigenfunction $\widehat{f}$. There are d orthogonal choices for the initial choice of $\widehat{x} = \widehat{f}(u)$. Therefore we obtain d orthogonal eigenfunctions $\widehat{f}_1, \ldots, \widehat{f}_d$ corresponding to the eigenvalue 0.

(ii) $\Longrightarrow$ *(iii)*. Let us consider the underlying graph G. Let $f_i : V \to \mathbf{R}$ denote the eigenfunctions of L corresponding to the eigenvalue λ_i for $i \in [n]$ respectively. Let $\widehat{f}_k$, for $k \in [d]$, be orthogonal eigenfunctions of $\mathbb{L}$ for the eigenvalue 0. By Lemma 1, each $\widehat{f}_k$ satisfies $\widehat{f}_k(u)O_{uv} = \widehat{f}_k(v)$. Our proof of this part follows directly from the following claim.

Claim. *Functions $f_i \otimes \widehat{f}_k : V \to \mathbf{R}^d$ for $i \in [n], k \in [d]$ are the orthogonal eigenfunctions of $\mathbb{L}$ corresponding to eigenvalue λ_i where $f_i \otimes \widehat{f}_k(v) = f_i(v)\widehat{f}_k(v)$.*

Proof. First, we need to verify that functions $f_i \otimes \widehat{f}_k$ are eigenfunctions of $\mathbb{L}$. We note that

$$\begin{aligned}
[f_i \otimes \widehat{f}_k \mathbb{L}](u) &= d(u) f_i \otimes \widehat{f}_k(u) - \sum_{v\sim u} w_{vu} f_i \otimes \widehat{f}_k(v) O_{vu} \\
&= d(u) f_i(u) \widehat{f}_k(u) - \sum_{v\sim u} w_{vu} f_i(v) \widehat{f}_k(v) O_{vu} \\
&= d(u) f_i(u) \widehat{f}_k(u) - \sum_{v\sim u} w_{vu} f_i(v) \widehat{f}_k(u) \\
&= \left(d(u) f_i(u) - \sum_{v\sim u} w_{vu} f_i(v) \right) \widehat{f}_k(u).
\end{aligned}$$

Since f_i is an eigenfunction of L corresponding to the eigenvalue λ_i, we have $f_i L = \lambda_i f_i$, i.e.

$$\left(d(u) f_i(u) - \sum_{v\sim u} w_{vu} f_i(v) \right) = \lambda_i f_i(u).$$

Thus,

$$[f_i \otimes \widehat{f}_k \mathbb{L}](u) = \lambda_i f_i(u) \widehat{f}_k(u) = \lambda_i f_i \otimes \widehat{f}_k(u)$$

and $f_i \otimes \widehat{f}_k, 1 \le i \le n, 1 \le k \le d$ are the eigenfunctions of $\mathbb{L}$ with eigenvalue λ_i.

To prove the orthogonality of $f_i \otimes \widehat{f}_k$'s, we note that if $k \neq l$,

$$\begin{aligned}
\langle f_i \otimes \widehat{f}_k, f_j \otimes \widehat{f}_l \rangle &= \sum_v \langle f_i \otimes \widehat{f}_k(v), f_j \otimes \widehat{f}_l(v) \rangle \\
&= \sum_v f_i(v) f_j(v) \langle \widehat{f}_k(v), \widehat{f}_l(v) \rangle \\
&= 0
\end{aligned}$$

since $\langle \widehat{f}_k(v), \widehat{f}_l(v) \rangle = 0$ for $k \neq l$. For the case of $k = l$ but $i \neq j$, we have

$$\begin{aligned}
\langle f_i \otimes \widehat{f}_k, f_j \otimes \widehat{f}_l \rangle &= \sum_v f_i(v) f_j(v) \langle \widehat{f}_k(v), \widehat{f}_k(v) \rangle \\
&= \sum_v f_i(v) f_j(v) \\
&= 0
\end{aligned}$$

because of $\langle f_i, f_j \rangle = 0$ for $i \neq j$. The claim is proved. □

(iii) $\Longrightarrow$ *(iv)*. Since 0 is an eigenvalue of L, we can let $\widehat{f}_1, \ldots, \widehat{f}_d$ be d orthogonal eigenfunctions of $\mathbb{L}$ corresponding to the eigenvalue 0. By Lemma 1, $\widehat{f}_k(u) O_{uv} = \widehat{f}_k(v)$ for all $k \in [d]$, $uv \in E$. For two adjacent vertices u and v, we have, for $i, j = 1, ..., d$,

$$\langle \widehat{f}_i(u), \widehat{f}_j(u) \rangle = \langle \widehat{f}_i(u) O_{uv}, \widehat{f}_j(u) O_{uv} \rangle = \langle \widehat{f}_i(v), \widehat{f}_j(v) \rangle$$

Therefore, $\widehat{f}_1(v), \ldots, \widehat{f}_d(v)$ must form an orthogonal basis of $\mathbf{R}^d$ for all $v \in V$. So for $v \in V$, define O_v to be the matrix with rows $\widehat{f}_1(v), ..., \widehat{f}_d(v)$, and if necessary normalize and adjust the signs of these vectors to guarantee that $O_v \in \mathsf{SO}(d)$. Then O_v is an orthogonal matrix for each d, and for an edge $uv \in E$, $O_u O_{uv} = O_v$, which implies $O_{uv} = O_u^{-1} O_v$.

(iv) $\Longrightarrow$ *(i)*. Let $C = (v_1, v_2, ..., v_k, v_1)$ be a cycle in G. Then

$$O_{v_k v_1} \prod_{i=1}^{k-1} O_{v_i v_{i+1}} = O_{v_k}^{-1} O_{v_1} \prod_{i=1}^{k-1} O_{v_i}^{-1} O_{v_{i+1}} = I_{d\times d}.$$

Therefore $\mathbb{G}$ is consistent. This completes the proof of the theorem. □

We note that item (iv) in the previous result is related to the $\mathsf{O}(d)$ synchronization problem studied by Bandeira, Singer, and Spielman in [5]. This problem consists of finding a function $O : V(G) \to \mathsf{O}(d)$ such that given the offsets O_{uv} in the edges, the function satisfies $O_{uv} = O_u^{-1} O_v$. The previous theorem shows that this has an exact solution if $\mathbb{G}$ is consistent. Particularly, [5] investigates how well a solution can be approximated even when the connection graph is not consistent. Their formulation gives a measure of how close a connection graph is to being consistent by looking at the operator on the space of functions $O : V(G) \longrightarrow \mathsf{O}(d)$ given by $\sum_{u \sim v} w_{uv} ||O_u O_{uv} - O_v||_2^2$. In order to investigate this, they also consider the operator on the space of vector valued functions $f : V(G) \longrightarrow \mathbb{R}^d$ given by $\sum_{u \sim v} w_{uv} ||f_u O_{uv} - f_v||_2^2$, which is what we are using to investigate the connection Laplacian.

2.3 Random walks on a connection graph

Consider the underlying graph G of a connection graph $\mathbb{G} = (V, E, O, w)$. A random walk on G is defined by the transition probability matrix P where $P_{uv} = w_{uv}/d_u$ denotes the probability of moving to a neighbor v at a vertex u. We can write $P = D^{-1}A$, where A is the weighted adjacency matrix of G and D is the diagonal matrix of weighted degree.

In a similar way, we can define a random walk on the connection graph $\mathbb{G}$ by setting the transition probability matrix $\mathbb{P} = \mathbb{D}^{-1}\mathbb{A}$. While P acts on the space of real-valued functions, $\mathbb{P}$ acts on the space of vector-valued functions $f : V \to \mathbf{R}^d$.

Theorem 2. *Suppose $\mathbb{G}$ is consistent. Then for any positive integer t, any vertex $u \in V$ and any function $\widehat{s} : V \to \mathbf{R}^d$ satisfying $\widehat{s}(v) = 0$ for all $v \in V \backslash \{u\}$, we have $\|\widehat{s}(u)\|_2 = \sum_v \|\widehat{s}\, \mathbb{P}^t(v)\|_2$.*

Proof. The proof of this theorem is straightforward from the assumption that $\mathbb{G}$ is consistent. For $\widehat{p} = \widehat{s}\, \mathbb{P}^t$, note that $\widehat{p}(v)$ is the summation of all d dimensional vectors resulted from rotating $\widehat{s}(u)$ via rotations along all possible paths of length t from u to v. Since $\mathbb{G}$ is consistent, the rotated vectors arrive at v via different paths are positive multiples of the same vector. Also the rotations maintain the 2-norm of vectors. Thus, $\frac{\|\widehat{p}(v)\|_2}{\|\widehat{s}(u)\|_2}$ is simply the probability that a random walk in G arriving at v from u after t steps. The theorem follows. □

3 PageRank Vectors in a Connection Graph

The PageRank vector is based on random walks. Here we consider a lazy walk on G with the transition probability matrix $Z = \frac{I+P}{2}$. In [3], a PageRank vector $\mathrm{pr}_{\alpha,s}$ is defined by a recurrence relation involving a seed vector s (as a probability distribution) and a positive jumping constant $\alpha < 1$ (or transportation constant). Namely, $\mathrm{pr}_{\alpha,s} = \alpha s + \mathrm{pr}_{\alpha,s}(1-\alpha)Z$.

For the connection graph $\mathbb{G}$, the PageRank vector $\widehat{\mathrm{pr}}_{\alpha,\widehat{s}} : V \to \mathbf{R}^d$ is defined by the same recurrence relation involving a seed vector $\widehat{s} : V \to \mathbf{R}^d$ and a positive jumping constant $\alpha < 1$:

$$\widehat{\mathrm{pr}}_{\alpha,\widehat{s}} = \alpha \widehat{s} + (1-\alpha)\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}\mathbb{Z}$$

where $\mathbb{Z} = \frac{1}{2}\left(I_{nd \times nd} + \mathbb{P}\right)$ is the transition probability matrix of a lazy random walk on $\mathbb{G}$. An alternative definition of the PageRank vector is the following geometric sum of random walks:

$$\widehat{\mathrm{pr}}_{\alpha,\widehat{s}} = \alpha \sum_{t=0}^{\infty} (1-\alpha)^t\, \widehat{s}\, \mathbb{Z}^t = \alpha \widehat{s} + (1-\alpha)\widehat{\mathrm{pr}}_{\alpha,\widehat{s}\mathbb{Z}}. \tag{2}$$

By Theorem 2 and Equation (2), we here state the following useful fact concerning PageRank vectors for a consistent connection graph.

Proposition 1. *Suppose that a connection graph $\mathbb{G}$ is consistent. Then for any $u \in V$, $\alpha \in (0,1)$ and any function $\widehat{s} : V \to \mathbf{R}^d$ satisfying $\|\widehat{s}(u)\|_2 = 1$ and $\widehat{s}(v) = 0$ for $v \neq u$, we have $\left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(v)\right\|_2 = \mathrm{pr}_{\alpha,\chi_u}(v)$. Here, $\chi_u : V \to \mathbb{R}$ denotes the characteristic function for the vertex u, so $\chi_u(v) = 1$ for $v = u$, and $\chi_u(v) = 0$ otherwise. In particular, $\sum_{v \in V} \left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(v)\right\|_2 = \|\mathrm{pr}_{\alpha,\chi_u}\|_1 = 1$.*

Proof. Since function $\widehat{s}$ satisfies $\|\widehat{s}(u)\|_2 = 1$ and $\widehat{s}(v) = 0$ for $v \neq u$, by Theorem 2, for a fixed $v \in V$, $[\widehat{s}\mathbb{Z}^t](v)$ are all equal to each other for all $t > 0$. By the geometric sum expression of PageRank vector, we have

$$\begin{aligned} \left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(v)\right\|_2 &= \left\|\alpha \sum_{t=0}^{\infty} (1-\alpha)^t \, [\widehat{s}\mathbb{Z}^t](v)\right\|_2 \\ &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \left\|[\widehat{s}\mathbb{Z}^t](v)\right\|_2 \\ &= \alpha \sum_{t=0}^{\infty} (1-\alpha)^t \, [\chi_u Z^t](v) \\ &= \mathrm{pr}_{\alpha,\chi_u}(v). \end{aligned}$$

Thus,

$$\sum_{v \in V} \left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(v)\right\|_2 = \|\mathrm{pr}_{\alpha,\chi_u}\|_1 = 1.$$

□

We will call such a PageRank vector $\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}$ a connection PageRank vector on u.

We next examine the problem of efficiently computing connection PageRank vectors. For graphs, an efficient sublinear algorithm is given in [8], in which PageRank vectors are approximated by realizing random walks of some bounded length. We here develop a version of their algorithm to apply to connection graphs. Our proof follows the template of their analysis, but uses the connection random walk.

$\widehat{p} =$ ApproximatePR$(v, \widehat{s}, \alpha, \epsilon, \rho)$

1. Initialze $\widehat{p} = 0$ and set $k = \log_{\frac{1}{1-\alpha}}(\frac{4}{\epsilon})$ and $r = \frac{1}{\epsilon\rho^2} 32 d \log(n\sqrt{d})$.
2. For r times do:
   a. Run one realization of the lazy random walk on $\mathbb{G}$ starting at the vertex v : At each step, with probablilty α, take a 'termination' step by returning to v and terminating, and with probability $1-\alpha$, randomly choose among the neighbors of the current vertex. At each step in the random walk, rotate $\widehat{s}(v)$ by the rotation matrix along the edge. The walk is artificially stopped after k steps if it has not terminated already.
   b. If the walk visited a node u just before making a termination step, then set $\widehat{p}(u) = \widehat{p}(u) + \widehat{s}(v) \prod_{i=1}^{j} O_{v_i v_{i+1}}$, where $(v = v_1, v_2, ..., v_{j-1}, v_j = u)$ is the path taken in the random walk.
3. Replace $\widehat{p}$ with $\frac{1}{r}\widehat{p}$.
4. Return $\widehat{p}$.

For our analysis of the algorithm, we will need the following well known concentration inequalities.

Lemma 2. *(Multiplicative Chernoff Bounds) Let X_i be i.i.d. Bernoulli random variable with expectation μ each. Define $X = \sum_{i=1}^{n} X_i$. Then*

- *For* $0 < \lambda < 1$, $\Pr(X < (1-\lambda)\mu n) < \exp(-\mu n \lambda^2/2)$.
- *For* $0 < \lambda < 1$, $\Pr(X > (1+\lambda)\mu n) < \exp(-\mu n \lambda^2/4)$.
- *For* $\lambda \geq 1$, $\Pr(X > (1+\lambda)\mu n) < \exp(-\mu n \lambda/2)$.

Theorem 3. *Let $\mathbb{G} = (V, E, O, w)$ be a connection graph and fix a vertex $v \in V$. Let $0 < \epsilon < 1$ be an additive error parameter, $0 < \rho < 1$ a multiplicative approximation parameter, and $0 < \alpha < 1$ a teleportation probablilty. Let $\widehat{s} : V \to \mathbf{R}^d$ be a function satisfying $||\widehat{s}(v)||_2 = 1$ and $\widehat{s}(u) = 0$ for $u \neq v$. Then with probablility at least $1 - \Theta\left(\frac{1}{n^2}\right)$, the algorithm ApproximatePR produces a vector $\widehat{p}$ that satisfies*

$$\left\|\widehat{p}(u) - \widehat{\text{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 < \rho \left\|\widehat{\text{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 + \frac{\epsilon}{4}$$

for vertices u of V for which $\left\|\widehat{\text{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 \geq \frac{\epsilon}{4}$, and satisfying $\|\widehat{p}(u)\|_2 < \frac{\epsilon}{2}$ for vertices u for which $\left\|\widehat{\text{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 \leq \frac{\epsilon}{4}$. The running time of the algorithm is $O\left(\frac{d^3 \log(n\sqrt{d}) \log(1/\epsilon)}{\epsilon\rho^2 \log(1/(1-\alpha))}\right)$.

Proof. We have from Equation 2 that

$$\widehat{\text{pr}}_{\alpha,\widehat{s}} = \alpha \widehat{s} \sum_{t=0}^{\infty} (1-\alpha)^t \mathbb{Z}^t.$$

We observe the the tth term in this sum is the contribution to the PageRank vector given by the walks of length t. We will approximate this by looking at walks of length at most k. Define

$$\widehat{p}_{\alpha,\widehat{s}}^{(k)} = \alpha \widehat{s} \sum_{t=0}^{k} (1-\alpha)^t \mathbb{Z}^t.$$

We then observe that by choosing k large enough so that $(1-\alpha)^k < \frac{\epsilon}{4}$, we have $\left\|\widehat{\text{pr}}_{\alpha,\widehat{s}} - \widehat{p}_{\alpha,\widehat{s}}^{(k)}\right\|_2 < \frac{\epsilon}{4}$. The choice of $k = \log_{\frac{1}{1-\alpha}}\left(\frac{4}{\epsilon}\right)$ will guarantee this.

The output of the algorithm $\widehat{p}$ gives an approximation to $\widehat{p}_{\alpha,\widehat{s}}^{(k)}$ by realizing walks of length at most k. The algorithm does so by taking the average count over $\frac{1}{\epsilon\rho^2} 32d \log(n\sqrt{d})$ trials. Note that $\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)$ is the expected value of the contribution of an instance of the random walk of length k. We will take an arbitrary entry of $\widehat{p}(u)$, say $\widehat{p}(u)(j)$, and compare it to $\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j)$. Assuming that for at least one j we have $\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j) > \epsilon/4d$, then we get by the multaplicative Chernoff bound that

$$\Pr\left(\widehat{p}(u)(j) < (1+\rho)\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j)\right) < \exp(-2\log(n\sqrt{d}))$$

and

$$\Pr\left(\widehat{p}(u)(j) < (1-\rho)\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j)\right) < \exp(-2\log(n\sqrt{d})).$$

which implies

$$\Pr\left(|\widehat{p}(u)(j) - \widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j)| > \rho \widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j)\right) < 2\exp(-2\log(n\sqrt{d})).$$

Note that this difference will be the same for all the entries of $\widehat{p}(u)$, therefore,

$$\Pr\left(\left\|\widehat{p}(u) - \widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)\right\|_2 > \rho \left\|\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)\right\|_2\right) < 2d\exp(-2\log(n\sqrt{d})) = \frac{2}{n^2}.$$

In a similar manner, if $\widehat{p}_{\alpha,\widehat{s}}^{(k)}(u)(j) \leq \frac{\epsilon}{4d}$ then by the Chernoff bound $\Pr\left(\widehat{p}(u)(j) > \frac{\epsilon}{2d}\right) < \exp\left(-2\log(n\sqrt{d})\right)$, so $\Pr\left(\|\widehat{p}(u)\|_2 > \frac{\epsilon}{2}\right) < d\exp\left(-2\log(n\sqrt{d})\right) = \frac{1}{n^2}$.

For the running time, note that the algorithm performs $\frac{1}{\epsilon\rho^2} 32d\log(n\sqrt{d})$ rounds, where each round simulates a walk of length at most $\log_{\frac{1}{1-\alpha}}\left(\frac{4}{\epsilon}\right)$, where each walk multiplies $\widehat{s}(v)$ by the $d \times d$ rotation matrices. Thus the running time is $O\left(\frac{d^3 \log(n\sqrt{d}) \log(1/\epsilon)}{\epsilon\rho^2 \log(1/(1-\alpha))}\right)$. □

4 The Connection Resistance

Motivated by the definition of effective resistance in electrical network theory, we consider the following block matrix $\Psi = \mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T \in \mathcal{F}(m, m, d; \mathbf{R})$ where $\mathbb{L}^+$ is the pseudo-inverse of $\mathbb{L}$. Note that for a matrix M, the *pseudo-inverse* of M is defined as the unique matrix M^+ satisfying the following four criteria [17, 29]: *(i)* $MM^+M = M$; *(ii)* $M^+MM^+ = M^+$; *(iii)* $(MM^+)^* = (MM^+)$; and *(iv)* $(M^+M)^* = M^+M$.

We define the connection resistance $\mathbb{R}_{\mathsf{eff}}(e)$ as $\mathbb{R}_{\mathsf{eff}}(v,u) = \|\Psi(e,e)\|_2$. Note that block $\Psi(e,e)$ is a $d \times d$ matrix. We will show that in the case that the connection graph $\mathbb{G}$ is consistent $\mathbb{R}_{\mathsf{eff}}(u,v)$ is reduced to the usual effective resistance $R_{\mathsf{eff}}(u,v)$ of the underlying graph G. In general, if the connection graph is not consistent, the connection resistance is not necessarily equal to its effective resistance in the underlying graph G.

Our first observation is the following Lemma.

Lemma 3. *Suppose $\mathbb{G}$ is a consistent connection graph, where the underlying graph is connected. For two vertices u, v of $\mathbb{G}$, let $p_{uv} = (v_1 = u, v_2, ..., v_k = v)$ be any path from u to v in $\mathbb{G}$. Define $O_{p_{uv}} = \prod_{j=1}^{k-1} O_{v_j v_{j+1}}$. Let $\mathbb{L}$ be the connection Laplacian of $\mathbb{G}$ and L be the discrete Laplacian of G respectively. Then*

$$\mathbb{L}^+(u,v) = \begin{cases} L^+(u,v)O_{p_{uv}} & i \neq j, \\ L^+(u,v)I_{d\times d} & i = j. \end{cases}$$

Proof. We first note that the matrix $O_{p_{uv}}$ is well-defined since G is consistent. Also note that if u and v are adjacent, then $O_{p_{uv}} = O_{uv}$. Also observe that for $\mathbb{L}(u,v) = L(u,v)O_{p_{uv}}$ since if uv is not an edge, $L(u,v) = 0$, and if u,v is an edge, $O_{p_{uv}} = O_{uv}$. To verify $\mathbb{L}^+$ is the pseudoinverse of $\mathbb{L}$, we just need to verify that $\mathbb{L}^+(u,v)$ satisfies all of the four criteria above.

To see *(i)* $\mathbb{L}\mathbb{L}^+\mathbb{L} = \mathbb{L}$, we consider two vertices u and v and note that

$$\begin{aligned}(\mathbb{L}\mathbb{L}^+\mathbb{L})(u,v) &= \sum_{x,y} \mathbb{L}(u,x)\mathbb{L}^+(x,y)\mathbb{L}(y,v) \\ &= \sum_{x,y} L(u,x)L^+(x,y)L(y,v)O_{p_{ux}}O_{p_{xy}}O_{p_{yv}} \\ &= \sum_{x,y} L(u,x)L^+(x,y)L(y,v)O_{p_{uv}}\end{aligned}$$

where the last equality follows by consistency. Since L^+ is the pseudoinverse of L, we also have $LL^+L = L$ which implies that

$$L(u,v) = \sum_{x,y} L(u,x)L^+(x,y)L(y,v).$$

Thus,

$$(\mathbb{L}\mathbb{L}^+\mathbb{L})(u,v) = L(u,v)O_{p_{uv}} = \mathbb{L}(u,v)$$

and the verification of *(i)* is completed.

The verification of *(ii)* is quite similar to that of *(i)*, and we omit it here.

To see *(iii)* $(\mathbb{L}\mathbb{L}^+)^* = (\mathbb{L}\mathbb{L}^+)$, we also consider two fixed vertices v_i and v_j. Note that

$$\begin{aligned}(\mathbb{L}\mathbb{L}^+)(u,v) &= \sum_{x} \mathbb{L}(u,x)\mathbb{L}^+(x,v) \\ &= \sum_{x} L(u,x)L^+(x,v)O_{p_{ux}}O_{p_{xv}} \\ &= \sum_{x} L(u,x)L^+(x,v)O_{p_{uv}}.\end{aligned}$$

On the other side,

$$\begin{aligned}(\mathbb{L}\mathbb{L}^+)(v,u) &= \sum_x L(v,x)L^+(x,u)O_{p_{vu}} \\ &= \sum_x L(v,x)L^+(x,u)O^T_{p_{uv}}.\end{aligned}$$

Since L^+ is the pseudoinverse of L, we also have $(LL^+)^* = LL^+$ which implies that

$$\sum_x L(u,x)L^+(x,v) = \sum_x L(v,x)L^+(x,u)$$

and thus $(\mathbb{L}\mathbb{L}^+)^* = (\mathbb{L}\mathbb{L}^+)$.

The verification of *(iv)* $(\mathbb{L}^+\mathbb{L})^* = \mathbb{L}^+\mathbb{L}$ is also similar to *(iii)*, we omit here. For all above, the lemma follows. □

By using the above lemma, we examine the relation between the connection resistance and the effective resistance for a consistent connection graph by the following theorem.

Theorem 4. *Suppose $\mathbb{G} = (V, E, O, w)$ is a consistent connection graph whose underlying graph G is connected. Then for any edge $(u, v) \in \mathbb{G}$, we have*

$$\mathbb{R}_{\mathit{eff}}(u,v) = R_{\mathit{eff}}(u,v).$$

Proof. Let $\mathbb{L}$ be the connection Laplacian of $\mathbb{G}$ and L the Laplacian of the underlying graph G. Let us fix an edge $e = (u, v) \in \mathbb{G}$. By the definition of effective resistance, $R_{\mathsf{eff}}(u, v)$ is the maximum eigenvalue of the following matrix

$$\Psi(e,e) = \begin{bmatrix} O_{vu} & -I_{d\times d}\end{bmatrix}\begin{bmatrix}\mathbb{L}^+(u,u) & \mathbb{L}^+(u,v)\\ \mathbb{L}^+(v,u) & \mathbb{L}^+(v,v)\end{bmatrix}\begin{bmatrix}O_{uv}\\ -I_{d\times d}\end{bmatrix}$$

where O_{uv} is the rotation from u to v. By Lemma 3, we have

$$\begin{aligned}\mathbb{L}^+(u,u) &= L^+(u,u)I_{d\times d},\\ \mathbb{L}^+(u,v) &= L^+(u,v)O_{p_{uv}},\\ \mathbb{L}^+(v,v) &= L^+(v,v)I_{d\times d},\\ \mathbb{L}^+(v,u) &= L^+(v,u)O_{p_{vu}} = L^+(u,v)O_{p_{vu}}.\end{aligned}$$

Thus, by the definition of matrix Ψ,

$$\Psi(e,e) = \left(L^+_{u,u} + L^+_{v,v}\right) I_{d\times d} - L^+_{u,v}\left(O_{p_{vu}}O_{uv} + O_{vu}O_{p_{uv}}\right).$$

Note that $O_{p_{uv}}O_{vu} = O_{uv}O^T_{uv} = I$ and similarly $O_{vu}O_{p_{vu}} = I$, so

$$\Psi(e,e) = \left(L^+(u,u) + L^+(v,v) - 2L^+(u,v)\right) I_{d\times d}.$$

Note that $(L^+(u,u) + L^+(v,v) - 2L^+(u,v))$ is exactly the effective resistance of e, so

$$\|\Psi(e,e)\|_2 = L^+(u,u) + L^+(v,v) - 2L^+(u,v) = R_{\mathsf{eff}}(u,v).$$

Thus, the theorem is proved. □

5 Ranking edges by using the connection Resistance

A central part of a graph sparsification algorithm is the sampling technique for selecting edges. It is crucial to choose the appropriate probabilistic distribution which can lead to a sparsifier preserving *every* cut in the original graph. In [37], the measure of how well the sparsifier preserves the cuts is given according to how well the sparsifier preserves the spectral properties of the original graph. We follow the template of [37] to present a sampling algorithm that will accomplish this. The following algorithm Sample is a generic sampling algorithm for a graph sparsification problem. We will sample edges using the distribution proportional to the weighted connection resistances.

$(\widetilde{\mathbb{G}} = (V, \widetilde{E}, O, \widetilde{w})) = \mathsf{Sample}(\mathbb{G} = (V, E, O, w), p', q)$

1. For every edge $e \in E$, set p_e proportional to p'_e.
2. Choose a random edge e of $\mathbb{G}$ with probability p_e, and add e to $\widetilde{\mathbb{G}}$ with edge weight $\widetilde{w}_e = \frac{w_e}{qp_e}$. Take q samples independently with replacement, summing weights if an edge is chosen more than once.
3. Return $\widetilde{\mathbb{G}}$.

Theorem 5. *For a given connection graph $\mathbb{G}$ and some positive $\xi > 0$, we consider $\widetilde{\mathbb{G}} = \mathsf{Sample}(G, p', q)$, where $p'_e = w_e \mathbb{R}_{\mathsf{eff}}(e)$ and $q = \frac{4nd(\log(nd)+\log(1/\xi))}{\epsilon^2}$. Suppose $\mathbb{G}$ and $\widetilde{\mathbb{G}}$ have connection Laplacian $\mathbb{L}_{\mathbb{G}}$ and $\mathbb{L}_{\widetilde{\mathbb{G}}}$ respectively. Then with probability at least $1-\xi$, for any function $f : V \to \mathbf{R}^d$, we have*

$$(1-\epsilon) f \mathbb{L}_{\mathbb{G}} f^T \leq f \mathbb{L}_{\widetilde{\mathbb{G}}} f^T \leq (1+\epsilon) f \mathbb{L}_{\mathbb{G}} f^T. \tag{3}$$

Before proving Theorem 5, we need the following two lemmas, in particular concerning the matrix $\Lambda = \mathbb{W}^{1/2}\mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T\mathbb{W}^{1/2}$.

Lemma 4. *The matrix Λ is a projection matrix, i.e. $\Lambda^2 = \Lambda$.*

Proof. Observe that

$$\begin{aligned} \Lambda^2 &= (\mathbb{W}^{\frac{1}{2}}\mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T\mathbb{W}^{1/2})(\mathbb{W}^{1/2}\mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T\mathbb{W}^{1/2}) \\ &= \mathbb{W}^{1/2}\mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{L}_{\mathbb{G}}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T\mathbb{W}^{1/2} \\ &= \mathbb{W}^{1/2}\mathbb{B}\mathbb{L}_{\mathbb{G}}^{+}\mathbb{B}^T\mathbb{W}^{1/2} \\ &= \Lambda. \end{aligned}$$

Thus, the lemma follows. □

To show that $\widetilde{\mathbb{G}} = (V, \widetilde{E}, O, \widetilde{w})$ is a good sparsifier for $\mathbb{G}$ satisfying (3), we need to show that the quadratic forms $f\mathbb{L}_{\widetilde{\mathbb{G}}}f^T$ and $f\mathbb{L}_{\mathbb{G}}f^T$ are close. By applying similar methods as in [37], we reduce the problem of preserving $f\mathbb{L}_{\mathbb{G}}f^T$ to that of $g\Lambda g^T$ for some function g. We consider a diagonal matrix $\mathbb{S} \in \mathcal{F}(m, m, d; \mathbf{R})$, where the diagonal blocks are scalar matrices given by $\mathbb{S}(e,e) = \frac{\widetilde{w}_e}{w_e} I_{d\times d} = \frac{N_e}{qp_e} I_{d\times d}$ and N_e is the number of times an edge e is sampled.

Lemma 5. *Suppose $\mathbb{S}$ is a nonnegative diagonal matrix such that $\|\Lambda\mathbb{S}\Lambda - \Lambda\Lambda\|_2 \leq \epsilon$. Then, $\forall f : V \to \mathbf{R}^d, (1-\epsilon) f\mathbb{L}_{\mathbb{G}}f^T \leq f\mathbb{L}_{\widetilde{\mathbb{G}}}f^T \leq (1+\epsilon) f\mathbb{L}_{\mathbb{G}}f^T$, where $\mathbb{L}_{\widetilde{\mathbb{G}}} = \mathbb{B}^T\mathbb{W}^{1/2}\mathbb{S}\mathbb{W}^{1/2}\mathbb{B}$.*

Proof. The assumption is equivalent to

$$\sup_{f\in\mathbb{R}^{md}, f\neq 0} \frac{\left|f\Lambda(\mathbb{S}-I)\Lambda f^T\right|}{ff^T} \leq \epsilon$$

Restricting our attention to vectors in $\mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right)$,

$$\sup_{f\in\mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right), f\neq 0} \frac{\left|f\Lambda(\mathbb{S}-I)\Lambda f^T\right|}{ff^T} \leq \epsilon$$

Since Λ is the identity on $\mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right)$, $f\Lambda = f$ for all $f \in \mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right)$. Also, every such f can be written as $f = g\mathbb{B}^T\mathbb{W}^{1/2}$ for $g \in \mathbb{R}^{nd}$. Thus,

$$\begin{aligned}
& \sup_{f\in \mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right), f\neq 0} \frac{\left|f\Lambda(\mathbb{S}-I)\Lambda f^T\right|}{ff^T} \\
= & \sup_{f\in \mathsf{im}\left(\mathbb{B}^T\mathbb{W}^{1/2}\right), f\neq 0} \frac{\left|f(\mathbb{S}-I)f^T\right|}{ff^T} \\
= & \sup_{g\in\mathbb{R}^{nd}, g\mathbb{B}^T\mathbb{W}^{1/2}\neq 0} \frac{\left|g\mathbb{B}^T\mathbb{W}^{1/2}\mathbb{S}\mathbb{W}^{1/2}\mathbb{B}g^T - g\mathbb{B}^T\mathbb{W}\mathbb{B}g^T\right|}{g\mathbb{B}^T\mathbb{W}\mathbb{B}g^T} \\
= & \sup_{g\in\mathbb{R}^{nd}, g\mathbb{B}^T\mathbb{W}^{1/2}\neq 0} \frac{\left|g\mathbb{L}_{\widetilde{\mathbb{G}}}g^T - g\mathbb{L}_{\mathbb{G}}g^T\right|}{g\mathbb{L}_{\mathbb{G}}g^T} \leq \epsilon
\end{aligned}$$

Rearranging yields the desired conclusion for all $g \in \mathbb{R}^{nd}$. □

We also require the following concentration inequality in order to prove our main theorems. Previously, various matrix concentration inequalities have been derived by many authors including Achiloptas [1], Cristofies-Markström [13], Recht [30], and Tropp [38]. Here we will use the simple version that is proved in [39].

Theorem 6. *Let $X_1, X_2, \ldots, X_q$ be independent symmetric random $k \times k$ matrices with zero means, $S_q = \sum_i X_i$, $\|X_i\|_2 \leq 1$ for all i a.s. Then for every $t > 0$ we have*

$$\Pr\left[\|S_q\|_2 > t\right] \leq k \max\left(\exp\left(-\frac{t^2}{4\sum_i \|\mathit{Var}(X_i)\|_2}\right), \exp\left(-\frac{t}{2}\right)\right).$$

A direct consequence of Theorem 6 is the following corollary.

Corollary 1. *Suppose $X_1, X_2, \ldots, X_q$ are independent random symmetric $k \times k$ matrices satisfying*

1. *for all $1 \leq i \leq q$, $\|X_i\|_2 \leq M$ a.s.,*
2. *for all $1 \leq i \leq q$, $\|\mathit{Var}(X_i)\|_2 \leq M\, \|\mathbf{E}[X_i]\|_2$.*

Then for any $\epsilon \in (0,1)$ we have

$$\Pr\left[\left\|\sum_i X_i - \sum_i \mathbf{E}[X_i]\right\|_2 > \epsilon \sum_i \|\mathbf{E}[X_i]\|_2\right] \leq k \exp\left(-\frac{\epsilon^2 \sum_i \|\mathbf{E}[X_i]\|_2}{4M}\right).$$

Proof. Let us consider the following independent random symmetric matrices

$$\frac{X_i - \mathbf{E}[X_i]}{M}$$

for $1 \leq i \leq q$. Clearly they are independent symmetric random $k \times k$ matrices with zero means satisfying

$$\left\|\frac{X_i - \mathbf{E}[X_i]}{M}\right\|_2 \leq 1$$

for $1 \leq i \leq q$. Also we note that

$$\mathsf{Var}\left(\frac{X_i - \mathbf{E}[X_i]}{M}\right) = \mathsf{Var}\left(\frac{X_i}{M}\right) = \frac{\mathsf{Var}(X_i)}{M^2}.$$

Thus, by applying the Theorem 6 we have

$$\begin{aligned}
& \Pr\left[\left\|\frac{\sum_i X_i - \mathbf{E}[X_i]}{M}\right\|_2 > t\right] \\
= & \Pr\left[\left\|\sum_i X_i - \sum_i \mathbf{E}[X_i]\right\|_2 > tM\right] \\
\leq & k \max\left(\exp\left(-\frac{t^2M^2}{4\sum_i \|\mathsf{Var}(X_i)\|_2}\right), \exp\left(-\frac{t}{2}\right)\right). \qquad (4)
\end{aligned}$$

Note that by condition (2) we obtain

$$\sum_i \|\mathsf{Var}(X_i)\|_2 \leq M \sum_i \|\mathbf{E}[X_i]\|_2 .$$

Thus if we set

$$t = \frac{\epsilon \sum_i \|\mathbf{E}[X_i]\|_2}{M},$$

the left term in the right hand side of Equation (4) can be bounded as follows.

$$\begin{aligned} \frac{t^2 M^2}{4\sum_{i=1}^q \|\mathsf{Var}(X_i)\|_2} &\geq \frac{\left(\epsilon \sum_{i=1}^q \|\mathbf{E}[X_i]\|_2\right)^2}{4M\sum_{i=1}^q \|\mathbf{E}[X_i]\|_2} \\ &= \frac{\epsilon^2 \sum_{i=1}^q \|\mathbf{E}[X_i]\|_2}{4M}. \end{aligned}$$

Thus, the corollary follows. □

Proof (of Theorem 5). Our algorithm samples edges from $\mathbb{G}$ independently with replacement, with probabilities p_e proportional to $w_e \mathbb{R}_{\mathsf{eff}}(e)$. Note that sampling q edges from $\mathbb{G}$ corresponds to sampling q columns from Λ. So we can write

$$\Lambda \mathbb{S} \Lambda = \sum_e \Lambda(\cdot, e)\mathbb{S}(e,e)\Lambda(\cdot,e)^T = \sum_e \frac{N_e}{q p_e}\Lambda(\cdot,e)\Lambda(\cdot,e)^T = \frac{1}{q}\sum_{i=1}^q y_i y_i^T$$

for block matrices $y_1, \ldots, y_q \in R^{nd\times d}$ drawn independently with replacements from the distribution $y = \frac{1}{\sqrt{p_e}}\Lambda(\cdot,e)$ with probability p_e. Now, we can apply Corollary 1. The expectation of yy^T is given by $\mathbf{E}\left[yy^T\right] = \sum_e p_e \frac{1}{p_e}\Lambda(\cdot,e)\Lambda(\cdot,e)^T = \Lambda$ which implies that $\left\|\mathbf{E}\left[yy^T\right]\right\|_2 = \|\Lambda\|_2 = 1$. We also have a bound on the norm of $y_i y_i^T$: $\left\|y_i y_i^T\right\|_2 \leq \max_e \left(\frac{\left\|\Lambda(\cdot,e)^T\Lambda(\cdot,e)\right\|_2}{p_e}\right) = \max_e \left(\frac{w_e \mathbb{R}_{\mathsf{eff}}(e)}{p_e}\right)$. Since the probability p_e is proportional to $w_e \mathbb{R}_{\mathsf{eff}}(e)$, i.e. $p_e = \frac{w_e \mathbb{R}_{\mathsf{eff}}(e)}{\sum_e w_e \mathbb{R}_{\mathsf{eff}}(e)} = \frac{\|\Lambda(e,e)\|_2}{\sum_e \|\Lambda(e,e)\|_2}$, we have $\left\|y_i y_i^T\right\|_2 \leq \sum_e \|\Lambda(e,e)\|_2 \leq \sum_e \mathrm{Tr}\left(\Lambda(e,e)\right) = \mathrm{Tr}\left(\Lambda\right) \leq nd$. To bound the variance observe that

$$\begin{aligned} \left\|\mathsf{Var}\left(yy^T\right)\right\|_2 &= \left\|\mathbf{E}\left[yy^T yy^T\right] - \left(\mathbf{E}\left[yy^T\right]\right)^2\right\|_2 \\ &\leq \left\|\mathbf{E}\left[yy^T yy^T\right]\right\|_2 + \left\|\left(\mathbf{E}\left[yy^T\right]\right)^2\right\|_2 . \end{aligned}$$

Since the second term of the right hand of above inequality can be bounded by

$$\begin{aligned} \left\|\left(\mathbf{E}\left[yy^T\right]\right)^2\right\|_2 &= \left\|\Lambda^2\right\|_2 \text{ (as property (1))} \\ &= \|\Lambda\|_2 \\ &= 1, \end{aligned}$$

it is sufficient to bound the term $\left\|\mathbf{E}\left[yy^T yy^T\right]\right\|_2$. By the definition of expectation, we observe that

$$\begin{aligned} \left\|\mathbf{E}\left[yy^T yy^T\right]\right\|_2 &= \left\|\sum_e p_e \frac{1}{p_e^2}\Lambda(\cdot,e)\Lambda(\cdot,e)^T\Lambda(\cdot,e)\Lambda(\cdot,e)^T\right\|_2 \\ &= \left\|\sum_e \frac{1}{p_e}\Lambda(\cdot,e)\Lambda(e,e)\Lambda(\cdot,e)^T\right\|_2 . \end{aligned}$$

This implies that

$$\begin{aligned}
&\left\|\mathbf{E}\left[yy^Tyy^T\right]\right\|_2 \\
= \quad & \max_{f\in \mathsf{im}\left(\mathbb{W}^{1/2}\mathbb{B}\right)} \sum_e \frac{1}{p_e}\frac{f^T\Lambda(\cdot,e)\Lambda(e,e)\Lambda(\cdot,e)^Tf}{f^Tf} \\
= \quad & \max_{f\in \mathsf{im}\left(\mathbb{W}^{1/2}\mathbb{B}\right)} \sum_e \frac{1}{p_e}\frac{f^T\Lambda(\cdot,e)\Lambda(e,e)\Lambda(\cdot,e)^Tf}{f^T\Lambda(\cdot,e)\Lambda(\cdot,e)^Tf}\frac{f^T\Lambda(\cdot,e)\Lambda(\cdot,e)^Tf}{f^Tf} \\
\leq \quad & \max_{f\in \mathsf{im}\left(\mathbb{W}^{1/2}\mathbb{B}\right)} \sum_e \frac{\|\Lambda(e,e)\|_2}{p_e}\frac{f^T\Lambda(\cdot,e)\Lambda(\cdot,e)^Tf}{f^Tf}.
\end{aligned}$$

Recall that the probability p_e is proportional to $w_e\mathbb{R}_{\mathsf{eff}}(e)$, i.e.

$$p_e = \frac{w_e\mathbb{R}_{\mathsf{eff}}(e)}{\sum_e w_e\mathbb{R}_{\mathsf{eff}}(e)} = \frac{\|\Lambda(e,e)\|_2}{\sum_e \|\Lambda(e,e)\|_2},$$

we have

$$\begin{aligned}
\left\|\mathbf{E}\left[yy^Tyy^T\right]\right\|_2 \quad & \leq \quad \sum_e \|\Lambda(e,e)\|_2 \left(\max_{f\in \mathsf{im}\left(\mathbb{W}^{1/2}\mathbb{B}\right)} \sum_e \frac{f^T\Lambda(\cdot,e)\Lambda(\cdot,e)^Tf}{f^Tf}\right) \\
& = \quad \sum_e \|\Lambda(e,e)\|_2 \|\Lambda\|_2 \\
& = \quad \sum_e \|\Lambda(e,e)\|_2 \\
& \leq \quad \sum_e \mathrm{Tr}\left(\Lambda(e,e)\right) \\
& = \quad \mathrm{Tr}\left(\Lambda\right) \\
& \leq \quad nd.
\end{aligned}$$

Thus,

$$\left\|\mathsf{Var}\left(yy^T\right)\right\|_2 \quad \leq \quad nd+1 \leq 2nd\left\|\mathbf{E}\left[yy^T\right]\right\|_2.$$

To complete the proof, by setting $q = \frac{4nd(\log(nd)+\log(1/\xi))}{\epsilon^2}$ and the fact that dimension of yy^T is nd, we have

$$\begin{aligned}
\Pr\left[\left\|\frac{1}{q}\sum_{i=1}^q y_iy_i^T - \mathbf{E}\left[yy^T\right]\right\|_2 > \epsilon\right] \quad & \leq \quad nd\ \exp\left(-\frac{\epsilon^2\sum_{i=1}^q\left\|\mathbf{E}\left[y_iy_i^T\right]\right\|_2}{4nd}\right) \\
& \leq \quad nd\ \exp\left(-\frac{\epsilon^2 q}{4nd}\right) \leq \xi
\end{aligned}$$

for some constant $0<\xi<1$. Thus, the theorem follows. □

In [27], a modification of the algorithm from [37] is presented. The oversampling Theorem in [27] can further be modified for connection graphs and stated as follows.

Theorem 7 (Oversampling)**.** *For a given connection graph $\mathbb{G}$ and some positive $\xi>0$, we consider $\widetilde{\mathbb{G}} = \mathsf{Sample}(G,p',q)$, where $p'_e = w_e\mathbb{R}_{\mathsf{eff}}(e), t = \sum_{e\in E}p'_e$ and $q = \frac{4t(\log(t)+\log(1/\xi))}{\epsilon^2}$. Suppose $\mathbb{G}$ and $\widetilde{\mathbb{G}}$ have connection Laplacian $\mathbb{L}_{\mathbb{G}}$ and $\mathbb{L}_{\widetilde{\mathbb{G}}}$ respectively. Then with probability at least $1-\xi$, for all $f:V\to\mathbf{R}^d$, we have $(1-\epsilon)f\mathbb{L}_{\mathbb{G}}f^T \leq f\mathbb{L}_{\widetilde{\mathbb{G}}}f^T \leq (1+\epsilon)f\mathbb{L}_{\mathbb{G}}f^T$.*

Proof. In the proof of Theorem 5, the key is the bound on the norm $\left\|y_iy_i^T\right\|_2$. If $p'_e \geq w_e\mathbb{R}_{\mathsf{eff}}(e)$, the norm $\left\|y_iy_i^T\right\|_2$ is bounded by $\sum_{e\in E}p'_e$. Thus, the theorem follows. □

Now let us consider a variation of the connection resistance denoted by $\overline{\mathbb{R}_{\mathsf{eff}}}(e) = \mathrm{Tr}\,(\Psi(e,e))$. Clearly, we have $\overline{\mathbb{R}_{\mathsf{eff}}}(e) = \mathrm{Tr}\,(\Psi(e,e)) \geq \|\Psi(e,e)\|_2 = \mathbb{R}_{\mathsf{eff}}(e)$ and $\sum_e w_e \overline{\mathbb{R}_{\mathsf{eff}}}(e) = \sum_e \mathrm{Tr}\,(\Lambda(e,e)) = \mathrm{Tr}\,(\Lambda) \leq nd$. Using Theorem 7, we have the following.

Corollary 2. *For a given connection graph $\mathbb{G}$ and some positive $\xi > 0$, we consider $\widetilde{\mathbb{G}} = \mathsf{Sample}(G, p', q)$, where $p'_e = w_e \overline{\mathbb{R}_{\mathsf{eff}}}(e)$ and $q = \frac{4nd(\log(nd)+\log(1/\xi))}{\epsilon^2}$. Suppose $\mathbb{G}$ and $\widetilde{\mathbb{G}} = \mathsf{Sample}(G, p', q)$ have connection Laplacian $\mathbb{L}_{\mathbb{G}}$ and $\mathbb{L}_{\widetilde{\mathbb{G}}}$ respectively. Then with probability at least 1-ξ, for all $f : V \to \mathbf{R}^d$, we have $(1-\epsilon) f \mathbb{L}_{\mathbb{G}} f^T \leq f \mathbb{L}_{\widetilde{\mathbb{G}}} f^T \leq (1+\epsilon) f \mathbb{L}_{\mathbb{G}} f^T$.*

We note that edge ranking can be accomplished using the quantities known as *Green's values*, which generalize the notion of effective resistance by allowing a damping constant. An edge ranking algorithm for graphs using Green's values was studied extensively in [11]. Here we will define a generalization of Green's values for connection graphs.

For $i = 0, ..., nd-1$, let $\widehat{\phi}_i$ be the ith eigenfunction of the *normalized* connection Laplacian $\mathbb{D}^{-1/2}\mathbb{L}\mathbb{D}^{-1/2}$ corresponding to eigenvalue λ_i. Define

$$\mathbb{G}_\beta = \sum_{i=0}^{nd-1} \frac{1}{\lambda_i + \beta} \widehat{\phi}_i^T \widehat{\phi}_i.$$

We remark that $\mathbb{G}_\beta$ can be viewed as a generalization of the pseudo-inverse of the normalized connection Laplacian. Define the *PageRank vector with a jumping constant* α as the solution to the equation

$$\widehat{\mathrm{pr}}_{\beta,\widehat{s}} = \frac{\beta}{2+\beta}\widehat{s} + \frac{2}{2+\beta}\widehat{\mathrm{pr}}_{\beta,\widehat{s}}\mathbb{Z}.$$

with $\beta = 2\alpha/(1-\alpha)$. These PageRank vectors are related to the matrix $\mathbb{G}_\beta$ via the following formula that is straightforward to check,

$$\frac{\widehat{\mathrm{pr}}_{\beta,\widehat{s}}}{\beta} = s\mathbb{D}^{-1/2}\mathbb{G}_\beta\mathbb{D}^{1/2}.$$

Now for each edge $e = \{u, v\} \in E$, we define the *connection Green's value* $\widehat{g}_\beta(u, v)$ of e to be the following combination of PageRank vectors:

$$\begin{aligned}\widehat{g}_\beta(u,v) &= \beta(\chi_u - \chi_v)\mathbb{D}^{-1/2}\mathbb{G}_\beta\mathbb{D}^{-1/2}(\chi_u - \chi_v)^T \\ &= \frac{\widehat{\mathrm{pr}}_{\beta,\chi_u}(u)}{d_u} - \frac{\widehat{\mathrm{pr}}_{\beta,\chi_u}(v)}{d_v} + \frac{\widehat{\mathrm{pr}}_{\beta,\chi_v}(v)}{d_v} - \frac{\widehat{\mathrm{pr}}_{\beta,\chi_v}(u)}{d_u}.\end{aligned}$$

This gives an alternative to the effective resistance as a technique for ranking edges. It could be used in place of the effective resistance in the edge sparsification algorithm.

6 Eliminating noise in data sets by deleting edges of high rank

In forming a connection graph, the possibility arises of there being erroneous data or errors in measurments, or other forms of "noise." This may be manifested in a resulting connection graph that is not consistent, where it is expected that it would be. It is therefore desirable to be able to identify edges whose rotations are causing the connection graph to be inconsistent. We propose that a possible solution to this problem is to randomly delete edges of high rank in the sense of the edge ranking. In this section we will obtain bounds on the eigenvalues of the connection Laplacian resulting from the deletion of edges of high rank. This will have the effect of reducing the smallest eigenvalue, thus making the connection graph "closer" to being consistent, as seen in Theorem 1.

To begin, we will derive a result on the spectrum of the connection Laplacian analogous to the result of Chung and Radcliffe in [10] on the adjacency matrix of a random graph.

Theorem 8. *Let $\mathbb{G}$ be a given fixed connection graph with Laplacian $\mathbb{L}$. Delete edges $ij \in E(\mathbb{G})$ with probability p_{ij}. Let $\widehat{\mathbb{G}}$ be the resulting connection graph, and $\widehat{\mathbb{L}}$ its connection Laplacian, and $\overline{\mathbb{L}} = E(\widehat{\mathbb{L}})$. Then for $\epsilon \in (0,1)$, with probability at least $1-\epsilon$*

$$|\lambda_i(\widehat{\mathbb{L}}) - \lambda_i(\overline{\mathbb{L}})| \leq \sqrt{6\Delta \ln(2nd/\epsilon)}$$

where Δ is the maximum degree, assuming $\Delta \geq \frac{2}{3}\ln(2nd/\epsilon)$.

To prove this we need the concentration inequality from [10].

Lemma 6. *Let $X_1, \ldots X_m$ be independent random $n \times n$ Hermitian matrices. Moreover, assume that $\|X_i - E(X_i)\|_2 \leq M$ for all i, and put $v^2 = \|\sum \mathsf{Var}(X_i)\|_2$. Let $X = \sum X_i$. Then for any $a > 0$,*

$$\Pr(\|X - E(X)\|_2 > a) \leq 2n \exp\left(-\frac{a^2}{2v^2 + 2Ma/3}\right).$$

Proof of Theorem 8. Our proof follows ideas from [10]. For $ij \in E(\mathbb{G})$ define $\mathbb{A}^{ij}$ to be the matrix with the rotation O_{ij} in the i,j position, and $O_{ji} = O_{ij}^T$ in the j,i position, and 0 elsewere. Define random variables $h_{ij} = 1$ if the edge ij is deleted, and 0 otherwise. Let $\mathbb{A}^{ii}$ be the diagonal matrix with $I_{d\times d}$ in the ith diagonal position and 0 elsewere. Then note that $\widehat{\mathbb{L}} = \mathbb{L} + \sum_{i,j\in E} h_{ij}\mathbb{A}^{ij} - \sum_{i=1}^n \sum_{j\sim i} h_{ij}\mathbb{A}^{ii}$ and $\overline{\mathbb{L}} = \mathbb{L} + \sum_{i,j\in E} p_{ij}\mathbb{A}^{ij} - \sum_{i=1}^n \sum_{j\sim i} p_{ij}\mathbb{A}^{ii}$, therefore

$$\widehat{\mathbb{L}} - \overline{\mathbb{L}} = \sum_{i,j\in E}(h_{ij} - p_{ij})\mathbb{A}^{ij} - \sum_{i=1}^n \sum_{j\sim i}(h_{ij} - p_{ij})\mathbb{A}^{ii}$$

To use Lemma 6 we must compute the variances. We have

$$\begin{aligned}\mathsf{Var}\left((h_{ij} - p_{ij})\mathbb{A}^{ij}\right) &= E\left((h_{ij} - p_{ij})^2(\mathbb{A}^{ij})^2\right)\\ &= \mathsf{Var}\left(h_{ij} - p_{ij}\right)\left(\mathbb{A}^{ii} + \mathbb{A}^{jj}\right)\\ &= p_{ij}(1 - p_{ij})(\mathbb{A}^{ii} + \mathbb{A}^{jj})\end{aligned}$$

and in a similar manner

$$\mathsf{Var}\left((h_{ij} - p_{ij})\mathbb{A}^{ii}\right) = p_{ij}(1 - p_{ij})\mathbb{A}^{ii}.$$

Therefore

$$\begin{aligned}v^2 &= \left\| \sum_{i,j\in E} p_{ij}(1 - p_{ij})(\mathbb{A}^{ii} + \mathbb{A}^{jj}) + \sum_{i=1}^n \sum_{i\sim j} p_{ij}(1 - p_{ij})\mathbb{A}^{ii} \right\|_2\\ &\leq 2\left\| \sum_{i=1}^n \left(\sum_{j=1}^n p_{ij}(1 - p_{ij}) \right) \mathbb{A}^{ii} \right\|_2\\ &= 2\max_i \sum_{j=1}^n p_{ij}(1 - p_{ij})\\ &\leq 2\max_i \sum_{j=1}^n p_{ij} \leq 2\Delta.\end{aligned}$$

Each $\mathbb{A}^{ij}$ clearly has norm 1, so we can take $M = 1$. Therefore by Lemma 6, taking $a = \sqrt{6\Delta\ln(2nd/\epsilon)}$, we see that

$$\Pr\left(\left\|\widehat{\mathbb{L}} - \overline{\mathbb{L}}\right\|_2 > a\right) \leq 2nd\exp\left(-\frac{a^2}{2v^2 + 2Ma/3}\right) \leq 2nd\exp\left(-\frac{6\Delta\ln(2nd/\epsilon)}{6\Delta}\right) = \epsilon$$

By a consequence of Weyl's Theorem (see, for example, [20]), since $\widehat{\mathbb{L}}$ and $\overline{\mathbb{L}}$ are Hermitian, we have $\left|\lambda_i(\widehat{\mathbb{L}}) - \lambda_i(\overline{\mathbb{L}})\right| \leq \left\|\widehat{\mathbb{L}} - \overline{\mathbb{L}}\right\|_2$. The result then follows. □

We now present an algorithm to delete edges of a connection graph with the goal of decrerasing the smallest eigenvalue of the connection Laplacian.

Our analysis of this algorithm will combine Theorem 5 and Theorem 8. Given a connection graph $\mathbb{G}$, define $\lambda_{\mathbb{G}}$ to be the smallest eigenvalue of its conncection Laplacian.

$(\mathbb{H} = (V, E', O, w')) = \mathsf{ReduceNoise}(\mathbb{G} = (V, E, O, w), p', q, \alpha)$

1. Select q edges in q rounds. In each round one edge is selected.

   Each edge e is chosen with probability p_e proportional to its effective resistence. Then the chosen edge is assigned a weight $w'_e = w_e/(qp_e)$.

2. Delete $\alpha q = q'$ edges in q' rounds. In each round one edge is deleted. Each edge e is chosen with probability p'_e proportional to the weight w'_e.

3. Return $\mathbb{H}$, the connection graph resulting after the edges are deleted.

Theorem 9. *Let $\xi, \epsilon, \delta \in (0,1)$ be given. Given a connection graph $\mathbb{G}$ with m edges, $m > q = \frac{4nd(\log(nd)+\log(1/\xi))}{\epsilon^2}$, $\alpha <$, let $\mathbb{H}$ be the connection graph resulting from the ReduceNoise algorithm. Then with probability at least $(1-\xi)(1-\delta)$ the subgraph $\mathbb{H}$ satisfies*

$$\lambda_{\mathbb{H}} \le (1-\alpha+\epsilon)\lambda_{\mathbb{G}} + \sqrt{6\Delta \ln(2nd/\delta)}$$

provided the maximum degree Δ satisfies $\Delta \ge \frac{2}{3}\ln(2nd/\delta)$.

Proof. We first note that with ξ, ϵ, and q as specified, the edge selection procedure described in step 1 of the algorithm is the same procedure as described in the algorithm Sample and in Theorem 5. Let $\widetilde{\mathbb{G}}$ be the weighted graph resulting from the edge selection, and let $\mathbb{L}_{\widetilde{\mathbb{G}}}$ be its connection Laplacian. Then by Theorem 5 we know that with probability at least ξ, for any $f : V \to \mathbb{R}^d$ we have

$$(1-\epsilon) f\mathbb{L}_{\mathbb{G}} f^T \le f\mathbb{L}_{\widetilde{\mathbb{G}}} f^T \le (1+\epsilon) f\mathbb{L}_{\mathbb{G}} f^T. \tag{5}$$

Now let $\mathbb{H}$ be the connection graph resulting after the deletion process in step 2 of the algorithm, and let $\mathbb{L}_{\mathbb{H}}$ be its connection Laplacian. We note the $\mathbb{H}$ is a random conncetion graph resulting from the deletion of edges of a fixed connection graph, as described in Theorem 8. Let $\overline{\mathbb{L}}_{\mathbb{H}}$ be the matrix of expected values of the entries of $\mathbb{L}_{\mathbb{H}}$, $\overline{\mathbb{L}}_{\mathbb{H}} = E(\mathbb{L}_{\mathbb{H}})$. Note that the deletion procedure deletes αq of the q edges from $\widetilde{G}$ with probability proportional to the weight on each edge, so that the expected value $\overline{\mathbb{L}}_{\mathbb{H}} = \mathbb{L}_{\mathbb{G}} - \alpha \mathbb{L}_{\widetilde{\mathbb{G}}}$. From equation 5 it follows that

$$f\mathbb{L}_{\mathbb{G}} f^T - (1+\epsilon)\alpha f\mathbb{L}_{\mathbb{G}} f^T \le f(\mathbb{L}_{\mathbb{G}} - \alpha\mathbb{L}_{\widetilde{\mathbb{G}}}) f^T \le f\mathbb{L}_{\mathbb{G}} f^T - (1-\epsilon)\alpha f\mathbb{L}_{\mathbb{G}} f^T$$

and thus

$$f\mathbb{L}_{\mathbb{G}} f^T - (1+\epsilon)\alpha f\mathbb{L}_{\mathbb{G}} f^T \le f\overline{\mathbb{L}}_{\mathbb{H}} f^T \le f\mathbb{L}_{\mathbb{G}} f^T - (1-\epsilon)\alpha f\mathbb{L}_{\mathbb{G}} f^T.$$

In particular, it follows that

$$\frac{f\overline{\mathbb{L}}_{\mathbb{H}} f^T}{ff^T} \le (1-\alpha+\epsilon)\frac{f\mathbb{L}_{\mathbb{G}} f^T}{ff^T}$$

for any $f : V \to \mathbb{R}^d$, and therefore that

$$\lambda_0(\overline{\mathbb{L}}_{\mathbb{H}}) \le (1-\alpha+\epsilon)\lambda_0(\mathbb{L}_{\mathbb{G}}).$$

Finally, by Theorem 8, we have, given any $\delta > 0$, with probability at least $\xi(1-\delta)$,

$$\lambda_{\mathbb{H}} < (1-\alpha+\epsilon)\lambda_{\mathbb{G}} + \sqrt{6\Delta \ln(2nd/\delta)}.$$

□

References

[1] D. Achlioptas, Database-friendly random projections, *Proceedings of the 20th ACM symposium on Principles of database systems*, (2001) 274–281.

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz and R. Szeliski, Building Rome in a Day, *Proceedings of the 12th IEEE International Conference on Computer Vision*, (2009) 72–79.

[3] R. Andersen, F. Chung, and K. Lang, Local graph partitioning using pagerank vectors. *Proceedings of the 47th IEEE Symposium on Founation of Computer Science*, (2006) 475–486.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni, An introduction to random matrices. *Cambridge University Press*, 2010.

[5] A. S. Bandeira, A. Singer and D. A. Spielman, A Cheeger Inequality for the Graph Connection Laplacian, 2012, Available at `http://arxiv.org/pdf/1204.3873v1.pdf`.

[6] A. A. Benczúr and D. R. Karger, Approximating *s*-*t* minimum cuts in $\tilde{O}(n^2)$ time. *Proceedings of the 28th ACM Symposium on Theory of Computing*, (1996) 47–55.

[7] P. Berkhin, Bookmark-coloring approach to personalized pagerank computing. *Internet Mathematics*, 3, (2006) 41–62.

[8] C. Borgs, M. Brautbar, J. Chayes, and S.-H. Teng, A sublinear time algorithm for PageRank computations, *Preceedings of the 9th International Workshop on Algorithnms and Models for the Web Graph*, (2012) 49–53.

[9] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, 30 (1-7), (1998) 107–117.

[10] F. Chung and M. Radcliffe, On the spectra of general random graphs, *Electronic Journal of Combinatorics*, 18(1), (2011) 215–229.

[11] F. Chung and W. Zhao, A sharp PageRank algorithm with applications to edge ranking and graph sparsification. *Proceedings of Workshop on Algorithms and Models for the Web Graph (WAW) 2010.* Stanford, California, 2010, 2–14.

[12] F. Chung and S. Sternberg, Laplacian and vibrational spectra for homogeneous graphs. *J. Graph Theory*, 16, (1992) 605–627.

[13] D. Cristofides and K. Markström, Expansion properties of random Cayley graphs and vertex transitive graphs via matrix martingales. *Random Structures Algorithms*, 32(8), (2008) 88–100.

[14] M. Cucuringu, Y. Lipman, and A. Singer, Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Transactions on Sensor Networks*, 8 (3) (2012), No. 19.

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

[16] 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*, 19(3), (2007) 355–369.

[17] G. H. Golub and C. F. Van Loan, *Matrix computations* (3rd ed.). Baltimore: Johns Hopkins, 1996.

[18] R. Hadani and A. Singer, Representation theoretic patterns in three dimensional cryo-electron microscopy I - the intrinsic reconstitution algorithm. *Annals of Mathematics*, 174(2), (2011),1219–1241.

[19] M. Herbster, M. Pontil and Sergio Rojas, Fast Prediction on a Tree. *Proceedings of the Neural Information Processing Systems Foundation*, (2008) 657–664.

[20] R. Horn and C. Johnson, *Matrix Analysis*, Campbridge University Press, 1985.

[21] G. Jeh and J. Widom, Scaling personalized web search. *Proceedings of the 12th World Wide Web Conference WWW*, (2003) 271–279.

[22] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, 2nd ed., 2002.

[23] D. R. Karger, Random sampling in cut, flow, and network design problems. *Mathematics of Operations Research*, 24(2), (1999) 383–413.

[24] D. R. Karger, Using randomized sparsification to approximate minimum cuts. *Proceedings of the 15th ACM symposium on Discrete algorithms*, (1994) 424–432.

[25] D. R. Karger, Minimum cuts in near-linear time. *Journal of the ACM*, 47(1), (2000), 46–76.

[26] F. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, *Ann. Phys. chem.* 72, (1847) 497–508.

[27] I. Koutis, G. L. Miller, and R. Peng, Approaching Optimality for Solving SDD Linear Systems. *Proceedings of 51st IEEE Symposium on Foundations of Computer Science*, (2010) 235–244.

[28] D. G. Lowe, Object recognition from local scale-invariant features, *Proceedings of the 7th IEEE International Conference on Computer Vision*, (1999) 1150–1157.

[29] R. Penrose, A generalized inverse for matrices. *Cambridge Philosophical Society*, 51, (1955) 406–413.

[30] B. Recht, Simpler approach to matrix completion. *Journal of Machine Learning Research*, 12, (2011) 3413-3430.

[31] A. Singer, Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis*, 30(1), (2011) 20–36.

[32] A. Singer, Z. Zhao Y. Shkolnisky, and R. Hadani, Viewing angle classification of cryo-electron microscopy images using eigenvectors. *SIAM Journal on Imaging Sciences*, 4(2), (2011) 723–759.

[33] A. Singer and H.-T. Wu, Vector Diffusion Maps and the Connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8), (2012) 1067-1144.

[34] D. A. Spielman and S.-H. Teng, Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. *Proceedings of the 36th ACM Symposium on Theory of Computing*, (2004) 81–90.

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

[36] D. A. Spielman and S.-H. Teng, Spectral Sparsification of Graphs. *SIAM Journal on Computing*, 40, (2011) 981–1025.

[37] D. A. Spielman and N. Srivastava, Graph sparsification by effective resistances. *Proceedings of 40th ACM symposium on Theory of Computing*, (2008) 563–568.

[38] J. Tropp, *User-Friendly Tail Bounds for Sums of Random Matrices.* Available at `http://arxiv.org/abs/1004.4389`

[39] Roman Vershynin, A note on sums of independent random matrices after ahlswede-winter, 2008. Available at `http://www-personal.umich.edu/~romanv/teaching/reading-group/ahlswede-winter.pdf`.

[40] V. Vu, Spectral norm of random matrices. *Combinatorica*, 27(6), (2007) 721–736.

[41] A. Wigderson and D. Xiao, Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory of Computing*, 4(1), (2008) 53–76.