UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

High Dimensional Spectral Graph Theory and Non- backtracking Random Walks on Graphs

Permalink

https://escholarship.org/uc/item/46f8f88f

Author

Kempton, Mark C.

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

High Dimensional Spectral Graph Theory and Non-backtracking Random Walks on Graphs

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Mark Kempton

Committee in charge:

Professor Fan Chung Graham, Chair
Professor Ronald Graham
Professor Shachar Lovett
Professor Jeffrey Remmel
Professor Jacques Verstraëte

2015

Copyright
Mark Kempton, 2015
All rights reserved.

The dissertation of Mark Kempton is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2015

TABLE OF CONTENTS

Signature Page iii

Table of Contents iv

Acknowledgements vi

Vita vii

Abstract of the Dissertation viii

Chapter 1 Introduction 1
1.1 Preliminaries and Notation 1
1.2 Spectral Graph Theory 2
1.2.1 Matrices and Eigenvalues Associated with Graphs 2
1.2.2 Random Walks on Graphs 6
1.3 Overview and Main Results 8

Chapter 2 Connection Graphs 10
2.1 Introduction 10
2.2 The Connection Laplacian 12
2.2.1 The Connection Laplacian 13
2.2.2 Consistency 15
2.2.3 Random Walks on Connection Graphs 18
2.3 Connection PageRank and Connection Resistance 19
2.3.1 Connection PageRank 19
2.3.2 Connection Resistance 24
2.4 Sparsification and Noise Reduction 27
2.4.1 Edge Ranking Using Effective Resistance 27
2.4.2 Noise Reduction in Connection Graphs 34

Chapter 3 A Clustering Algorithm for Connection Graphs 39
3.1 Introduction 39
3.2 Generalizing Consistency 40
3.2.1 ϵ-consistency 41
3.2.2 Consistent and ϵ-consistent Subsets 44
3.3 Identifying Subsets 48
3.3.1 PageRank Vectors and ϵ-consistent Subsets 48
3.3.2 A Local Partitioning Algorithm 51

Chapter 4 Non-backtracking Random Walks on Graphs 59
4.1 Non-backtracking Random Walks 59
4.1.1 Walks on Directed Edges 60
4.1.2 The Directed Laplacian 63
4.2 A Weighted Ihara's Theorem 66
4.2.1 Ihara's Theorem 66
4.2.2 A Weighted Ihara's Theorem 66
4.2.3 Regular Graphs 69
4.2.4 Biregular Graphs 71
4.3 A Non-backtracking Pólya's Theorem 75
4.3.1 Pólya's Theorem 75
4.3.2 Non-backtracking Walks on Infinite Regular Graphs 76
4.3.3 The Infinite Grid $\mathbb{Z}^2$ 77

Bibliography 80

ACKNOWLEDGEMENTS

This thesis has come as the result of the support and encouragement of numerous people. First, and most notably, I would like to thank my advisor, Fan Chung, for her encouragement, patience, and enthusiasm. Her supportive guidance has made a huge impact in shaping me as a researcher, and I am deeply grateful for the role she has played in my academic development. I would also like to thank the many quality professors in the UCSD Math Department who have helped and taught me so much. Many thanks also to Wayne Barrett, who first introduced me to combinatorics, and played a major role in setting me on the path that I am now on.

Special thanks to my classmates at UCSD: David, Jeremy, Andy, Anson, James, Christian, Jake, Franklin, Sinan, Josh, and all the many friends who have made UCSD such a great place to be, and who have made these years very enjoyable.

Thanks to my parents for their constant love and support. They have been with me the whole way, and have provided so much that is invaluable. And finally, to my wonderful fiancé Katherine, who has given me something to look forward to.

Chapter 2 is based on the paper "Ranking and sparsifying a connection graph," written jointly with Fan Chung and Wenbo Zhao [17]. It appeared in *Journal of Internet Mathematics*, 10 (2014), 87-115. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3 is based on the paper "A local partitioning algorithm for connection graphs," written jointly with Fan Chung [18]. It appeared in the Proceedings of the 10th International Workshop on Algorithms and Models for the Web Graph (WAW 2013), LNCS 8305, 26-43. The dissertation author was one of the primary investigators and authors of this paper.

VITA

2008	B. S. in Mathematics *magna cum laude*, Brigham Young University
2010	M. S. in Mathematics, Brigham Young University
2015	Ph. D. in Mathematics, University of California, San Diego

PUBLICATIONS

W. Barret, S. Gibelyou, M. Kempton, N. Malloy, C. Nelson, and J. Sinkovic, "The inverse eigenvalue and inertia problems for minimum rank two graphs," *Electronic Journal of Linear Algebra*, 22 (2011), 389-418.

M. Kempton and J. Sinkovic, "Minimum rank of outerplanar graphs," *Linear Algebra and its Applications*, 436 (2012), no. 9, 3701-3720.

W. Barret, M. Kempton, N. Malloy, C. Nelson, W. Sexton, and J. Sinkovic, "Decompostions of minimum rank matrices," *Linear Algebra and its Applications*, 438 (2013), no. 10, 3913-1938.

F. Chung, M. Kempton, and W. Zhao, "Ranking and sparsifying a connection graph," *Journal of Internet Mathematics*, 10 (2014), 87-115.

F. Chung and M. Kempton, "A local clustering algorithm for connection graphs," *Proceedings of WAW 2013*, LNCS 8305, 26-43.

ABSTRACT OF THE DISSERTATION

High Dimensional Spectral Graph Theory and Non-backtracking Random Walks on Graphs

by

Mark Kempton

Doctor of Philosophy in Mathematics

University of California, San Diego, 2015

Professor Fan Chung Graham, Chair

This thesis has two primary areas of focus. First we study connection graphs, which are weighted graphs in which each edge is associated with a d-dimensional rotation matrix for some fixed dimension d, in addition to a scalar weight. Second, we study non-backtracking random walks on graphs, which are random walks with the additional constraint that they cannot return to the immediately previous state at any given step.

Our work in connection graphs is centered on the notion of consistency, that is, the product of rotations moving from one vertex to another is independent of the path taken, and a generalization called ϵ-consistency. We present higher dimensional versions of the combinatorial Laplacian matrix and normalized Laplacian

matrix from spectral graph theory, and give results characterizing the consistency of a connection graph in terms of the spectra of these matrices. We generalize several tools from classical spectral graph theory, such as PageRank and effective resistance, to apply to connection graphs. We use these tools to give algorithms for sparsification, clustering, and noise reduction on connection graphs.

In non-backtracking random walks, we address the question raised by Alon et. al. ([3]) concerning how the mixing rate of a non-backtracking random walk to its stationary distribution compares to the mixing rate for an ordinary random walk. Alon et. al. address this question for regular graphs. We take a different approach, and use a generalization of Ihara's Theorem to give a new proof of Alon's result for regular graphs, and to extend the result to biregular graphs. Finally, we give a non-backtracking version of Pólya's Random Walk Theorem for 2-dimensional grids.

Chapter 1

Introduction

1.1 Preliminaries and Notation

In this section, we will introduce the preliminary definitions and ideas that will be used through the thesis. We will use notation that is standard in graph theory. We will begin by defining a *graph* G, which is a pair of sets (V, E), where $V = V(G)$ is called the *vertex set*, and $E = E(G)$ is some subset of the collection of two element subsets of V. The set E is referred to as the *edge set* of G. Less formally, the vertex set V is some set of objects, and the edge set E specifies which of these objects is connected by an edge. Unless otherwise stated, we will assume throughout that V and E are finite sets, and we will denote the number of vertices with n, and the number of edges with m. We will typically denote an edge $\{u, v\} \in E$ simply as uv. If $uv \in E$, we say that the vertices u and v are *adjacent* and write $u \sim v$, and we say that the edge uv is *incident* to the vertices u and v. Unless otherwise specified, all graphs used are *simple* (meaning no multiple edges between two vertices, and no loops on single vertices) and *undirected*. When we refer to *directed* graphs, then the edge set E is a set of ordered pairs of vertices (u, v).

We will also have occasion to discuss *weighted* graphs, which are graphs in which each edge uv is associated with a *weight* w_{uv}. For our purposes, the weight w_{uv} will be a positive real number, and we require that $w_{uv} = w_{vu}$. A simple (unweighted) graph can be thought of as a weighted graph in which $w_{uv} = 1$ for

all $uv \in E$. We define the *degree* of a vertex v, denoted d_v, by $d_v = \sum_{u \sim v} w_{uv}$. For an unweighted graph this is simply the number of edges incident to v. If the degree of every vertex in the graph G is the same, say $d_v = k$ for all $v \in V$, then we say that V is k-regular. For a subset $S \subset V$, we define the *volume* of S as $\text{vol}(S) = \sum_{v \in S} d_v$.

A *walk* on a graph is a sequence of vertices $(v_1, v_2, \cdots, v_k)$ where $v_i \sim v_{i+1}$ for $i = 1, ..., k-1$. A *path* is a walk in which every vertex is distinct. The *distance* between two vertices u and v in G is the number of edges in a shortest path between u and v. The *diameter* of a graph G is the largest distance between any pair of distinct vertices in G. We say that G is *connected* if there is a path between any two distinct vertices of G. A graph is called *bipartite* if there is a partition of the vertex set into two subsets $V = A \cup B$ such that A and B are disjoint, and every edge $uv \in E$ has one vertex in A and one vertex in B.

1.2 Spectral Graph Theory

1.2.1 Matrices and Eigenvalues Associated with Graphs

The goal of spectral graph theory is to understand properties of graphs using tools from linear algebra, particularly using eigenvalues and eigenvectors of various matrices associated with graphs.

For a graph G on n vertices, define the *adjacency matrix* to be the symmetric $n \times n$ matrix, with rows and columns indexed by $V(G)$, given by

$$A(u,v) = \begin{cases} 1 & \text{if } u \sim v \\ 0 & \text{if } u \not\sim v \end{cases}$$

for $u, v \in V$. For a weighted graph, we define $A(u,v) = w_{uv}$ for each edge. To illustrate the principle of how a matrix can give information about a graph, we will state the following well-known result from spectral graph theory.

Lemma 1.2.1. *Let G be a graph with adjacency matrix A. Then for any two vertices $u, v \in V(G)$, the number of walks of length k from u to v is given by $A^k(u,v)$, the (u,v) entry of the kth power of the adjacency matrix.*

Define the diagonal *degree matrix* D to be the $n \times n$ diagonal matrix given by $D(v, v) = d_v$ for $v \in V$. The *combinatorial Laplacian* matrix is

$$L = D - A.$$

The *normalized Laplacian* is

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$$

and so the entries of $\mathcal{L}$ are given by

$$\mathcal{L}(u, v) = \begin{cases} 1 & \text{if } u = v \\ \frac{1}{\sqrt{d_u d_v}} & \text{if } u \sim v \\ 0 & \text{otherwise.} \end{cases}$$

Observe that each of the matrices defined above are symmetric, and thus have real eigenvalues. One of the main tools for investigating eigenvalues of a symmetric matrix is the *Rayleigh quotient*: given a symmetric $n \times n$ matrix M and $x \neq 0 \in \mathbb{R}^n$, define

$$R(x) = \frac{x^T M x}{x^T x}.$$

Observe that if x is an eigenvector of M for eigenvalue λ then $R(x) = \lambda$.

Eigenvalues can be computed by investigating the Rayleigh quotient by way of the well-known Courant-Fischer Min-Max Theorem.

Theorem 1.2.2 (Courant-Fischer, [28])**.** *Let M be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then for $1 \leq k \leq n$,*

$$\lambda_k = \max_{\substack{W \\ \dim(W) = k-1}} \min_{\substack{x \in W^\perp \\ x \neq 0}} R(x)$$

where the maximum ranges over all subspaces W of $\mathbb{R}^n$ of dimension $k - 1$.

We remark that the vector achieving the optimization in the Courant-Fischer Theorem is an eigenvector associated to λ_k.

We will now present several facts about these matrices. In general, we will omit proofs of these well-known facts. Further details can be found in [14]. When

considering matrices associated with a graph G on n vertices, we will think of a vector in $\mathbb{R}^n$ as a function $f : V(G) \to \mathbb{R}$, and A, D, L, and $\mathcal{L}$ are operators on the space of such functions. Observe that the action of A as an operator on this space is given by $(Af)(u) = \sum_{v\sim u} w_{uv} f(v)$, the action of L is given by $(Lf)(u) = \sum_{v\sim u} w_{uv}(f(u) - f(v))$, and the action of $\mathcal{L}$ is given by $(\mathcal{L}f)(u) = \frac{1}{\sqrt{d_u}} \sum_{v\sim u} w_{uv} \left(\frac{f(u)}{\sqrt{d_u}} - \frac{f(v)}{\sqrt{d_v}}\right)$.

We will now turn our attention to the normalized Laplacian matrix. Let $g : V(G) \to \mathbb{R}$. The Rayleigh quotient for the normalized Laplacian is

$$R(g) = \frac{g^T \mathcal{L} g}{g^T g} = \frac{g^T D^{-1/2} L D^{-1/2} g}{g^T g}.$$

Define $f = D^{-1/2} g$, and this becomes

$$R(g) = \frac{f^T L f}{f^T D f}.$$

When g is an eigenvector of $\mathcal{L}$, then f is called a *harmonic eigenfunction* of $\mathcal{L}$, and we will often refer to the above quantity as simply the Rayleigh quotient of f. Direct computation shows that

$$f^T L f = \sum_{uv\in E(G)} w_{uv}(f(u) - f(v))^2$$

and thus we can see that the Rayleigh quotient for a harmonic eigenfunction f for the normalized Laplacian is given by

$$R(f) = \frac{\sum_{uv\in E(G)} w_{uv}(f(u) - f(v))^2}{\sum_{v\in V(G)} f(v)^2 d_v}.$$

Let the eigenvalues of $\mathcal{L}$ be $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$. It can be easily verified (see [14]) that $\lambda_0 = 0$ and $\lambda_{n-1} \leq 2$. Thus $\mathcal{L}$ (as well as L) is positive-semidefinite. The Courant-Fischer Theorem yields the following important characterization of λ_1.

$$\lambda_1 = \inf_{\substack{f \perp D\mathbf{1} \\ f \neq 0}} \frac{\sum_{uv\in E(G)} w_{uv}(f(u) - f(v))^2}{\sum_{v\in V(G)} f(v)^2 d_v}$$

where $\mathbf{1}$ denotes the constant vector all of whose entries are 1. It is easily seen that $\lambda_1 > 0$ if and only if G is connected. The fact can be generalized in the sense that

λ_1 gives an estimate on the connectivity of G, by way of the classic result known as the Cheeger Inequality, which we now present.

For a subset $S \subset V(G)$, let $E(S, \bar{S})|$ denote the set of edges in $E(G)$ that have one vertex in S and one vertex in $\bar{S}$, the complement of S. Define the *Cheeger ratio* of S, denoted $h_G(S)$, by

$$h_G(S) = \frac{|E(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}}.$$

The *Cheeger constant* of the graph G is defined to be

$$h_G = \min_{\substack{S \subset V \\ \mathrm{vol}(S) \leq \frac{1}{2}\mathrm{vol}(G)}} h_G(S).$$

The Cheeger constant of G is also sometimes called the *conductance* of G, or the *isoperimetric constant* of G. Finding the Cheeger constant of a graph is analogous to the classical isoperimetric problem of geometry, in the sense that it measures the boundary of a subset of the vertex set (the set of edges leaving that subset) against the measure of the subset itself (the total volume of the subset). In this way, the Cheeger constant gives a notion of how well connected the graph is: a small Cheeger constant means there is some subset with relatively few edges leaving it, and so a community could be disconnected from the rest of the graph by the removal of a few edges, and a high Cheeger constant means that no such clustered community exists, so that the edges are fairly evenly spread out, and the graph is well connected.

The Cheeger Inequality states that λ_1 gives an estimate of the Cheeger constant. It was first proved for regular graphs, using the adjacency matrix by Tanner in [52] and Alon and Milman in [4]. A general proof using the normalized Laplacian, and without any regularity condition can be found in [14].

Theorem 1.2.3 (Cheeger Inequality)**.** *Let G be a graph, let h_G denote the Cheeger constant of G, and let λ_1 be the second smallest eigenvalue of the normalized Laplacian matrix of G. Then*

$$\frac{h_G^2}{2} \leq \lambda_1 \leq 2h_G.$$

We end this section with an important result known as the Alon-Boppana bound which gives a lower bound for the second largest eigenvalue of the adjacency matrix of a regular graph. The version we give here is due to Nilli in [40].

Theorem 1.2.4 ([40]). *If G is a d-regular graph with diameter $2(k+1)$, then the second largest eigenvalue λ of the adjacency matrix of G satisfies*

$$\lambda \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{k+1}.$$

1.2.2 Random Walks on Graphs

Several important applications of spectral graph theory are in the study of random walks on graphs. A *random walk* on a graph G is a walk $(v_1, ..., v_k)$ in which v_{i+1} is chosen uniformly at random from among the neighbors of v_i. Random walks on graphs have been studied extensively, and [38] and [14] provide good surveys of what is known.

Given a graph G, define the $n \times n$ matrix $P = D^{-1}A$, where D is the diagonal degree matrix of G and A is the adjacency matrix of G as defined in the previous section. So

$$P(u,v) = \begin{cases} \frac{1}{d_u} & \text{if } u \sim v \\ 0 & \text{if } u \not\sim v. \end{cases}$$

Then it is clear the P is the transition probability matrix for a random walk on the graph G, that is, $P(u,v)$ is the probability of moving from vertex u to vertex v in one step of the random walk. Thus, given a probability distribution on the vertices of G, $f : V(G) \to \mathbb{R}$ (thought of as a row vector in $\mathbb{R}^n$) satisfying, $f(v) \geq 0$ for all v and $\sum_{v \in V(G)} f(v) = 1$, then the product fP gives the expected distribution after one step of the random walk. Note that the transition probability matrix for k steps of a random walk is simply given by P^k. Note that if P is not symmetric, but is similar to a symmetric matrix $D^{-1/2}AD^{-1/2} = I - \mathcal{L}$. Thus the spectral properties of the normalized Laplacian $\mathcal{L}$ are directly related to the spectral properties of P. Indeed, the eigenvalues of P are real, and if we order them as $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$, then it is easy to see that $\mu_1 = 1$ with eigenvector $\mathbf{1}$, and $\mu_n \geq -1$. By Perron-Frobenius theory, if the matrix P is irreducible, then we have that $\mu_2 < 1$, and

if P is aperiodic, then $\mu_n > -1$. The matrix P being irreducible and aperiodic corresponds to the graph G being connected and non-bipartite.

We define the *stationary distribution* for a random walk on G by

$$\pi(v) = \frac{d_v}{\mathrm{vol}(G)}.$$

The stationary distribution has the important property the $\pi P = \pi$, so that a random walk with initial distribution π will stay at π at each step. An important fact about the stationary distribution is the following lemma.

Lemma 1.2.5. *If G is a finite connected graph that is not bipartite, then for any initial distribution f_0 on $V(G)$, we have*

$$\lim_{t\to\infty}(f_0P^t)(v) = \pi(v)$$

for all v.

In other words, as long as G is connected and non-bipartite, a random walk will always converge to the graph's stationary distribution. A proof of this lemma can be found in [38]. Knowing that a random walk will converge to some stationary distribution, a fundamental question to consider is to determine how quickly the random walk approaches the stationary distribution, or in other words, to determine the *mixing rate*. In order to make this question precise, we need to consider how to measure the distance between two distribution vectors.

Several measures for defining the mixing rate of a random walk have been given (see [14]). Classically, the mixing rate is defined in terms of the pointwise distance (see [38]). That is, the mixing rate is

$$\mu = \limsup_{t\to\infty}\max_{u,v}\left|P^t(u,v) - \pi(v)\right|^{1/t}.$$

Note that a small mixing rate corresponds to fast mixing. Alternatively, the mixing rate can be considered in terms of the standard L_2 (Euclidean) norm,

$$\|fP^t - \pi\|,$$

the *relative pointwise distance*

$$\Delta(t) = \max_{u,v}\frac{|P^t(u,v) - \pi(v)|}{\pi(v)},$$

the *total variation distance*

$$\Delta_{TV}(t) = \max_{A \subset V(G)} \max_{u \in V(G)} \left| \sum_{v \in A} (P^t(u,v) - \pi(v) \right|,$$

or the *χ-squared distance*

$$\Delta'(t) = \max_{u \in V(G)} \left(\sum_{v \in V(G)} \frac{(P^t(u,v) - \pi(v))^2}{\pi(v)} \right)^{1/2}.$$

In general, each of these measures can yield different distances, but spectral bounds on the mixing rate are essentially the same for each. See [14] for a detailed comparison of each.

We will end this section with a known result tying the mixing rate to the eigenvalues of P.

Theorem 1.2.6 ([38])**.** *Let G be a connected non-bipartite graph with transition probability matrix P, and let the eigenvalues of P be $1 = \mu_1 > \mu_2 \geq \cdots \geq \mu_n > -1$. Then the mixing rate is $\max\{\mu_2, |\mu_n|\}$.*

Thus, the smaller the eigenvalues of P, the faster the random walk converges to its stationary distribution.

1.3 Overview and Main Results

This thesis touches on two main topics: connection graphs as well as non-backtracking random walks on graphs. A connection graph is a weighted graph in which each edge is associated with a rotation matrix, in addition to its scalar weight. A non-backtracking random walk is a random walk with the additional restriction that we are not allowed to return to the immediately previous vertex at any given step.

In Chapter 2, we introduce connection graphs and give higher dimensional versions of the various matrices associated to a graph. We describe the important notion of consistency in a connection graph and prove a spectral characterization of consistency. We give analogues for connection graphs to classical tools in spectral graph theory. In particular we define connection PageRank and connection

resistance in a graph. We generalize a known graph sparsification algorithm to the case of connection graphs, and use this to also give a noise reduction algorithm for an inconsistent connection graph.

Chapter 3 generalizes the notion of consistency from Chapter 2 to that of ϵ-consistency. Bounds on the spectrum of the connection Laplacian are given for ϵ-consistent connection graphs. We tie this to clustering, and use connection PageRank vectors to generalize a clustering algorithm to connection graphs.

Chapter 4 addresses questions concerning non-backtracking random walks on graphs. In particular we investigate the mixing rate of non-backtracking random walks. We present the idea of turning a non-backtracking random walk into a Markov chain by walking along directed edges of a graph. We prove convergence to a stationary distribution in this case, and connect the mixing rate to the eigenvalues of an edge transition probability matrix. We give a non-backtracking Laplacian, and compare its spectrum to the classical Laplacian. We also discuss a result called Ihara's Theorem, and give a weighted version which allows us to compare the spectrum of the transition matrix for a non-backtracking random walk to that of an ordinary random walk. In this way, we give an alternate proof to a result of Alon et. al. ([3]) that in many cases, a non-backtracking random walk on a regular graph has a faster mixing rate than an ordinary random walk. We further generalize this result to a class of graphs called biregular graphs. Finally we give a non-backtracking version of a classical result known as Pólya's Random Walk Theorem for an infinite two-dimensional grid.

Chapter 2

Connection Graphs

2.1 Introduction

In this chapter, we consider a generalization of the notion of a graph, called a connection graph, 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 be related by certain rotations [45]. There are many recent developments of related research in cryo-electron microscopy [26, 44], angular synchronization of eigenvectors [21, 43] and vector diffusion maps [45]. In many areas of machine learning, high-dimensional data points in general can be treated by various methods, such as the Principle Component Analysis [30], 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 [39] 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 [19].

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 [45] 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 [12] 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 [5, 10, 29, 11]. 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 [11].

The effective resistance plays a major role in electrical network theory and can be traced back to the classical work of Kirchhoff [34]. 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 [9, 31, 32, 33] 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 [47] constructed a spectral sparsifier with $O(n \log^c n)$ edges for some large constant c. In [50], 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 [50].

In recent work of Bandeira, Singer, and Spielman in [8], 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 2.2.2). 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.

2.2 The Connection Laplacian

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 [25]).

2.2.1 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 2.2.1. *For any function $f : V \to \mathbf{R}^d$, we have*

$$f\mathbb{L}f^T = \sum_{(u,v)\in E} w_{uv} \left\| f(u)O_{uv} - f(v) \right\|_2^2 \tag{2.1}$$

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 2.1, 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 (2.1), $\lambda_1 \geq 0$ 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. □

2.2.2 Consistency

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 2.2.2. *Let $\mathbb{G}$ be a connected connection graph on n vertices having connection 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 $O_u \in \mathsf{SO}(d)$ 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 $\widehat{f_1}, ..., \widehat{f_d}$ 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 2.2.1, 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 2.2.3. *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 $f_iL = \lambda_i f_i$, 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 2.2.1, $\widehat{f_k}(u) O_{uv} = \widehat{f_k}(v)$ 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 [8]. 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, [8] 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.

2.2.3 Random Walks on Connection Graphs

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 2.2.4. *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. □

2.3 Connection PageRank and Connection Resistance

2.3.1 Connection PageRank

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 [5], 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:

$$\widehat{\mathrm{pr}}_{\alpha,\widehat{s}} = \alpha \sum_{t=0}^{\infty} (1-\alpha)^t\, \widehat{s}\, \mathbb{Z}^t = \alpha\widehat{s} + (1-\alpha)\widehat{\mathrm{pr}}_{\alpha,\widehat{s}\mathbb{Z}}. \tag{2.2}$$

By Theorem 2.2.4 and Equation (2.2), we here state the following useful fact concerning PageRank vectors for a consistent connection graph.

Proposition 2.3.1. *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 2.2.4, 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 [11], 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.

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

Lemma 2.3.2. *(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 2.3.3. *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*

$\widehat{p} = \mathsf{ApproximatePR1}(v, \widehat{s}, \alpha, \epsilon, \rho)$

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 $\widehat{p}$.

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* ApproximatePR1 *produces a vector* $\widehat{p}$ *that satisfies*

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

for vertices u *of* V *for which* $\left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 \geq \frac{\epsilon}{4}$, *and satisfying* $\left\|\widehat{p}(u)\right\|_2 < \frac{\epsilon}{2}$ *for vertices* u *for which* $\left\|\widehat{\mathrm{pr}}_{\alpha,\widehat{s}}(u)\right\|_2 \leq \frac{\epsilon}{4}$. *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 2.2 that

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

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

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

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

The output of the algorithm $\widehat{p}$ gives an approximation to $\widehat{p}^{(k)}_{\alpha,\widehat{s}}$ 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 $\widehat{p}^{(k)}_{\alpha,\widehat{s}}(u)$ 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 $\widehat{p}^{(k)}_{\alpha,\widehat{s}}(u)(j)$. Assuming that for at least one j we have $\widehat{p}^{(k)}_{\alpha,\widehat{s}}(u)(j) > \epsilon/4d$, then we get by the multiplicative Chernoff bound that

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

and

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

which implies

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

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

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

In a similar manner, if $\widehat{p}^{(k)}_{\alpha,\widehat{s}}(u)(j) \leq \frac{\epsilon}{4d}$ 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 $\log_{\frac{1}{1-\alpha}}(\frac{4}{\epsilon})$, where each walk multiplies $\widehat{s}(v)$ by the $d\times d$ rotation matrices. Thus the running time is $O\left(\frac{d^3\log(n\sqrt{d})\log(1/\epsilon)}{\epsilon\rho^2\log(1/(1-\alpha))}\right)$. □

We remark that there is another algorithm that computes an approximate PageRank vector called ApproximatePR that has a different type of error bound. The specifics of the algorithm as well as its run-time analysis can be found in [16] and a version for connection graphs is found in [59]. For completeness, we will state their algorithm here, as well as the theorem providing its analysis, as this algorithm will be used in the next chapter. For the algorithm we need the following subroutine called Push and Lemma 2.3.4.

Push(u,α) :

Let $\widehat{p}' = \widehat{p}$ and $\widehat{r}' = \widehat{r}$, except for these changes:

1. Let $\widehat{p}'(u) = \widehat{p}(u) + \alpha\widehat{r}(u)$ and $\widehat{r}'(u) = \frac{1-\alpha}{2}\widehat{r}(u)$.
2. For each vertex v such that $(u,v)\in E$:

$$\widehat{r}'(v) = \widehat{r}(v) + \frac{(1-\alpha)w_{uv}}{2d_u}\widehat{r}(u)\ O_{uv}.$$

Lemma 2.3.4. *Let $\widehat{p}'$ and $\widehat{r}'$ denote the resulting vectors after performing operation Push(u) with $\widehat{p}$ and $\widehat{r}$. Then $\widehat{p}' + \widehat{\mathrm{pr}}_{\alpha,\widehat{r}'} = \widehat{p} + \widehat{\mathrm{pr}}_{\alpha,\widehat{r}}$.*

$(\widehat{p}, \widehat{r}) = \mathsf{ApproximatePR}(\widehat{s}, \alpha, \epsilon)$

1. Let $\widehat{p}(v) = \mathbf{0}$ and $\widehat{r}(v) = \widehat{s}(v)$ for all $v \in V$.
2. While $\|\widehat{r}(u)\|_2 \geq \epsilon d_u$ for some vertex u:

   Pick any vertex u where $\|\widehat{r}(u)\|_2 \geq \epsilon d_u$ and apply operation $\mathsf{Push}(u, \alpha)$.
3. Return $\widehat{p}$ and $\widehat{r}$.

Theorem 2.3.5. *For a vector $\widehat{s}$ with $\sum_{v \in V} \|\widehat{s}(v)\|_2 \leq 1$, and a constant $0 < \epsilon < 1$, the algorithm $\mathsf{ApproximatePR}(\widehat{s}, \alpha, \epsilon)$ computes an approximate PageRank vector $\widehat{p} = \widehat{\mathrm{pr}}_{\alpha, \widehat{s} - \widehat{r}}$ such that the residual vector $\widehat{r}$ satisfies $\frac{\|\widehat{r}(v)\|_2}{d_v} \leq \epsilon$, for all $v \in V$ and $\sum_{v : \|\widehat{p}(v)\|_2 > 0} d_v \leq \frac{1}{\epsilon \alpha}$. The running time for the algorithm is $O\left(\frac{d^2}{\epsilon \alpha}\right)$.*

2.3.2 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 [25, 41]: *(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 2.3.6. *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)*, and we omit it 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 2.3.7. *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 2.3.6, 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 $O_{p_{uv}}O_{vu} = O_{uv}O_{uv}^T = I$ 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. □

2.4 Sparsification and Noise Reduction

2.4.1 Edge Ranking Using Effective 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 [50], 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 [50] 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.

Theorem 2.4.1. *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*

$$(1-\epsilon)f\mathbb{L}_{\mathbb{G}}f^T \leq f\mathbb{L}_{\widetilde{\mathbb{G}}}f^T \leq (1+\epsilon)f\mathbb{L}_{\mathbb{G}}f^T. \tag{2.3}$$

Before proving Theorem 2.4.1, we need the following two lemmas, in particular concerning the matrix $\Lambda = \mathbb{W}^{1/2}\mathbb{B}\mathbb{L}^+_{\mathbb{G}}\mathbb{B}^T\mathbb{W}^{1/2}$.

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

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 $\widetilde{\mathbb{G}}$.

Lemma 2.4.2. *The matrix Λ is a projection matrix, i.e. $\Lambda^2 = \Lambda$.*

Proof. Observe that

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

Thus, the lemma follows. □

To show that $\widetilde{\mathbb{G}} = (V, \widetilde{E}, O, \widetilde{w})$ is a good sparsifier for $\mathbb{G}$ satisfying (2.3), 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 [50], 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 2.4.3. *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 [20], Recht [42], and Tropp [53]. Here we will use the simple version that is proved in [54].

Theorem 2.4.4. *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*

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

A direct consequence of Theorem 2.4.4 is the following corollary.

Corollary 2.4.5. *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 $1 \leq i \leq q, \|\mathsf{Var}(X_i)\|_2 \leq M \|\mathbf{E}[X_i]\|_2$.

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 2.4.4 we have

$$\begin{aligned}
& \Pr\left[\left\|\frac{\sum_i X_i - \mathbf{E}[X_i]}{M}\right\|_2 > t\right] \\
= \; & \Pr\left[\left\|\sum_i X_i - \sum_i \mathbf{E}[X_i]\right\|_2 > tM\right] \\
\leq \; & k \max\left(\exp\left(-\frac{t^2M^2}{4\sum_i \|\mathsf{Var}(X_i)\|_2}\right), \exp\left(-\frac{t}{2}\right)\right). \qquad (2.4)
\end{aligned}$$

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 (2.4) can be bounded as follows.

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

Thus, the corollary follows. □

Proof (of Theorem 2.4.1). 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

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

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 2.4.5. 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 $y_iy_i^T$: $\left\|y_iy_i^T\right\|_2 \leq \max_e\left(\frac{\left\|\Lambda(\cdot,e)^T\Lambda(\cdot,e)\right\|_2}{p_e}\right) = \max_e\left(\frac{w_e\mathbb{R}_{\mathsf{eff}}(e)}{p_e}\right)$. 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 $\left\|y_iy_i^T\right\|_2 \leq \sum_e\|\Lambda(e,e)\|_2 \leq \sum_e \mathrm{Tr}\left(\Lambda(e,e)\right) = \mathrm{Tr}\left(\Lambda\right) \leq nd$. To bound the variance observe that

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

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 $\left\|\mathbf{E}\left[yy^Tyy^T\right]\right\|_2$. By the definition of expectation, we observe that

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

This implies that

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

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

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

Thus,

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

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

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

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

In [36], a modification of the algorithm from [50] is presented. The oversampling Theorem in [36] can further be modified for connection graphs and stated as follows.

Theorem 2.4.6 (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 2.4.1, the key is the bound on the norm $\left\| y_i y_i^T \right\|_2$. If $p'_e \geq w_e \mathbb{R}_{\mathsf{eff}}(e)$, the norm $\left\| y_i y_i^T \right\|_2$ 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 $\overline{\mathbb{R}_{\mathsf{eff}}}(e) = \operatorname{Tr}(\Psi(e,e))$. Clearly, we have $\overline{\mathbb{R}_{\mathsf{eff}}}(e) = \operatorname{Tr}(\Psi(e,e)) \geq \|\Psi(e,e)\|_2 = \mathbb{R}_{\mathsf{eff}}(e)$ and $\sum_e w_e \overline{\mathbb{R}_{\mathsf{eff}}}(e) = \sum_e \operatorname{Tr}(\Lambda(e,e)) = \operatorname{Tr}(\Lambda) \leq nd$. Using Theorem 2.4.6, we have the following.

Corollary 2.4.7. *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 [16]. 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.

2.4.2 Noise Reduction in Connection Graphs

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 2.2.2.

To begin, we will derive a result on the spectrum of the connection Laplacian analogous to the result of Chung and Radcliffe in [15] on the adjacency matrix of a random graph.

Theorem 2.4.8. *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 [15].

Lemma 2.4.9. *Let $X_1, ...X_m$ 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 2.4.8. Our proof follows ideas from [15]. 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 2.4.9 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 2.4.9, 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, [28]), 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.

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

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

Our analysis of this algorithm will combine Theorem 2.4.1 and Theorem 2.4.8. Given a connection graph $\mathbb{G}$, define $\lambda_{\mathbb{G}}$ to be the smallest eigenvalue of its conncection Laplacian.

Theorem 2.4.10. *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*

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

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

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 2.4.1. 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 2.4.1 we know that with probability at least ξ, for any $f : V \to \mathbb{R}^d$ we have

$$(1-\epsilon) f\mathbb{L}_{\mathbb{G}} f^T \leq f\mathbb{L}_{\widetilde{\mathbb{G}}} f^T \leq (1+\epsilon) f\mathbb{L}_{\mathbb{G}} f^T. \tag{2.5}$$

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 2.4.8. 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 2.5 it follows that

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

and thus

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

In particular, it follows that

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

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

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

Finally, by Theorem 2.4.8, we have, given any $\delta > 0$, with probability at least $\xi(1 - \delta)$,

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

□

Acknowledgment: This chapter is based on the paper "Ranking and sparsifying a connection graph," written jointly with Fan Chung and Wenbo Zhao [17]. It appeared in *Journal of Internet Mathematics*, 10 (2014), 87-115. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3

A Clustering Algorithm for Connection Graphs

3.1 Introduction

For high dimensional data sets, a central problem is to uncover lower dimensional structures in spite of possible errors or noises. An approach for reducing the effect of errors is to consider the notion of inconsistency, which quantifies the difference of accumulated rotations while traveling along distinct paths between two vertices. In many applications, it is desirable to identify edges causing the inconsistencies, or to identify portions of the graph that have relatively small inconsistency. In [17], an algorithm is given, utilizing a version of effective resistance from electrical network theory, that deletes edges of a connection graph in such a way that reduces inconsistencies. In this paper, rather than deleting edges, our focus is on identifying subsets of a connection graph with small inconsistency. The notion of ϵ-consistency of a subset of the vertex set of a connection graph will be introduced, which quantifies the amount of inconsistency for the subset to within an error ϵ. This can be viewed as a generalization of the notion of consistency.

One of the major problems in computing is to design efficient clustering algorithms for finding a good cut in a graph. That is, it is desirable to identify a subset of the graph with small edge boundary in comparison to the overall volume

of the subset. Many clustering algorithms have been derived including some with quantitative analysis (e.g., [5, 6]). As we are looking for ϵ-consistent subsets, it is natural that clustering and the Cheeger ratio should arise in examining local subsets of a graph. In this paper, we will combine the clustering problem and the problem of identifying ϵ-consistent subsets. In particular, we will give an algorithm that uses PageRank vectors to identify a subset of a connection graph which has a small cut, given that there is a subset with small cut that is ϵ-consistent.

The notion of PageRank was first introduced by Brin and Page [12] in 1998 for Google's web search algorithms. It has since proven useful in graph theory for quantifying relationships between vertices in a graph. Algorithms from [5] and [6] utilize PageRank vectors to locally identify good cuts in a graph. In [17], a vectorized version of PageRank is given for connection graphs. Here we use these connection PageRank vectors in a manner similar to [6] to find good cuts under the assumption of an ϵ-consistent subset.

3.2 Generalizing Consistency

We define the *normalized connection Laplacian* $\hat{\mathcal{L}}$ to be the operator on $\mathcal{F}(V, \mathbb{R}^d)$ given by

$$\hat{\mathcal{L}} = \mathbb{D}^{-1/2}\mathbb{L}\mathbb{D}^{-1/2} = I_{nd\times nd} - \mathbb{D}^{-1/2}\mathbb{A}\mathbb{D}^{-1/2}.$$

We remark that $\mathbb{L}$ and $\hat{\mathcal{L}}$ are symmetric, positive semi-definite matrices. Using the Courant-Fischer Theorem (see, for example, [28]), we can investigate the eigenvalues of $\hat{\mathcal{L}}$ by examining the *Rayleigh quotient*

$$\mathcal{R}(g) = \frac{g\hat{\mathcal{L}}g^T}{gg^T}$$

where $g : V \to \mathbb{R}^d$ is thought of as a $1 \times nd$ row vector. Defining $f = g\mathbb{D}^{-1/2}$, we see that

$$\mathcal{R}(g) = \frac{f\mathbb{L}f^T}{f\mathbb{D}f^T} = \frac{\sum\limits_{(u,v)\in E} w_{uv}\left\|f(u)O_{uv} - f(v)\right\|_2^2}{\sum\limits_{v\in V} d_v\left\|f(v)\right\|_2^2}.$$

It is not hard to see that $\mathcal{R}(f) \leq 2$. In particular, letting $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{nd}$ denote the eigenvalues of $\hat{\mathcal{L}}$, we see that $\lambda_k \leq 2$ for all k.

3.2.1 ϵ-consistency

For a connection graph $\mathbb{G} = (V, E, O, w)$, we say that $\mathbb{G}$ is *consistent* if

$$\inf_{\substack{f:V\to\mathbb{R}^d \\ \|f\|_2=1}} \sum_{(u,v)\in E} w_{uv}\|f(u)O_{uv} - f(v)\|_2^2 = 0.$$

As seen in the previous chapter, an equivalent definition for consistency is that there exists a function $f : V \to \mathbb{R}^d$ assigning a vector $f(u) \in \mathbb{R}^d$ to each vertex $u \in V$ such that for all edges $uv \in E$, $f(v) = f(u)O_{uv}$. Therefore for any two vertices u, v in a consistent graph, any two distinct paths starting and ending at u and v, $P_1 = (u = u_1, u_2, ..., u_k = v)$ and $P_2 = (u = v_1, v_2, ..., v_l = v)$, then the product of rotations along either path is the same. That is,

$$\prod_{i=1}^{k-1} O_{u_i u_{i+1}} = \prod_{j=1}^{l-1} O_{v_j v_{j+1}}.$$

For any cycle $C = (v_1, v_2, ..., v_k, v_{k+1} = v_1)$ of the underlying graph, the product of rotations along the cycle C is the identity, i.e. $\prod_{i=1}^{k} O_{v_i v_{i+1}} = I_{d\times d}$.

For ease of notation, given a cycle $C = (v_1, v_2, ..., v_k, v_{k+1} = v_1)$, define $O_C = \prod_{i=1}^{k} O_{v_i v_{i+1}}$, and for a path joining distinct vertices u and v, $P_{uv} = (u = v_1, v_2, ..., v_k = v)$, define $O_{P_{uv}} = \prod_{i=1}^{k-1} O_{v_i v_{i+1}}$. Therefore consistency can be characterized by saying $O_C = I_{d\times d}$ for any cycle C, or given any two vertices u and v of $\mathbb{G}$, then $O_{P_{uv}} = O_{P'_{uv}}$ for any two paths P_{uv}, P'_{uv} connecting u and v.

In this section, we will generalize the notion of consistency, and generalize the theorem from the previous chapter giving a spectral characterization of consistency.

We say a connection graph $\mathbb{G}$ is *ϵ-consistent* if, for every simple cycle $C = (v_1, v_2, ..., v_k, v_{k+1} = v_1)$ of the underlying graph G, we have $\|O_C - I_{d\times d}\|_2 \leq \epsilon$ where $O_C = \prod_{i=1}^{k} O_{v_i v_{i+1}}$. That is, the product of rotations along any cycle is within ϵ of the identity in the 2-norm. An equivalent formulation is as follows. Given vertices u and v, and two distinct paths from u to v, $P_1 = (v_1 = u, v_2, ..., v_k = v)$ and $P_2 = (u_1 = u, u_2, ..., u_l = v)$, define $O_{P_1} = \prod_{i=1}^{k-1} O_{v_i v_{i+1}}$ and $O_{P_2} = \prod_{i=1}^{l-1} O_{u_i u_{i+1}}$. Then $\mathbb{G}$ is ϵ-consistent if and only if $\|O_{P_1} - O_{P_2}\|_2 \leq \epsilon$. This follows from the observation that $O_C = O_{P_1} O_{P_2}^{-1} = O_{P_1} O_{P_2}^T$ and the fact that the

2-norm of a rotation matrix is 1. For ease of notation, we will simply use $\|\cdot\|$ to denote the ℓ_2 norm $\|\cdot\|_2$.

We observe that the triangle inequality implies that any connection graph is 2-consistent, and that a consistent connection graph is 0-consistent. We generalize the first part of the above mentioned result from [17] with the following theorem, which bounds the d smallest eigenvalues of the normalized connection Laplacian for an ϵ-consistent connection graph.

Theorem 3.2.1. *Let $\mathbb{G}$ be an ϵ-consistent connection graph whose underlying graph is connected. Let $\hat{\mathcal{L}}$ be the normalized connection Laplacian and let $0 \leq \lambda_1 \leq \cdots \leq \lambda_{nd}$ be the eigenvalues of $\hat{\mathcal{L}}$. Then for $i = 1, ..., d$,*

$$\lambda_i \leq \frac{\epsilon^2}{2}.$$

Proof. We will define a function $f : V \to \mathbb{R}^d$ whose Rayleigh quotient will bound the smallest eigenvalue. For a fixed vertex $z \in V$, we assign $f(z) = x$, where x is a unit vector in $\mathbb{R}^d$. Fix a spanning tree T of G, and define f to be consistent with T. That is, for any vertex v of G assign $f(v)$ as follows. Let $P_{zv} = (z = v_1 v_2 ... v_k = v)$ be the path from z to v in T. Then let $f(v) = f(z)O_{P_{zv}}$. Notice that $\|f(v)\| = 1$ for all $v \in V$. We will examine the Rayleigh quotient of this function. Notice that for uv an edge of T, we have

$$\|f(u)O_{uv} - f(v)\| = \|f(v) - f(v)\| = 0$$

by construction. For any other edge uv of G, consider the cycle obtained by taking the path $P_{vu} = (v = v_1 v_2 ... v_k = u)$ in T, and adding in the edge uv. Then by construction of f and the ϵ-consistency condition, we have

$$\begin{aligned}
\|f(u)O_{uv} - f(v)\| &= \|f(v)O_{P_{vu}}O_{uv} - f(v)\| \\
&= \left\| f(v) \left(\prod_{i=1}^{k-1} O_{v_i v_{i+1}} O_{v_k v_1} - I \right) \right\| \\
&\leq \epsilon \, \|f(v)\| = \epsilon.
\end{aligned}$$

Therefore

$$\lambda_1 \leq \mathcal{R}(f) = \frac{\sum\limits_{(u,v)\in E} w_{uv}||f(u)O_{uv} - f(v)||^2}{\sum\limits_{v} d_v||f(v)||^2}$$
$$\leq \frac{\sum\limits_{(u,v)\in E} w_{uv}\epsilon^2}{\sum\limits_{v} d_v} = \frac{\epsilon^2}{2}.$$

The initial choice of the unit vector $x \in \mathbb{R}^d$ in the construction of f was arbitrary. We thus have d orthogonal choices for the initial assignment of x, which leads to d orthogonal functions satisfying this inequality. Therefore, by the Courant-Fischer Theorem, $\lambda_1, ..., \lambda_d$ all satisfy this bound. □

The following result concerns the second block of d eigenvalues of $\hat{\mathcal{L}}$ for an ϵ-consistent connection graph, and gives an analog to the upper bound in the Cheeger inequality.

Theorem 3.2.2. *Let $\hat{\mathcal{L}}$ be the normalized connection Laplacian of the ϵ-consistent connection graph $\mathbb{G}$, with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_{nd}$, and let h_G denote the Cheeger constant of the underlying graph. Then for $i = d+1, ..., 2d$,*

$$\lambda_i \leq 2h_G + \frac{\epsilon^2}{2}.$$

Proof. Let $f_1, ..., f_d$ be the orthogonal set of vectors defined in the proof of Theorem 3.2.1, each with $\mathcal{R}(f) \leq \epsilon^2/2$. Then $\|f(v)\|^2 = 1$ for all v. Given $A \subset V$ and $B = \bar{A}$, define $g_i : V \to \mathbb{R}^d$ by

$$g_i(v) = \begin{cases} \frac{1}{\operatorname{vol} A} f_i(v) & \text{for } v \in A \\ -\frac{1}{\operatorname{vol} B} f_i(v) & \text{for } v \in B \end{cases}$$

For ease of notation we will simply write g and f for g_i and f_i. Note that if both $u, v \in A$, then $\|g(u)O_{uv} - g(v)\|^2 = \left\|\frac{1}{\operatorname{vol} A} f(u)O_{uv} - \frac{1}{\operatorname{vol} A} f(v)\right\|^2 \leq \frac{1}{(\operatorname{vol} A)^2}\epsilon^2$. Similarly, if both $u, v \in B$, $\|g(u)O_{uv} - g(v)\|^2 \leq \frac{1}{(\operatorname{vol} B)^2}\epsilon^2$. For $u \in A$ and $v \in B$, we have $\|g(u)O_{uv} - g(v)\|^2 = \left\|\frac{1}{\operatorname{vol} A} f(u)O_{uv} + \frac{1}{\operatorname{vol} B} f(v)\right\|^2 \leq \left(\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}\right)^2$ by the triangle inequality.

Therefore

$$\begin{aligned}
\mathcal{R}(g) &= \frac{\sum\limits_{(u,v)\in E} w_{uv}\|g(u)O_{uv} - g(v)\|^2}{\sum\limits_{v\in V}\|g(v)\|^2 d_v} \\
&\leq \frac{\frac{1}{2}\operatorname{vol} A\frac{1}{(\operatorname{vol} A)^2}\epsilon^2 + \frac{1}{2}\operatorname{vol} B\frac{1}{(\operatorname{vol} B)^2}\epsilon^2 + \left(\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}\right)^2 |E(A,B)|}{\sum\limits_{v\in A}\frac{1}{(\operatorname{vol} A)^2}d_v + \sum\limits_{v\in B}\frac{1}{(\operatorname{vol} B)^2}d_v} \\
&= \frac{\frac{1}{2}\epsilon^2\left(\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}\right) + \left(\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}\right)^2 |E(A,B)|}{\frac{1}{\operatorname{vol} A} + \frac{1}{\operatorname{vol} B}} \\
&\leq \frac{1}{2}\epsilon^2 + 2h_G(A).
\end{aligned}$$

Therefore we have d orthogonal vectors $g_1, ..., g_d$ satisfying this bound, each orthogonal to $f_1, ..., f_d$ which clearly satisfy the bound, so the result follows. □

We remark that the paper of Bandeira, Singer, and Spielman [8] gives a different, but related notion of "almost consistent" for a connection graph which they call the *frustration constant*, denoted $\eta_{\mathbb{G}}$, defined by

$$\eta_{\mathbb{G}} = \min_{f:V\to\mathbb{S}^{d-1}} \frac{\sum\limits_{(u,v)\in E} w_{uv}||f(u)O_{uv} - f(v)||^2}{\sum\limits_{v} d_v||f(v)||^2}$$

where $\mathbb{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$. So the frustration constant restricts only to functions whose entries have norm 1, and as remarked in [8], computation of $\lambda_1(\hat{\mathcal{L}})$ is a relaxation of the computation of $\eta_{\mathbb{G}}$. The proof of Theorem 3.2.1 only utilized functions $f : V \to \mathbb{R}^d$ whose entries have norm 1, so the proof shows that if $\mathbb{G}$ is an ϵ-consistent connection graph, then

$$\eta_{\mathbb{G}} \leq \frac{\epsilon^2}{2}.$$

3.2.2 Consistent and ϵ-consistent Subsets

In this section, we will consider the case where a connection graph has been created in which some subset of the data is error-free (or close to it), leading to a consistent or ϵ-consistent induced subgraph. We will define functions on the vertex set in such a way that the Rayleigh quotient will keep track of the edges leaving

the consistent subset. In this way, we will obtain bounds on the spectrum of the normalized connection Laplacian involving the Cheeger ratio of such subsets.

Theorem 3.2.3. *Let $\mathbb{G}$ be a connection graph of dimension d with normalized connection Laplacian $\hat{\mathcal{L}}$, and $S \subset V$ a subset of the vertex set that is ϵ-consistent for given $\epsilon \geq 0$. Then for $i = 1, ..., d$,*

$$\lambda_i(\hat{\mathcal{L}}) \leq \frac{\epsilon^2}{2} + h_G(S).$$

Proof. Fix a spanning tree T of the subgraph induced by S. Define f as follows. For a fixed vertex u of S, define $f(u) = x$ where $||x|| = 1$, and for $v \in S$, define f to be consistent with the subtree T. For $v \notin S$, define $f(v) = 0$. Fix an edge $uv \in E$ and note that for $u, v \notin S$, $||f(u)O_{uv} - f(v)|| = 0$, for $u, v \in S$, $||f(u)O_{uv} - f(v)|| = ||f(v)\left(O_{P_{vu}}O_{uv} - I\right)|| < \epsilon$, and for $u \in S$, $v \notin S$, $||f(u)O_{uv} - f(v)|| = 1$. Therefore

$$\begin{aligned}
\mathcal{R}(f) &= \frac{\sum\limits_{(u,v)\in E} w_{uv}||f(u)O_{uv} - f(v)||^2}{\sum\limits_{v} d_v||f(v)||^2} \\
&< \frac{\sum\limits_{\substack{uv\in E \\ u,v\in S}} w_{uv}\epsilon^2}{\mathrm{vol}(S)} + \frac{\sum\limits_{\substack{uv\in E \\ u\in S, v\notin S}} w_{uv}}{\mathrm{vol}(S)} \\
&\leq \frac{\epsilon^2}{2} + h_G(S).
\end{aligned}$$

There are d orthogonal choices for the initial choice of x leading to d orthogonal vectors satisfying this bound, so by the Courant-Fisher Theorem, the result follows. □

In the next result, we consider the situation where most of the edges are close to being consistent except for some edges in the edge boundary of a subset.

Theorem 3.2.4. *Suppose G is an ϵ_1-consistent graph for some $\epsilon_1 > 0$, and suppose that $S \subset V$ is a set such that the subgraphs induced by S and $\bar{S}$ are both ϵ_2-consistent, with $0 \leq \epsilon_2 < \epsilon_1$, and $\mathrm{vol}(S) \leq \frac{1}{2}\mathrm{vol}(G)$. Let $\hat{\mathcal{L}}$ be the normalized connection Laplacian. Then for $i = 1, ..., d$,*

$$\lambda_i(\hat{\mathcal{L}}) < \frac{\epsilon_2^2}{2} + \frac{\epsilon_1^2}{2}h_G(S)$$

Proof. We will construct a function $f : V \to \mathbb{R}^d$ whose Rayleigh quotient will bound λ_1. Fix a spanning tree T of S and T' of $\bar{S}$, and fix a vertex $w \in S$. Choose a unit vector $x \in \mathbb{R}^d$, and assign $f(w) = x$. For $v \in S$, assign $f(v)$ for each vertex $v \in S$ such that $f(v) = f(u)O_{uv}$ moving along edges uv of T. Now choose an arbitrary edge $e = yz \in E(S, \bar{S})$ such that $y \in S$ and $z \in \bar{S}$. Assign $f(z) = f(y)O_{yz}$. Assign the remaining vertices of $\bar{S}$ so that $f(v) = f(u)O_{uv}$ moving along edges uv of T'. Note that f is consistent with both T and T'.

Let us examine the Dirichlet sum $\sum_{uv \in E} w_{uv}||f(u)O_{uv} - f(v)||^2$. Consider an edge $f = uv \in E(S, \bar{S})$, $f \neq e$. We may, without loss of generality, assume that both S and $\bar{S}$ are connected. (If one or both is not, then we may alter our definition of f to be consistent along even more edges). Therefore, there is a cycle, $C = v_1v_2...v_kv_1$ where $v_1 = u$, $v_k = v$, C contains the edges e and f, and all other edges have endpoints lying in either S or $\bar{S}$. By construction, $f(v) = f(u)O_{P_{uv}}$, so by the ϵ-consistency condition, we have

$$\begin{aligned}
||f(u)O_{uv} - f(v)|| &= \|f(v)O_{P_{vu}}O_{uv} - f(v)\| \\
&= \left\| f(v)\left(\prod_{i=1}^{k-1} O_{v_iv_{i+1}}O_{v_kv_1} - I\right)\right\| \\
&\leq \epsilon_1||f(v)|| = \epsilon_1.
\end{aligned}$$

In a similar manner, we have that $||f(u)O_{uv} - f(v)|| \leq \epsilon_2$ for each edge uv with both u and v in S or both in $\bar{S}$.

Therefore

$$\begin{aligned}
\lambda_1 \leq \mathcal{R}(f) &= \frac{\sum\limits_{(u,v)\in E} w_{uv}||f(u)O_{uv} - f(v)||^2}{\sum\limits_v d_v||f(v)||^2} \\
&\leq \frac{\sum\limits_{(u,v)\in E} w_{uv}\epsilon_2^2}{\sum\limits_v d_v} + \frac{\sum\limits_{\substack{u\sim v \\ u\in S, v\in\bar{S}}} w_{uv}\epsilon_1^2}{\sum\limits_v d_v} \\
&\leq \frac{\epsilon_2^2|E(G)|}{\mathrm{vol}(G)} + \frac{\epsilon_1^2|E(S,\bar{S})|}{2\,\mathrm{vol}(S)} \\
&= \frac{\epsilon_2^2}{2} + \frac{\epsilon_1^2}{2}h_G(S).
\end{aligned}$$

We have d orthogonal choices for the initial assignment of x, which leads to d

orthogonal vectors satisfying this inequality. Therefore $\lambda_1, ..., \lambda_d$ all satisfy this bound. □

Our next result is similar to Theorem 3.2.2, but in a setting similar to the previous theorem.

Theorem 3.2.5. *Let $\mathbb{G}$ be a connection graph, and suppose $S \subset V$ is a set such that the subgraphs induced by S and $\bar{S}$ are ϵ-consistent, with $\mathrm{vol}(S) \leq \frac{1}{2}\mathrm{vol}(G)$. Let $\hat{\mathcal{L}}$ be the normalized connection Laplacian with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_{nd}$. Then for $i = d+1, ..., 2d$,*

$$\lambda_i \leq \frac{\epsilon^2}{2} + 2h_G(S).$$

Proof. Let $f_1, ...f_d$ be d orthogonal vectors defined as in the proof of the preceding theorem. Each of these has $\mathcal{R}(f_i) \leq \frac{\epsilon^2}{2} + 2h_G(S)$ and $||f(v)||^2 = 1$ for all v. Define $g_i : V \to \mathbb{R}^d$ by

$$g_i(v) = \begin{cases} \frac{1}{\mathrm{vol}\, S} f_i(v) & \text{for } v \in S \\ -\frac{1}{\mathrm{vol}\, \bar{S}} f_i(v) & \text{for } v \in \bar{S}. \end{cases}$$

For ease of notation we will simply write g and f for g_i and f_i. Then

$$\begin{aligned} \mathcal{R}(g) &= \frac{\sum_{u\sim v} w_{uv} \|g(u)O_{uv} - g(v)\|_2^2}{\sum_{v\in V} \|g(v)\|^2 d_v} \\ &\leq \frac{\frac{1}{2}\left(\frac{1}{\mathrm{vol}\, S} + \frac{1}{\mathrm{vol}\, \bar{S}}\right)\epsilon^2 + \sum\limits_{\substack{u\sim v \\ u\in S, v\in\bar{S}}} w_{uv} \|\frac{1}{\mathrm{vol}\, S} f(u)O_{uv} + \frac{1}{\mathrm{vol}\, \bar{S}} f(v)\|}{\frac{1}{\mathrm{vol}\, S} + \frac{1}{\mathrm{vol}\, \bar{S}}} \\ &\leq \frac{\epsilon^2}{2} + \left(\frac{1}{\mathrm{vol}\, S} + \frac{1}{\mathrm{vol}\, \bar{S}}\right) |E(S, \bar{S})| \leq \frac{\epsilon^2}{2} + 2h_G(S). \end{aligned}$$

We have d orthogonal vectors $g_1, ..., g_d$ satisfying this bound, and observe that each is orthogonal to the vectors $f_1, ..., f_d$. Therefore the result follows.

□

We remark that this theorem is a stronger result than Theorem 3.2.2, as the hypothesis does not require that the full graph be ϵ-consistent. That is, the result still holds even if the edges going from S to $\bar{S}$ involve inconsistencies that cause the full graph to fail to be ϵ-consistent.

3.3 Identifying Subsets

In this section, we follow ideas from [5] and [6] to relate connection PageRank vectors to the Cheeger ratio of ϵ-consistent subsets of a connection graph. We will give an algorithm, which runs in time nearly linear in the size of the vertex set, which outputs a subset of the vertex set (if one exists) which has small Cheeger ratio and is ϵ-consistent.

3.3.1 PageRank Vectors and ϵ-consistent Subsets

We define, for $S \subset V$, $f(S) = \sum_{v \in S} ||f(v)||_2$. Given a vertex v of $\mathbb{G}$, define a connection characteristic function χ_v to be any vector satisfying $||\chi_v(v)||_2 = 1$ and $\chi_v(u) = 0$ for $u \neq v$. Likewise, for a subset S of V, define a characteristic function χ_S to be a function such that $||\chi_S(v)||_2 = 1$ for $v \in S$, and $\chi_S(v) = 0$ for $v \notin S$.

Recall the definition of connection PageRank (see [17]). Given a seed vector $\hat{s} : V \to \mathbb{R}^d$ is the vector $\mathrm{pr}(\alpha, \hat{s}) : V \to \mathbb{R}^d$ that satisfies

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

where $\mathbb{Z} = \frac{1}{2}(I + \mathbb{D}^{-1}\mathbb{A})$ is the matrix for the random walk. Define $\mathbb{R}_\alpha = \alpha(I - (1 - \alpha)\mathbb{Z})^{-1} = \alpha \sum_{t=0}^{\infty} (1 - \alpha)^t \mathbb{Z}^t$ and note that $\mathrm{pr}(\alpha, \hat{s}) = \hat{s}\mathbb{R}_\alpha$.

Lemma 3.3.1. *Let $S \subset V$ be a subset of the vertex set of a connection graph, and let χ_S be a characteristic function for S. Then*

$$\|\chi_S \mathbb{D}\mathbb{R}_\alpha(v)\| \leq d_v$$

for all $v \in V$

Proof. First, we will show that

$$\left\|\chi_S \mathbb{D}\mathbb{Z}^k(v)\right\| \leq d_v$$

for all k by induction. For $k = 1$,

$$\|\chi_S \mathbb{D}\mathbb{Z}(v)\| = \frac{1}{2}\|\chi_S \mathbb{D}(I + \mathbb{D}^{-1}\mathbb{A})(v)\| \leq \frac{1}{2}\left(d_v + \sum_{\substack{u \in S \\ u \sim v}} w_{uv}\|\chi_S(u)O_{uv}\|\right) \leq d_v.$$

By the induction hypothesis

$$\begin{aligned}
\left\|\chi_S \mathbb{D}\mathbb{Z}^{k+1}(v)\right\| &= \left\|\chi_S \mathbb{D}\mathbb{Z}^k\mathbb{Z}(v)\right\| \\
&= \left\|\sum_{u\in V} \chi_S \mathbb{D}\mathbb{Z}^k(u)\mathbb{Z}(u,v)\right\| \\
&\leq \sum_{u\in V} \left\|\chi_S \mathbb{D}\mathbb{Z}^k(u)\right\|_2 \|\mathbb{Z}(u,v)\| \\
&\leq \sum_{u\in V} d_u \frac{1}{2}\|I(u,v) + \mathbb{D}^{-1}\mathbb{A}(u,v)\| \\
&\leq \frac{d_v}{2} + \frac{1}{2}\sum_{u\in V} d_u \left\|\frac{1}{d_u} w_{uv} O_{uv}\right\| \\
&= \frac{d_v}{2} + \frac{1}{2}\sum_{u\in V} w_{uv} = d_v
\end{aligned}$$

so this claim follows by induction.

Then from this claim,

$$\|\chi_S \mathbb{D}\mathbb{R}_\alpha(v)\| = \left\|\chi_S \mathbb{D}\alpha \sum_{k=0}^{\infty}(1-\alpha)^k \mathbb{Z}^k\right\| \leq \alpha \sum_{k=0}^{\infty}(1-\alpha)^k \left\|\chi_S \mathbb{D}\mathbb{Z}^k(v)\right\| \leq d_v.$$

□

Lemma 3.3.2. *Let $S \subset V$ be a subset of the vertices such that the subgraph of $\mathbb{G}$ induced by S is ϵ-consistent. Let χ_S be some connection characteristic function for S that is consistent with some spanning subtree T of S. Define $\hat{f}_S$ by $\hat{f}_S(v) = \frac{d_v}{\operatorname{vol}(S)}\chi_S(v)$. The function $\hat{f}_S$ is the expected value for a characteristic function χ_u when a vertex u is chosen from S at random with probability $d_u/\operatorname{vol}(S)$. Then*

$$\operatorname{pr}(\alpha, \hat{f}_S)(S) \geq 1 - \frac{1-\alpha}{\alpha}(h(S) + \epsilon).$$

Proof. We have

$$\begin{aligned}
\mathrm{pr}(\alpha, \hat{f}_S)(S) =& \sum_{v\in S} ||\mathrm{pr}(\alpha, \hat{f}_S)(v)|| = \sum_{v\in S} \|\mathrm{pr}(\alpha, \hat{f}_S)(v)\|\|\chi_S(v)\| \\
\geq& \sum_{v\in S} \mathrm{pr}(\alpha, \hat{f}_S)(v)\chi_S(v)^T = \mathrm{pr}(\alpha, \hat{f}_S)\chi_S^T = \hat{f}_S\mathbb{R}_\alpha\chi_S^T \\
=& \hat{f}_S\left(I - \frac{(1-\alpha)(I-\mathbb{Z})}{I-(1-\alpha)\mathbb{Z}}\right)\chi_S^T = 1 - \left(\hat{f}_S\frac{(1-\alpha)(I-\mathbb{Z})}{I-(1-\alpha)\mathbb{Z}}\right)\chi_S^T \\
=& 1 - \left(\frac{(1-\alpha)\chi_S\mathbb{D}}{\alpha\,\mathrm{vol}(S)}\frac{\alpha I}{I-(1-\alpha)\mathbb{Z}}(I-\mathbb{Z})\right)\chi_S^T \\
=& 1 - \frac{1-\alpha}{\alpha\,\mathrm{vol}(S)}\left(\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}\frac{(\mathbb{D}-\mathbb{A})}{2}\right)\chi_S^T \\
=& 1 - \frac{1-\alpha}{2\alpha\,\mathrm{vol}(S)}\sum_{uv\in E}\Big(w_{uv}\left(\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(u)O_{uv} - \chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(v)\right) \\
& \cdot\left((\chi_S(u)O_{uv})^T - \chi_S(v)^T\right)\Big).
\end{aligned}$$

Here the first inequality follows from the Cauchy-Schwarz Inequality. Note that χ_S is a characteristic function, so all the terms in the sum corresponding to $u, v \notin S$ are 0, for $v \in S$ and $u \notin S$ we are left with just $\chi_s(v)$, and for $u, v \in S$, since S is ϵ-consistent and χ_S was chosen to be consistent with a spanning subtree of S, then we have $\chi_S(u)O_{uv} - \chi_S(v)$ has norm less than ϵ. Applying this, the Cauchy-Schwarz Inequality, and the triangle inequality to the above, we have

$$\begin{aligned}
&\mathrm{pr}(\alpha, \hat{f}_S)(S) \\
&\quad\geq 1 - \frac{1-\alpha}{2\alpha\,\mathrm{vol}(S)}\Big(\sum_{\substack{u\sim v\\ v\in S, u\in\bar{S}}} w_{uv}\left\|\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(u)O_{uv} - \chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(v)\right\| \\
&\quad+ \sum_{\substack{u\sim v\\ u,v\in S}} w_{uv}\left\|\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(u)O_{uv} - \chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(v)\right\|\|\chi_S(u)O_{uv} - \chi_S(v)\|\Big) \\
&\quad\geq 1 - \frac{1-\alpha}{2\alpha\,\mathrm{vol}(S)}\Big(\sum_{\substack{u\sim v\\ v\in S, u\in\bar{S}}} w_{uv}\left(||\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(u)O_{uv}|| + ||\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(v)||\right) \\
&\qquad+ \sum_{\substack{u\sim v\\ u,v\in S}} w_{uv}\left(||\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(u)O_{uv}|| + ||\chi_S\mathbb{D}\mathbb{R}_\alpha\mathbb{D}^{-1}(v)||\right)\epsilon\Big).
\end{aligned}$$

Using Lemma 3.3.1 we can conclude that

$$\mathrm{pr}(\alpha, \hat{f}_S)(S) \geq 1 - \frac{1-\alpha}{\alpha\,\mathrm{vol}(S)}\left(|\partial S| + \epsilon|E(S,S)|\right) \geq 1 - \frac{1-\alpha}{\alpha}(h(S)+\epsilon).$$

□

Theorem 3.3.3. *Let $S \subset V$ be a subset of the vertex set such that the subgraph induced by S is ϵ-consistent. Let χ_S be some connection characteristic function for S that is consistent with some spanning subtree T of S. For each vertex $v \in S$, define $\chi_v : V \to \mathbb{R}^d$ by $\chi_v(v) = \chi_S(v)$ and $\chi_v(u) = 0$ for $u \neq v$. Then for any $\alpha \in (0, 1]$, there is a subset $S_\alpha \subset S$ with volume $\operatorname{vol}(S_\alpha) \geq \operatorname{vol}(S)/2$ such that for any vertex $v \in S_\alpha$, the PageRank vector $\operatorname{pr}(\alpha, \chi_v)$ satisfies*

$$\operatorname{pr}(\alpha, \chi_v)(S) \geq 1 - \frac{2(h(S) + \epsilon)}{\alpha}.$$

Proof. Let v be a vertex of S chosen randomly from the distribution given by $\hat{f}_S$ of the previous result. Define the random variable $X = \operatorname{pr}(\alpha, \chi_v)(\bar{S})$ and note that the definition of PageRank and linearity of expectation implies that $E[X] = \operatorname{pr}(\alpha, \hat{f}_S)$. Therefore, by the preceding result,

$$E[X] = \operatorname{pr}(\alpha, \hat{f}_S)(\bar{S}) \leq \frac{1 - \alpha}{\alpha \operatorname{vol}(S)}(h(S) + \epsilon) \leq \frac{h(S) + \epsilon}{\alpha}.$$

Define

$$S_\alpha = \left\{ v : \operatorname{pr}(\alpha, \chi_v)(S) \geq 1 - \frac{2(h(S) + \epsilon)}{\alpha} \right\}.$$

Then Markov's inequality implies

$$\Pr[v \notin S_\alpha] \leq \Pr[X > 2E[X]] \leq \frac{1}{2}.$$

Therefore $\Pr[v \in S_\alpha] \geq \frac{1}{2}$, so $\operatorname{vol}(S_\alpha) \geq \frac{1}{2} \operatorname{vol}(S)$.

□

3.3.2 A Local Partitioning Algorithm

We will follow ideas from [6] to produce an analogue of the Sharp Drop Lemma. Given any function $p : V \to \mathbb{R}^d$, define $q^{(p)} : V \to \mathbb{R}^d$ by $q^{(p)}(u) = p(u)/d_u$ for all $u \in V$. Order the vertices such that $\|q^{(p)}(v_1)\| \geq \|q^{(p)}(v_2)\| \geq \cdots \geq \|q^{(p)}(v_n)\|$. Define $S_j = \{v_1, ..., v_j\}$. The following lemma will be the basis of our algorithm.

Lemma 3.3.4 (Sharp Drop Lemma)**.** *Let $v \in V(\mathbb{G})$ and let $p = \operatorname{pr}(\alpha, \chi_v)$ for some $\alpha \in (0, 1)$, let $q = q^{(p)}$ and let $\phi \in (0, 1)$ be a real number. Then for any index*

$j \in [1, n]$, either S_j satisfies

$$h(S_j) < 2\phi,$$

or there exists some index $k > j$ such that

$$\operatorname{vol}(S_k) \geq \operatorname{vol}(S_j)(1+\phi) \text{ and } \|q(v_k)\| \geq \|q(v_j)\| - \frac{2\alpha}{\phi \operatorname{vol}(S_j)}.$$

Proof. Let $S \subset V$ be a subset of the vertex set that contains v. We have

$$\begin{aligned}
p\mathbb{Z}(S) &= \sum_{u \in S} \|p\mathbb{Z}(u)\| \\
&= \sum_{u \in S} \left\| \frac{1}{2} p(u) + \frac{1}{2} q\mathbb{A}(u) \right\| \\
&\leq \frac{1}{2} \left(\sum_{u \in S} \|p(u)\| + \sum_{u \in S} \left\| \sum_{v \sim u} q(v) O_{uv} \right\| \right) \\
&\leq \frac{1}{2} \left(\sum_{u \in S} \|p(u)\| + \sum_{u \in S} \sum_{v \sim u} \|q(v)\| \right) \\
&= \frac{1}{2} \left(2 \sum_{u \in S} \|p(u)\| - \sum_{(u,v) \in E(S, \bar{S})} (\|q(u)\| - \|q(v)\|) \right) \\
&= p(S) - \frac{1}{2} \sum_{(u,v) \in E(S, \bar{S})} (\|q(u)\| - \|q(v)\|) .
\end{aligned}$$

Since $p = \operatorname{pr}(\alpha, \chi_v)$, we have that p satisfies $p\mathbb{Z} = \alpha\chi_v + (1-\alpha)p\mathbb{Z}$, therefore

$$\|p\mathbb{Z}(u)\| = \frac{1}{1-\alpha} \|p(u) - \alpha\chi_v(u)\| \geq \|p(u)\| - \alpha\|\chi_v(u)\|$$

for any u. Therefore

$$p\mathbb{Z}(S) \geq p(S) - \alpha.$$

Combining these, we see that

$$\sum_{(u,v) \in E(S, \bar{S})} (\|q(u)\| - \|q(v)\|) \leq 2\alpha. \tag{3.1}$$

Now we will consider S_j. If $\operatorname{vol}(S_j)(1+\phi) > \operatorname{vol}(G)$, then

$$|E(S_j, \bar{S}_j)| \leq \operatorname{vol}(\bar{S}_j) \leq \operatorname{vol}(G) \left(1 + \frac{1}{1+\phi} \right) \leq \phi \operatorname{vol}(S_j)$$

and the result holds. Assume $\text{vol}(S_j)(1+\phi) \leq \text{vol}(G)$. Then there exists a unique index $k > j$ such that

$$\text{vol}(S_{k-1}) \leq \text{vol}(S_j)(1+\phi) \leq \text{vol}(S_k).$$

If $e(S_j, \bar{S}_j) < 2\phi \,\text{vol}(S_j)$, then we are done. If $e(S_j, \bar{S}_j) \geq 2\phi \,\text{vol}(S_j)$, then we note that we can also get a lower bound on $e(S_j, \bar{S}_{k-1})$, namely

$$e(S_j, \bar{S}_{k-1}) \geq e(S_j, \bar{S}_j) - \text{vol}(S_{k-1} \setminus S_j) \geq 2\phi \,\text{vol}(S_j) - \phi \,\text{vol}(S_j) = \phi \,\text{vol}(S_j).$$

Therefore, by equation (3.1)

$$\begin{aligned} 2\alpha &\geq \sum_{(u,v)\in E(S_j,\bar{S}_j)} (\|q(u)\| - \|q(v)\|) \\ &\geq \sum_{(u,v)\in E(S_j,\bar{S}_{k-1})} (\|q(u)\| - \|q(v)\|) \\ &\geq e(S_j, \bar{S}_{k-1})(\|q(v_j)\| - \|q(v_k)\|) \\ &\geq \phi \,\text{vol}(S_j)(\|q(v_j)\| - \|q(v_k)\|). \end{aligned}$$

This implies that $\|q(v_j)\| - \|q(v_k)\| \leq 2\alpha/\phi \,\text{vol}(S_j)$, and the result follows.

□

For our algorithm, we will employ the algorithm ApproximatePR from Chapter 2 to compute an approximate connection PageRank vector. We note that if $\hat{p}$ is the approximate connection PageRank vector resulting form ApproximatePR, then

$$\frac{\|\hat{p}(u)\|}{d_u} \geq \frac{\|\text{pr}(\alpha, \chi_v)(u)\|}{d_u} - \epsilon$$

for all u.

We are now ready to present the algorithm ConnectionPartition that utilizes PageRank vectors to come up with an ϵ-consistent subset of small Cheeger ratio.

Theorem 3.3.5. *Suppose $\mathbb{G}$ is a connection graph with a subset C such that $\text{vol}(C) \leq \frac{1}{2}\text{vol}(\mathbb{G})$, and $h(C) \leq \alpha/64\gamma$ with α as chosen in the algorithm. Assume further that C is ϵ-consistent for some $\epsilon < h(C)$. Let*

$$C_\alpha = \left\{ v \in C : \text{pr}(\alpha, \chi_v)(\bar{C}) \leq \frac{2(h(C)+\epsilon)}{\alpha} \right\}.$$

ConnectionPartition(v, ϕ, x):

The input into the algorithm is a vertex $v \in V$, a target Cheeger ratio $\phi \in (0, 1)$, and a target volume $x \in [0, 2m]$.

1. Set $\gamma = \frac{1}{8} + \sum_{k=1}^{2m} \frac{1}{k}$ where m is the number of edges, $\alpha = \frac{\phi^2}{8\gamma}$, and $\delta = \frac{1}{16\gamma x}$.

2. Compute $p = \mathsf{ApproximatePR}(v, \alpha, \delta)$ (which approximates $\mathrm{pr}(\alpha, \chi_v)$).

   Set $q(u) = p(u)/d_u$ for each u and order the vertices $v_1, ..., v_n$ so that $\|q(v_1)\| \geq \|q(v_2)\| \geq \cdots \geq \|q(v_n)\|$ and for each $j \in [1, n]$ define $S_j = \{v_1, ..., v_j\}$.

3. Choose a starting index k_0 such that $\|q(v_{k_0})\| \geq \frac{1}{\gamma \operatorname{vol}(S_{k_0})}$.

   If no such starting vertex exists, output Fail: No starting vertex.

4. While the algorithm is running:

   (a) If $(1 + \phi) \operatorname{vol}(S_{k_i}) > \operatorname{vol}(G)$, output Fail: No cut found.

   (b) Otherwise, let k_{i+1} be the smallest index such that $\operatorname{vol}(S_{k_{i+1}}) \geq (1 + \phi) \operatorname{vol}(S_{k_i})$.

   (c) If $\|q(v_{k_{i+1}})\| \leq \|q(v_{k_i})\| - 2\alpha/\phi \operatorname{vol}(S_{k_i})$, then output $S = S_{k_i}$ and quit.

   Otherwise repeat the loop.

Then for $v \in C_\alpha$, $\phi < 1$, and $x \geq \mathrm{vol}(C)$, the algorithm ConnectionPartition *outputs a set S satisfying the following properties:*

1. $h(S) \leq 2\phi$.
2. $\mathrm{vol}(S) \leq (2/3)\,\mathrm{vol}(\mathbb{G})$.
3. $\mathrm{vol}(S \cap C) \geq (3/4)\,\mathrm{vol}(S)$.

Proof.

Claim 3.3.6. *There exist an index j such that $\|q(v_j)\| \geq \frac{1}{\gamma\,\mathrm{vol}(S_j)}$.*

Proof. Suppose that $\|q(v_j)\| < \frac{1}{\gamma\,\mathrm{vol}(S_j)}$ for every index j. Since $v \in C_\alpha$, $\epsilon < h(C)$, and $h(C) \leq \alpha/64\gamma$ then we know that

$$\begin{aligned} p(C) &\geq \mathrm{pr}(\alpha, \chi_v)(C) - \delta\,\mathrm{vol}(C) \\ &\geq 1 - \frac{2(h(C)+\epsilon)}{\alpha} - \frac{1}{16\gamma x}\,\mathrm{vol}(C) \\ &\geq 1 - \frac{1}{16\gamma} - \frac{1}{16\gamma} \\ &= 1 - \frac{1}{8\gamma} \end{aligned}$$

since $x \geq \mathrm{vol}(C)$.

On the other hand, under our assumption,

$$\begin{aligned} p(C) \leq p(V) = \sum_{i=1}^{n} \|p(v_i)\| &= \sum_{i=1}^{n} \|q(v_i)\| d_{v_i} \\ &< \sum_{i=1}^{n} \frac{d_{v_i}}{\gamma\,\mathrm{vol}(S_j)} \\ &\leq \frac{1}{\gamma} \sum_{k=1}^{2m} \frac{1}{k}. \end{aligned}$$

Putting these together, we have

$$1 - \frac{1}{8\gamma} < \frac{1}{\gamma} \sum_{k=1}^{2m} \frac{1}{k}.$$

With the choice of $\gamma = \frac{1}{8} + \sum_{k=1}^{2m} \frac{1}{k}$ as in the algorithm, this yields a contradiction. Therefore there exists some index j with $\|q(v_j)\| \geq \frac{1}{\gamma\,\mathrm{vol}(S_j)}$ and the claim is proved.

□

It follows from Claim 3.3.6, that the algorithm will not fail to find a starting vertex.

Let k_f be the final vertex considered by the algorithm.

Claim 3.3.7. *If $k_0, ..., k_f$ is a sequence of vertices satisfying both*

- $\|q(v_{k_{i+1}})\| \geq \|q(v_{k_i})\| - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_i})}$
- $\operatorname{vol}(S_{k_{i+1}}) \geq (1+\phi)\operatorname{vol}(S_{k_i})$

then

$$\|q(k_f)\| \geq \|q(k_0)\| - \frac{4\alpha}{\phi^2 \operatorname{vol}(S_{k_0})}.$$

Proof. We note that $\operatorname{vol}(S_{k_{i+1}}) \geq (1+\phi)^i \operatorname{vol}(S_{k_0})$, and so we have

$$\begin{aligned}
\|q(k_f)\| &\geq \|q(k_0)\| - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_0})} - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_1})} - \cdots - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_{f-1}})} \\
&\geq \|q(k_0)\| - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_0})}\left(1 + \frac{1}{1+\phi} + \cdots + \frac{1}{(1+\phi)^{f-1}}\right) \\
&\geq \|q(k_0)\| - \frac{2\alpha}{\phi \operatorname{vol}(S_{k_0})}\frac{1+\phi}{\phi} \\
&\geq \|q(k_0)\| - \frac{4\alpha}{\phi^2 \operatorname{vol}(S_{k_0})}
\end{aligned}$$

and the claim follows. □

Now we will use this claim, the choice of $\alpha = \phi^2/8\gamma$, and the condition on the starting vertex $\|q(k_0)\| \geq 1/\gamma \operatorname{vol}(S_{k_0})$ to obtain a lower bound on $\|q(k_f)\|$,

$$\begin{aligned}
\|q(k_f)\| &\geq \|q(k_0)\| - \frac{4\alpha}{\phi^2 \operatorname{vol}(S_{k_0})} \\
&\geq \frac{1}{\gamma \operatorname{vol}(S_{k_0})} - \frac{1}{2\gamma \operatorname{vol}(S_{k_0})} \\
&\geq \frac{1}{2\gamma \operatorname{vol}(S_{k_0})}.
\end{aligned}$$

As in the proof of Claim 3.3.6, we have that $p(C) \geq 1 - \frac{1}{8\gamma}$, and therefore $p(\bar{C}) \leq \frac{1}{8\gamma}$.

Now observe that $\operatorname{vol}(S_{k_f} \cap \bar{C}) \leq \dfrac{p(\bar{C})}{\|q(k_f)\|}$. This follows since

$$\|q(k_f)\| \operatorname{vol}(S_{k_f} \cap \bar{C}) = \sum_{v \in S_{k_f} \cap \bar{C}} \|q(k_f)\| d_v \leq \sum_{v \in S_{k_f} \cap \bar{C}} \|q(v)\| d_v \leq \sum_{v \in \bar{C}} \|p(v)\| = p(\bar{C}).$$

Thus

$$\begin{aligned}\operatorname{vol}(S_{k_f} \cap \bar{C}) &\leq \frac{p(\bar{C})}{\|q(k_f)\|} \\ &\leq \frac{2\gamma \operatorname{vol}(S_{k_f})}{8\gamma} \\ &= \frac{1}{4}\operatorname{vol}(S_{k_f}).\end{aligned}$$

Therefore $\operatorname{vol}(S_{k_f}) \leq \operatorname{vol}(C) + \operatorname{vol}(S_{k_f} \cap \bar{C}) \leq \operatorname{vol}(C) + \frac{1}{4}\operatorname{vol}(S_{k_f})$, implying that $\operatorname{vol}(S_{k_f}) \leq \frac{4}{3}\operatorname{vol}(C)$. Using that fact that $\operatorname{vol}(C) \leq \frac{1}{2}\operatorname{vol}(G)$,

$$\operatorname{vol}(S_{k_f}) \leq \frac{4}{3}\operatorname{vol}(C) \leq \frac{2}{3}\operatorname{vol}(G) \leq \frac{\operatorname{vol}(G)}{1+\phi}.$$

This last step follows under the assumption that $\phi \leq 1/2$. We can do this without loss of generality since the guarantee on $h(S)$ in the theorem is trivial for $\phi > 1/2$.

The above shows that the algorithm will not experience a failure due to the volume becoming too large, and we have seen that conditions (2) and (3) will be satisfied by the output.

Finally, to show condition (1), we apply the Sharp Drop Lemma. We know that k_f is the smallest index such that $\operatorname{vol}(S_{k_f+1}) \geq (1+\phi)\operatorname{vol}(S_{k_f})$, and $\|q(v_{k_f+1})\| \leq \|q(v_{k_f})\| - 2\alpha/\phi \operatorname{vol}(S_{k_i})$. Therefore the Sharp Drop Lemma guarantees that $h(S_{k_f}) < 2\phi$, and the proof is complete.

□

Theorem 3.3.8. *The running time for the algorithm* ConnectionPartition *is*

$$O\left(d^2 x \frac{\log^2 m}{\phi^2}\right).$$

Proof. The running time is dominated by the computation of the PageRank vector. According to Theorem **??**, the running time for this is $O\left(\frac{d^2}{\delta\alpha}\right)$. In the algorithm, we have $\alpha = \frac{\phi^2}{8\gamma}$, $\delta = \frac{1}{16\gamma x}$, and $\gamma = \frac{1}{8} + \sum_{k=1}^{2m} \frac{1}{k} = \Theta(\log m)$. Therefore $\alpha = O(\frac{\phi^2}{\log m})$ and $\delta = O(\frac{1}{x \log m})$. Therefore the running time is as claimed.

□

Acknowledgment: This chapter is based on the paper "A local partitioning algorithm for connection graphs," written jointly with Fan Chung [18]. It appeared

in the Proceedings of the 10th International Workshop on Algorithms and Models for the Web Graph (WAW 2013), LNCS 8305, 26-43. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 4

Non-backtracking Random Walks on Graphs

4.1 Non-backtracking Random Walks

Recall that a *random walk* on a graph is a sequence $(v_0, v_1, ..., v_k)$ of vertices $v_i \in V$ where v_i is chosen uniformly at random among the neighbors of v_{i-1}. Random walks on graphs are well-studied, and considerable literature exists about them. See in particular [14] and [38] for good surveys, especially in the use of spectral techniques in studying random walks on graphs.

A random walk on a graph G is a Markov process with transition probability matrix $P = D^{-1}A$, where A denotes the adjacency matrix of G, and D is the diagonal matrix whose diagonal entries are the degrees of the vertices of G. Given any starting probability distribution f_0 on the vertex set V, the resulting probability distribution f_k after applying k random walk steps is given by $f_k = f_0 P^k$. Here we are considering f_0 and f_k as row vectors in $\mathbb{R}^n$.

A *non-backtracking random walk* on G is a sequence $(v_0, v_1, ..., v_k)$ of vertices $v_i \in V$ where v_{i+1} is chosen randomly among the neighbors of v_i such that $v_{i+1} \neq v_{i-1}$ for $i = 1, ..., k-1$. In other words, a non-backtracking random walk is a random walk in which a step is not allowed to go back to the immediately previous state. A non-backtracking random walk on a graph is not a Markov chain since, in

any given state, we need to remember the previous step in order to take the next step.

Define $P^{(k)}$ to be the $n \times n$ transition probability matrix for a k-step non-backtracking random walk on the vertices. That is $P^{(k)}(u, v)$ is the probability that a non-backtracking random walk starting at vertex u ends up at vertex v after k steps. Note that $P^{(1)} = P$ where $P = D^{-1}A$ is the transition matrix for an ordinary random walk on G. However, $P^{(k)}$ is not simply P^k since a non-backtracking random walk is not a Markov chain.

The focus of this chapter will be to analyze various aspects of non-backtracking random walks in comparison to ordinary random walks. In particular, we will compare the mixing rate for a non-backtracking random walk to a classical random walk. This problem was addressed for regular graphs in [3]. Their main result is that, in many case, a the mixing rate for a non-backtracking random walk is faster than for an ordinary random walk. Their proof involves the enumeration of non-backtracking walks on a graph via a recurrence relation involving the adjacency matrix. They thus obtain an expression for the transition probability matrix for a non-backtracking random walk as a polynomial in the adjacency matrix of the graph. In this way, they analyze directly the mixing rate. We will take a different approach, viewing the problem in terms of walks along edges of the graph, and give an alternate proof of the result on regular graphs in [3], and generalize the result to a wider class of graphs. In the final section, we will review the classical Pólya's Theorem for random walk on grids, and give a non-backtracking version.

4.1.1 Walks on Directed Edges

The difficulty in the analysis of non-backtracking random walks is that this process is not a Markov chain, making the k step transition probability matrix harder to determine. However, this process can be turned into a Markov chain by replacing each edge in E with two directed edges (one in each direction), and given a state at a directed edge (u, v), choose the next state uniformly among directed edges (v, x) where $x \neq u$. Denote the set of directed edges with $\overrightarrow{E}$. The transition

probability matrix for this process we will call $\tilde{P}$. Observe that

$$\tilde{P}((u,v),(x,y)) = \begin{cases} \frac{1}{d_v - 1} & \text{if } v = x \text{ and } y \neq u \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\tilde{P}$ is a $2m \times 2m$ matrix. Note also that $\tilde{P}^k$ is the transition matrix for a walk with k steps on the directed edges.

Lemma 4.1.1. *Given any graph G, the matrix $\tilde{P}$ as defined above is doubly stochastic.*

Proof. Observe first that the rows of the matrix $\tilde{P}$ sum to 1, as it is a transition probability matrix. In addition, the columns of $\tilde{P}$ sum to 1. To see this, consider the column indexed by the directed edge (u,v). The entry of this column corresponding to the row indexed by (x,y) is $\frac{1}{d_y - 1}$ if $y = u$ and if $v \neq x$. Since $y = u$ this is equal to $\frac{1}{d_u - 1}$. Otherwise, the entry is 0. Thus the column sum is

$$\sum_{\substack{x \sim u \\ x \neq v}} \frac{1}{d_u - 1} = \frac{d_u - 1}{d_u - 1} = 1$$

as claimed. □

Define the distribution $\tilde{\pi} : \overrightarrow{E} \to \mathbb{R}$ by

$$\tilde{\pi} = \frac{\mathbf{1}}{\text{vol}(G)}$$

where $\mathbf{1}$ is the vector of length $2m$ with each entry equal to 1.

Proposition 4.1.2. *Let $\tilde{f}_0 : \overrightarrow{E} \to \mathbb{R}$ be any distribution on the directed edges of G. If the matrix $\tilde{P}$ is irreducible and aperiodic, then*

$$\tilde{f}_0 \tilde{P}^k \longrightarrow \tilde{\pi}$$

as $k \to \infty$.

Proof. It follows from Lemma 1 that $\tilde{\pi}$ is a stationary distribution for $\tilde{P}$. This follows because, since the columns of $\tilde{P}$ sum to 1, we have

$$\tilde{\pi}\tilde{P} = \tilde{\pi}.$$

Therefore, if the sequence $\tilde{f}_0 \tilde{P}^k$ converges, it must converge to $\tilde{\pi}$. Now, $\tilde{P}$ being irreducible and aperiodic are precisely the conditions for this to converge. □

Let f be a probability distribution on the vertices of G. Then f can be turned into a distribution $\tilde{f}$ on $\overrightarrow{E}$ as follows. Define

$$\tilde{f}((u,v)) = \frac{1}{d_u} f(u).$$

Conversely, given a distribution $\tilde{g}$ on $\overrightarrow{E}$, define a distribution g on the vertices by

$$g(u) = \sum_{(u,v)\in\overrightarrow{E}} \tilde{g}(u,v).$$

Thus, given any starting distribution $f_0 : V \to \mathbb{R}$ on the vertex set of G, we can compute the distribution after k non-backtracking random walk steps $f_k : V \to \mathbb{R}$ as follows. First compute the distribution $\tilde{f}_0$ on the directed edges as above, then compute $\tilde{f}_k = \tilde{f}_0 \tilde{P}^k$, then f_k is given by $f_k(u) = \sum_{v\sim u} \tilde{f}_k(u,v)$. The following proposition tells us that this converges to the same stationary distribution as an ordinary random walk on a graph.

Proposition 4.1.3. *Given a graph G and a starting distribution $f_0 : V \to \mathbb{R}$ on the vertices of G, define $f_k = f_0 P^{(k)}$ to be the distribution on the vertices after k non-backtracking random walk steps. Define the distribution $\pi : V \to \mathbb{R}$ by $\pi(v) = \frac{d_v}{\mathrm{vol}(G)}$ (note that this is the stationary distribution for an ordinary random walk on G). Then if the matrix $\tilde{P}$ is irreducible and aperiodic, then for any starting distribution f_0 on V, we have*

$$f_k \longrightarrow \pi \text{ as } k \to \infty.$$

Proof. As described above, take the distribution f_0 on vertices to the corresponding distribution $\tilde{f}_0$ on directed edges. Then define $\tilde{f}_k = \tilde{f}_0 \tilde{P}^k$. Then by the proposition above, $\tilde{f}_k$ converges to $\tilde{\pi}$. Now $\tilde{\pi} = \frac{\mathbf{1}}{\mathrm{vol}\, G}$, and observe that $\pi(u) = \sum_{v\sim u} \frac{1}{\mathrm{vol}(G)} = \sum_{v\sim u} \tilde{\pi}((u,v))$. So pulling the distribution $\tilde{\pi}$ on directed edges back to a distribution on the vertices yields π. Thus the result follows. □

Definition 4.1.4. The *χ-squared distance* for measuring convergence of a random walk is defined by

$$\Delta'(t) = \max_{y\in V(G)} \left(\sum_{x\in V(G)} \frac{(\tilde{P}^t(y,x) - \tilde{\pi}(x))^2}{\tilde{\pi}(x)} \right)^{1/2}.$$

Notice that since $\tilde{\pi} = \mathbf{1}/\operatorname{vol}(G)$,

$$\begin{aligned}\Delta'(t)^2 &= \max_y \frac{1}{2m}\|(\chi_y \tilde{P}^t - \tilde{\pi})\|^2 \\ &= \max_y \frac{1}{2m}\|(\chi_y - \tilde{\pi})\tilde{P}^t\|^2\end{aligned}$$

Theorem 4.1.5. *Let $\mu_1 = 1, \mu_2, ...\mu_{2m}$ be the eigenvalues of $\tilde{P}$. Then the convergence rate for the non-backtracking random walk with respect to the χ-squared distance is bounded above by $\max_{i\neq 1} |\mu_i|$.*

Proof. We have

$$\Delta'(t)^2 = \max_y \frac{1}{2m}\|(\chi_y - \tilde{\pi})\tilde{P}^t\|^2.$$

Observe that $\chi_u - \tilde{\pi}$ is orthogonal $\tilde{\pi}$, which is the eigenvector for μ_1, so we see that

$$\Delta'(t) \leq \frac{1}{2m} \max_{i\neq 1} |\mu_i|^t.$$

Therefore,

$$\lim_{t\to\infty} (\Delta'(t))^{1/t} \leq \max_{i\neq 1} |\mu_i|.$$

□

4.1.2 The Directed Laplacian

The transition probability matrix $\tilde{P}$ for the walk on directed edges can be thought of as a transition matrix for a random walk on a directed line graph of the graph G. In this way, theory for random walks on directed graphs can be applied to analyze non-backtracking random walks. Random walks on directed graphs have been studied by Chung in [13] by way of a directed version of the normalized graph Laplacian matrix. In [13], the Laplacian for a directed graph is defined as follows. Let P be the transition probability matrix for a random walk on the directed graph, and let ϕ be its Perron vector, that is, $\phi P = \phi$. Then let Φ be the diagonal matrix with the entries of ϕ along the diagonal. Then the *Laplacian* for the directed graph is defined as

$$\mathcal{L} = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}}{2}.$$

This produces a symmetric matrix that thus has real eigenvalues. Those eigenvalues are then related to the convergence rate of a random walk on the directed graph. In particular, the convergence rate is bounded above by $2\lambda_1^{-1}(-\log \min_x \phi(x))$, where λ_1 is the second smallest eigenvalue of $\mathcal{L}$ (see Theorem 7 of [13]).

Applying this now to non-backtracking random walks, define $\tilde{P}$ as before. Then as seen above, ϕ is the constant vector with $\phi(v) = 1/\operatorname{vol}(G)$ for all v. Then the directed Laplacian for a non-backtracking walk becomes

$$\tilde{\mathcal{L}} = I_{2m} = \frac{\tilde{P} + \tilde{P}^*}{2}.$$

Then Theorem 1 of [13], applied to the matrix $\tilde{\mathcal{L}}$ as defined, gives the Rayleigh quotient for a function $f : \overrightarrow{E} \to \mathbb{C}$ by

$$\tilde{R}(f) = \frac{f^*\tilde{\mathcal{L}}f}{f^*f} = \frac{1}{2}\frac{\sum\limits_{(u,v)\in\overrightarrow{E}(G)} \sum\limits_{\substack{(v,w)\\ w\neq u}} (f(u,v) - f(v,w))^2 \, \tilde{P}((u,v),(v,w))}{\sum\limits_{(u,v)\in\overrightarrow{E}(G)} f(u,v)^2}.$$

From this it is clear that $\tilde{\mathcal{L}}$ is positive semidefinite with smallest eigenvalue $\lambda_0 = 0$. If $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{2m-1}$ are the eigenvalues of $\tilde{\mathcal{L}}$, then Theorem 7 from [13] implies that the convergence rate for the corresponding random walk is bounded above by

$$\frac{2\log \operatorname{vol}(G)}{\lambda_1}.$$

We remark that for an ordinary random walk on an undirected graph G, the convergence rate is also on the order of $1/\lambda_1(\mathcal{L})$, where $\mathcal{L}$ now denotes the normalized Laplacian of the undirected graph G. Note that

$$\lambda_1(\mathcal{L}) = \inf_{\substack{f:V(G)\to\mathbb{R}\\ f\perp D\mathbf{1}}} R(f)$$

where $R(f) = \dfrac{\sum_{uv\in E(G)}(f(u) - f(v))^2}{\sum_{v\in V(G)} f(v)^2 d_v}$ denotes the Rayleigh quotient with respect to $\mathcal{L}$, and

$$\lambda_1(\tilde{\mathcal{L}}) = \inf_{\substack{f:\overrightarrow{E}(G)\to\mathbb{R}\\ f\perp\mathbf{1}}} \tilde{R}(f)$$

with $\tilde{R}$ given above.

The following result shows that the Laplacian bound does not give an improvement for non-backtracking random walks over ordinary random walks.

Proposition 4.1.6. *Let G be any graph, and let $\mathcal{L}$ be the normalized graph Laplacian and $\tilde{\mathcal{L}}$ the non-backtracking Laplacian defined above. Then we have*

$$\lambda_1(\tilde{\mathcal{L}}) \leq \lambda_1(\mathcal{L}).$$

Proof. Let $f : V(G) \to \mathbb{R}$ be the function orthogonal to $D\mathbf{1}$ that achieves the minimum in the Rayleigh quotient for $\mathcal{L}$. So

$$\sum_{v \in V(G)} f(v) d_v = 0 \text{ and } \lambda_1(\mathcal{L}) = \frac{\sum_{uv \in E(G)} (f(u) - f(v))^2}{\sum_{v \in V(G)} f(v)^2 d_v}.$$

Define $f' : \overrightarrow{E} \to \mathbb{R}$ by $f'(u,v) = f(u)$. Observe that

$$\sum_{(u,v) \in \overrightarrow{E}(G)} f'(u,v) = \sum_{(u,v) \in \overrightarrow{E}(G)} f(u) = \sum_{u \in V(G)} f(u) d_u = 0.$$

So f' is orthogonal to $\mathbf{1}$. Therefore

$$\begin{aligned}
\lambda_1(\tilde{\mathcal{L}}) \leq \tilde{R}(f') &= \frac{1}{2} \frac{\sum\limits_{(u,v) \in \overrightarrow{E}(G)} \sum\limits_{\substack{(v,w) \\ w \neq u}} (f'(u,v) - f'(v,w))^2 \, \tilde{P}((u,v),(v,w))}{\sum\limits_{(u,v) \in \overrightarrow{E}(G)} f'(u,v)^2} \\
&= \frac{1}{2} \frac{\sum\limits_{(u,v)} \sum\limits_{\substack{(v,w) \\ w \neq u}} (f(u) - f(v))^2 \frac{1}{d_v - 1}}{\sum\limits_{(u,v)} f(u)^2} \\
&= \frac{1}{2} \frac{\sum\limits_{(u,v)} (f(u) - f(v))^2}{\sum\limits_{u \in V(G)} f(u)^2 d_u} \\
&= \frac{\sum\limits_{\{u,v\} \in E(G)} (f(u) - f(v))^2}{\sum\limits_{u \in V(G)} f(u)^2 d_u} = R(f) = \lambda.
\end{aligned}$$

□

4.2 A Weighted Ihara's Theorem

4.2.1 Ihara's Theorem

The transition probability matrix $\tilde{P}$ defined above is a weighted version of an important matrix that comes up in the study of zeta functions on finite graphs. We define B to be the $2m \times 2m$ matrix with rows and columns indexed by the set of directed edges of G as follows.

$$B((u,v),(x,y)) = \begin{cases} 1 & \text{if } v = x \text{ and } y \neq u \\ 0 & \text{otherwise.} \end{cases}$$

The matrix B can be thought of as a non-backtracking edge adjacency matrix, and the entries of B^k describe the number of non-backtracking walks of length k from one directed edge to another, in the same way that the entries of powers of the adjacency matrix, A^k, count the number of walks of length k from one vertex to another. The expression $\det(I - uB)$ is closely related to zeta functions on finite graphs which. A result known as Ihara's Theorem further relates such zeta functions to a determinant expression involving the adjacency matrix. While we will not go into zeta functions on finite graphs in this paper, the following result equivalent to Ihara's theorem will be of interest to us.

Ihara's Theorem. For a graph G, let B be the matrix defined above, let A denote the adjacency matrix, D the diagonal degree matrix, and I the identity. Then

$$\det(I - uB) = (1 - u^2)^{m-n} \det(I - uA + u^2(D - I)).$$

Numerous proofs of this result exist in the literature. We remark that the expressions $\det(I - uB)$ is the characteristic polynomial of B evaluated at $1/u$. In this way the complete spectrum of the matrix B is given by the reciprocals of the roots of the polynomial $(1 - u^2)^{m-n} \det(I - uA + u^2(D - I))$.

4.2.2 A Weighted Ihara's Theorem

In this section, we will follow the main ideas of the proof of Ihara's theorem found in [35] to try to obtain a weighted version of this result.

To each vertex $x \in V(G)$ we assign a weight $w(x) \neq 0$, and let W be the $n \times n$ diagonal matrix given by $W(x,x) = w(x)$. Now define S to be the $2m \times n$ matrix whose rows are indexed by the directed edges of G and whose columns are indexed by the vertices of G, given by

$$S((u,v),x) = \begin{cases} 1 & \text{if } v = x \\ 0 & \text{otherwise} \end{cases}$$

and define $\tilde{S} = SW$. So S is the endpoint incidence operator, and $\tilde{S}$ is the weighted version of S. Define T to be the $n \times 2m$ matrix given by

$$T(x,(u,v)) = \begin{cases} 1 & \text{if } u = x \\ 0 & \text{otherwise} \end{cases}$$

and define $\tilde{T} = WT$. So T is the starting point incidence operator. We will also define τ to be the $2m \times 2m$ matrix giving the reversal operator that switches a directed edge with its opposite. That is,

$$\tau((a,b),(c,d)) = \begin{cases} 1 & \text{if } b = c, a = d \\ 0 & \text{otherwise} \end{cases}$$

and define $\tilde{\tau}$ to be the weighted version of τ, that is

$$\tilde{\tau}((a,b),(c,d)) = \begin{cases} w(b)^2 & \text{if } b = c, a = d \\ 0 & \text{otherwise} \end{cases}$$

Finally, define the $2m \times 2m$ matrix P by

$$\tilde{P}((a,b),(c,d)) = \begin{cases} w(b)^2 & \text{if } b = c, a \neq d \\ 0 & \text{otherwise.} \end{cases}$$

We remark that if we take $w(x) = 1/\sqrt{d_x - 1}$ for each $x \in V(G)$, then $\tilde{P}$ is exactly the transition probability matrix for a non-backtracking random walk on the directed edges of G.

Now, a straightforward computation verifies that

$$\tilde{P} = \tilde{S}\tilde{T} - \tilde{\tau} \tag{4.1}$$

and

$$\tilde{T}\tilde{S} = WAW. \tag{4.2}$$

We will define $\tilde{A} = WAW$.

From (4.1) and (4.2) we obtain the following equations.

$$(I - u\tilde{P})(I - u\tilde{\tau}) = I - u\tilde{S}\tilde{T} + u^2\tilde{S}\tilde{T}\tilde{\tau} - u^2\tilde{\tau}^2 \tag{4.3}$$

$$(I - u\tilde{\tau})(I - u\tilde{P}) = I - u\tilde{S}\tilde{T} + u^2\tilde{\tau}\tilde{S}\tilde{T} - u^2\tilde{\tau}^2 \tag{4.4}$$

We define $\tilde{D}$ to be the diagonal $n \times n$ matrix $\tilde{D}(x,x) = \sum_{v\sim x} w(x)^2 w(v)^2$ and observe that $\tilde{T}\tilde{\tau}\tilde{S} = \tilde{D}$. It then follows that

$$\left((I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2\right)\tilde{S} = \tilde{S}\left(I - u\tilde{A} + u^2\tilde{D}\right) \tag{4.5}$$

$$\tilde{T}\left((I - u\tilde{\tau})(I - u\tilde{P}) + u^2\tilde{\tau}^2\right) = \left(I - u\tilde{A} + u^2\tilde{D}\right)\tilde{T} \tag{4.6}$$

Remark. In the proof in [35], they use τ rather than $\tilde{\tau}$, and $\tau^2 = I$, so that S and T will factor through τ^2, so that the $u^2\tau^2$ term stays on the right hand side of the above equations. Here we have $\tilde{\tau}^2$ is a $2m \times 2m$ diagonal matrix with $\tilde{\tau}^2((u,v),(u,v)) = w(u)^2w(v)^2$. This difference from [35] is one of the primary difficulties in generalizing this result.

We will now perform a change of basis to see how the operator $(I - u\tilde{P})(I - u\tilde{\tau}) + u^2\tilde{\tau}^2$ behaves with respect to the decomposition of the space of functions $f : \overrightarrow{E} \to \mathbb{C}$ as the direct sum of Image $\tilde{S}$ and Ker $\tilde{S}^T$. To this end, fix any basis of the subspace Ker $\tilde{S}^T$, and let R be the $2m \times (2m - n)$ matrix whose columns are the vectors of that basis (note that $\tilde{S}$ has rank n). Define $M = \begin{bmatrix}\tilde{S} & R\end{bmatrix}$. This will be our change of basis matrix. To obtain the inverse of M, form the matrix $\begin{bmatrix}(\tilde{S}^T\tilde{S})^{-1}\tilde{S}^T \\ (R^TR)^{-1}R^T\end{bmatrix}$ and observe that

$$\begin{bmatrix}(\tilde{S}^T\tilde{S})^{-1}\tilde{S}^T \\ (R^TR)^{-1}R^T\end{bmatrix}\begin{bmatrix}\tilde{S} & R\end{bmatrix} = \begin{bmatrix}(\tilde{S}^T\tilde{S})^{-1}\tilde{S}^T\tilde{S} & (\tilde{S}^T\tilde{S})^{-1}\tilde{S}^TR \\ (R^TR)^{-1}R^T\tilde{S} & (R^TR)^{-1}R^TR\end{bmatrix} = \begin{bmatrix}I_n & 0 \\ 0 & I_{2m-n}\end{bmatrix}.$$

Therefore we have that $M^{-1} = \begin{bmatrix} (\tilde{S}^T\tilde{S})^{-1}\tilde{S}^T \\ (R^TR)^{-1}R^T \end{bmatrix}$.

Applying this change of basis, direct computation, applying (4.3) and (4.5), yields

$$\begin{aligned} &\begin{bmatrix} (\tilde{S}^T\tilde{S})^{-1}\tilde{S}^T \\ (R^TR)^{-1}R^T \end{bmatrix} \left((I-u\tilde{P})(I-u\tilde{\tau})+u^2\tilde{\tau}^2\right) \begin{bmatrix} \tilde{S} & R \end{bmatrix} \\ &\qquad = \begin{bmatrix} I-u\tilde{A}+u^2\tilde{D} & -u\tilde{T}R+u^2\tilde{T}\tilde{\tau}R \\ 0 & I \end{bmatrix}. \end{aligned} \tag{4.7}$$

Thus, a weighted form of Ihara's Theorem can be stated as follows.

Theorem 4.2.1. *Given a graph G and a positive weight $w(x) > 0$ assigned to each vertex x, then with $\tilde{P}$, $\tilde{\tau}$, $\tilde{A}$, and $\tilde{D}$ as defined above, we have*

$$\det\left((I-u\tilde{P})(I-u\tilde{\tau})+u^2\tilde{\tau}^2\right) = \det(I-u\tilde{A}+u^2\tilde{D}).$$

4.2.3 Regular Graphs

Applying the results of the previous section to regular graphs yields a different proof of the results from [3] on the mixing rate of non-backtracking random walks on regular graphs.

Let G be a regular graph where each vertex has degree d. Then choosing $w(x) = 1/\sqrt{d-1}$ for all x yields gives us that $\tilde{P}$ is the transition probability matrix for the non-backtracking random walk on G. We remark that, from the previous section, we have $\tilde{\tau} = \frac{1}{d-1}\tau$, $\tilde{\tau}^2 = \frac{1}{(d-1)^2}I$, $\tilde{A} = \frac{1}{d-1}A$, and $\tilde{D} = \frac{d}{(d-1)^2}I$. Therefore, the decomposition in (4.7) becomes

$$(I-u\tilde{P})(I-u\tilde{\tau}) \sim \begin{bmatrix} I-\frac{u}{d-1}A+\frac{u^2}{d-1}I & * \\ 0 & \left(1-\frac{u^2}{(d-1)^2}\right)I \end{bmatrix}.$$

Noting that $\tilde{\tau}$ can be thought of as block diagonal with m blocks of the form $\begin{bmatrix} 0 & 1/(d-1) \\ 1/(d-1) & 0 \end{bmatrix}$, then taking determinants, we find that

$$\det(I-u\tilde{P})\left(1-\frac{u^2}{(d-1)^2}\right)^m = \left(1-\frac{u^2}{(d-1)^2}\right)^{2m-n}\det\left(I-\frac{u}{d-1}A+\frac{u^2}{d-1}I\right)$$

and hence

$$\det(I - u\tilde{P}) = \left(I - \left(\frac{u}{d-1}\right)^2\right)^{m-n} \prod_{\lambda} \left(1 - \frac{\lambda}{d-1}u + \frac{1}{d-1}u^2\right)$$

where the product ranges over all the eigenvalues λ of the adjacency matrix A. As remarked above, $\det(I - u\tilde{P})$ is the characteristic polynomial of $\tilde{P}$ evaluated at $1/u$, so setting this to zero and taking reciprocals, we see that the eigenvalues of $\tilde{P}$ are

$$\pm\frac{1}{d-1}, \frac{\lambda \pm \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)}$$

for λ ranging over the eigenvalues of A and $\pm 1/(d-1)$ each having multiplicity $m - n$. We remark that the expression $\frac{\lambda+\sqrt{\lambda^2-4(d-1)}}{2(d-1)}$ is precisely the expression derived by Alon et al. in [3] for the mixing rate of a non-backtracking random walk on a regular graph, and we may proceed with the analysis of the convergence rate in the same way they do. The convergence rate is given by the second largest eigenvalue of $\tilde{P}$, which will be obtained setting λ to be the second largest eigenvalue of A. Let μ be this eigenvalue. Note that for $2\sqrt{d-1} \leq \lambda \leq d$ we have

$$\frac{\lambda}{2(d-1)} < \frac{\lambda + \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)} \leq \frac{\lambda}{d}.$$

For $\lambda < 2\sqrt{d-1}$, μ is complex, and we obtain

$$|\mu|^2 = \left|\frac{\lambda + \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)}\right|^2 = \left(\frac{\lambda}{2(d-1)}\right)^2 + \left(\frac{\sqrt{4(d-1) - \lambda^2}}{2(d-1)}\right)^2 = \frac{1}{d-1}$$

so $|\mu| = \frac{1}{\sqrt{d-1}}$.

We remark that in this case that $\lambda < 2\sqrt{d-1}$, a classic result of Nilli ([40]) related to the Alon-Boppana Theorem implies that we are never too far below this bound. Indeed, the result states that if G is d-regular with diameter at least $2(k+1)$, then $\lambda \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{k+1}$. We can thus state the result from [3].

Theorem 4.2.2. *Let G be a non-bipartite, connected d-regular graph on n vertices for $d \geq 3$, and let ρ and $\tilde{\rho}$ denote the mixing rates of simple and non-backtracking random walk on G, respectively. Let λ be the second largest eigenvalue of the adjacency matrix of G in absolute value.*

If $\lambda \geq 2\sqrt{d-1}$, *then*

$$\frac{d}{2(d-1)} \leq \frac{\tilde{\rho}}{\rho} \leq 1.$$

If $\lambda < 2\sqrt{d-1}$ *and* $d = n^{o(1)}$, *then*

$$\frac{\tilde{\rho}}{\rho} = \frac{d}{2(d-1)} + o(1).$$

4.2.4 Biregular Graphs

A graph G is called (c,d)-*biregular* if it is bipartite and each vertex in one part of the bipartition has degree c, and each vertex of the other part has degree d. In the weighted Ihara's Theorem, we have $\tilde{\tau}^2((u,v),(u,v) = \frac{1}{(d_u-1)(d_v-1)}$, so in the case where G is (c,d)-biregular, then we have $\tilde{\tau}^2 = \frac{1}{(c-1)(d-1)}I$. So since $\tilde{\tau}^2$ is a multiple of the identity, as with regular graphs, in the decomposition (4.7), the $u^2\tilde{\tau}^2$ term can be taken to the other side of the equation. Note that $\tilde{D}$ is diagonal with $\tilde{D}(u,u) = \sum_{v\sim u} \frac{1}{(d_u-1)(d_v-1)} = \frac{c}{(c-1)(d-1)}$ if u has degree c, or $\frac{d}{(c-1)(d-1)}$ if u has degree d. Then $\tilde{D} - \tilde{\tau}^2$ is diagonal with entry $\frac{c}{(c-1)(d-1)} - \frac{1}{(c-1)(d-1)} = \frac{1}{d-1}$ or $\frac{d}{(c-1))(d-1)} - \frac{1}{(c-1)(d-1)} = \frac{1}{c-1}$. Hence the decomposition (4.7) becomes

$$(I - u\tilde{P})(I - u\tilde{\tau}) \sim \begin{bmatrix} I - u\begin{bmatrix} 0 & \frac{1}{c-1}M \\ \frac{1}{d-1}M^T & 0 \end{bmatrix} + u^2\begin{bmatrix} \frac{1}{d-1}I & 0 \\ 0 & \frac{1}{c-1}I \end{bmatrix} & * \\ 0 & \left(1 - \frac{u^2}{(c-1)(d-1)}\right)I \end{bmatrix}$$

where $A = \begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix}$ is the adjacency matrix of G.

Note that $\tilde{\tau}$ is similar to a block diagonal matrix with blocks of the form $\begin{bmatrix} 0 & 1/(c-1) \\ 1/(d-1) & 0 \end{bmatrix}$, so taking the determinant above we obtain

$$\det(I - u\tilde{P})\left(1 - \frac{u^2}{(c-1)(d-1)}\right)^m = \left(1 - \frac{u^2}{(c-1)(d-1)}\right)^{2m-n} \times \det\left(I - u\begin{bmatrix} 0 & \frac{1}{c-1}M \\ \frac{1}{d-1}M^T & 0 \end{bmatrix} + u^2\begin{bmatrix} \frac{1}{d-1}I & 0 \\ 0 & \frac{1}{c-1}I \end{bmatrix}\right)$$

so

$$\det(I - u\tilde{P}) = \left(1 + \frac{u^2}{(c-1)(d-1)}\right)^{m-n} \det \begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right) I & \frac{u}{c-1} M \\ \frac{u}{d-1} M^T & \left(1 - \frac{u^2}{c-1}\right) I \end{bmatrix}$$

We will look at the matrix $\begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right) I & \frac{u}{c-1} M \\ \frac{u}{d-1} M^T & \left(1 + \frac{u^2}{c-1}\right) I \end{bmatrix}$. Suppose the first part in the bipartition of G has size r, and the second part has size s, where without loss of generality, $r > s$. By row reduction, this has the same determinant as the matrix

$$\begin{bmatrix} \left(1 + \frac{u^2}{d-1}\right) I & \frac{u}{c-1} M \\ 0 & \left(1 - \frac{u^2}{c-1}\right) I - \frac{1}{1 + \frac{u^2}{d-1}} \frac{u^2}{(c-1)(d-1)} M^T M \end{bmatrix}$$

which is

$$\left(1 + \frac{u^2}{d-1}\right)^r \det \left(\left(1 + \frac{u^2}{c-1}\right) I - \frac{1}{1 + \frac{u^2}{d-1}} \frac{1}{(c-1)(d-1)} M^T M \right)$$
$$= \left(1 + \frac{u^2}{d-1}\right)^{r-s} \det \left(\left(1 + \frac{u^2}{c-1}\right) \left(1 + \frac{u^2}{d-1}\right) I - \frac{u^2}{(c-1)(d-1)} M^T M \right).$$

Now, the above determinant is given by the product of the eigenvalues of the matrix. Observe that if λ is an eigenvalue of the adjacency matrix A, then λ^2 is an eigenvalue of $M^T M$. Therefore, in all we have

$$\det(I - u\tilde{P}) = \left(1 - \frac{u^2}{(c-1)(d-1)}\right)^{m-n} \left(1 + \frac{u^2}{d-1}\right)^{r-s} \times \prod_{\lambda} \left(\left(1 + \frac{u^2}{c-1}\right) \left(1 + \frac{u^2}{d-1}\right) - \frac{\lambda^2 u^2}{(c-1)(d-1)} \right)$$

where the product ranges over the s largest eigenvalues of A (or in other words, λ^2 ranges of the s eigenvalues of $M^T M$). Therefore the characteristic polynomial is given by

$$\det(uI - \tilde{P}) = \left(u^2 - \frac{1}{(c-1)(d-1)}\right)^{m-n} \left(u^2 + \frac{1}{d-1}\right)^{r-s} \times \prod_{\lambda} \left(\left(u^2 + \frac{1}{c-1}\right) \left(u^2 + \frac{1}{d-1}\right) - \frac{\lambda^2 u^2}{(c-1)(d-1)} \right).$$

Thus we can explicitly obtain the eigenvalues of $\tilde{P}$ which are $\pm\frac{1}{\sqrt{(c-1)(d-1)}}$ with multiplicity $m-n$ each, $\pm\frac{1}{\sqrt{d-1}}i$ with multiplicity $r-s$ each, as well as the 4 roots of the polynomial

$$u^4 + \left(\frac{1}{(c-1)} + \frac{1}{(d-1)} - \frac{\lambda^2}{(c-1)(d-1)}\right)u^2 + \frac{1}{(c-1)(d-1)}$$

for each of the s values of λ. These roots are

$$\pm\sqrt{\frac{\lambda^2-(c-1)-(d-1) \pm \sqrt{(\lambda^2-(c-1)-(d-1))^2-4(c-1)(d-1)}}{2(c-1)(d-1)}}. \quad (4.8)$$

We next ask how these eigenvalues compare to the eigenvalues of $P = D^{-1}A = \begin{bmatrix} 0 & \frac{1}{c}M \\ \frac{1}{d}M^T & 0 \end{bmatrix}$. Note that for λ an eigenvalue of A, we have

$$\begin{bmatrix} 0 & M \\ M^T & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \lambda\begin{bmatrix} x \\ y \end{bmatrix}$$

which implies $My = \lambda x$ and $M^T x = \lambda y$. Then observe

$$\begin{bmatrix} 0 & \frac{1}{c}M \\ \frac{1}{d}M^T & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{c}}x \\ \frac{1}{\sqrt{d}}y \end{bmatrix} = \begin{bmatrix} \frac{1}{c\sqrt{d}}My \\ \frac{1}{d\sqrt{c}}M^T x \end{bmatrix} = \frac{\lambda}{\sqrt{cd}}\begin{bmatrix} \frac{1}{\sqrt{c}}x \\ \frac{1}{\sqrt{d}}y \end{bmatrix},$$

so the eigenvalues of P are $\lambda/\sqrt{cd}$ where λ ranges over the eigenvalues of A. Note that the largest eigenvalue of A is $\sqrt{cd}$.

Let μ equal the expression (4.8) taking the + signs. So μ is the second largest (in modulus) eigenvalue not of modulus 1. Note that the value of λ at which μ transitions from real to complex is $\lambda = \sqrt{(c-1)+(d-1)+2\sqrt{(c-1)(d-1)}} = \sqrt{c-1}+\sqrt{d-1}$. Thus, consider the following cases.

If $\sqrt{c-1}+\sqrt{d-1} \leq \lambda \leq \sqrt{cd}$, then μ is real. Direct computation verifies that, evaluating the expression (4.8) at $\lambda = \sqrt{cd}$ yields $\mu = 1 = \lambda/\sqrt{cd}$ and $\mu < \lambda/\sqrt{cd}$ for λ in this range. Therefore, in this case the eigenvalue of $\tilde{P}$ always has smaller absolute value than the corresponding eigenvalue of P.

If $\lambda < \sqrt{c-1}+\sqrt{d-1}$, then μ is complex, and direct computation shows, for any λ in this range,

$$|\mu|^2 = \frac{1}{\sqrt{(c-1)(d-1)}},$$

so

$$|\mu| = \frac{1}{((c-1)(d-1))^{1/4}}.$$

A version of the Alon-Boppana Theorem exists for (c,d)-biregular graphs as well, proven by Feng and Li in [22] (see also [37]).

Theorem 4.2.3 ([22])**.** *Let G be a (c,d)-biregular graph, and let λ be the second largest eigenvalue of the adjacency matrix A of G. Then*

$$\lambda^2 \geq \left(\sqrt{c-1} + \sqrt{d-1}\right)^2 - \frac{2\sqrt{(c-1)(d-1)} - 1}{k}$$

where the diameter of G is greater than $2(k+1)$.

We can now give a version of Theorem 4.2.2 for (c,d)-biregular graphs.

Theorem 4.2.4. *Let G be a (c,d)-biregular graph with $c,d \geq 2$. Let $\rho = \lambda^2/cd$ be the square of the second largest eigenvalue of the transition probability matrix P for a random walk on G, and let $\tilde{\rho} = |\mu|^2$ be the square of the second largest modulus of an eigenvalue of $\tilde{P}$. Let λ be the second largest eigenvalue of the adjacency matrix of G. Then we have the following cases.*

If $\lambda > \sqrt{c-1} + \sqrt{d-1}$, then

$$\frac{cd}{2(c-1)(d-1)}\left(1 - \frac{c-1+d-1}{c-1+2\sqrt{(c-1)(d-1)}+d-1}\right) \leq \frac{\tilde{\rho}}{\rho} \leq 1.$$

If $\lambda < \sqrt{c-1} + \sqrt{d-1}$ and both c and d are $n^{o(1)}$, then

$$\frac{\tilde{\rho}}{\rho} \leq \frac{cd}{2(c-1)(d-1)} + o(1).$$

Proof. For the first case, for the upper bound, we already remarked above that $\mu \leq \lambda/\sqrt{cd}$ implying $\tilde{\rho} \leq \rho$. The lower bound follows from 4.8 ignoring the square root inside.

For the second case, observe that certainly the diameter is at least $\log_{cd} n$, so that the condition on the degrees and Theorem 4.2.3 imply that

$$\lambda^2 \geq 2\sqrt{(c-1)(d-1)}(1 - o(1)).$$

We remarked above that in this case, $|\mu|^2 = \frac{1}{\sqrt{(c-1)(d-1)}}$, so this gives the result. □

4.3 A Non-backtracking Pólya's Theorem

4.3.1 Pólya's Theorem

We will now turn our attention to infinite graphs. Suppose G is a graph with infinitely many vertices. Consider a random walk on G starting at some initial vertex v_0. The random walk on G is called *recurrent* if the probability that the walk eventually returns to v_0 is 1. If this probability is less than one, the random walk is called *transient*. A famous result on recurrence and transience is referred to as Pólya's Theorem.

Theorem 4.3.1 (Pólya's Theorem)**.** *A random walk on the infinite grid $\mathbb{Z}^d$ is recurrent for $d = 1, 2$ and transient for $d \geq 3$.*

Pólya's Theorem is well know, and numerous proofs exist in the literature. See for example [46]. The goal of this section will be to obtain a non-backtracking version for the case $d = 2$. We remark that for the $d = 1$ case, a non-backtracking walk on $\mathbb{Z}$ is clearly transient, as a walk returning to its starting vertex on $\mathbb{Z}$ requires backtracking.

Let $p(t)$ denote the probability that a walk returns to its starting vertex after t steps. The key to the proof of Pólya's Theorem is to investigate the series of $p(t)$.

Proposition 4.3.2 (Theorem 1.2 in [46])**.** *If the sum*

$$\sum_{t=0}^{\infty} p(t)$$

is convergent, then the random walk is transient. Otherwise, it is recurrent.

Therefore, to prove recurrence or transience of a random walk, one approach is to enumerate the total number of walks of length t on the graph, and enumerate the total number of walks of length t that return to the initial vertex at step t, then from this obtain the probability $p(t)$, and analyze the series. For the grid $\mathbb{Z}^d$, it turns out that $p(t) \sim \frac{c}{(\pi t)^{d/2}}$, (see [46]), and from this, Pólya's Theorem follows.

With this in mind, we will start by enumerating the total number of closed non-backtracking walks of a given length.

4.3.2 Non-backtracking Walks on Infinite Regular Graphs

We consider the matrix $A^{(k)} = A^{(k)}(G)$ by defining $A^{(k)}(u, v)$ to be the number of non-backtracking walks of length k from vertex u to vertex v. As described in [3], the matrices $A^{(k)}$ satisfy the recurrence

$$\begin{cases} A^{(1)} = A \\ A^{(2)} = A^2 - D \\ A^{(k+2)} = AA^{(k+1)} - (D - I)A^{(k)} \end{cases}$$

where A denotes the adjacency matrix of G. For convenience we will define $A^{(0)} = I$. Define the generating function

$$F(x) = \sum_{k=0}^{\infty} A^{(k)} x^k,$$

then from this recurrence we can determine the generating function

$$F(x) = (1 - x^2)\left(I - xA + x^2(D - I)\right)^{-1}.$$

Expanding this as a geometric sum, we obtain

$$F(x) = (1 - x^2)\sum_{k=0}^{\infty} (A - (D - I)x)^k x^k.$$

Now, if G is d-regular, then $D - I = (d-1)I$, and the above can be further expanded, yielding

$$F(x) = (1 - x^2)\sum_{k=0}^{\infty}\sum_{i=0}^{k}(-1)^i \binom{k}{i}(d-1)^i A^{k-i}x^{k+i}.$$

Thus, a general formula for $A^{(n)}$ can be obtained by extracting the x^n coefficient.

$$\begin{aligned} A^{(n)} =& [x^n]F(x) \\ =& \sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i \binom{n-i}{i}(d-1)^i A^{n-2i} \\ & - \sum_{i=0}^{\lfloor n/2-1 \rfloor} (-1)^i \binom{n-i-2}{i}(d-1)^i A^{n-2i-2}. \end{aligned}$$

Therefore, the number of non-backtracking random walks of length n from a vertex u to a vertex v in a regular graph can be expressed explicitly as

$$\begin{aligned} A^{(n)}(u,v) = &\sum_{i=0}^{\lfloor n/2 \rfloor} (-1)^i \binom{n-i}{i} (d-1)^i A^{n-2i}(u,v) \\ &- \sum_{i=0}^{\lfloor n/2-1 \rfloor} (-1)^i \binom{n-i-2}{i} (d-1)^i A^{n-2i-2}(u,v). \end{aligned} \tag{4.9}$$

We remark that the expression $A^k(u,v)$ is simply the total number of walks of length k from u to v, so we have expressed the number of non-backtracking walks in terms of the total number of walks.

4.3.3 The Infinite Grid $\mathbb{Z}^2$

In this section, we will use the above tools to obtain a non-backtracking version of Pólya's Random Walk Theorem for the two-dimensional grid $\mathbb{Z}^2$.

Lemma 4.3.3 ((1.68) in [46])**.** *The total number of closed walks of length $2n$ from a vertex to itself in $\mathbb{Z}^2$ is $\binom{2n}{n}^2$.*

We therefore know that, if A is the adjacency operator on $\mathbb{Z}^2$, then any diagonal entry of A^{2n} is $\binom{2n}{n}^2$. Thus if we wish to count the number of closed non-backtracking walks of length $2n$ on $\mathbb{Z}^2$ from a vertex to itself, then by way of (4.9), setting $d = 4$ since $\mathbb{Z}^2$ is 4-regular, we obtain the diagonal entry of $A^{(2n)}$.

Proposition 4.3.4. *The total number of closed non-backtracking walks of length $2n$ from a vertex to itself in $\mathbb{Z}^2$ is*

$$\sum_{i=0}^{n} (-3)^i \binom{2n-i}{i} \binom{2n-2i}{n-i}^2 - \sum_{i=0}^{n-1} (-3)^i \binom{2n-i-2}{i} \binom{2n-2i-2}{n-i-1}^2.$$

Changing the indices, this can alternatively be expressed as

$$\sum_{k=0}^{n} (-3)^{n-k} \binom{n+k}{2k} \binom{2k}{k}^2 - \sum_{i=0}^{n-1} (-3)^{n-1-k} \binom{n+k-2}{2k-2} \binom{2(k-1)}{k-1}^2.$$

It so happens that the expression $\sum_{k=0}^{n}(-3)^{n-k}\binom{n+k}{2k}\binom{2k}{k}^2$ shows up in the study of the *central trinomial coefficients*, T_n, which are defined to be the largest coefficient in the expansion of $(1+x+x^2)^n$. Formally, that is

$$T_n = [x^n](1+x+x^2)^n.$$

From the definition, one can derive the formula

$$T_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k}\binom{2k}{k}.$$

In a paper of Zhi-Wei Sun ([51]), it is proven that T_n satisfy the following relationship with the above sum.

Lemma 4.3.5 (Lemma 4.1 of [51])**.** *For any $n \in \mathbb{N}$ we have*

$$T_n^2 = \sum_{k=0}^{n} \binom{n+k}{2k}\binom{2k}{k}^2 (-3)^{n-k}.$$

From this we obtain an expression for the number of closed walks from a vertex to itself on $\mathbb{Z}^2$ in terms of the squares of the central trinomial coefficients.

Corollary 4.3.6. *For any $n \in \mathbb{N}$ and any vertex $v \in \mathbb{Z}^2$, we have*

$$A^{(2n)}(v,v) = T_n^2 - T_{n-1}^2.$$

The asymptotics of the numbers T_n are investigated in [56].

Lemma 4.3.7 ([56])**.** *The asymptotics for the numbers T_n are given by*

$$T_n = \frac{\sqrt{3}}{2\sqrt{n\pi}} 3^n \left(1 - \frac{3}{16n} + O\left(\frac{1}{n^2}\right)\right).$$

Corollary 4.3.8. *Asymptotically, the number of closed non-backtracking walks from a vertex to itself on the grid $\mathbb{Z}^2$ is given by*

$$A^{(2n)}(v,v) \sim \frac{2}{\pi n} 3^{2n-1}.$$

Proof. Using Corollary 4.3.6 and Lemma 4.3.7, we have

$$\begin{aligned} A^{(2n)} &= T_n^2 - T_{n-1}^2 \\ &= \frac{3}{4\pi n} 3^{2n} \left(1 + O\left(\frac{1}{n}\right)\right) - \frac{3}{4\pi(n-1)} 3^{2n-2} \left(1 + O\left(\frac{1}{n}\right)\right) \\ &= \frac{24n - 27}{4\pi n(n-1)} 3^{2n-2} \left(1 + O\left(\frac{1}{n}\right)\right) \\ &= \frac{2}{\pi n} 3^{2n-1} \left(1 + O\left(\frac{1}{n}\right)\right) \end{aligned}$$

and the result follows. □

We are now ready to give a non-backtracking version of Polya's Theorem for $\mathbb{Z}^2$.

Theorem 4.3.9. *A non-backtracking random walk on the infinite grid $\mathbb{Z}^2$ is recurrent.*

Proof. Let $p(t)$ denote the probability that a non-backtracking random walk on $\mathbb{Z}^2$ returns to its starting point after t steps. Note that the total number of non-backtracking random walks of length t is

$$4 \cdot 3^{t-1}$$

since there are 4 choices for the first step, and then 3 choices for each subsequent step since we must exclude the edge that would backtrack. Note also that $p(t) = 0$ for t odd since a walk on $\mathbb{Z}^2$ returning to its starting point must contain an equal number of steps up as down, and an equal number of steps to the left as to the right. So we need only consider $p(2t)$. The total number of non-backtracking walks of length $2t$ returning to their starting vertex v is given by $A^{(2t)}(v, v)$ so by Corollary 4.3.8, we obtain $p(2t)$ asymptotically is given by

$$p(2t) \sim \frac{\frac{2}{\pi t} 3^{2t-1}}{4 \cdot 3^{2t-1}} = \frac{1}{2\pi t}.$$

Therefore

$$\sum_{t=0}^{\infty} p(2t) = \sum_{t=0}^{\infty} \frac{1}{2\pi t}$$

which is divergent. Therefore, by Proposition 4.3.2, the random walk is recurrent.

□

Bibliography

[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] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, Non-backtracking random walks mix faster, *Communications in Contemporary Mathematics*, **09**, (2007) 585.

[4] N. Alon and V. D. Milman, λ_1 isoperimetric inequalities for graphs and superconcentrators, *J. Comb. Theory* B 38, (1985), 73–88.

[5] 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.

[6] R. Andersen and F. Chung, Detecting sharp drops in PageRank and a simplified local partitioning algorithm, *Theory and Applications of Models of Computation*, Proceedings of TAMC 2007, LNCS 4484, Springer, (2007), 1–12.

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

[8] A. S. Bandeira, A. Singer and D. A. Spielman, A Cheeger Inequality for the Graph Connection Laplacian. *SIAM. J. Matrix Anal. & Appl.*, 34(4), (2013) 1611–1630.

[9] 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.

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

[11] 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.

[12] 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.

[13] F. Chung, Laplacians and the Cheeger inequality for directed graphs, *Annals of Combinatorics*, **9** (2005), 1–19.

[14] F. Chung, *Spectral Graph Theory,* AMS Publications, 1997.

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

[16] 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.

[17] F. Chung, M. Kempton, and W. Zhao, Ranking and sparsifying a connection graph. *Journal of Internet Mathematics*, 10 (2014), 87-115.

[18] F. Chung and M. Kempton, A local clustering algorithm for connection graphs. *Proceedings of WAW 2013*, LNCS 8305, 26-43.

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

[20] 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.

[21] 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.

[22] K. Feng and W.-C. W. Li, Spectra of hypergraphs and applications. *Journal of Number Theory*, 60(1), (1996) 1–22.

[23] 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.

[24] 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.

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

[26] 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.

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

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

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

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

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

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

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

[34] 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.

[35] M. Kotani and T. Sunada, Zeta functions of finite graphs, *J. Math. Sci. Univ. Tokyo* , **7** (2000), 7-25.

[36] 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.

[37] W.-C. W. Li and P. Solé, Spectra of regular graphs and hypergraphs and orthogonal polynomials. *European Journal of Combinatorics*, 17(5), (1996) 461–477.

[38] L. Lovász, Random walks on graphs: a survey, *Combinatorics, Paul Erdős is Eighty (Volume 2),* Keszthely (Hungary) (1993) pp.1-46.

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

[40] A. Nilli, On the second eigenvallue of a graph. *Discrete Math.*, 91 (1991), 207-210.

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

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

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

[44] 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.

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

[46] J. Spencer and L. Florescu, *Asymptopia*, Student Mathematical Library, Volume 71, AMS Publications, 2014.

[47] 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.

[48] 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.

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

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

[51] Z.-W. Sun, Congruences involving generalized central trinomial coefficients, *Sci. China Math.*, **57** (2014), no. 7, 1375-1400.

[52] R.M. Tanner, Explicit construction of concentrators from generalized N-gons, *SIAM J. Algebraic Discrete Methods*, **5** (1984), 287–294.

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

[54] 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.

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

[56] S. Wagner, Asymptotics of generalized trinomial coefficients, 2012, Available online at http://arxiv.org/abs/1205.5402v3.

[57] 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.

[58] D. West, *Introduction to Graph Theory*, 2nd Edition, 2001 Prentice Hall, Inc. Upper Saddle River, NJ 07458.

[59] W. Zhao, Ranking and sparsifying edges of a graph. Ph.D. Thesis, University of California, San Diego, 2012.