UNIVERSITY OF CALIFORNIA, SAN DIEGO

Eigenvalues and Structures of Graphs

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

in

Mathematics

by

Steven Kay Butler

Committee in charge:

Professor Fan Chung Graham, Chair
Professor Samuel Buss
Professor Ronald Graham
Professor Russell Impagliazzo
Professor Jacques Verstraete

2008

Copyright
Steven Kay Butler, 2008
All rights reserved.

The dissertation of Steven Kay Butler is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Fan Chung Graham and Ron Graham who have given me so much of their time, support, knowledge and occasional use of their basement. I also express my thanks to Wayne Barrett who first introduced me to the subject of spectral graph theory. I am also grateful to my family for their patience over the years, and particularly my wife Zou Yang, who perhaps has been the most patient of all. Since my wife and I met eight years ago I have constantly been in school at one level or another leading her to believe I had become a perpetual student. This dissertation will finally mark the end of being a perpetual student and the beginning of being a perpetual student who also happens to teach.

Part of this dissertation (particularly Chapter 1) is based on unpublished lecture notes for a series of talks I gave at the Center for Combinatorics at Nankai University in Tianjin, China. I thank the center and in particular my host Bill Chen for giving me the opportunity to visit the center and interact with the students.

Chapter 3 is based on my paper "Relating singular values and discrepancy of weighted directed graphs" published in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (Miami, FL, 2006).* An improved version was subsequently published as "Using discrepancy to control singular values for nonnegative matrices" in *Linear Algebra and its Applications* (2006).

Chapter 4 is based on two papers, "Eigenvalues of 2-edge-coverings" and "Cospectral graphs for both the adjacency and normalized Laplacian matrices" which are both currently submitted.

The first half of Chapter 5 is based on my paper "Interlacing for weighted graphs using the normalized Laplacian" in *Electronic Journal of Linear Algebra* (2007).

VITA

2001	B. S. in Mathematics and Economics, Brigham Young University
2001-2003	Instructor, Brigham Young University
2003	M. S. in Mathematics, Brigham Young University
2003-2008	Graduate Teaching Assistant/Research Assistant, University of California, San Diego
2007	Instructor, University of California, San Diego
2008	Ph. D. in Mathematics, University of California, San Diego

PUBLICATIONS

S. Butler and F. Chung, "Small spectral gap in the combinatorial Laplacian implies Hamiltonian", to appear in *Annals of Combinatorics*.

S. Butler and R. Graham, "Enumerating (multiplex) juggling sequences", to appear in *Annals of Combinatorics*.

S. Butler, M. Hajiaghayi, R. Kleinberg and T. Leighton, "Hat guessing games", *SIAM Journal on Discrete Mathematics*, 22, 2008, pp. 592–605.

M. Bousquet-Melou and S. Butler , "Forest-like permutations", *Annals of Combinatorics*, 11, 2007, pp. 335–354.

S. Butler, R. Graham and J. Mao, "How to play the majority game with liars", *AAIM 2007, Lecture Notes in Computer Science 4508*, Springer-Verlag, Heidelberg, M.-Y. Kao and X.-Y. Li, eds., 2007, pp. 221–230.

S. Butler, "Interlacing for weighted graphs using the normalized Laplacian", *Electronic Journal of Linear Algebra*, 16, 2007, 90–98.

S. Butler, "Using discrepancy to control singular values for nonnegative matrices", *Linear Algebra and its Applications*, 419, 2006, pp. 486–493.

S. Butler, "Relating singular values and discrepancy of weighted directed graphs", *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (Miami, FL, 2006)*, SIAM, Philadelphia, PA, 2006, pp. 1112–1116.

S. Butler, "Tangent line transformations", *The College Mathematics Journal*, 34, 2003, pp. 105–106.

ABSTRACT OF THE DISSERTATION

Eigenvalues and Structures of Graphs

by

Steven Kay Butler
Doctor of Philosophy in Mathematics
University of California San Diego, 2008
Professor Fan Chung Graham, Chair

Given a graph we can associate several matrices which record information about vertices and how they are interconnected. The question then arises, given that you know the eigenvalues of some matrix associated with the graph, what can you say about the graph? Spectral graph theory looks at answering questions of this type.

In this dissertation we will be focusing on the eigenvalues of the normalized Laplacian of a matrix which is defined as $\mathcal{L} = D^{-1/2}(D - A)D^{-1/2}$ where D is the diagonal matrix of degrees and A is the adjacency matrix of the graph. In addition to some background material on spectral graph theory we will be looking at three main results about how eigenvalues and structures of graphs are interrelated. These are as follows.

- For any graph (including directed graphs) the edge discrepancy is a measurement of how randomly the edges are placed. While it has been known for some time that for undirected graphs that a tight clustering of eigenvalues around 1 implies a good measure of discrepancy, only recently has some progress been made in the other direction. We will show that for any graph (including directed graphs) that a small discrepancy implies a tight clustering of singular values of the normalized adjacency matrix. This shows that having small discrepancy and a tight clustering of singular values are in the same quasirandom class of properties for directed graphs.

- Graphs which share common local structure tend to share eigenvalues. We will consider one type of covering that preserves local structures, namely 2-edge-coverings

which, as the name strongly suggests, is a mapping from a graph G to a graph H so that each edge in H is twice covered. We show how to compute the eigenvalues of G from the eigenvalues of two modified forms of H. As an application we give a construction of two graphs which are not regular but are cospectral with respect to both the adjacency and normalized Laplacian matrix.

- Given a graph G, the removal of a small graph will have an effect on the eigenvalues of the graph. We will show that the new eigenvalues will interlace the old eigenvalues (with the size of the interlacing dependent on the number of vertices in the graph which is removed). We will also mention some negative results about interlacing and a normalized Laplacian which has been introduced for directed graphs.

1 The basic matrices of spectral graph theory

1.1 Introduction

Given a large graph it would be useful to be able to take a small snapshot that can concisely capture information about the graph. One of the most useful ways of doing this has been by studying the various spectra of matrices (i.e., the eigenvalues of the matrices) that can be associated with the graph. By looking at these eigenvalues it is possible to get information about a graph that might otherwise be difficult to obtain.

The study of the relations between eigenvalues and structures in graphs is the heart of spectral graph theory. Thus someone interested in using spectral graph theory needs to be familiar both with graph theory and the basic tools of linear algebra including eigenvalues, eigenvectors, determinants, the Courant-Fischer Theorem, the Perron-Frobenius Theorem and so on. We will make use of all these tools throughout the following chapters.

In this chapter we will introduce the three most common matrices associated with graphs (namely the adjacency matrix, the combinatorial Laplacian, and the normalized Laplacian), and give some simple examples for each about how the eigenvalues can be used to give some information about the graph. We will also give some comparison between the combinatorial Laplacian and the normalized Laplacian. In the following chapters we will mainly focus on the normalized Laplacian and give some more specific results about how to use that spectrum to find properties of the graph, and, on the other hand how we can sometimes use the structure of a graph to help find eigenvalues.

Throughout this chapter we will make use of the graph in Figure 1.1 as a basic example, and will give the spectrum of this graph for each of the three matrices.

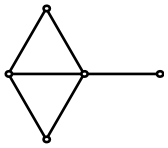

Figure 1.1: A simple example of a graph.

1.2 The adjacency matrix

Given a graph G we can form a matrix A, called the adjacency matrix, by letting the vertices index the columns and rows, and then letting

$$A_{i,j} = \begin{cases} 1 & \text{if } i \text{ is adjacent to } j; \\ 0 & \text{if } i \text{ is not adjacent to } j. \end{cases}$$

For example one adjacency matrix for the graph in Figure 1.1 is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

In some sense the adjacency matrix is not unique, because we can relabel the vertices of the graph which would cause a simultaneous permutation of the rows and columns. So for example we could also have gotten the following matrix as an adjacency matrix of the graph,

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

It should be noted however that the eigenvalues of the two matrices will always be the same for any relabeling. And if we tie eigenfunctions (i.e., the eigenvectors which can be thought of as functions on the vertices) to the vertices it will also follow that eigenfunctions are independent of the choice of labeling. In particular, we have the following.

Observation. *If the eigenvalues of two graphs do not match, then the graphs are not isomorphic.*

The converse is not true. That is there exists graphs which are not isomorphic but do have the same eigenvalues (in general showing two graphs are not isomorphic is a nontrivial problem, or so we currently believe). One such example is the Saltire pair shown in Figure 1.2. We will look at cospectral graphs in more detail in Section 4.6.

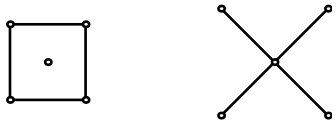

Figure 1.2: Two nonisomorphic graphs whose adjacency matrices have eigenvalues $-2, 0, 0, 0, 2$.

For the graph in Figure 1.2 a calculation (for which computers are especially well suited for) shows that the eigenvalues for the graph are:

$$2.68554393\ldots,\ 0.33490398\ldots,\ 0,\ -1.27133037\ldots,\ -1.74911754\ldots. \tag{1.1}$$

When dealing with the spectrum of a graph we are dealing with the set of eigenvalues. Most frequently, it is more correct to say that we are dealing with a multi-set, i.e., we allow for repetitions. So for example in the graphs given in Figure 1.1 we have 0 listed as an eigenvalue three times.

All the eigenvalues given in (1.1) are real. This follows from the fact that the adjacency matrix is symmetric which in turn follows from the fact that the graph is undirected. In addition, since the adjacency matrix is symmetric, the matrix has a full set of eigenvectors which are mutually orthogonal. This latter fact is useful in some applications. These are the main reasons that the vast majority of results in spectral graph theory deal with undirected graphs. We will look at directed graphs in Section 1.7 and Chapter 3.

In the adjacency matrix we have used 0s and 1s. This is useful for emphasizing the discrete nature of the graph. However, more generally we can allow for other entries besides 0s and 1s. The differing entries can than be thought of as weights on the edges, usually denoted $w(u, v)$, so that $A_{u,v} = w(u, v)$. Many results relating to the spectra can be easily generalized to their "weighted" versions (which includes, for example, multigraphs).

1.2.1 Estimating the number of walks of length k

We can use the eigenvalues of the adjacency matrix to count the number of walks of length k. To do this, first notice that by the rules of matrix multiplication we have

$$(A^k)_{i,j} = \sum_{i_1,i_2,\ldots,i_{k-1}} A_{i,i_1} A_{i_1,i_2} \cdots A_{i_{k-2},i_{k-1}} A_{i_{k-1},j}.$$

Looking at the term in the summand we see that $A_{i,i_1} A_{i_1,i_2} \cdots A_{i_{k-2},i_{k-1}} A_{i_{k-1},j}$ will be 1 if and only if vertex i is adjacent to i_1 which is adjacent to i_2 and so on until we get to j. In other words, the term in the summand is 1 if and only if it corresponds to a walk of length k starting at vertex i and ending at vertex j. So we have the following.

Lemma 1. *Let A be the adjacency matrix of G. Then $(A^k)_{i,j}$ is the number of walks of length k starting at vertex i and ending at vertex j.*

Let us now use the lemma with the eigenvalues of the adjacency matrix. We will make use of the following two facts. First, the trace of the determinant is the sum of the eigenvalues of the matrix. Secondly, the eigenvalues of A^k are the eigenvalues of A raised to the kth power. In what follows we will let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of A.

- $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 0$. This follows by noting that the sum of the eigenvalues is the trace of the adjacency matrix which is 0 since A is 0 on the diagonal.
- $\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_n^2 = 2E(G)$, where $E(G)$ is the number of edges of the graph. This follows by noting that the sum of the squares of the eigenvalues is the same as the trace of A^2. The diagonal entries of A^2 count the number of closed walks of length 2 (a closed walk is a walk that starts and ends at the same vertex; since we are on the diagonal the starting and ending vertices are the same), for which each edge is counted exactly twice.
- $\lambda_1^3 + \lambda_2^3 + \cdots + \lambda_n^3 = 6T(G)$, where $T(G)$ is the number of triangles of the graph. This follows by noting that the sum of the cubes of the eigenvalues is the same as the trace of A^3, i.e., the same as the number of closed walks of length 3. Each triangle will be counted exactly six times (i.e., a choice of 3 initial vertices and 2 directions for each triangle).

This process can be continued, but starts to become impractical for getting useful information about a graph.

Adding up the eigenvalues, the eigenvalues squared, and the eigenvalues cubed as given in (1.1) gives respectively, 0, 12, and 12 as expected from the above and the structure of the graph in Figure 1.1.

Next suppose that we wanted to count the total number of walks of length k. This can be achieved by considering $\mathbf{1}^T A^k \mathbf{1}$, where $\mathbf{1}$ is an all 1s vector of size n and we use T in an exponent to denote the transpose of a matrix. Since A is symmetric then, as noted above, we have a full set of (real) orthonormal eigenvectors. So let ϕ_i be the eigenvector associated with λ_i. Then for some appropriate constants a_i we have that $\mathbf{1} = \sum_i a_i \phi_i$. Putting this in for $\mathbf{1}$ we have that the total number of walks of length k is

$$\left(\sum_i a_i \phi_i^T\right) A^k \left(\sum_i a_i \phi_i\right) = \left(\sum_i a_i \phi_i^T\right)\left(\sum_i a_i \lambda_i^k \phi_i\right) = \sum_i a_i^2 \lambda_i^k.$$

Lemma 2. *Given a connected, non-bipartite graph G, the number of walks of length k (for k very large) is $\approx a_n^2 \lambda_n^k$.*

Without loss of generality we can assume that the graph has an edge, from which it easily follows that $\lambda_n > 0$. Now we have

$$\lim_{k\to\infty} \frac{\mathbf{1}^T A^k \mathbf{1}}{\lambda_n^k} = \lim_{k\to\infty} \sum_i a_i^2 \frac{\lambda_i^k}{\lambda_n^k} = a_n^2.$$

In the last step we used that $|\lambda_i| < \lambda_n$ for $i \neq n$. The latter statement is an easy consequence of the Perron-Frobenius Theorem. This shows that λ_n tells us the growth rate for the number of walks of length k in non-bipartite graphs (this constant is sometimes referred to as "capacity").

Note in the above derivation that the only eigenvalue which was important was the largest eigenvalue. In spectral graph theory almost all of the focus and energy has been put into only the few largest and few lowest eigenvalues, the middle range of the spectra being usually neglected. This is an indication of how little we understand about the spectrum of graphs. On the other hand, as we will see in Section 5.5 this is not unreasonable for properties which can easily be changed by the addition or deletion of a few edges (i.e., being bipartite).

1.3 The combinatorial Laplacian

The second type of matrix that we will consider is the combinatorial Laplacian matrix, denoted as L. In some settings this is referred to as the Laplacian, however unless otherwise specified when we refer to the Laplacian we will mean the normalized Laplacian. Here we have added the term combinatorial to help distinguish the two types of Laplacian matrices that we will consider. We will see that the combinatorial Laplacian is involved in an interesting enumeration problem (some partial justification for the term combinatorial).

Again we let the vertices index the columns and rows and define L entrywise as follows:

$$L_{i,j} = \begin{cases} d_i & \text{if } i = j; \\ -1 & \text{if } i \text{ is adjacent to } j; \\ 0 & \text{otherwise,} \end{cases}$$

where d_i is the degree of the ith vertex. This is closely related to the adjacency matrix and is sometimes written as $L = D - A$, where D is the diagonal matrix with the degree on the diagonals and A is the adjacency matrix.

For the graph in Figure 1.1 the combinatorial Laplacian will be

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & -1 & 0 & -1 & 2 \end{pmatrix},$$

with eigenvalues

$$5,\ 4,\ 2,\ 1,\ 0. \tag{1.2}$$

The fact that all the eigenvalues are integers is coincidence and does not hold in general. However there is one special eigenvalue, namely 0. The fact that 0 is always an eigenvalue is easy to see by noting that all of the row sums are 0, i.e., $\mathbf{1}$ is an eigenvector for the eigenvalue 0. All the other eigenvalues are nonnegative, so in other words the combinatorial Laplacian is positive semi-definite. This follows from, for example, the Gershgorin Disc Theorem. This can also be proven using the incidence matrix which we will introduce below.

That $\mathbf{1}$ is an eigenvector turns out to be very useful in some applications. We will make use of a similar idea for the normalized Laplacian in Chapter 3. For the

adjacency matrix, **1** is an eigenvector if and only if the graph is regular. This is one reason why spectral results are usually first proved for regular graphs (i.e., for regular graphs we have good control on the eigenvectors, because we know the "most important" eigenvector).

The combinatorial Laplacian is associated with the incidence matrix. The incidence matrix, which we will denote C, has rows indexed by the vertices and columns indexed by the edges. To define the entries, given an edge $e = \{i, j\}$ then in the column corresponding to e we put in 1 and -1 for the vertices to which the edge is incident and 0s otherwise, i.e.,

$$C = \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} \begin{array}{c} e \\ \vdots \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ \vdots \end{array} \begin{pmatrix} & & \\ & 1 & \cdots\cdots \\ & & \\ & -1 & \cdots\cdots \\ & & \end{pmatrix},$$

which entry is negative turns out to be unimportant for our purposes.

The relation between C and L is that $L = CC^T$. To see this we note that $(CC^T)_{i,j}$ can be found by taking the inner product of the ith and jth row of C. It is easy to check that for $i \neq j$ that this inner product is 0 if there is no edge and -1 otherwise, while for $i = j$ we add 1 for each edge incident to i, i.e., we get d_i.

If instead of using ± 1 in C we had only used 1, the resulting matrix would be the unsigned Laplacian, also known as the quasi-Laplacian, which differs from the combinatorial Laplacian in that the off-diagonal entries are positive. This matrix has not been as extensively studied.

We can use this representation of L to now show that the all of the eigenvalues are nonnegative. Suppose that σ is an eigenvalue with a (real) normal eigenvector ϕ. Then

$$\sigma = \phi^T(\sigma\phi) = \phi^T L\phi = \phi^T CC^T\phi = (C^T\phi)^T C^T\phi = \|C^T\phi\|^2 \geq 0.$$

1.3.1 The Matrix Tree Theorem

The most interesting result related to the combinatorial Laplacian (and perhaps its most important tool) is in counting the number of spanning trees for connected graphs. A spanning tree of a graph G is, as its name strongly suggests, a subgraph of G which

is a tree and is incident to all the vertices.

Theorem 3 (Matrix Tree Theorem)**.** *Let G be a graph, and $0 = \sigma_0 \leq \sigma_1 \leq \cdots \leq \sigma_{n-1}$ be the eigenvalues of the combinatorial Laplacian of G. Then the number of spanning trees of G is given by*

$$\frac{\sigma_1 \sigma_2 \cdots \sigma_{n-1}}{n}.$$

As an example, taking the eigenvalues given in (1.2) for the graph in Figure 1.1 we would expect $5 \cdot 4 \cdot 2 \cdot 1 / 5 = 8$ spanning trees. It is easy to check that this is in the case and we show them below.

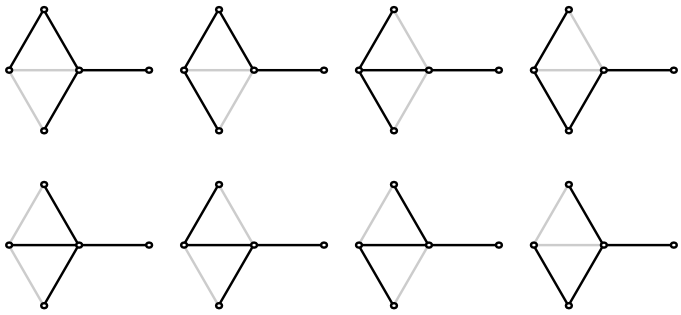

Figure 1.3: The eight spanning trees of the graph in Figure 1.1.

We will not include a proof of the Matrix Tree Theorem here but will instead give a sketch of the ideas. The first thing we note is that the Matrix Tree Theorem is usually stated as saying that any cofactor of the matrix is the number of spanning trees. Knowing this it can be shown that the coefficient of the first degree term in the characteristic polynomial is n times the number of spanning trees (since this coefficient is found by adding the n cofactors from the diagonal). On the other hand the coefficient of the first degree term is also found by summing all possible products of $n-1$ of the n eigenvalues. Since 0 is an eigenvalue only one of these will be nonzero and the result follows.

So now let us compute the cofactor for the ith diagonal term. This involves taking the determinant of L_0 where we have removed the ith row and column of L. It is easy to check that $L_0 = C_0 C_0^T$ where C_0 is C with the ith row removed. Using the Cauchy-Binet formula we have

$$\det(L_0) = \det\left(C_0 C_0^T\right) = \sum_{\substack{X \subseteq E \\ |X| = n-1}} \det\left(C_X C_X^T\right) = \sum_{\substack{X \subseteq E \\ |X| = n-1}} \left(\det(C_X)\right)^2,$$

where C_X is the matrix formed by the columns in C_0 corresponding to X. The remainder of the proof is to show that

$$\det(C_X) = \begin{cases} 0 & \text{if the edges corresponding to } X \text{ have a cycle;} \\ \pm 1 & \text{if the edges corresponding to } X \text{ form a tree.} \end{cases}$$

The first statement easily follows by finding a linear dependence in the columns of B_X. The second one takes more work and we will not attempt it here, but proofs can be found in several books on the subject such as Brualdi and Ryser [4].

1.4 The normalized Laplacian

The final type of matrix that we will consider is the normalized Laplacian matrix, denoted $\mathcal{L}$. As the name suggests this is closely related to the combinatorial Laplacian that we have just looked at. For graphs with no isolated vertices the relationship is given by $\mathcal{L} = D^{-1/2}LD^{-1/2} = D^{-1/2}(D-A)D^{-1/2} = I - D^{-1/2}AD^{-1/2}$. (Throughout the rest of this and ensuing chapters we will usually assume no isolated vertices since they contribute little more than technicalities to the arguments.) Entrywise we have,

$$\mathcal{L}_{i,j} = \begin{cases} 1 & \text{if } i = j; \\ \dfrac{-1}{\sqrt{d_i d_j}} & \text{if } i \text{ is adjacent to } j; \\ 0 & \text{otherwise.} \end{cases}$$

For graphs with isolated vertices we let the diagonal entries of that vertex be 0. This gives the nice property that the multiplicity of the eigenvalue 0 is the number of connected components of the graph.

For the graph in Figure 1.1 the normalized Laplacian will be

$$\mathcal{L} = \begin{pmatrix} 1 & \frac{-1}{2} & 0 & 0 & 0 \\ \frac{-1}{2} & 1 & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{8}} \\ 0 & \frac{-1}{\sqrt{8}} & 1 & \frac{-1}{\sqrt{6}} & 0 \\ 0 & \frac{-1}{\sqrt{12}} & \frac{-1}{\sqrt{6}} & 1 & \frac{-1}{\sqrt{6}} \\ 0 & \frac{-1}{\sqrt{8}} & 0 & \frac{-1}{\sqrt{6}} & 1 \end{pmatrix},$$

with eigenvalues

$$1.72871355\ldots,\ 1.5,\ 1,\ 0.77128644\ldots,\ 0. \tag{1.3}$$

As before we have that 0 is an eigenvalue (now with eigenvector $D^{1/2}\mathbf{1}$) and the remaining eigenvalues are nonnegative. A major difference between the two spectra though is that while for the combinatorial Laplacian the eigenvalues can be essentially as large as desired (in particular between 0 and twice the maximum degree), the normalized Laplacian has eigenvalues always lying in the range between 0 and 2 inclusive as shown by Chung [11].

One advantage to this is that it makes it easier to compare the distribution of the eigenvalues for two different graphs, especially if there is a large difference in the "size" of the graphs.

The normalized Laplacian has connections with many interesting properties of graphs. For now we will consider the problem of random walks.

1.4.1 Random walks—A rambling introduction

A random walk on a graph G can be thought of as a walk where we start at a vertex on the graph and at each time step pick randomly (in our case uniformly) one of the edges incident to the current vertex and go along that edge to the next vertex, repeating as often as desired.

As an example, consider the problem of shuffling cards. In this setting the graph is all possible ways to arrange a deck (a large graph!) and the edges represent shuffles, i.e., starting with a deck of cards which orderings can be reached using one shuffle. In this case a random walk corresponds to doing a random sequence of shuffles.

One problem of interest for people shuffling cards is how many times do we need to shuffle until the cards are sufficiently "random". In this setting "random" can be taken to mean as saying that knowing the initial configuration of cards before starting the shuffling will not give you any significant information about the current placement of cards (or put another way, all of the initial information has been lost). The study of random walks on graphs can help answer such questions.

To study this problem we can keep track of the probability distribution of the various destinations after k steps. In other words, what is the probability that we are at some particular vertex in the graph after k steps.

Pictorially, imagine that we have a cup full of water at our initial vertex, and

empty cups everywhere else (see Figure 1.4). At each step we will simultaneously redistribute the water at each vertex to the vertices neighbors, and thus overtime the water should diffuse throughout the graph. In our problem the water represents the probability distribution and the fractional amount of water at a vertex at k steps is the probability that we are at that vertex in k steps.

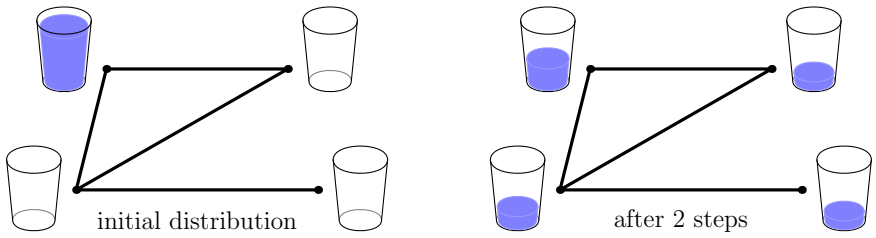

Figure 1.4: A "watered" down approach to random walks.

The question that we want to ask is how many steps does it take before we are sufficiently random. As a first step we should decide what is meant by random. We will say that a probability distribution is random if the probability of being at any vertex is proportional to its degree. More explicitly, we are random if the probability of being at vertex i is exactly equal to $d_i / \sum_{\ell \in G} d_\ell$.

1.4.2 Convergence of random walks—The technical approach

In what follows we will change our convention slightly in that all of the vectors will be treated as *row* vectors (so now we will be focusing on multiplication on the left hand side).

Putting the above discussion in terms of matrices, if A is the adjacency matrix, then $D^{-1}A$ is the "probability transition matrix". That is,

$$(D^{-1}A)_{i,j} = \begin{cases} 0 & \text{if } i \text{ is not adjacent to } j; \\ \dfrac{1}{d_i} & \text{if } i \text{ is adjacent to } j; \end{cases}$$

is the probability that given you are at vertex i you move to vertex j. An initial probability distribution will be a vector f, and the probability distribution after k steps will be $f(D^{-1}A)^k$.

The two requirements to be a probability distribution is that all of the entries are nonnegative and sum to 1. To see that the matrix $D^{-1}A$ takes one probability distribution to another it suffices to check that these two properties are always maintained.

Nonnegativity follows easily since all of the terms are nonnegative. To show that the sum of the entries of a vector g is 1 it suffices to show that $g\mathbf{1}^T = 1$. In our case we have that $fD^{-1}A\mathbf{1}^T = fD^{-1}D\mathbf{1}^T = f\mathbf{1}^T = 1$. The last step follows since we started with a probability distribution.

To see the connection between the probability transition matrix and the normalized Laplacian, note that

$$D^{-1/2}(I - \mathcal{L})D^{1/2} = D^{-1/2}(D^{-1/2}AD^{-1/2})D^{1/2} = D^{-1}A,$$

showing that $D^{-1}A$ and $I - \mathcal{L}$ are similar. As a consequence if λ is an eigenvalue of $\mathcal{L}$ then $1 - \lambda$ is an eigenvalue of $D^{-1}A$. In particular, 1 is always an eigenvalue of $D^{-1}A$ (since 0 is always an eigenvalue of $\mathcal{L}$) and its *left* eigenvector is easily shown to be $\mathbf{1}D$. If we normalize the vector to a probability distribution the resulting vector is

$$\frac{\mathbf{1}D}{\sum_\ell d_\ell},$$

which we refer to as the stationary distribution. This is also the distribution which, from the above discussion, we want our random walk to converge to.

Now let ϕ_i be an orthonormal set of eigenvectors associated with λ_i for $\mathcal{L}$ (here we will follow the convention that $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$). By the above ϕ_i is also associated with $1 - \lambda_i$ for $D^{-1/2}AD^{-1/2}$. It is easy to check that $\phi_0 = \mathbf{1}D^{1/2}/\sqrt{\sum_\ell d_\ell}$. Finally, since we have the full set of eigenvalues and orthonormal eigenvectors, we can use the idea of projections onto eigenspaces to write

$$D^{-1/2}AD^{-1/2} = \sum_i (1 - \lambda_i)\phi_i^T\phi_i.$$

To check how close our random walk is after k steps to the stationary distribution we can use several types of measurements. Here we will choose the L^2-norm, i.e., $\|g\| = \sqrt{\sum_i |g_i|^2}$ (different types of measurements will give different types of bounds for the rate of convergence). In particular, the L^2-norm distance between the random walk

after k steps and the stationary distribution is

$$\begin{aligned}
\left\| f\big(D^{-1}A\big)^k - \frac{\mathbf{1}D}{\sum_\ell d_\ell} \right\| &= \left\| fD^{-1/2}\big(D^{-1/2}AD^{-1/2}\big)^k D^{1/2} - \frac{\mathbf{1}D}{\sum_\ell d_\ell} \right\| \\
&= \left\| fD^{-1/2}\Big(\sum_i (1-\lambda_i)\phi_i^T\phi_i \Big)^k D^{1/2} - \frac{\mathbf{1}D}{\sum_\ell d_\ell} \right\| \\
&= \left\| fD^{-1/2}\big(\sum_i (1-\lambda_i)^k\phi_i^T\phi_i \big) D^{1/2} - \frac{\mathbf{1}D}{\sum_\ell d_\ell} \right\| \\
&= \left\| fD^{-1/2}\big(\sum_{i\neq 0} (1-\lambda_i)^k\phi_i^T\phi_i \big) D^{1/2} \right\| \\
&\leq \max_{i\neq 0} |1-\lambda_i|^k \frac{\max_i \sqrt{d_i}}{\min_j \sqrt{d_j}}.
\end{aligned}$$

Going from the third to the fourth line is an easy calculation to show that the $i = 0$ term in the summand will cancel out with the other expression. In the last step we used matrix norms. In particular, we used that $\|fB\| \leq \|f\|\|B\|$ three times, where $\|f\|$ is the L^2-norm and $\|B\|$ is the operator norm. For the case of symmetric matrices (which all three are) $\|A\|$ is the maximal absolute value of an eigenvalue, and with the matrices as above it is easy to find these values and then get the bound (of course $\|f\| \leq 1$ since it is a probability distribution).

In particular, this tells us that we can use eigenvalues to get an estimate on the rate of convergence of a random walk. The more closely the eigenvalues are gathered around 1 for the normalized Laplacian the faster we should expect to converge to the stationary distribution. This allows us to give an estimate on the number of steps needed to produce random-like results.

As an example, in the graph of Figure 1.1 if we start at any particular vertex and take 10 steps the probability distribution to where we end up will be at most `0.08444976`... away from the stationary distribution in the L^2-norm (though we are most likely much closer). If we wanted to be within `0.000001` of the stationary distribution then using the above results it is easy to check that it will take at most 46 steps.

Since the eigenvalues of the normalized Laplacian are between 0 and 2, we have $\max_{i\neq 0} |1-\lambda_i| \leq 1$. When can this equal 1? The first possibility is that we have 0 as an eigenvalue multiple times. Chung [11] showed that this means that the graph is not connected. The second possibility is that we have 2 as an eigenvalue. Again, Chung [11] showed that this means that the graph has a bipartite component.

Therefore, for a random walk to converge to the stationary distribution it suf-

fices to be on a graph which is connected and not bipartite. In terms of ergodic theory (which can be thought of as the study of mixing systems) the first condition that the graph be connected is equivalent to the requirement that starting at any point in the system any other point can be reached (i.e., transitive) while the second condition that the graph not be bipartite is equivalent to the requirement that the system be aperiodic.

1.5 Comparing different spectrums

As a general rule of which matrix to try we have the following: if the important property that you are using is related to adjacency and enumerating walks then use the adjacency matrix; if you can relate the problem to spanning trees or incidence of vertices and edges use the combinatorial Laplacian; if you can relate the problem to random walks use the normalized Laplacian. Given the recent interest in using random walks to make efficient searches of large databases, the normalized Laplacian has received increased attention in the last few years.

Comparing the eigenvalues in (1.1), (1.2) and (1.3) we see that they can be quite different. In general it will make a big difference as to which spectrum is used, and some results which might hold for one spectrum may not hold for another. There are a few exceptions, the most well known being that for regular graphs the spectrum of all three graphs are shifted/scaled versions of one another. This follows since for regular graphs $L = k\mathcal{L} = k\big(I - \frac{1}{k}A\big)$. (This relationship between the three matrices can make it difficult to know which matrix to use when trying to generalize a result from regular graphs to general graphs.)

Less well known is the fact that for a graph without isolated vertices the multiplicity of 0 as an eigenvalue of the adjacency matrix is the same as the multiplicity of 1 as an eigenvalue of the normalized Laplacian. While the number of negative eigenvalues for the adjacency matrix is the same as the number of eigenvalues of the normalized Laplacian greater than 1, and the number of positive eigenvalues for the adjacency matrix is the same as the number of eigenvalues of the normalized eigenvalues less than 1. This follows by a simple application of Sylvester's Law of Inertia.

While for general graphs there is no simple scaling between the three spectra (and so the spectra can have different "shape"), it can be shown that when the graph is almost regular, the combinatorial Laplacian and normalized Laplacian have similar spectra. More specifically we have the following result.

Theorem 4. *Let G be a graph with $d_{\max}$ the maximum degree of a vertex in G and $d_{\min}$ the minimum degree of a non-isolated vertex in G. Further let ϕ_i and λ_i denote the eigenvalues of the combinatorial and normalized Laplacians (respectively). Then for $0 \leq i \leq n-1$,*

$$\frac{1}{d_{\max}}\phi_i \leq \lambda_i \leq \frac{1}{d_{\min}}\phi_i.$$

As an immediate consequence we have the following result for almost regular graphs.

Corollary 5. *Let G be a graph and d be such that for each vertex the degree d_i satisfies $|d_i - d| = o(d)$. Then $\phi_i = \big(d + o(d)\big)\lambda_i$.*

This can be used to extend some results for regular graphs to results for almost regular graphs.

Proof of Theorem 4. Since isolated vertices contribute 0 to the spectrum for both the combinatorial and normalized Laplacian the result trivially holds for those corresponding 0 eigenvalues. Therefore without loss of generality we may assume that G has no isolated vertices.

The proof of the theorem follows from an easy application of the Courant-Fischer Theorem which will be discussed in more detail in Chapter 5. We first note that

$$\phi_i = \max_{\mathcal{X}^i}\left(\min_{x\perp\mathcal{X}^i, x\neq 0}\frac{x^TLx}{x^Tx}\right) \tag{1.4}$$

while (with a little work)

$$\lambda_i = \max_{\mathcal{X}^i}\left(\min_{x\perp\mathcal{X}^i, x\neq 0}\frac{x^TLx}{x^TDx}\right). \tag{1.5}$$

(Here $\mathcal{X}^i$ denotes an i-dimensional subspace of $\mathbb{R}^n$.)

The other ingredient is to note that $x^Tx = \sum_i |x_i|^2$ and $x^TDx = \sum_i |x_i|^2 d_i$, from which it easily follows that

$$d_{\min}x^Tx \leq x^TDx \leq d_{\max}x^Tx.$$

We now get the key relationship that for any $x \neq 0$

$$\frac{1}{d_{\max}}\frac{x^TLx}{x^Tx} \leq \frac{x^TLx}{x^TDx} \leq \frac{1}{d_{\min}}\frac{x^TLx}{x^Tx}. \tag{1.6}$$

(Here we also used the fact that L is positive semidefinite so that $x^TLx \geq 0$ for all x and thus preserving the inequalities.)

We now show how to establish that $(1/d_{\max})\phi_i \leq \lambda_i$. Fix an i-dimensional subspace $\mathcal{X}^i$. Then there exists some vector y achieving the minimum in the middle optimization of (1.5) (such a y exists by a compactness argument), and using that y it follows that:

$$\min_{x \perp \mathcal{X}_i, x \neq 0} \frac{x^T L x}{x^T D x} = \frac{y^T L y}{y^T D y} \geq \frac{1}{d_{\max}} \frac{y^T L y}{y^T y} \geq \frac{1}{d_{\max}} \min_{x \perp \mathcal{X}_i, x \neq 0} \frac{x^T L x}{x^T x}. \tag{1.7}$$

Since this holds for *every* $\mathcal{X}^i$, let $\mathcal{Y}$ be an i-dimensional subspace which maximizes the right hand side of (1.7). Then we have

$$\begin{aligned} \frac{1}{d_{\max}} \phi_i = \frac{1}{d_{\max}} \max_{\mathcal{X}^i} \left(\min_{x \perp \mathcal{X}^i, x \neq 0} \frac{x^T L x}{x^T x} \right) &= \frac{1}{d_{\max}} \min_{x \perp \mathcal{Y}, x \neq 0} \frac{x^T L x}{x^T x} \\ &\leq \min_{x \perp \mathcal{Y}, x \neq 0} \frac{x^T L x}{x^T D x} \leq \max_{\mathcal{X}^i} \left(\min_{x \perp \mathcal{X}^i, x \neq 0} \frac{x^T L x}{x^T D x} \right) = \lambda_i. \end{aligned}$$

The proof for the other inequality is handled similarly. □

1.6 Generalizing from regular to general graphs

In the literature, the combinatorial Laplacian has been more widely used than the normalized Laplacian. There are several reasons for this; the first is that the normalized Laplacian is a rather new tool (popularized by Chung [11] in the early 1990s). The second is the somewhat (at first glance) unnatural definition of the normalized Laplacian. The third, as already mentioned, is that most new theories start out by considering regular graphs for which it does not matter much which definition is used as the spectrum of the matrices differ only by a scaling factor.

The last reason though is perhaps the most dangerous, for when you try to generalize results to non-regular graphs it becomes very important which spectrum you are considering, since you lose symmetry in the degree sequence which can cripple some tools. One of the original motivations for the definition of the normalized Laplacian was to give a more natural way of weighing the vertices of non-regular graphs.

Roughly speaking, in the combinatorial Laplacian each vertex is given equal weight (and so many results related to the combinatorial Laplacian involve terms like $|X|$, the number of vertices in a subset X of the vertices). On the other hand the normalized Laplacian gives each vertex a weight proportional to its degree (and so many results related to the normalized Laplacian involve $\operatorname{vol} X = \sum_{x \in X} d_x$ which we call the volume of the subset X). For many applications weighing vertices by their degrees is

more natural and it is in this setting that the normalized Laplacian can give better results.

As an example, Goldberg [22] considered the problem of bounding the gap between the nontrivial largest and smallest eigenvalues of the combinatorial Laplacian. His approach was based on the following two "reverse Cauchy-Schwarz" inequalities.

Theorem 6 (Pólya-Szegö [30]). *Let $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_m)$ be positive m-tuples satisfying $0 < \alpha \leq a_i \leq A$ and $0 < \beta \leq b_i \leq B$ for all i. Then*

$$\frac{\sum a_i^2 \sum b_i^2}{\big(\sum a_i b_i\big)^2} \leq \frac{1}{4}\big(\sqrt{\tfrac{AB}{\alpha\beta}} + \sqrt{\tfrac{\alpha\beta}{AB}}\,\big)^2.$$

Theorem 7 (Ozeki [29]). *Let $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_m)$ be positive m-tuples satisfying $0 < \alpha \leq a_i \leq A$ and $0 < \beta \leq b_i \leq B$ for all i. Then*

$$\sum a_i^2 \sum b_j^2 - \big(\sum a_i b_i\big)^2 \leq \frac{1}{4} m^2 (AB - \alpha\beta)^2.$$

By using the preceding two theorems on a k-regular graph with the two sequences $(k, k, \ldots, k)$ (which corresponds to the degree sequence with a vertex removed) and $(\lambda_1, \lambda_2, \ldots, \lambda_{n-1})$ (the nontrivial eigenvalues), with a little bit of bookkeeping, Goldberg established the following result.

Corollary 8. *Let G be a connected k-regular graph (so that $\lambda_1 > 0$) and $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of the Laplacian. Then the following holds:*

(a) $\sqrt{\frac{\lambda_{n-1}}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_{n-1}}} \geq 2\sqrt{\big(1 - \frac{1}{n}\big)\big(1 + \frac{1}{k}\big)}$.

(b) $\lambda_{n-1} - \lambda_1 \geq 2\sqrt{\frac{n-k-1}{nk}}$.

The problem in generalizing this approach to general graphs is that if the graph is not regular the degree sequence is not regular and there is no clear way to relate the degree and eigenvalue sequences together.

On the other hand by working with a normalized Laplacian we can avoid the irregularity in the degree sequence altogether. We have the following general result which for the case of the graph being regular reduces to that given by Goldberg.

Corollary 9. *Let G be a connected graph (so that $\lambda_1 > 0$) and $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of the Laplacian. Then the following holds:*

(a) $\sqrt{\frac{\lambda_{n-1}}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_{n-1}}} \geq 2\sqrt{\big(1 - \frac{1}{n}\big)\big(1 + \frac{1}{n}\mathbf{1}^* D^{-1} A D^{-1} \mathbf{1}\big)}$.

(b) $\lambda_{n-1} - \lambda_1 \geq \frac{2}{n-1}\sqrt{(n-1)\mathbf{1}^* D^{-1} A D^{-1}\mathbf{1} - n}$.

Proof. We first note that $\sum \lambda_i = \operatorname{trace} \mathcal{L} = n$ and

$$\sum_{i=1}^{n-1} \lambda_i^2 = \sum_{i=0}^{n-1} \lambda_i^2 = \operatorname{Tr}(\mathcal{L}^2) = n + \sum_x \sum_{y \sim x} \frac{1}{d_y d_x} = n + \mathbf{1}^* D^{-1} A D^{-1} \mathbf{1}.$$

Now (a) follows by using Theorem 6 and (b) follows by using Theorem 7 with the sequences $(\lambda_1, \ldots, \lambda_{n-1})$ and $(1, \ldots, 1)$ (the "normalized" degree sequence). □

1.7 Associating graphs with matrices for directed graphs

While the majority of spectral graph theory has dealt with undirected graphs (due to having real eigenvalues and a complete set of eigenvectors), matrices can be associated with directed graphs and information about the graph can be found by studying the properties of these matrices.

For the adjacency matrix we still have a similar definition except now we have $A_{u,v} = 1$ if and only if there is a directed edge from u to v (denoted $u \to v$). A generalization of the Laplacian (combinatorial or normalized) to directed graphs has not yet been well realized. The best attempt was given by Chung [12], and we will look at this in more detail in Section 5.4.

Since the eigenvalues can now be complex it is sometimes preferable to work with the singular values of the matrix which will always be real and nonnegative. (For symmetric matrices the singular values are the absolute values of the eigenvalues and so the two are essentially equivalent. Looking at some proofs stated in terms of eigenvalues it is clear that it is singular values which are being used.)

As a simple example comparing and contrasting eigenvalues and singular values we will count two types of walks on directed graphs. We will denote a walk by $W = (v_1, v_2, \ldots, v_{k+1})$ (where the v_i need not all be distinct). The first walk is a directed walk where $v_i \to v_{i+1}$ for $i = 1, \ldots, k$ (i.e., the intuitive definition for a directed walk). The second type of walk is an alternating directed walk where $v_i \to v_{i+1}$ if i is odd and $v_i \leftarrow v_{i+1}$ if i is even (i.e., a walk where at each step we reverse the direction we walk). For undirected graphs there is no distinction between the two but for directed graphs we will see that there is a distinction.

1.7.1 Counting the number of directed walks

Our tool for counting the number of directed walks is to start with the adjacency matrix A, defined as earlier. We will let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of the adjacency matrix.

As in the undirected case, we have that $\left(A^k\right)_{u,v}$ is the number of directed walks of length k starting at vertex u and ending at vertex v. Since the terms on the diagonal count the number of closed walks of length k, it follows that the total number of closed walks of length k is $\lambda_1^k + \cdots + \lambda_n^k$. But not only can we use the eigenvalues to count the number of closed walks, we can also use the eigenvalues to approximate the total number of walks.

Theorem 10. *Let G be a strongly connected acyclic directed graph (or digraph) and let λ_1 be the unique largest (in norm) eigenvalue of the adjacency matrix of G. Then there exists a constant c so that the total number of walks of length k is $\approx c\lambda_1^k$.*

A graph is strongly connected if there is a directed walk joining any two vertices and is acyclic if the greatest common divisor of all closed walks is 1. These two assumptions correspond to a system which is transitive and aperiodic (in other words an ergodic system). Moreover, it is known (by use of the Perron-Frobenius Theorem) that for such a graph there exists a largest *real* eigenvalue λ_1 so that $\lambda_1 > |\lambda_i|$ for $i \neq 1$.

Proof. By an application of Schur's Theorem, along with the preceding comments, there exists a matrix S so that $A = S^{-1}TS$, where

$$T = \begin{pmatrix} \lambda_1 & O \\ O & T' \end{pmatrix},$$

with T' triangular with the diagonal entries corresponding to the eigenvalues (and in particular in size strictly bounded by λ_1). We now note that the total number of walks of length k is $\mathbf{1}^* A^k \mathbf{1}$ where $\mathbf{1}$ is the all 1s vector of size n.

We now have the following:

$$\frac{\mathbf{1}^* A^k \mathbf{1}}{\lambda_1^k} = \mathbf{1}^* S^{-1} \left(\frac{1}{\lambda_1} T\right)^k S\mathbf{1} \approx \mathbf{1}^* S^{-1} \begin{pmatrix} 1 & O \\ O & O \end{pmatrix} S\mathbf{1} = c$$

where c is the product of the sum of the entries on the first row of S and the sum of the entries on the first column of S^{-1}. (Note that the constant c by the description of the problem is both real and independent of S.)

Finally it remains to justify the "≈" given above. This follows by noting that

$$\left(\frac{1}{\lambda_1}T\right)^k = \begin{pmatrix} 1 & O \\ O & (1/\lambda_1)T' \end{pmatrix}^k = \begin{pmatrix} 1 & O \\ O & \left((1/\lambda_1)T'\right)^k \end{pmatrix}.$$

Since T' is an upper triangular matrix with all entries strictly bounded by λ_1 then $(1/\lambda_1)T'$ has diagonal entries strictly bounded in size by 1. It is an easy exercise to then check that such a matrix converges entrywise to the 0 matrix when taking high powers, concluding the proof. □

1.7.2 Counting the number of alternating walks

For alternating walks we deal with matrices of the form $AA^*AA^*AA^*\cdots$. Since the singular values can be found by computing the eigenvalues of AA^*, the singular values work quite nicely for this kind of walk. As an example, to count the number of closed alternating walks of length $2k$ this is the trace of $(AA^*)^k$ and so in particular is $\sigma_1^{2k} + \sigma_2^{2k} + \cdots + \sigma_n^{2k}$, where σ_i are the singular values of A. One open problem is how to count the number of closed alternating walks of length $2k+1$, since in such a case the sum of the powers of singular values might not be integer and so do not give the correct result.

As with eigenvalues we can use singular values to count the number of alternating walks. In this case however the proof is simpler.

Theorem 11. *Let G be a digraph and let σ_1 be the largest singular value of the adjacency matrix of G. Then there exists constants c_{odd} and c_{even} so that the total number of alternating walks of length k is $\approx c_{odd}\sigma_1^k$ if k is odd and $\approx c_{even}\sigma_1^k$ if k is even.*

Note here we do not restrict our digraphs. This is partially because the singular values are always nonnegative. But also partially because even with the assumption of strongly connected and acyclic there is no guarantee for uniqueness in the largest singular value. (As an example, the graph shown in Figure 1.5 is strongly connected and acyclic but has repeated largest singular value of $\sqrt{(3+\sqrt{5})/2}$.)

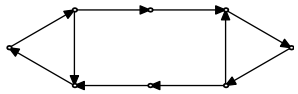

Figure 1.5: A directed graph with repeated largest singular value.

Proof. By the singular value decomposition theorem we can write $A = UDV$ where U and V are unitary and D is a diagonal matrix with entries the singular values of A. Let $u_1, \ldots, u_n$ be the orthonormal columns of U and similarly $v_1, \ldots, v_n$ be the orthonormal rows of V.

By matrix multiplication we have that the number of alternating walks of length k is

$$\mathbf{1}^* \underbrace{AA^*AA^* \cdots A}_{k \text{ terms, } k \text{ odd}} \mathbf{1} = \mathbf{1}^* U D^k V \mathbf{1} = \sum_i \mathbf{1}^* u_i \sigma_i^k v_i \mathbf{1}.$$

Suppose that $\sigma_1 = \sigma_2 = \cdots = \sigma_\ell > \sigma_\ell \geq \cdots$. Then for large k

$$\frac{\sum_i \mathbf{1}^* u_i \sigma_i^k v_i \mathbf{1}}{\sigma_1^k} \approx \sum_{i=1}^{\ell} \mathbf{1}^* u_i v_i \mathbf{1} = c_{odd}.$$

Similarly, we have that

$$\mathbf{1}^* \underbrace{AA^*AA^* \cdots A^*}_{k \text{ terms, } k \text{ even}} \mathbf{1} = \mathbf{1}^* U D^k U^* \mathbf{1} = \sum_i |\mathbf{1}^* u_i|^2 \sigma_i^k,$$

so that for large k

$$\frac{\sum_i |\mathbf{1}^* u_i|^2 \sigma_i^k}{\sigma_1^k} \approx \sum_{i=1}^{\ell} |\mathbf{1}^* u_i|^2 = c_{even}. \qquad \square$$

The constants c_{odd} and c_{even} may not be the same.

Acknowledgement

Sections 1.1 through 1.4 are based on unpublished lecture notes for a series of talks given at the Center for Combinatorics at Nankai University in Tianjin, China. I thank the center and in particular my host Bill Chen for giving me the opportunity to visit the center and interact with the students.

2 Using eigenvectors to compute eigenvalues

The eigenvalues of the graph are closely tied to the eigenvectors of a graph. While you do not always need to find the eigenvectors to find eigenvalues, in some cases using the structure of a graph we can construct the eigenvectors and thus find the eigenvalues. In this chapter we will lightly touch upon this theme and return to it in Chapter 4.

It should also be noted that eigenvectors can sometimes contain useful information. For instance Chung [11] used the eigenvector corresponding to the first nontrivial eigenvalue to help find a "good" cut of the vertices. There also is a recent monograph dealing with eigenvectors for the combinatorial Laplacian (see [3]).

2.1 Relationship of eigenvectors to eigenvalues

When we think of an eigenvector it is good to think of it as a function on the vertices. We can then translate the matrix definition for the eigenvector-eigenvalue relationship to a (local) function on graphs. For instance if we start with the adjacency matrix (where we now allow for weighted graphs so that $A_{u,v} = A_{v,u} = w(u,v)$) then the relationship $A\mathbf{x} = \lambda\mathbf{x}$ translates into the condition that at each vertex v

$$\sum_{u:u\sim v} w(u,v)\mathbf{x}(u) = \lambda\mathbf{x}(v). \tag{2.1}$$

For the normalized Laplacian instead of focusing on the eigenvectors it is easier to use the *harmonic* eigenvectors. Namely if $\mathbf{y}$ is an eigenvector of $\mathcal{L}$ associated with λ then the harmonic eigenvector is $\mathbf{x} = D^{-1/2}\mathbf{y}$. (The harmonic eigenvectors allow us to "hide" the square root terms inside the eigenvector and so simplify some of the expressions.)

The relationship $\mathcal{L}\mathbf{y} = \lambda\mathbf{y}$ then becomes $D^{-1/2}L\mathbf{x} = \lambda D^{1/2}\mathbf{x}$. Multiplying both sides by $D^{1/2}$ it then becomes $L\mathbf{x} = \lambda D\mathbf{x}$. Looking at the vth coordinate of both sides we get the following condition at each vertex v:

$$d(v)\mathbf{x}(v) - \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) = \lambda d(v)\mathbf{x}(v). \tag{2.2}$$

By showing that one of these relationships holds for a given value of λ (and a nonzero $\mathbf{x}$) then we have that λ is an eigenvalue (for the appropriate matrix) of the graph.

2.2 Cyclic graphs

For graphs with "nice" structure we can often find the eigenvalues by exploiting the graph's structure. Much work in spectral graph theory has focused on Cayley graphs and related structures which are derived from group relationships. Using the structure of these groups it is possible to find the eigenvectors explicitly; this area of spectral graph theory would be best described as algebraic graph theory (see [20]).

One very simple example in this direction is cyclic graphs. A cyclic graph on n vertices can be generated by taking $\mathbf{Z}_n$ and taking a generator set $S = \{a_1, a_2, \ldots, a_k\}$ which is closed under inversion (to make sure that the graph is undirected). The vertex set are the elements of the group and we connect elements a and b if $a - b \in S$. So for instance if we let $S = \{1, -1\}$ then the resulting graph is a cycle on n vertices denoted C_n, while if we let $S = \{1, 2, \ldots, n-1\}$ then the resulting graph is the complete graph on n vertices denoted K_n.

For a cyclic graph the eigenvectors take the form $(1, \theta, \theta^2, \ldots, \theta^{n-1})^*$ where θ is an nth root of unity, i.e., $\theta = e^{2\pi ik/n}$ for some $k \in \{0, 1, \ldots, n-1\}$. Then applying relationship (2.1) for the adjacency matrix we have that at vertex v that

$$\sum_{u:u\sim v} \mathbf{x}(u) = \sum_{t\in S} \mathbf{x}(v+t) = \sum_{t\in S} \theta^{v+t} = \big(\sum_{t\in S} \theta^t\big)\mathbf{x}(v),$$

showing that $\sum_{t\in S} \theta^t$ is an eigenvalue for each root of unity.

So if our graph is a cycle on n vertices then the eigenvalues are

$$\theta + \theta^{-1} = e^{2\pi ik/n} + e^{-2\pi ik/n} = 2\cos\big(\frac{2\pi k}{n}\big) \qquad \text{for } k = 0, 1, \ldots, n-1.$$

Similarly if our graph is a complete graph on n vertices then the eigenvalues are

$$\theta + \theta^2 + \cdots + \theta^{n-1} = \begin{cases} n-1 & \text{if } \theta = 1; \\ -1 & \text{otherwise.} \end{cases}$$

This uses the fact that the sum of all nth roots of unity is 0.

Since cyclic graphs are all regular of degree $|S|$ it is easy to translate the eigenvalues of such graphs into any of the three general matrices. So for instance the eigenvalues of the cycle on n vertices for the normalized Laplacian are

$$1 - \cos\left(\frac{2\pi k}{n}\right) \qquad \text{for } k = 0, 1, \ldots, n-1,$$

while the eigenvalues of the complete graph on n vertices for the normalized Laplacian are

$$0, \underbrace{\frac{n}{n-1}, \ldots, \frac{n}{n-1}}_{\times(n-1)}.$$

2.3 Constructing larger graphs

There are several ways to combine smaller graphs to form larger graphs. The eigenvalues of the smaller graphs then can (sometimes) be used to find the eigenvalues of the newly constructed graph. A well known example is given two graphs $G = (V_1, E_1)$ and $H = (V_2, E_2)$ to take the Cartesian product, denoted $G \,\square\, H$. The graph $G \,\square\, H$ has as its vertex set $\{(u, v) : u \in V_1, v \in V_2\}$ and (u_1, v_1) is adjacent to (u_2, v_2) if either $u_1 = u_2$ and v_1 is adjacent to v_2 in H or $v_1 = v_2$ and u_1 is adjacent to u_2 in G. One simple example of a graph that can be formed using Cartesian products is the hypercube Q_n, which can be defined inductively as $Q_1 = K_2$ and $Q_n = Q_{n-1} \,\square\, K_2$.

It is well known that if $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of the adjacency matrix of G and $\beta_1, \ldots, \beta_m$ are the eigenvalues of the adjacency matrix of H then the eigenvalues of $G \,\square\, H$ are $\alpha_i + \beta_j$ over all possible i and j. The proof of this is to use the eigenvectors of G and H to construct new eigenvectors for $G \,\square\, H$.

Unfortunately, the same result does not hold for the normalized Laplacian. For instance it is well known (and we will show shortly) that $C_4 = K_{2,2}$ and $K_{1,3}$ are cospectral (i.e., share the same set of eigenvalues), but the spectrum of the normalized Laplacian of $K_2 \,\square\, C_4$ is $\{0, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{4}{3}, \frac{4}{3}, \frac{4}{3}, 2\}$ while the spectrum of the normalized Laplacian of $K_2 \,\square\, K_{1,3}$ is $\{0, \frac{1}{2}, \frac{1}{2}, \frac{3}{4}, \frac{5}{4}, \frac{3}{2}, \frac{3}{2}, 2\}$. In particular the eigenvalues of the normalized

Laplacian for a Cartesian product cannot be determined solely by the eigenvalues of the two graphs of the product.

One obvious exception to this is for the Cartesian product of two regular graphs. Since the Cartesian product will again produce a regular graph then the spectra of the adjacency and normalized Laplacian have clear relationships (as mentioned in Section 1.5). So in that sense it is not particularly interesting.

There is a construction combining two graphs that works for both the adjacency matrix and the normalized Laplacian. Given graphs $G = (V_1, E_1)$ and $H = (V_2, E_2)$ let $G \equiv H$ be the graph with vertex set $V_1 \cup V_2$ and edge set $E_1 \cup E_2 \cup \big\{\{u, v\} : u \in V_1, v \in V_2\big\}$. Pictorially we take a copy of G and a copy of H and then connect each vertex of G to each vertex of H.

Theorem 12. *Let G be an r-regular graph on n vertices and H an s-regular graph on m vertices, neither with loops.*

- *If $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}, \alpha_n = r$ are the eigenvalues of the adjacency matrix of G and $\beta_1, \beta_2, \ldots, \beta_{m-1}, \beta_m = s$ are the eigenvalues of the adjacency matrix of H, then the eigenvalues of the adjacency matrix of $G \equiv H$ are*

$$\alpha_1, \ldots, \alpha_{n-1}, \beta_1, \ldots, \beta_{m-1}, \frac{(r+s) \pm \sqrt{(r-s)^2 + 4mn}}{2}.$$

- *If $0 = \lambda_0, \lambda_1, \ldots, \lambda_{n-1}$ are the eigenvalues of the normalized Laplacian matrix of G and $0 = \varphi_0, \varphi_1, \ldots, \varphi_{m-1}$ are the eigenvalues of the normalized Laplacian matrix of H, then the eigenvalues of the normalized Laplacian of $G \equiv H$ are*

$$0, \frac{m + r\lambda_1}{m+r}, \ldots, \frac{m + r\lambda_{n-1}}{m+r}, \frac{n + s\varphi_1}{n+s}, \ldots, \frac{n + s\varphi_{m-1}}{n+s}, 2 - \frac{r}{m+r} - \frac{s}{n+s}.$$

Before giving the proof let us make two observations. First, while G and H are regular the graph $G \equiv H$ is typically not. So the two spectrums are not related by the usual trivial scale and shift as happens in regular graphs. In other words we will need to prove the result for both spectrums. The other thing is that it is important that the graphs G and H are regular. This is because we know the most important eigenvector (for both spectrums) is the all 1s eigenvector which we denote $\mathbf{1}$. Since the matrices are symmetric we can assume that every other eigenvector for G or H is orthogonal to this one. In particular, the sum of all of the entries of every other eigenvector is 0. This forms the heart of the argument.

Proof. Let us first work through the case of the adjacency matrix. Given $\mathbf{x}_i$ an eigenvector of α_i for the graph G (with $i < n$) we extend this to a vector $\mathbf{x}'_i$ on the graph $G \equiv H$ by letting

$$\mathbf{x}'_i(v) = \begin{cases} \mathbf{x}_i(v) & \text{if } v \in V(G); \\ 0 & \text{if } v \in V(H). \end{cases}$$

We now claim that this is an eigenvector associated with eigenvalue α for $G \equiv H$. To see this we have that for a vertex v in $V(G)$ that

$$\sum_{u:u\sim v} \mathbf{x}'_i(u) = \sum_{\substack{u\sim v \\ u\in V(G)}} \mathbf{x}'_i(u) + \sum_{u\in V(H)} \mathbf{x}'_i(u) = \sum_{\substack{u\sim v \\ u\in V(G)}} \mathbf{x}_i(u) = \alpha_i \mathbf{x}_i(v) = \alpha_i \mathbf{x}'_i(v).$$

While for a vertex v in $V(H)$ we have

$$\sum_{u:u\sim v} \mathbf{x}'_i(u) = \sum_{u\in V(G)} \mathbf{x}'_i(u) + \sum_{\substack{u\sim v \\ u\in V(H)}} \mathbf{x}'_i(u) = 0 + 0 = \alpha_i \mathbf{x}'_i(v).$$

(Here we used the observation above that the sum of the entries in the eigenvector is 0.) This shows that $\alpha_1, \ldots, \alpha_{n-1}$ are also eigenvalues of $G \equiv H$. By the exact same construction we also have that $\beta_1, \ldots, \beta_{m-1}$ are also eigenvalues of $G \equiv H$. That leaves us with two eigenvalues left to determine, for these we can use the basic properties of eigenvalues given in Section 1.2.1, namely that the sum of all the eigenvalues is 0 and the sum of the square of the eigenvalues is twice the number of edges.

In particular if τ and ρ are the remaining two eigenvalues to determine then we have

$$\underbrace{\alpha_1 + \ldots + \alpha_{n-1}}_{=-r} + \underbrace{\beta_1 + \ldots + \beta_{m-1}}_{=-s} + \rho + \tau = 0$$

so that $\rho + \tau = r + s$. Similarly we have

$$\underbrace{\alpha_1^2 + \ldots + \alpha_{n-1}^2}_{=nr-r^2} + \underbrace{\beta_1^2 + \ldots + \beta_{m-1}^2}_{=ms-s^2} + \rho^2 + \tau^2 = nr + ms + 2mn$$

so that $\rho^2 + \tau^2 = r^2 + s^2 + 2mn$. Combining these two relationships it is easy to solve for ρ and τ to get that

$$\rho, \tau = \frac{(r+s) \pm \sqrt{(r-s)^2 + 4mn}}{2},$$

which completes the spectrum in the case of the adjacency matrix.

We now turn to the case of the normalized Laplacian. We can use the exact same eigenvector construction as in the adjacency matrix case. So suppose that $\mathbf{x}_i$ is an eigenvector for λ_i (where $i > 0$). So we have that in the graph G

$$-\sum_{u:u\sim v} \mathbf{x}_i(u) = (\lambda_i - 1)r\mathbf{x}_i(v).$$

Then for a vertex v in $V(G)$ (which now has degree $m + r$) we have that

$$\begin{aligned} d(v)\mathbf{x}'_i(v) - \sum_{u:u\sim v} \mathbf{x}'_i(u) &= (m+r)\mathbf{x}'_i(v) - \sum_{\substack{u\sim v \\ u\in V(G)}} \mathbf{x}'_i(u) - \sum_{u\in V(H)} \mathbf{x}'_i(u) \\ &= (m+r)\mathbf{x}_i(v) + (\lambda_i - 1)r\mathbf{x}_i(v) - 0 = \left(\frac{m + r\lambda_i}{m+r}\right) d(v)\mathbf{x}'_i(v). \end{aligned}$$

This shows that $(m + r\lambda_i)/(m + r)$ is an eigenvalue of $G \equiv H$ for $i = 1, \ldots, n - 1$. The same argument will also show that $(n + s\varphi_i)/(n + s)$ are also eigenvalues of $G \equiv H$. That leaves us with two eigenvalues left to determine. One of them must be 0 (since 0 is always an eigenvalue and none of the ones that we have listed above are 0). Then since the sum of all the eigenvalues must be $m + n$ it is easy to check that the remaining eigenvalue must be $2 - r/(m + r) - s/(n + s)$. This completes the spectrum for the normalized Laplacian. □

2.3.1 Some additional graph spectrums

We can use Theorem 12 to find the spectrum of several graphs. For instance we have that the complete bipartite graph $K_{m,n}$ is $(nK_1) \equiv (mK_1)$ where nK_1 is a graph with n isolated vertices (so that it is regular of degree 0). The spectrums of nK_1 and mK_1 consist entirely of 0. So applying the result we have that the spectrum of $K_{m,n}$ is $\{0, 1, \ldots, 1, 2\}$.

As another example of a graph that can be constructed consider $K_{n-2} \equiv (2K_1)$. The resulting graph is K_n with a single edge removed. Applying the result we have that the spectrum is

$$\begin{aligned} &\left\{0, \underbrace{\frac{2 + (n-3)\left(\frac{n-2}{n-3}\right)}{2 + (n-3)}, \ldots, \frac{2 + (n-3)\left(\frac{n-2}{n-3}\right)}{2 + (n-3)}}_{\times(n-3)}, \frac{n-2}{n-2}, 2 - \frac{n-3}{n-1}\right\} \\ &\qquad = \left\{0, 1, \underbrace{\frac{n}{n-1}, \ldots, \frac{n}{n-1}}_{\times(n-3)}, \frac{n+1}{n-1}\right\}. \end{aligned}$$

Similarly one could construct the cocktail party graph on $2n$ vertices, denoted CP_{2n}, which is K_{2n} with a perfect matching removed. (The name of the graph comes from the idea of a set of n couples going to a cocktail party and everyone shakes hands with the people they did not come with. The resulting handshake graph is the cocktail party graph.) It is easy to see that CP_{2n} is $CP_{2(n-1)} \equiv (2K_1)$. It can then be checked with induction that the spectrum of CP_{2n} is

$$\left\{0, \underbrace{1, \ldots, 1}_{\times n}, \underbrace{\frac{n+1}{n}, \ldots, \frac{n+1}{n}}_{\times(n-2)}, \frac{n+2}{n}\right\}.$$

Finally, consider W_{n+1}, the wheel graph on $n+1$ vertices, which is $C_n \equiv K_1$. Namely take an n-cycle and connect each vertex to a new central vertex (pictorially forming something akin to a bicycle tire, hence the name wheel graph). Since we have already calculated the spectrum of the cycle in Section 2.2 we can now easily find the spectrum of the wheel graph using the above result. Namely we have that the spectrum is

$$\left\{0, \underbrace{1 - \frac{2}{3}\cos\frac{2\pi k}{n}}_{k=1,\ldots,n-1}, \frac{4}{3}\right\}.$$

3 Discrepancy of graphs

3.1 Introduction

Random graphs have proven to be a powerful tool in graph theory, and have been used in many cases to show existence of certain graphs without actually producing a graph. But while random graphs can be shown to have many "nice" properties, when we are given a specific graph how do we know that it shares any of these properties? Put in another way, how do we know that a specific graph behaves like a random graph?

This requires that we understand how a random graph behaves. One of the most useful properties of random graphs is that there is independence among the edges. So we can try to measure the distribution of edges and try to measure how randomly they were distributed. The discrepancy of a graph is one such way to measure this.

To define discrepancy we begin by counting the number of edges between subsets of vertices. Given subsets X, Y of the vertices of G we have that

$$e(X,Y) = \big|\{(x,y) : x \in X, y \in Y \text{ and } x \sim y\}\big|,$$

where $x \sim y$ means that x and y are adjacent in the graph. Note that $e(x,y)$ can be thought of as an edge indicator function. By convention, edges in $X \cap Y$ are counted twice. We can use the adjacency matrix A to calculate $e(X,Y)$, to do this we let ψ_X denote the characteristic vector of X, i.e.,

$$\psi_X(u) = \begin{cases} 1 & \text{if } u \text{ in } X; \\ 0 & \text{else;} \end{cases}$$

then $e(X,Y) = \langle \psi_X, A\psi_Y \rangle$. Recall that the volume of a subset of vertices X, denoted by $\operatorname{vol} X$, is the sum of its degrees, i.e., $\operatorname{vol} X := \sum_{u \in X} d_u = \|D^{1/2}\psi_X\|^2$. (Throughout this chapter we will use $\|\cdot\|$ to denote the L_2-norm.)

The discrepancy between two subsets X and Y is defined by

$$\operatorname{disc}(X,Y) = \left| e(X,Y) - \frac{\operatorname{vol} X \ \operatorname{vol} Y}{\operatorname{vol} G} \right|.$$

Intuitively this is a measure of the difference between the actual number of edges and the expected number of edges. One way to see where the $(\operatorname{vol} X \ \operatorname{vol} Y)/\operatorname{vol} G$ term comes from is to consider a *weighted* random graph model in which we give each vertex u weight d_u so the probability of including edge $\{u,v\}$ is $d_u d_v/\operatorname{vol} G$. With this model it can be checked (under some simple assumptions) that the expected number of edges between X and Y is $(\operatorname{vol} X \ \operatorname{vol} Y)/\operatorname{vol} G$.

For regular graphs this becomes $\operatorname{disc}(X,Y) = \left| e(X,Y) - \rho |X||Y| \right|$ where ρ is the edge density of the graph. The definition we have given here provides a generalization to non-regular graphs where we put our measure on edges instead of on vertices.

The discrepancy of G, which we denote $\operatorname{disc} G$, will be the minimal β such that

$$\left| e(X,Y) - \frac{\operatorname{vol} X \ \operatorname{vol} Y}{\operatorname{vol} G} \right| \leq \beta \sqrt{\operatorname{vol} X \ \operatorname{vol} Y}.$$

Counting the number of edges joining vertices in X to vertices in Y is the same as counting the number of walks of length one joining a vertex in X to a vertex in Y. This suggests a generalization to looking at the number of (weighted) walks of length t that join vertices in X to vertices in Y. This will lead to the t-discrepancy of G which we will denote by $\operatorname{disc}_t G$.

We first define a weight function for walks (following the convention of [14]). If $P = (x_0, x_1, \ldots, x_t)$ is a walk of length t (i.e., we have $x_i \sim x_{i+1}$ for $0 \leq i < t$) we define the weight of such a walk by $w(P) = \prod_{0<i<t} (1/d_{x_i})$; in the case that $P = (x_0, x_1)$ (i.e., the walk is a single edge) we define $w(P) = 1$. Let $P_t(X,Y)$ be the set of all walks of length t starting at a vertex in X and ending at a vertex in Y. Then we define

$$e_t(X,Y) = \sum_{P \in P_t(X,Y)} w(P).$$

In the case that G is a d-regular graph we have that

$$e_t(X,Y) = \frac{\left| \left\{ \begin{array}{c} \text{walks of length } t \text{ joining a} \\ \text{vertex in } X \text{ to a vertex in } Y \end{array} \right\} \right|}{d^{t-1}}.$$

Given a graph G we will let $\operatorname{disc}_t(G)$ denote the minimal β such that

$$\left| e_t(X,Y) - \frac{\operatorname{vol} X \ \operatorname{vol} Y}{\operatorname{vol} G} \right| \leq \beta \sqrt{\operatorname{vol} X \ \operatorname{vol} Y}.$$

Note that $\operatorname{disc} G = \operatorname{disc}_1 G$.

There are many ways that we could have defined the weight of a path which would lead to different definitions of the t-discrepancy. The weight we have chosen here is useful in that it allows for an easy generalization of proofs involving discrepancy to t-discrepancy. This is a consequence of the following fact that is easily checked,

$$e_t(X,Y) = \langle \psi_X, D^{1/2}(D^{-1/2}AD^{-1/2})^t D^{1/2}\psi_Y \rangle,$$

where here and throughout we will let $\langle x, y\rangle$ denote the inner product of two vectors.

The definition of $e_t(X,Y)$ has another interpretation. Let the probabilities on G be distributed proportionally, i.e., the probability of randomly choosing vertex u is $d_u/\operatorname{vol} G$ and the probability of moving from vertex u to vertex v is $e(u,v)/d_u$. Then $e_t(X,Y)/\operatorname{vol} G$ is the probability that a randomly generated walk of length t joins a vertex in X to a vertex in Y. As a consequence we have that $e_t(X,V) = \operatorname{vol} X$ (i.e., $e_t(X,V)/\operatorname{vol} G$ reduces to the probability of starting in X which is $\operatorname{vol} X/\operatorname{vol} G$). This can also be shown directly with the relationship given above.

3.2 Matrix form of discrepancy

The result for discrepancy on graphs will follow from a discrepancy result on matrices, which we now turn our attention to. Bollobás and Nikiforov [5] showed that there is a constant C so that for any Hermitian matrix $A = (a_{ij})_{n\times n}$, $\sigma_2(A) \le C \operatorname{Disc}(A) \log n$ where $\sigma_2(A)$ is the second singular value of A and $\operatorname{Disc}(A)$ is the minimal α so that for all $S, T \subseteq [n]$

$$\left| \Big(\sum_{i\in S}\sum_{j\in T} a_{ij}\Big) - \rho|S|\,|T| \right| \le \alpha\sqrt{|S||T|} \quad \text{where} \quad \rho = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij}. \tag{3.1}$$

Their approach was to approximate the vector associated with $\sigma_2(A)$ as a linear combination of at most $C' \log n$ 0-1 vectors for a constant C' depending only on how close the approximation needs to be. They then used this approximation to collapse the matrix A and get the desired result.

Seperately, Bilu and Linial [2] showed (among other things) that for the special case when A is the adjacency matrix of a d-regular (undirected) graph that $\sigma_2(A) \le O\big(\alpha\big(1 + \log(d/\alpha)\big)\big)$. Their approach also involved an approximation of the vector, but this time the entries of the approximation were powers of 2, and instead of collapsing the matrix they used some clever manipulation of the sums.

We will combine the approximation ideas of Bollobás-Nikiforov [5] and the manipulation of the resulting sums as in Bilu-Linial [2] to obtain a discrepancy result which we state below for nonnegative (not necessarily square) matrices. The proofs of these results will be given in Section 3.3 and the interpretation to directed graphs will be given in Section 3.4.

We will let J denote the matrix of all 1s.

Theorem 13. *Let $B \in M_{m\times n}$ be a matrix with nonnegative entries and no zero rows or columns. Also, let $R \in M_{m\times m}$ and $C \in M_{n\times n}$ be the unique diagonal matrices such that $B\mathbf{1} = R\mathbf{1}$ and $\mathbf{1}B = \mathbf{1}C$. Then for all $S \subseteq [m]$ and $T \subseteq [n]$*

$$\left|\langle \psi_S, B\psi_T\rangle - \frac{\langle \psi_S, B\mathbf{1}\rangle \langle \mathbf{1}, B\psi_T\rangle}{\langle \mathbf{1}, B\mathbf{1}\rangle}\right| \leq \sigma_2(R^{-1/2}BC^{-1/2})\sqrt{\langle \psi_S, B\mathbf{1}\rangle \langle \mathbf{1}, B\psi_T\rangle}.$$

Note that the diagonal entries of R and C are the row sums and column sums (respectively) of B.

Theorem 14. *Let B, R, C be as above. If for all $S \subseteq [m]$ and $T \subseteq [n]$*

$$\left|\langle \psi_S, B\psi_T\rangle - \frac{\langle \psi_S, B\mathbf{1}\rangle \langle \mathbf{1}, B\psi_T\rangle}{\langle \mathbf{1}, B\mathbf{1}\rangle}\right| \leq \alpha\sqrt{\langle \psi_S, B\mathbf{1}\rangle \langle \mathbf{1}, B\psi_T\rangle} \tag{3.2}$$

(we can and will assume that $\alpha \leq 1$), then

$$\sigma_2(R^{-1/2}BC^{-1/2}) \leq 150\alpha(1 - 8\log\alpha).$$

The minimal α satisfying equation (3.2) is a discrepancy of A which we denote by $\mathrm{disc}(A)$. For nonnegative Hermitian matrices the difference between (3.1) and (3.2) can be viewed as how rows/columns are weighted. In (3.1) each row is given equal weight, and so the important measurement is the number of rows, while in (3.2) each row is weighted according to its row sum, and so the important measurement is the sum of the row sums (similarly for the columns). It is because of this different approach that we need to normalize the matrix A by multiplying on the left by $R^{-1/2}$ and on the right by $C^{-1/2}$.

3.3 Proofs of matrix form of discrepancy

Before beginning our proofs we note the following:

(i) $(R^{-1/2}BC^{-1/2})C^{1/2}\mathbf{1} = R^{1/2}\mathbf{1}$.

(ii) $\mathbf{1}R^{1/2}(R^{-1/2}BC^{-1/2}) = \mathbf{1}C^{1/2}$.

(iii) $\sigma_1(R^{-1/2}BC^{-1/2}) = 1$.

(iv) $\sigma_2(R^{-1/2}BC^{-1/2}) = \sigma_1\big(R^{-1/2}BC^{-1/2} - \frac{1}{\langle \mathbf{1}, B\mathbf{1}\rangle} R^{1/2}JC^{1/2}\big)$.

Equalities (i) and (ii) are an easy calculation, while (iii) follows by the Perron-Frobenius Theorem on $(R^{-1/2}BC^{-1/2})^*(R^{-1/2}BC^{-1/2})$ which has eigenvector $C^{1/2}\mathbf{1}$ associated with eigenvalue 1. For (iv) we subtract out the largest singular value which by (i)-(iii) has left and right vectors $\mathbf{1}R^{1/2}$ and $C^{1/2}\mathbf{1}$ respectively, and noting that $\|\mathbf{1}R^{1/2}\|^2 = \|C^{1/2}\mathbf{1}\|^2 = \langle \mathbf{1}, B\mathbf{1}\rangle$.

Proof of Theorem 13. This follows from $|\langle x, My\rangle| \le \sigma_1(M)\|x\|\,\|y\|$ (see [31]), i.e.,

$$\begin{aligned}
\big|\langle \psi_S, B\psi_T\rangle - \frac{\langle \psi_S, B\mathbf{1}\rangle \langle \mathbf{1}, B\psi_T\rangle}{\langle \mathbf{1}, B\mathbf{1}\rangle}\big| \;&=\; \big|\big\langle \psi_S, \big(B - \frac{RJC}{\langle \mathbf{1}, B\mathbf{1}\rangle}\big)\psi_T\big\rangle\big| \\
&=\; \big|\big\langle R^{1/2}\psi_S, \big(R^{-1/2}BC^{-1/2} - \frac{R^{1/2}JC^{1/2}}{\langle \mathbf{1}, B\mathbf{1}\rangle}\big)C^{1/2}\psi_T\big\rangle\big| \\
&\le\; \sigma_1\big(R^{-1/2}BC^{-1/2} - \frac{R^{1/2}JC^{1/2}}{\langle \mathbf{1}, B\mathbf{1}\rangle}\big)\|R^{1/2}\psi_S\|\,\|C^{1/2}\psi_T\|.
\end{aligned}$$

A calculation shows that $\|R^{1/2}\psi_S\|^2 = \langle \psi_S, B\mathbf{1}\rangle$ and $\|C^{1/2}\psi_T\|^2 = \langle \mathbf{1}, B\psi_T\rangle$, which with the above comments concludes the proof. □

For Theorem 14 we need the following approximation lemmas.

Lemma 15. *Let $x \in \mathbb{C}^n$ with $\|x\| = 1$, and D a diagonal matrix with positive entries, d_t, on the diagonal. Then there is a vector $y \in \mathbb{C}^n$ such that $\|Dy\| \le 1$, $\|x - Dy\| \le \frac{1}{3}$ and the nonzero entries of y are of the form $\big(\frac{4}{5}\big)^j e^{2\pi ik/29}$ for j, k integers with $0 \le k < 29$.*

Proof. Let $x = (x_t)_{1\le t\le n}$ then we define $y = (y_t)_{1\le t\le n}$ entrywise. If $x_t = 0$ then set $y_t = 0$. Otherwise for some $r > 0$ and $0 \le \theta < 2\pi$, we have $x_t = re^{i\theta}$. For the unique integer j so that $\big(4/5\big)^j < r/d_t \le \big(4/5\big)^{j-1}$, set $y_t = \big(\frac{4}{5}\big)^j e^{2\pi i\lfloor 29\theta/2\pi\rfloor/29}$. By construction we have

$$0 < |x_t| - |d_t y_t| \le \Big(\big(\frac{4}{5}\big)^{j-1} - \big(\frac{4}{5}\big)^j\Big)d_t = \big(\frac{5}{4} - 1\big)\big(\frac{4}{5}\big)^j d_t < \frac{1}{4}|x_t|,$$

while the argument between x_t and y_t is bounded above by $2\pi/29$.

By use of the law of cosines it follows that $|x_t - d_t y_t|^2 \le |x_t|^2/9$, which implies $\|x - Dy\|^2 = \sum_t |x_t - d_t y_t|^2 \le \frac{1}{9}\sum_t |x_t|^2 = \frac{1}{9}$. □

Lemma 16. *Let M be a matrix and x', y' vectors such that $\|x'\| = \|y'\| = 1$ and $\sigma_1(M) = |\langle x', My'\rangle|$. If x, y are vectors such that $\|x\|, \|y\| \leq 1$ and $\|x'-x\|$, $\|y'-y\| \leq \frac{1}{3}$, then $\sigma_1(M) \leq \frac{9}{2}|\langle x, My\rangle|$.*

Proof. We again use $|\langle x, My\rangle| \leq \sigma_1(M)\|x\|\,\|y\|$.

$$\begin{aligned}\sigma_1(M) &= |\langle x', My'\rangle| = |\langle x + (x'-x), M\big(y + (y'-y)\big)\rangle| \\ &\leq |\langle x, My\rangle| + |\langle x, M(y'-y)\rangle| + |\langle (x'-x), My\rangle| + |\langle (x'-x), M(y'-y)\rangle| \\ &\leq |\langle x, My\rangle| + \frac{1}{3}\sigma_1(M) + \frac{1}{3}\sigma_1(M) + \frac{1}{9}\sigma_1(M),\end{aligned}$$

rearranging then gives the result. □

Proof of Theorem 14. Let $\mathcal{B} = B - \frac{1}{\langle \mathbf{1}, B\mathbf{1}\rangle} RJC$, so that

$$\sigma_2(R^{-1/2}BC^{-1/2}) = \sigma_1(R^{-1/2}\mathcal{B}C^{-1/2}).$$

There exists vectors x', y' such that $\|x'\| = 1$ and $\|y'\| = 1$ where

$$\sigma_1(R^{-1/2}\mathcal{B}C^{-1/2}) = |\langle x', R^{-1/2}\mathcal{B}C^{-1/2}y'\rangle|.$$

Applying Lemma 15 twice, there exist (step) vectors x, y with $\|x\|, \|y\| \leq 1$, and also satisfy $\|x' - R^{1/2}x\|$, $\|y' - C^{1/2}y\| \leq \frac{1}{3}$. It follows from Lemma 16 that

$$\sigma_1(R^{-1/2}\mathcal{B}C^{-1/2}) \leq \frac{9}{2}|\langle R^{1/2}x, (R^{-1/2}\mathcal{B}C^{-1/2})C^{1/2}y\rangle| = \frac{9}{2}|\langle x, \mathcal{B}y\rangle|.$$

We now partition $[m]$ according to the vector x. Let $X^{(t)} = \{j : |x_j| = \left(\frac{4}{5}\right)^t\}$, and let $x = \sum_t \left(\frac{4}{5}\right)^t x^{(t)}$, where $x^{(t)}$ is the "signed" indicator function of $X^{(t)}$, i.e.,

$$x_j^{(t)} = \begin{cases} x_j/|x_j| & \text{if } |x_j| = \left(\frac{4}{5}\right)^t; \\ 0 & \text{otherwise.}\end{cases}$$

We similarly partition $[n]$ to get $y = \sum_s \left(\frac{4}{5}\right)^s y^{(s)}$. We now have

$$\sigma_2(R^{-1/2}BC^{-1/2}) \leq \frac{9}{2}|\langle x, \mathcal{B}y\rangle| \leq \frac{9}{2}\sum_t\sum_s \left(\frac{4}{5}\right)^{t+s}|\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle|.$$

By assumption, we have for any 0-1 vectors w and z that

$$|\langle w, \mathcal{B}z\rangle| = \Big|\langle w, Bz\rangle - \frac{\langle w, B\mathbf{1}\rangle\,\langle \mathbf{1}, Bz\rangle}{\langle \mathbf{1}, B\mathbf{1}\rangle}\Big| \leq \alpha\sqrt{\langle w, B\mathbf{1}\rangle\,\langle \mathbf{1}, Bz\rangle}.$$

More generally, if $w = \sum_{k=0}^{28} e^{2\pi ik/29} w^{\langle k \rangle}$ and $z = \sum_{\ell=0}^{28} e^{2\pi i\ell/29} z^{\langle \ell \rangle}$ where $w^{\langle k \rangle}, z^{\langle \ell \rangle}$ are 0-1 vectors and the $w^{\langle k \rangle}$ ($z^{\langle \ell \rangle}$) are mutually orthogonal, then by the triangle and Cauchy-Schwarz inequalities we have

$$\begin{aligned}
|\langle w, \mathcal{B}z\rangle| &= |\langle \sum_{k=0}^{28} e^{2\pi ik/29} w^{\langle k \rangle}, \mathcal{B} \sum_{\ell=0}^{28} e^{2\pi i\ell/29} z^{\langle \ell \rangle}\rangle| \\
&\leq \sum_{k=0}^{28}\sum_{\ell=0}^{28} |\langle w^{\langle k \rangle}, \mathcal{B}z^{\langle \ell \rangle}\rangle| \\
&\leq \alpha \sum_{k=0}^{28}\sum_{\ell=0}^{28} \sqrt{\langle w^{\langle k \rangle}, B\mathbf{1}\rangle \langle \mathbf{1}, Bz^{\langle \ell \rangle}\rangle} \\
&\leq 29\alpha \sqrt{\sum_{k=0}^{28}\sum_{\ell=0}^{28} \langle w^{\langle k \rangle}, B\mathbf{1}\rangle \langle \mathbf{1}, Bz^{\langle \ell \rangle}\rangle} \\
&= 29\alpha \sqrt{\langle \sum_{k=0}^{28} w^{\langle k \rangle}, B\mathbf{1}\rangle \langle \mathbf{1}, B\sum_{\ell=0}^{28} z^{\langle \ell \rangle}\rangle} \\
&= 29\alpha \sqrt{\langle |w|, B\mathbf{1}\rangle \langle \mathbf{1}, B|z|\rangle},
\end{aligned}$$

where $|x|$ denotes the vector of the absolute value of the entries of x. Applying this to $x^{(t)}$ and $y^{(s)}$ we have

$$|\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \leq 29\alpha \sqrt{\langle |x^{(t)}|, B\mathbf{1}\rangle \langle \mathbf{1}, B|y^{(s)}|\rangle}. \tag{3.3}$$

We also have that

$$\sum_s |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \leq 2\langle |x^{(t)}|, B\mathbf{1}\rangle \quad \text{and} \quad \sum_t |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \leq 2\langle \mathbf{1}, B|y^{(s)}|\rangle. \tag{3.4}$$

To see this, by the triangle inequality we have $|\langle w, Mz\rangle| \leq \langle |w|, |M|\,|z|\rangle$, and so

$$\begin{aligned}
\sum_s |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| &\leq \langle |x^{(t)}|, |\mathcal{B}| \sum_s |y^{(s)}|\rangle \\
&\leq \langle |x^{(t)}|, \big(B + \frac{RJC}{\langle \mathbf{1}, B\mathbf{1}\rangle}\big)\mathbf{1}\rangle \\
&= 2\langle |x^{(t)}|, B\mathbf{1}\rangle.
\end{aligned}$$

The other result is proved similarly.

We now let $\gamma = \log_{4/5} \alpha$ and consider

$$\begin{aligned}
\sum_t \sum_s \big(\frac{4}{5}\big)^{t+s} |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \;\leq\; & \sum_{|s-t|\leq\gamma} \big(\frac{4}{5}\big)^{t+s} |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \\
& + \sum_t \big(\frac{4}{5}\big)^{2t+\gamma} \sum_s |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| + \sum_s \big(\frac{4}{5}\big)^{2s+\gamma} \sum_t |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle|.
\end{aligned} \tag{3.5}$$

The inequality can be verified by comparing the coefficient of $|\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle|$ on both sides. Clearly when $|s-t| \le \gamma$ the result holds, and when $t > s+\gamma$ then $s+t > 2s+\gamma$ so that $\left(\frac{4}{5}\right)^{s+t} < \left(\frac{4}{5}\right)^{2s+\gamma}$, and similarly when $s > t+\gamma$ then $\left(\frac{4}{5}\right)^{s+t} < \left(\frac{4}{5}\right)^{2t+\gamma}$ and the inequality follows.

We now bound the three terms on the right side of (3.5). For the first term we have

$$\begin{aligned}
&\sum_{|s-t|\le\gamma} \left(\frac{4}{5}\right)^{s+t} |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \\
&\quad\le \frac{29}{2}\alpha \sum_{|s-t|\le\gamma} 2\sqrt{\left(\frac{4}{5}\right)^{2t}\langle |x^{(t)}|, B\mathbf{1}\rangle \left(\frac{4}{5}\right)^{2s}\langle \mathbf{1}, B|y^{(s)}|\rangle} \\
&\quad\le \frac{29}{2}\alpha \sum_{|s-t|\le\gamma} \left(\left(\frac{4}{5}\right)^{2t}\langle |x^{(t)}|, B\mathbf{1}\rangle + \left(\frac{4}{5}\right)^{2s}\langle \mathbf{1}, B|y^{(s)}|\rangle\right) \\
&\quad\le \frac{29}{2}\alpha(2\gamma+1)\left(\sum_t \left(\frac{4}{5}\right)^{2t}\langle |x^{(t)}|, B\mathbf{1}\rangle + \sum_s \left(\frac{4}{5}\right)^{2s}\langle \mathbf{1}, B|y^{(s)}|\rangle\right) \\
&\quad\le 29\alpha(2\gamma+1).
\end{aligned}$$

The inequalities follow from (respectively) (3.3), the geometric-arithmetic mean inequality, the fact that any term can show up at *most* $2\gamma+1$ times, and

$$\begin{aligned}
\sum_t \left(\frac{4}{5}\right)^{2t}\langle |x^{(t)}|, B\mathbf{1}\rangle &= \|R^{1/2}x\|^2 \le 1 \quad \text{and} \\
\sum_s \left(\frac{4}{5}\right)^{2s}\langle \mathbf{1}, B|y^{(s)}|\rangle &= \|C^{1/2}y\|^2 \le 1.
\end{aligned}$$

For the second term we use (3.4) to get

$$\sum_t \left(\frac{4}{5}\right)^{2t+\gamma} \sum_s |\langle x^{(t)}, \mathcal{B}y^{(s)}\rangle| \le 2\left(\frac{4}{5}\right)^{\gamma} \sum_t \left(\frac{4}{5}\right)^{2t}\langle |x^{(t)}|, B\mathbf{1}\rangle \le 2\left(\frac{4}{5}\right)^{\gamma},$$

a similar statement holds for the third term.

Putting this together we have that

$$\sigma_2(R^{-1/2}BC^{-1/2}) \le \frac{9}{2}\left(29\alpha(2\gamma+1) + 4\left(\frac{4}{5}\right)^{\gamma}\right) \le 150\alpha(1-8\log\alpha). \qquad \square$$

3.4 Discrepancy for directed graphs

In this section we consider directed graphs which have a weight function w which assigns $w(u\to v) > 0$ to each edge $u\to v$. As in the undirected case, the weight

function is used to give the adjacency matrix $A = A(G)$ by $A_{u,v} = w(u \to v)$ for all edges $u \to v$ and 0 otherwise. The in- and out-degrees are $d_{in}(u) = \sum_v w(v \to u)$ and $d_{out}(u) = \sum_v w(u \to v)$, respectively the column and row sums of A, and form the entries of the diagonal matrices D_{in} and D_{out}. While the in- and out-volume of subsets X of vertices are $\mathrm{vol}_{in}(X) = \sum_{x \in X} d_{in}(x)$ and $\mathrm{vol}_{out}(X) = \sum_{x \in X} d_{out}(x)$.

We also have a discrepancy for directed graphs, denoted $\mathrm{disc}(G)$, which is the minimal α so that for any subsets X, Y of vertices

$$\left| \Big(\sum_{u \in X} \sum_{v \in Y} w(u \to v) \Big) - \frac{\mathrm{vol}_{out}(X)\, \mathrm{vol}_{in}(Y)}{\mathrm{vol}\,(G)} \right| \leq \alpha \sqrt{\mathrm{vol}_{out}\,(X)\, \mathrm{vol}_{in}(Y)}, \tag{3.6}$$

where $\mathrm{vol}\,(G) := \mathrm{vol}_{in}\,(V) = \mathrm{vol}_{out}(V)$. The discrepancy for a directed graph and of a matrix are related by $\mathrm{disc}(G) = \mathrm{disc}(A(G))$. Applying Theorems 13 and 14 we get the following result.

Theorem 17. *For G a weighted directed graph without sources or sinks,*

$$\mathrm{disc}(G) \leq \sigma_2(D_{out}^{-1/2} A D_{in}^{-1/2}) \leq 150\, \mathrm{disc}(G)\big(1 - 8 \log \mathrm{disc}(G)\big).$$

This shows that for a directed graph having a small second singular value gives control on discrepancy and vice versa. Since an undirected graph can be made into a directed graph by replacing each edge with a pair of directed edges this also shows a relationship for discrepancy for undirected graphs and the size of the second largest eigenvalue (in absolute value).

Another type of discrepancy for graphs is based on $\mathrm{Disc}(A(G))$, the difference between these two discrepancies can be viewed in how a set of vertices are weighted. While in $\mathrm{Disc}(A(G))$ each vertex is given equal weight so that the measure is the number of vertices, in $\mathrm{disc}(A(G))$ the vertices are weighted by their degree so that the measure is the sum of the degrees. This idea of normalizing the weights has been used with great success in spectral techniques by Chung [11].

3.4.1 Alternating walks

Chung and Graham [14] have generalized discrepancy for undirected graphs by considering the discrepancy of walks of length t (the case $t = 1$ gives the original form of discrepancy). There has been limited success in generalizing these results to directed graphs (see [9]). The difficulty seems to lie in that to count walks we look at a matrix such as $AA \cdots A$ (t terms) which works well with eigenvalues but not with

singular values. However, if we consider alternating walks, a walk where at every step we reverse direction, which are counted by a matrix such as $AA^*AA^*\cdots$ (t terms), these do work well with singular values. Here we will consider a discrepancy for alternating walks.

For an alternating walk $P = x_0 \to x_1 \leftarrow x_2 \to x_3 \leftarrow x_4 \cdots x_t$ we associate a weight

$$w(P) = \begin{cases} \dfrac{w(x_0\to x_1)w(x_1\leftarrow x_2)\cdots w(x_{t-1}\leftarrow x_t)}{d_{in}(x_1)d_{out}(x_2)d_{in}(x_3)\cdots d_{in}(x_{t-1})} & t \text{ even;} \\ \dfrac{w(x_0\to x_1)w(x_1\leftarrow x_2)\cdots w(x_{t-1}\to x_t)}{d_{in}(x_1)d_{out}(x_2)d_{in}(x_3)\cdots d_{out}(x_{t-1})} & t \text{ odd.} \end{cases}$$

There is a slight difference between the case t odd and t even, which corresponds to the direction of the last edge.

Let $\mathcal{P}_t(x\to y)$ denote the set of all alternating walks of length t starting at x and ending at y. Then define $w_t(x\to y) = \sum_{P\in\mathcal{P}_t(x\to y)} w(P)$, equivalently, $w_t(x\to y)/\operatorname{vol}(G)$ is the probability that a randomly generated alternating walk of length t starts at x and ends at y.

We now define the discrepancy of alternating t-walks, denoted $\operatorname{AltDisc}_t(G)$, to be the minimal β such that for all $X, Y \subseteq V$

$$\left|\sum_{x\in X}\sum_{y\in Y} w_t(x\to y) - \frac{\operatorname{vol}_{out}(X)\,\operatorname{vol}_{out}(Y)}{\operatorname{vol}(G)}\right| \le \beta\sqrt{\operatorname{vol}_{out}(X)\,\operatorname{vol}_{out}(Y)} \qquad t \text{ even;}$$

$$\left|\sum_{x\in X}\sum_{y\in Y} w_t(x\to y) - \frac{\operatorname{vol}_{out}(X)\,\operatorname{vol}_{in}(Y)}{\operatorname{vol}(G)}\right| \le \beta\sqrt{\operatorname{vol}_{out}(X)\,\operatorname{vol}_{in}(Y)} \qquad t \text{ odd.}$$

Theorem 18. *For G a weighted directed graph without sources or sinks,*

$$\operatorname{AltDisc}_t(G) \le \left(\sigma_2(D_{out}^{-1/2}AD_{in}^{-1/2})\right)^t \le 150\operatorname{AltDisc}_t(G)(1 - 8\log\operatorname{AltDisc}_t(G)).$$

Proof. We consider the case t odd (t even is handled similarly). Let

$$\begin{aligned} B &= AD_{in}^{-1}A^*D_{out}^{-1}AD_{in}^{-1}A^*D_{out}^{-1}\cdots D_{out}^{-1}A \\ &= D_{out}^{1/2}\underbrace{(D_{out}^{-1/2}AD_{in}^{-1/2})(D_{out}^{-1/2}AD_{in}^{-1/2})^*\cdots(D_{out}^{-1/2}AD_{in}^{-1/2})}_{t \text{ terms}}D_{in}^{1/2}. \end{aligned}$$

We have $D_{out} = R$, $D_{in} = C$, $\sum_{x\in X}\sum_{y\in Y} w_t(x\to y) = \langle\psi_X, B\psi_Y\rangle$, $\operatorname{vol}_{out}(X) = \langle\psi_X, B\mathbf{1}\rangle$ and $\operatorname{vol}_{in}(Y) = \langle\mathbf{1}, B\psi_Y\rangle$. From Theorems 13, 14 and the definition of discrepancy for alternating t-walks we have

$$\operatorname{AltDisc}_t(G) \le \sigma_2\big(D_{out}^{-1/2}BD_{in}^{-1/2}\big) \le 150\operatorname{AltDisc}_t(G)\big(1 - 8\log\operatorname{AltDisc}_t(G)\big).$$

It remains to show that $\sigma_2\big(D_{out}^{-1/2}BD_{in}^{-1/2}\big) = \big(\sigma_2(D_{out}^{-1/2}AD_{in}^{-1/2})\big)^t$. But this follows immediately from the definition of B and the fact that for a matrix F, $\sigma_2(\underbrace{FF^*FF^*F\cdots F}_{t \text{ terms}}) = \big(\sigma_2(F)\big)^t$. □

3.4.2 Directed walks for regular directed graphs

Alternating walks might not seem intuitive, and are chosen because of the nature of singular values. There is one case when we can relax our conditions and examine the traditional walk and that is the case of directed graphs where the in-degrees and out-degrees are equal at each vertex.

Let $P = x_0 \to x_1 \to x_2 \to \cdots \to x_t$ be a walk of length t from x_0 to x_t. We define the weight of the walk P by

$$w(P) = \frac{w(x_0 \to x_1)w(x_1 \to x_2)\cdots w(x_{t-1} \to x_t)}{d_{x_1} d_{x_2} \cdots d_{x_{t-1}}}.$$

Note that since the in- and out-degrees are equal we have that $d_u := d_{in}(u) = d_{out}(u)$, $\operatorname{vol} X := \operatorname{vol}_{in} X = \operatorname{vol}_{out} X$, and $D := D_{in} = D_{out}$. If we let $\mathcal{P}_t(X, Y)$ denote the set of all paths of length t joining a vertex in X to a vertex in Y, then the actual sum of weighted paths is

$$e_t(X \to Y) = \sum_{P \in \mathcal{P}_t(X \to Y)} w(P).$$

Note that $e_t(X \to Y)/\operatorname{vol} G$ is the probability that a randomly generated walk of length t starts in X and ends in Y. With $e_t(X \to Y)$ we now define $\operatorname{disc}_t G$ to be the minimal β such that for all $X, Y \subseteq V$

$$\Big|e_t(X \to Y) - \frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} G}\Big| \leq \beta\sqrt{\operatorname{vol} X \operatorname{vol} Y}.$$

With this definition we get the following theorem.

Theorem 19. *Let G be a directed graph where in-degree equals out-degree at each vertex. Then*

$$\operatorname{disc}_t G \leq \big(\sigma_2(D^{-1/2}AD^{-1/2})\big)^t.$$

Further, if the in-degree and out-degree are positive for each vertex in G then

$$|\lambda_2|^t \leq 150 \operatorname{disc}_t G(1 - 8 \log \operatorname{disc}_t G),$$

where $1 = \lambda_1 \geq |\lambda_2| \geq \cdots$ are the eigenvalues of $D^{-1/2}AD^{-1/2}$.

For undirected graphs $|\lambda_2| = \sigma_2(D^{-1/2}AD^{-1/2})$, showing that the distribution of t-walks and the first non-trivial eigenvalue are equivalent in that case.

Proof. From Theorems 13, 14 and the definition of discrepancy we have

$$\operatorname{disc}_t G \leq \sigma_2\big((D^{-1/2}AD^{-1/2})^t\big) \leq \ 150 \operatorname{disc}_t G(1 - 8 \log \operatorname{disc}_t G).$$

We have that $D^{1/2}\mathbf{1}$ is the left and the right vector corresponding to the largest singular value of $(D^{-1/2}AD^{-1/2})^t$ and in this case it is also the left and right eigenvector associated with the largest eigenvalue. From this it follows that

$$\begin{aligned}
\sigma_2\big((D^{-1/2}AD^{-1/2})^t\big) &= \sigma_1\big((D^{-1/2}AD^{-1/2})^t - \frac{1}{\operatorname{vol} G} D^{1/2}JD^{1/2}\big) \\
&= \sigma_1\big((D^{-1/2}AD^{-1/2} - \frac{1}{\operatorname{vol} G} D^{1/2}JD^{1/2})^t\big) \\
&\leq \big(\sigma_1(D^{-1/2}AD^{-1/2} - \frac{1}{\operatorname{vol} G} D^{1/2}JD^{1/2})\big)^t \\
&= \big(\sigma_2(D^{-1/2}AD^{-1/2})\big)^t.
\end{aligned}$$

On the other hand, by biorthogonality we have that the right eigenvector of $(D^{-1/2}AD^{-1/2})^t$ corresponding to λ_2, which we will denote by y, is orthogonal to the left eigenvector of 1, i.e., $D^{1/2}\mathbf{1}$. So we have that

$$\begin{aligned}
\sigma_2\big((D^{-1/2}AD^{-1/2})^t\big) &= \sup_{x:\langle x, D^{1/2}\mathbf{1}\rangle = 0} \frac{\|(D^{-1/2}AD^{-1/2})^t x\|}{\|x\|} \\
&\geq \frac{\|(D^{-1/2}AD^{-1/2})^t y\|}{\|y\|} \ = \ |\lambda_2|^t.
\end{aligned}$$

The result now follows. □

3.5 A discrepancy for hypergraphs

Hypergraphs take sets of vertices as edges. However where in graphs these sets have either cardinality one (loops) or two (edges) in hypergraphs they can have different cardinalities. We will give a small discrepancy result for k-graphs which are graphs where edges are k-elements sets. Thus, a simple graph without loops can be thought of as a 2-graph.

The difficulty with generalizing discrepancy to hypergraphs is how to handle "adjacent". For a k-graph one approach is to replace the matrix with a multi-dimensional matrix (i.e., a k-dimensional array) with 1 in an entry if and only if the union of the

indices corresponds to an edge. An obvious difficulty with this is the matrix is *very* sparse, but even worse there is no well defined notion of eigenvalues and/or eigenvectors. The approach we take here will be different.

For a k-graph we fix an i with $0 < i < k$. We then will define an adjacency matrix $A^{(i,k-i)}$ by indexing the rows by the i-element subsets of V, indexing the columns by the $(k-i)$-element subsets of V, and an entry of $A^{(i,k-i)}$ is 1 if the union of the sets indexing the row and column gives a k-edge, and 0 otherwise. Note that for a 2-graph that $A^{(1,1)}$ gives the adjacency matrix we have seen before. (We note that in general this matrix will not be square, so we will have to work with singular values instead of eigenvalues.)

Given an i element subset X of V we define the analogue of the degree by

$$\Delta^{(i)}(X) = \left| \left\{ Y : \begin{array}{c} Y \text{ a } k-i \text{ element subset of V} \\ \text{and } X \cup Y \text{ is a } k\text{-edge of } G \end{array} \right\} \right|$$

Note that $\Delta^{(i)}(X)$ is the row sum of $A^{(i,k-i)}$ which corresponds to X, while $\Delta^{(k-i)}(Y)$ is the column sum of $A^{(i,k-i)}$ which corresponds to Y. We also have the analogues of the diagonal degree matrix, $\Delta^{(i)}$ and $\Delta^{(k-i)}$. This gives us our analogue of the normalized adjacency matrix, namely,

$$(\Delta^{(i)})^{-1/2} A^{(i,k-i)} (\Delta^{(k-i)})^{-1/2}.$$

Given $\mathcal{X}^{(i)}$ and $\mathcal{Y}^{(k-i)}$, collections of i and $k-i$ element subsets respectively, then we have that the number of k-edges formed by these subsets is given by

$$e(\mathcal{X}^{(i)}, \mathcal{Y}^{(k-i)}) = \left| \left\{ (X,Y) : X \in \mathcal{X}^{(i)}, Y \in \mathcal{Y}^{(k-i)}, \text{ and } X \cup Y \text{ forms a } k \text{ edge} \right\} \right|.$$

We also need the other half of discrepancy, namely, the "expected" number of k-edges given a collection of i and $k-i$ element subsets. Let us suppose that $\mathcal{X}^{(i)}$ is a collection of i element subsets of the vertices, we define the volume of $\mathcal{X}^{(i)}$ analogously as before by

$$\mathrm{vol}^{(i)}(\mathcal{X}^{(i)}) = \sum_{X \in \mathcal{X}^{(i)}} \Delta^{(i)}(X),$$

and similarly if we have $\mathcal{Y}^{(k-i)}$ a collection of $k-i$ element subsets of the vertices, we have

$$\mathrm{vol}^{(k-i)}(\mathcal{Y}^{(k-i)}) = \sum_{Y \in \mathcal{Y}^{(k-i)}} \Delta^{(k-i)}(Y).$$

We note in passing that $\mathrm{vol}^{(i)}(G) = \mathrm{vol}^{(k-i)}(G) := \mathrm{vol}^{(i,k-i)}(G)$ denotes the total number of nonzero entries in $A^{(i,k-i)}$.

We are now ready to define the $(i, k-i)$-discrepancy of G, denoted $\mathrm{disc}^{(i,k-i)}(G)$. It is the minimum β such that

$$\left| e(\mathcal{X}^{(i)}, \mathcal{Y}^{(k-i)}) - \frac{\mathrm{vol}^{(i)}(\mathcal{X}^{(i)})\ \mathrm{vol}^{(k-i)}(\mathcal{Y}^{(k-i)})}{\mathrm{vol}^{(i,k-i)}(G)} \right| \leq \beta \sqrt{\mathrm{vol}^{(i)}(\mathcal{X}^{(i)})\ \mathrm{vol}^{(k-i)}(\mathcal{Y}^{(k-i)})}.$$

Similar to before by using Theorems 13 and 14 on the matrix $A^{(i,k-i)}$ we have

$$\begin{aligned} \mathrm{disc}^{(i,k-i)}(G) &\leq \sigma_2\big((\Delta^{(i)})^{-1/2} A^{(i,k-i)} (\Delta^{(k-i)})^{-1/2}\big) \\ &\leq 150\, \mathrm{disc}^{(i,k-i)}(G)(1 - 8 \log \mathrm{disc}^{(i,k-i)}(G)). \end{aligned}$$

We note that $\sigma_2\big((\Delta^{(i)})^{-1/2} A^{(i,k-i)} (\Delta^{(k-i)})^{-1/2}\big) \leq 1$. In particular, we see that again both of these properties are equivalent in that if one goes to zero then the other also goes to zero.

3.6 Comments about quasirandom directed graphs

We have seen that for directed graphs that there is a direct relationship between having small singular values and having small discrepancy. This is an example of two quasirandom properties for directed graphs.

Quasirandom graph properties (introduced in the work of Chung-Graham-Wilson [15]) are a collection of graph properties where if a graph has any one of the properties then it must have them all. The name quasirandom is that these properties are ones that we would expect a random graph to have, so that a graph having these properties would behave somewhat like a random graph should. While the theory for quasirandom graphs has been well developed for undirected graphs there has been little work done for directed graphs. One possible explanation is that the obvious generalization of the quasirandom (undirected) graph properties fail to be quasirandom (directed) graph properties. Some of the (undirected) quasirandom graph properties are listed below (p being the density of the edges, in the Erdős-Renyi model it is the probability of including an edge, n is the number of vertices).

P : There are at least $\big(1 + o(1)\big)\frac{p}{2}n^2$ edges in the graph and the eigenvalues of the adjacency matrix have $\lambda_1 = \big(1 + o(1)\big)pn$ and $\lambda_2 = o(n)$.

Q : For any graph H on s vertices the number of induced copies of H in the graph is $(1+o(1))p^{e(H)}(1-p)^{\binom{s}{2}-e(H)}n^s$, where $e(H)$ is the number of edges in H.

R : For any graph H on s vertices the number of (not necessarily induced) copies of H in the graph is $(1+o(1))p^{e(H)}n^s$.

S : For all but $o(n^2)$ pairs of vertices the number of common neighbors of vertices u and v (i.e., the vertices w with the same adjacency relationship between u and v) is $\big(1+o(1)\big)(1-2p+2p^2)n$.

In particular, we see that if all but the largest eigenvalue of the graph are small then the graph has rich structure, i.e., contains all small subgraphs the appropriate number of times. This is an example of where knowing the distribution of eigenvalues can give you a lot of information about the graph. The usual method of proving the equivalency of these relationships is to show that $P \Rightarrow S \Rightarrow (Q,R) \Rightarrow P$

If we look at the generalization of these properties to directed graphs we get the following natural candidates.

P' : There are at least $\big(1+o(1)\big)pn^2$ edges in the graph and the singular values of the adjacency matrix have $\sigma_1 = \big(1+o(1)\big)pn$ and $\sigma_2 = o(n)$.

Q' : For any directed graph H on s vertices the number of induced copies of H in G is $(1+o(1))p^{e(H)}(1-p)^{s(s-1)-e(H)}n^s$, where $e(H)$ is the number of arcs in H.

R' : For any directed graph H on s vertices the number of (not necessarily induced) copies of H in the graph is $(1+o(1))p^{e(H)}n^s$.

S' : For all but $o(n^2)$ pairs of vertices the number of common neighbors of vertices u and v (i.e., the vertices w with the same adjacency relationship between u and v) is $\big(1+o(1)\big)(1-4p+8p^2-8p^3+4p^4)n$.

The problem now arises in that $P \not\Rightarrow (Q,R)$. To see this take a quasirandom undirected graph on n vertices and replace each edge by a pair of directed edges. The resulting singular values are the same and we have double the number of edges so we satisfy P'. However if we count the number of copies of H where H is the graph on two vertices with a pair of bidrected edges we see that this shows up as a subgraph pn^2 times, but Q' and R' would predict p^2n^2 occurrences.

The problem seems to lie in the somewhat innocuous looking condition S and S'. In undirected graphs it can be shown that this property is a control on the entries of the matrix $A^2 + (J - A)^2$. Since quasirandom graphs are almost regular this can be easily translated into an eigenvalue condition and so we can see that $P \Rightarrow S$. On the other hand, for directed graphs we do not have this clear relationship between sameness and the adjacency matrix.

This suggests that for directed graphs that the quasirandom properties subdivide into smaller groups. There still remain many questions in this direction.

Acknowledgement

Sections 3.1 through 3.4 are based on the paper "Relating singular values and discrepancy of weighted directed graphs" published in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (Miami, FL, 2006)* and the subsequent improved version published as "Using discrepancy to control singular values for nonnegative matrices" in *Linear Algebra and its Applications* (2006).

4 Eigenvalues and 2-edge-coverings

4.1 Introduction

We have seen that the eigenvalues can be used to give information about the structure of a graph. We would then expect that if two graphs have many eigenvalues in common that they also share some structure. One example of this is when two graphs both "cover" a smaller graph.

If we look at the spectrum of the three graphs in Figure 4.1 (either using the adjacency matrix or the normalized Laplacian) we see that there are four eigenvalues which are common to all three graphs. In this chapter we will show how to compute the spectrum of graphs of this type by computing the spectrum of two smaller graphs, which we will call the (modified) cover and the anti-cover. In the example of these three graphs we will see that the common eigenvalues are traced to a shared anti-cover graph.

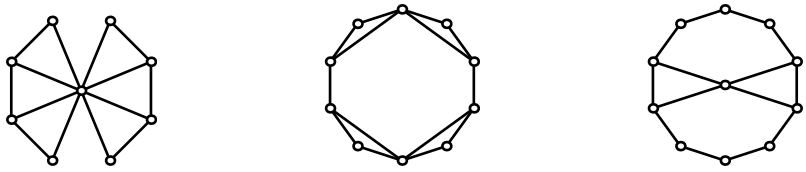

Figure 4.1: Three graphs sharing some common eigenvalues.

4.2 2-edge-coverings of graphs

We will be considering a special type of cover called a 2-edge-cover. We say that a graph G is a 2-edge-covering of a graph $\hat{G}$ if there is an onto map $\pi : V(G) \to V(\hat{G})$

satisfying the following conditions:

(i) if $u \sim v$ in G then $\pi(u) \sim \pi(v)$ in $\hat{G}$ and further $w\big(\pi(u), \pi(v)\big) = w(u, v)$;

(ii) if $\pi(u) \sim \hat{w}$ in $\hat{G}$ then there is some vertex v in G so that $u \sim v$ and $\pi(v) = \hat{w}$;

(iii) for each (ordered) edge $(\hat{w}, \hat{z})$ in $\hat{G}$ there are exactly two (ordered) edges (p, q), (r, s) in G so that $\big(\pi(p), \pi(q)\big) = \big(\pi(r), \pi(s)\big) = (\hat{w}, \hat{z})$.

Property (i) states that the graph is a weight preserving homomorphism (for more about graph homomorphisms and coverings the reader is referred to Godsil and Royle [20]). Property (ii) insures that we can lift edges from $\hat{G}$ back up to G, while property (iii) states that each edge is covered twice. The reason that we insist on having ordered pairs is to deal with the creation of loops, namely, if $u \neq v$, $u \sim v$ and $\pi(u) = \pi(v)$ then we would have a loop at $\pi(u)$ in $\hat{G}$; by our convention the loop is double covered by (u, v) and (v, u).

Some examples of 2-edge-coverings involving the 6-cycle are shown in Figure 4.2 (the labeling indicating how the vertices map).

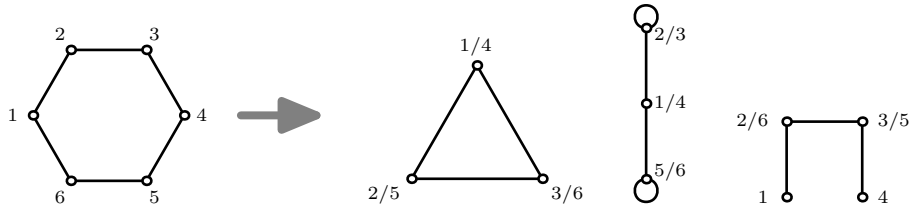

Figure 4.2: Examples of 2-edge-coverings involving the 6-cycle.

For our purposes, the most important feature of a 2-edge-covering is what happens at the vertices. It is easy to see that for a connected graph each vertex in the covered graph $\hat{G}$ can have either one or two preimages.

Lemma 20. *Let G be a nonempty connected graph which is a 2-edge-covering of $\hat{G}$ under the map π, and $\pi(v) = \hat{v}$.*

- *If $\big|\pi^{-1}(\hat{v})\big| = 2$ and $u \sim v \sim w$ with $u \neq w$ in G, then $\pi(u) \neq \pi(w)$.*
- *If $\big|\pi^{-1}(\hat{v})\big| = 1$ and $u \sim v$, then there is some w such that $v \sim w$, $u \neq w$ and $\pi(u) = \pi(w)$.*

In other words when a vertex has two preimages there is a 1-to-1 correspondence between the edges incident to v and $\hat{v}$. On the other hand, if a vertex has only a single preimage the edges incident to v in G map 2-to-1 to the edges incident to $\hat{v}$ in $\hat{G}$. Intuitively in the latter case the edges incident to v "fold" over and we will refer to such vertices in either G or $\hat{G}$ as folding vertices throughout.

Proof. First consider the case $\pi^{-1}(\hat{v}) = \{v, v'\}$. Suppose that $u \sim v \sim w$; then by property (i) $\pi(u) \sim \hat{v}$ so by property (ii) there exists some vertex z in G so that (z, v') and (u, v) are distinct edges in G both covering the same edge in $\hat{G}$. Now if $\pi(u) = \pi(w)$ then (w, v) would be a third edge in G which also covers but this contradicts (iii). Therefore we have that $\pi(u) \neq \pi(w)$.

Now consider the case $\pi^{-1}(\hat{v}) = \{v\}$. If $u \sim v$ then by property (i) $\pi(u) \sim \hat{v}$. By property (iii) this edge is double covered and since $\hat{v}$ has only one preimage the two edges in G which double cover it are (u, v) and (w, v) (for some w). But now note that $v \sim w$, $u \neq w$ and that $\pi(u) = \pi(w)$ as needed. □

Similarly one can show that it is not possible with our conventions for two folding vertices to be adjacent and we will implicitly assume this in our proofs.

4.3 Finding eigenvalues of 2-edge-coverings

4.3.1 2-edge-coverings and the adjacency matrix

In this subsection and the next we will illustrate the techniques of how to calculate the eigenvalues of a graph G which has a 2-edge-covering (i.e., there is some H so that $\pi : V(G) \to V(H)$ is a 2-edge-covering). We will use the graph in Figure 4.3 for an example of how to apply the techniques.

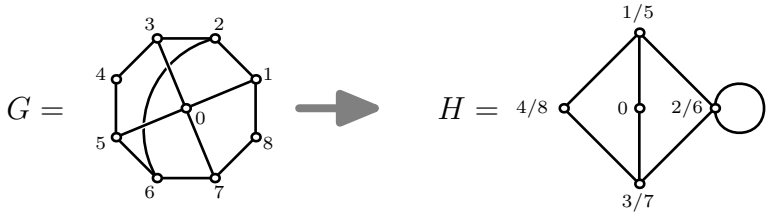

Figure 4.3: Our example 2-edge-covering. (All edge weights are 1.)

To find the eigenvalues of the adjacency matrix of G we will make use of two

graphs. The first is a modified cover graph, denoted $H_\circ$ (to keep track of which graph we will be dealing with; anytime we are referring to $H_\circ$ we will use the $\square_\circ$ notation). The graph will have the same vertices and the only modification will be to the weight function as follows:

$$w_\circ(u_\circ, v_\circ) = \begin{cases} \sqrt{2}w(u,v) & u \text{ or } v \text{ is a folding vertex;} \\ w(u,v) & \text{otherwise.} \end{cases}$$

The second graph will be an anti-cover graph and we denote it by H° (again anytime we refer to H° we use the $\square^\circ$ notation). The first step to defining the anti-cover is to give a sign function on the vertices of G so that for each v, $\operatorname{sgn}(v) \in \{-1, 0, 1\}$ where $\operatorname{sgn}(v) = 0$ if and only if the vertex folds, otherwise if $\pi(u) = \pi(v)$ for $u \neq v$ then $\operatorname{sgn}(u) = -\operatorname{sgn}(v)$.

Then an anti-cover H° is formed by removing all folding vertices and incident edges, for any remaining edge $u^\circ \sim v^\circ$ which is covered by edge $u \sim v$, the edge weight will be

$$w^\circ(u^\circ, v^\circ) = w(u,v) \operatorname{sgn}(u) \operatorname{sgn}(v). \tag{4.1}$$

Similar to Lemma 20 it can be shown that this weight function is well defined, i.e., choosing either of the two edges which cover an edge will give the same result. Also we note that a signing of the graph, and so also the anti-cover, is not unique, but it can be shown that the adjacency matrix of two different anti-cover graphs are similar by a diagonal matrix with ± 1 on the diagonal.

In the special case that there are no vertices which fold then $H_\circ = H$ and H° is a signed version of H. This case has been previously considered by D'Amato [17] and more recently Bilu and Linial [1].

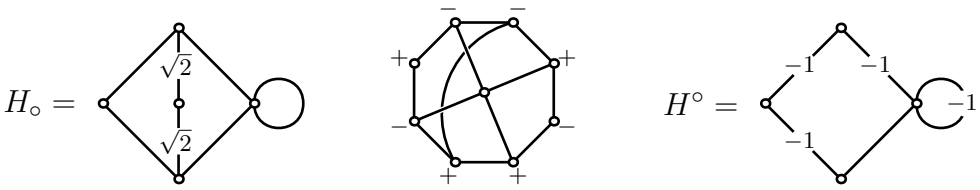

Figure 4.4: The graphs $H_\circ$, a signing of G, and corresponding H° for graphs in Figure 4.3. (Unmarked edges have weight 1.)

Theorem 21. *If G has a 2-edge-covering of H then the eigenvalues of the adjacency matrix of G is the union of the eigenvalues of the adjacency matrices of $H_\circ$ and H° (counting multiplicity).*

Proof. We will show how to lift eigenvectors of $H_\circ$ and H° to be eigenvectors of G. Then since the number of vertices (and hence) eigenvectors of $H_\circ$ and H° combined is equal to the number of vertices of G, and the independence of the vectors will follow from the construction the result will follow.

So suppose that λ is an eigenvalue of $H_\circ$ with eigenvector $\mathbf{x}_\circ$. Then consider the following vector defined for G by

$$\mathbf{x}(u) = \begin{cases} \sqrt{2}\mathbf{x}_\circ(u_\circ) & u \text{ is a folding vertex;} \\ \mathbf{x}_\circ(u_\circ) & \text{otherwise;} \end{cases}$$

where $u_\circ = \pi(u)$.

We now show that (2.1) holds for $\mathbf{x}$ showing that λ is an eigenvvalue of G; we do this by considering two cases.

(1) v is not a folding vertex:

$$\begin{aligned} \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) &= \sum_{\substack{u\sim v \\ u \text{ folds}}} w(u,v)\mathbf{x}(u) + \sum_{\substack{u\sim v \\ u \text{ not fold}}} w(u,v)\mathbf{x}(u) \\ &= \sum_{\substack{u_\circ\sim v_\circ \\ u_\circ \text{ folds}}} \frac{w_\circ(u_\circ,v_\circ)}{\sqrt{2}}\big(\sqrt{2}\mathbf{x}_\circ(u_\circ)\big) + \sum_{\substack{u_\circ\sim v_\circ \\ u_\circ \text{ not fold}}} w_\circ(u_\circ,v_\circ)\mathbf{x}_\circ(u_\circ) \\ &= \sum_{u_\circ:u_\circ\sim v_\circ} w_\circ(u_\circ,v_\circ)\mathbf{x}_\circ(u_\circ) \;=\; \lambda\mathbf{x}_\circ(v_\circ) \;=\; \lambda\mathbf{x}(v). \end{aligned}$$

(2) v is a folding vertex:

$$\sum_{u:u\sim v} w(u,v)\mathbf{x}(u) \;=\; 2\sum_{u_\circ:u_\circ\sim v_\circ} \frac{w_\circ(u_\circ,v_\circ)}{\sqrt{2}}\mathbf{x}_\circ(u_\circ) \;=\; \sqrt{2}\lambda\mathbf{x}_\circ(v_\circ) \;=\; \lambda\mathbf{x}(v).$$

Similarly, now suppose that λ is an eigenvalue of H° with eigenvector $\mathbf{x}^\circ$. Then consider the following vector defined for G by

$$\mathbf{x}(u) = \begin{cases} 0 & u \text{ is a folding vertex;} \\ \operatorname{sgn}(u)\mathbf{x}^\circ(u^\circ) & \text{otherwise.} \end{cases}$$

We again show that (2.1) holds for $\mathbf{x}$ showing that λ is an eigenvalue of G, by considering two cases.

(1) v is not a folding vertex:

$$\begin{aligned} \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) &= \sum_{\substack{u\sim v \\ u \text{ not folds}}} w(u,v)\mathbf{x}(u) + \sum_{\substack{u\sim v \\ u \text{ folds}}} w(u,v)\mathbf{x}(u) \\ &= \operatorname{sgn}(v) \sum_{u^\circ:u^\circ\sim v^\circ} w^\circ(u^\circ,v^\circ)\mathbf{x}^\circ(u^\circ) \\ &= \lambda \operatorname{sgn}(v)\mathbf{x}^\circ(v^\circ) = \lambda\mathbf{x}(v). \end{aligned} \tag{4.2}$$

(2) v is a folding vertex:

$$\begin{aligned}\sum_{u:u\sim v} w(u,v)\mathbf{x}(u) &= \sum_{\substack{u\sim v\\ \mathrm{sgn}(u)=1}} w(u,v)\mathbf{x}(u) + \sum_{\substack{u\sim v\\ \mathrm{sgn}(u)=-1}} w(u,v)\mathbf{x}(u)\\ &= \sum_{\substack{u\sim v\\ \mathrm{sgn}(u)=1}} w(u,v)\mathbf{x}^\circ(u^\circ) - \sum_{\substack{u\sim v\\ \mathrm{sgn}(u)=-1}} w(u,v)\mathbf{x}^\circ(u^\circ)\\ &= 0 \;=\; \lambda\mathbf{x}(v). \end{aligned} \tag{4.3}$$

Thus the eigenvalues of $H_\circ$ and H° are eigenvalues of G and, as noted above, this concludes the proof. □

4.3.2 2-edge-coverings and the normalized Laplacian

We again consider the problem of how to calculate the eigenvalues of G but this time for the normalized Laplacian. We will again make use of two graphs, but this time they will be slightly different. First, we will let $H_\Delta = H$.

The second graph H^Δ is again found by removing the folding vertices and incident vertices, and also taking a sign function on G as before and defining $w^\Delta(u^\Delta, v^\Delta)$ similarly to (4.1). There is one additional structure that we will need for H^Δ and that is a weight function on the vertices,

$$w^\Delta(v^\Delta) = \sum_{\substack{u\sim v\\ u \text{ folds}}} w(u,v).$$

This new weight function shows up in the degrees in H^Δ which are defined as follows,

$$d^\Delta(v^\Delta) = w^\Delta(v^\Delta) + \sum_{u^\Delta\sim v^\Delta} \left|w^\Delta(u^\Delta, v^\Delta)\right|.$$

Intuitively, $w^\Delta(v^\Delta)$ is used to correct for the change caused by the removal of edges incident to folding vertices so that now $d(v) = d^\Delta(v^\Delta)$. Pictorially, we will note a weight at vertex v^Δ of k by putting " $\boxed{k}$ " at the vertex.

Theorem 22. *If G has a 2-edge-covering of H then the set of eigenvalues of the normalized Laplacian matrix of G is the union of the eigenvalues of the normalized Laplacian matrices of $H = H_\Delta$ and H^Δ (counting multiplicity).*

Proof. We again show how to lift eigenvectors of H_Δ and H^Δ to be eigenvectors of G. Then since the number of vertices (and hence) eigenvectors of H_Δ and H^Δ combined

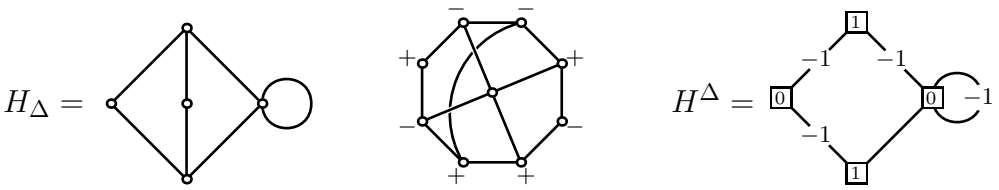

Figure 4.5: The graphs H_Δ, a signing of G, and corresponding H^Δ for graphs in Figure 4.3. (Unmarked edges have weight 1.)

is equal to the number of vertices of G, and the independence of the vectors will follow from the construction, the result will follow.

So suppose that λ is an eigenvalue of H_Δ with eigenvector $\mathbf{x}_\Delta$. Then consider the following vector defined for G by $\mathbf{x}(u) = \mathbf{x}_\Delta(u_\Delta)$, where $u_\Delta = \pi(u)$.

We now show that (2.2) holds for $\mathbf{x}$ showing that λ is an eigenvalue of G, we again have two cases.

(1) v is not a folding vertex:

$$\begin{aligned} d(v)\mathbf{x}(v) - \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) &= d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) - \sum_{u_\Delta:u_\Delta\sim v_\Delta} w_\Delta(u_\Delta,v_\Delta)\mathbf{x}_\Delta(u_\Delta) \\ &= \lambda d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) \;=\; \lambda d(v)\mathbf{x}(v). \end{aligned}$$

(2) v is a folding vertex:

$$\begin{aligned} d(v)\mathbf{x}(v) - \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) &= 2d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) - 2\sum_{u_\Delta:u_\Delta\sim v_\Delta} w_\Delta(u_\Delta,v_\Delta)\mathbf{x}_\Delta(u_\Delta) \\ &= 2\lambda d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) \;=\; \lambda d(v)\mathbf{x}(v). \end{aligned}$$

Now suppose that λ is an eigenvalue of H^Δ with eigenvector $\mathbf{x}^\Delta$, then consider the following vector defined for G by

$$\mathbf{x}(u) = \begin{cases} 0 & u \text{ is a folding vertex;} \\ \operatorname{sgn}(u)\mathbf{x}^\Delta(u^\Delta) & \text{otherwise;} \end{cases}$$

note that this is equivalent to the definition given in Theorem 21. We again show that (2.2) holds for $\mathbf{x}$ showing that λ is an eigenvalue of G. We again have two cases.

(1) v is not a folding vertex:

$$d(v)\mathbf{x}(v) - \sum_{u:u\sim v} w(u,v)\mathbf{x}(u) \;=\; \operatorname{sgn}(v)d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) - \underbrace{\operatorname{sgn}(v)\sum_{u_\Delta:u_\Delta\sim v_\Delta} w_\Delta(u_\Delta,v_\Delta)\mathbf{x}_\Delta(u_\Delta)}_{\text{same as in (4.2)}}$$
$$= \operatorname{sgn}(v)\lambda d_\Delta(v_\Delta)\mathbf{x}_\Delta(v_\Delta) \;=\; \lambda d(v)\mathbf{x}(v).$$

(2) v is a folding vertex:

$$d(v)\underbrace{\mathbf{x}(v)}_{=0} - \underbrace{\sum_{u:u\sim v} w(u,v)\mathbf{x}(u)}_{=0 \text{ same as in (4.3)}} \;=\; 0 \;=\; \lambda d(v)\underbrace{\mathbf{x}(v)}_{=0}.$$

Thus the eigenvalues of H_Δ and H^Δ are eigenvalues of G, and as noted above, this concludes the proof. □

4.3.3 2-edge-coverings and directed graphs

For directed graphs it is easy to adapt the definition of a 2-edge-covering and the constructions and proof given in Section 4.3.1 to establish the following theorem. (We omit the definitions, but they are the obvious generalizations of what we have already given.)

Theorem 23. *If $\vec{G}$ has a 2-edge-covering of $\vec{H}$ then the eigenvalues of the adjacency matrix of $\vec{G}$ contains the union of the eigenvalues of the adjacency matrices of $\vec{H}_\circ$ and $\vec{H}^\circ$ (counting multiplicity of the respective eigenspaces).*

It would be interesting to determine if the eigenvalues of $\vec{G}$ was the union of the eigenvalues of $\vec{H}_\circ$ and $\vec{H}^\circ$. The difficulty lies in that for directed graphs the adjacency matrices do not need to have a full set of eigenvectors, a key fact which was used in the earlier proof. A similar result for directed graphs for the results in Section 4.3.2 is more problematic as there is often no well defined normalized Laplacian for a directed graph, see for example Chung [12] and Butler [8].

4.4 Can you hear the shape of a graph?

Returning to the graphs in Figure 4.1 it is now easy to find the (modified) cover and anti-cover graphs of all three graphs (where the 2-edge-covering is given by folding in

half along the vertical axis). Moreover for both the adjacency matrix and the normalized Laplacian all three have the same anti-cover (the path of length four in the first case and the path of length four with all vertex weights 1 in the second) and thus they all share four eigenvalues in common. It is easy to expand on this example and start with a given anti-cover and construct many graphs which will share some nontrivial eigenvalues.

A famous question in spectral graph theory is "can you hear the shape of a graph?" That is, given the eigenvalues can you determine the graph that produced them. There are many examples of two graphs that share the same spectrum but are not the same graph, while there are also examples of graphs that are uniquely determined by the spectrum (see van Dam and Haemers [19]).

But when it comes to the normalized Laplacian of the anti-covers H^{Δ} the situation can be even worse. Consider the two weighted graphs in Figure 4.6. These

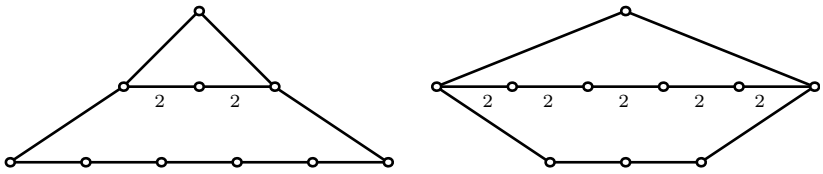

Figure 4.6: Two graphs with several shared eigenvalues. (Unmarked edges have weight 1.)

graphs share four nontrivial eigenvalues. They also have obvious left/right symmetry and the eigenvalues found by the 2-edge-covering when the graphs are folded in half are the ones which are not common. Thus the shared eigenvalues come from their anti-cover graphs. These are simple to construct and are shown in Figure 4.7. However it is not

Figure 4.7: Anti-covers of the graphs shown in Figure 4.6

possible to obtain one of these anti-coverings from the other by relabeling and/or scaling. So the shared eigenvalues appear to be the result of two cospectral anti-covers.

However the situation is a little more interesting than that. If we now compute

$\mathcal{L}^{\Delta}$ for both the graphs in Figure 4.7 they both give

$$\begin{pmatrix} 1 & \frac{-1}{2\sqrt{2}} & 0 & 0 \\ \frac{-1}{2\sqrt{2}} & 1 & \frac{-1}{2} & 0 \\ 0 & \frac{-1}{2} & 1 & \frac{-1}{2} \\ 0 & 0 & \frac{-1}{2} & \frac{3}{2} \end{pmatrix}.$$

In particular, the anti-cover graph cannot in general be uniquely determined from the corresponding normalized Laplacian of the anti-cover. So these graphs are not only co-spectral they also have the same normalized Laplacian.

4.5 Applications of 2-edge-coverings

If we now think of a real matrix as the anti-cover matrix of some graph then we can sometimes use information about graphs to help find eigenvalues and eigenvectors. As a simple example consider the following.

Corollary 24. *Let B_n be the following $n \times n$ matrix*

$$B_n = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 & -1 \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & 1 \\ -1 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Then the eigenvalues of B_n are $2\cos\left(\pi(2j+1)/n\right)$ for $j = 0, 1, \ldots, n-1$.

Proof. It is easy to check that B_n corresponds to the adjacency matrix of an anti-covering of the $2n$-cycle where we have wrapped the cycle on itself twice (an example for $n = 3$ is shown on the left in Figure 4.2). Therefore by Theorem 21 it follows that the spectrum of B_n is the spectrum of the $2n$-cycle with the spectrum of the n-cycle removed. Since the eigenvalues of the cycles are known the result follows. □

More generally any real symmetric matrix can be thought of as the adjacency matrix of some anti-cover. Given the anti-cover it is easy to construct a graph which

covered it, i.e., make vertices $\{1^+, 1^-, 2^+, 2^-, \ldots, n^+, n^-\}$ and then if $w(i,j) > 0$ put in edges $\{i^+, j^+\}$ and $\{i^-, j^-\}$ with weight $|w(i,j)|$, while if $w(i,j) < 0$ put in edges $\{i^+, j^-\}$ and $\{i^-, j^+\}$ with weight $|w(i,j)|$. This is only one of many possible graphs that generated the anti-cover, but it has the additional property that the matrix of the corresponding graph is found by taking the absolute values of the matrix we started with. This gives us an easy constructive proof for the following theorem.

Theorem 25. *Let A be a real $n\times n$ symmetric matrix and let $|A|$ be the $n\times n$ matrix defined entrywise by $|A|_{i,j} = |A_{i,j}|$ (i.e., the absolute value of the corresponding entry of A). Then there exists a nonnegative $2n\times 2n$ symmetric matrix B so that the eigenvalues of B are the union of the eigenvalues of A and $|A|$ (counting multiplicity).*

4.6 A simple construction of cospectral graphs

As another application for 2-edge-coverings we can give a construction for cospectral graphs. There have been several constructions given which produce cospectral graphs (see [21, 24]). However, these constructions have focused on the adjacency matrix, the (combinatorial) Laplacian and the signless (combinatorial) Laplacian. Our construction will generate a large class of graphs which are cospectral with respect to the normalized Laplacian and the adjacency matrix. While our method is to use 2-edge-coverings, we will include simplified proofs here to demonstrate the techniques used in this chapter.

We note that the construction gives pairs of graphs which are cospectral simultaneously with the adjacency and the normalized Laplacian and are also not regular. Examples were previously known of regular graphs which are cospectral with respect to the adjacency matrix and so also trivially cospectral with respect to the normalized Laplacian. This is the first known example of non-regular graphs with this property.

For a matrix M, let $\sigma(M)$ be the (multi-)set of eigenvalues of M, i.e., the spectrum of M. Further, for a nonnegative symmetric matrix M with positive row (and hence column) sums let $\mathcal{L}(M) = I - R^{-1/2}MR^{-1/2}$ where R is the diagonal matrix with diagonal entries composed of the row (or column) sums of M. Note that if A is the adjacency matrix then $\mathcal{L}(A)$ is the normalized Laplacian.

Lemma 26. *Let B be a $p\times q$ matrix. Then*

$$\sigma\left(\begin{bmatrix} O & B & B \\ B^* & O & O \\ B^* & O & O \end{bmatrix}\right) = \sqrt{2}\,\sigma\left(\begin{bmatrix} O & B \\ B^* & O \end{bmatrix}\right) \cup \{\underbrace{0,0,\ldots,0}_{\times q}\}, \quad \textit{and}$$

$$\sigma\left(\begin{bmatrix} O & B & O \\ B^* & O & B^* \\ O & B & O \end{bmatrix}\right) = \sqrt{2}\,\sigma\left(\begin{bmatrix} O & B \\ B^* & O \end{bmatrix}\right) \cup \{\underbrace{0,0,\ldots,0}_{\times p}\}.$$

Further, if B is nonnegative and has positive row and column sums then

$$\sigma\left(\mathcal{L}\left(\begin{bmatrix} O & B & B \\ B^* & O & O \\ B^* & O & O \end{bmatrix}\right)\right) = \sigma\left(\mathcal{L}\left(\begin{bmatrix} O & B \\ B^* & O \end{bmatrix}\right)\right) \cup \{\underbrace{1,1,\ldots,1}_{\times q}\}, \quad \textit{and}$$

$$\sigma\left(\mathcal{L}\left(\begin{bmatrix} O & B & O \\ B^* & O & B^* \\ O & B & O \end{bmatrix}\right)\right) = \sigma\left(\mathcal{L}\left(\begin{bmatrix} O & B \\ B^* & O \end{bmatrix}\right)\right) \cup \{\underbrace{1,1,\ldots,1}_{\times p}\}.$$

Proof. We will prove the first statement in both cases (the second statements are handled similarly). Let $\left[\begin{smallmatrix}\mathbf{x}\\ \mathbf{y}\end{smallmatrix}\right]$ be an eigenvector for the eigenvalue λ and the matrix $\left[\begin{smallmatrix} O & B \\ B^* & O\end{smallmatrix}\right]$ (i.e., so $B\mathbf{y} = \lambda\mathbf{x}$ and $B^*\mathbf{x} = \lambda\mathbf{y}$). Then consider the following,

$$\begin{bmatrix} O & B & B \\ B^* & O & O \\ B^* & O & O \end{bmatrix}\begin{bmatrix} \sqrt{2}\,\mathbf{x} \\ \mathbf{y} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} 2B\mathbf{y} \\ \sqrt{2}\,B^*\mathbf{x} \\ \sqrt{2}\,B^*\mathbf{x} \end{bmatrix} = \sqrt{2}\lambda\begin{bmatrix} \sqrt{2}\,\mathbf{x} \\ \mathbf{y} \\ \mathbf{y} \end{bmatrix},$$

so that $\sqrt{2}\,\lambda$ is an eigenvalue for the enlarged matrix. For the remaining eigenvalues let $\mathbf{e}_i$ denote the vector of length q which is 1 in the ith position and 0 otherwise. Then

$$\begin{bmatrix} O & B & B \\ B^* & O & O \\ B^* & O & O \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathbf{e}_i \\ -\mathbf{e}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{4.4}$$

since these vectors are orthogonal to the ones previously given, it shows that the remaining q eigenvalues are 0.

For the second part we first note that $\hat{\mathbf{x}}$ is an eigenvector of $\mathcal{L}(M)$ if and only if $\hat{\mathbf{y}} = R^{-1/2}\hat{\mathbf{x}}$ (known as the harmonic eigenvector) satisfies $(R - M)\hat{\mathbf{y}} = \lambda R\hat{\mathbf{y}}$. So for an eigenvalue λ let $\left[\begin{smallmatrix}\mathbf{x}\\ \mathbf{y}\end{smallmatrix}\right]$ be chosen to satisfy

$$\begin{bmatrix} R_1 & -B \\ -B^* & R_2 \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} R_1\mathbf{x} - B\mathbf{y} \\ -B^*\mathbf{x} + R_2\mathbf{y} \end{bmatrix} = \lambda\begin{bmatrix} R_1\mathbf{x} \\ R_2\mathbf{y} \end{bmatrix}$$

(where R_1 and R_2 are the partitioned parts of R). Then the following holds

$$\begin{bmatrix} 2R_1 & -B & -B \\ -B^* & R_2 & O \\ -B^* & O & R_2 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} 2R_1\mathbf{x} - 2B\mathbf{y} \\ -B^*\mathbf{x} + R_2\mathbf{y} \\ -B^*\mathbf{x} + R_2\mathbf{y} \end{bmatrix} = \lambda \begin{bmatrix} 2R_1\mathbf{x} \\ R_2\mathbf{y} \\ R_2\mathbf{y} \end{bmatrix},$$

so that λ is also an eigenvalue of the enlarged matrix. For the remaining eigenvalues we can use the same eigenvector and relationship given in (4.4). In particular since these vectors, $\hat{\mathbf{y}}$, satisfy $M\hat{\mathbf{y}} = \mathbf{0}$ then it follows that $(R - M)\hat{\mathbf{y}} = R\hat{\mathbf{y}}$ showing that these are q additional eigenvectors associated with an eigenvalue of 1, which completes the spectrum. □

4.6.1 The construction

Let G be a bipartite graph with vertex set $V(G) = V_1 \cup V_2$, and an edge set $E(G)$ where edges go between V_1 and V_2. Then we construct new graphs G_1 and G_2 where

$$\begin{aligned} V(G_1) &= V_1 \cup V_2 \cup V_2', \\ E(G_1) &= \big\{\{v_1(i), v_2(j)\}, \{v_1(i), v_2'(j)\} \,\big|\, \{v_1(i), v_2(j)\} \in E(G)\big\}; \\ V(G_2) &= V_1 \cup V_1' \cup V_2, \\ E(G_1) &= \big\{\{v_1(i), v_2(j)\}, \{v_1'(i), v_2(j)\} \,\big|\, \{v_1(i), v_2(j)\} \in E(G)\big\}. \end{aligned}$$

Intuitively this construction can be thought of as taking a bipartite graph and "unfolding" it either along V_1 or V_2 to construct the new graphs. An example of this construction is shown in Figure 4.8.

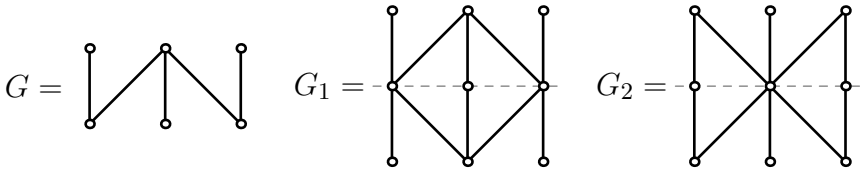

Figure 4.8: Examples of constructing G_1 and G_2.

Theorem 27. *Let G be a given bipartite graph with $|V_1| = p \leq |V_2| = q$, and let $A(G)$ denote the adjacency matrix of G. Then $\sigma\big(A(G_1)\big)$ and $\sigma\big(A(G_2)\big)$ differ by $(q - p)$ eigenvalues of 0. If further, G has no isolated vertices and $\mathcal{L}(G)$ denotes the normalized Laplacian matrix of G, then $\sigma\big(\mathcal{L}(G_1)\big)$ and $\sigma\big(\mathcal{L}(G_2)\big)$ differ by $(q - p)$ eigenvalues of 1.*

In particular, if G is a bipartite graph with $|V_1| = |V_2|$ and no isolated vertices then G_1 and G_2 are cospectral both with respect to the adjacency matrix and the normalized Laplacian.

Proof. The statement follows by noting that the form of the matrices $A(G)$, $A(G_1)$ and $A(G_2)$ are those in the statement of Lemma 26, and that the normalized Laplacian matrix of a graph is $\mathcal{L}(A(G))$. □

We now have a construction for cospectral graphs, and it remains to find examples where these graphs are non-isomorphic. It is easy to check that when the maximum degree of vertices in V_1 is not equal to the maximum degree of vertices in V_2, that the corresponding graphs G_1 and G_2 have different maximum degrees and hence are non-isomorphic. An example of this situation is shown in Figure 4.8.

In the case when $p < q$ the resulting graphs G_1 and G_2 can be made cospectral with respect to the adjacency matrix by the addition of isolated vertices to the smaller graph (there is not a similar operation for the normalized Laplacian). For instance if we start with the path on three vertices then G_1 and G_2 are a four-cycle and a star on five vertices, and by adding an isolated vertex to the four-cycle the two graphs are now cospectral and give the famous Saltire pair (these are shown in Figure 1.2).

4.7 Calculating Dirichlet eigenvalues

Dirichlet eigenvalues of a graph are related to the problem of random walks on a subset S of the vertices of a graph. When dealing with a random walk on a subset S the question becomes how to deal with boundary of S. There are two popular methods, the Dirichlet boundary condition and the Neumann boundary condition. More details about what these conditions are and how they relate to random walks can be found in Chung [11]. For our purposes it suffices to know that the eigenvalues associated with a random walk with Dirichlet boundary conditions can be found by looking at $\mathcal{L}_S$ which is the matrix $\mathcal{L}$ (for the whole graph) restricted to the columns and rows associated with S.

We can use 2-edge-coverings to compute Dirichlet eigenvalues. The key is to note that the difference between $\mathcal{L}_S$ and the normalized Laplacian that is formed by the graph restricted to S is the degree of points near the boundary; namely, in $\mathcal{L}_S$ the degree of vertices at the boundary counts the edges that leave S. So the difference is

that for $\mathcal{L}_S$ we have extra weight added to vertices incident to the boundary. This is the same as the extra vertex weight function we introduced for the anti-cover graphs. So calculating the Dirichlet eigenvalues of a subset S is the same as calculating eigenvalues of an anti-cover graph.

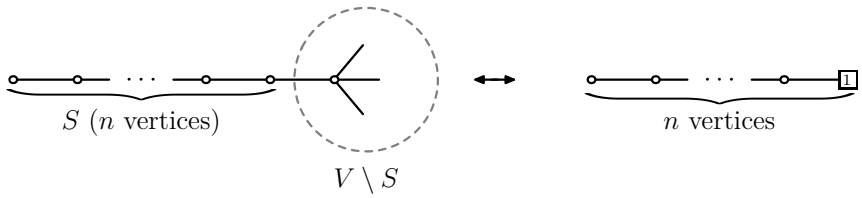

Figure 4.9: A subset S that we want to compute Dirichlet eigenvalues for and corresponding anti-cover graph.

For a more concrete example consider the graphs shown in Figure 4.9, where S is a path on n vertices connected by an edge to some graph (as shown on the left). Now translating the edges at the boundary as extra weight on vertices we have that computing the Dirichlet eigenvalues for S is the same as computing the eigenvalues of the graph on the right. As mentioned above this graph can be thought of as an anti-cover of some graph. In this case if we start with a path on $2n+1$ vertices and fold it in half at the middle vertex then the covered graph is the path on $n+1$ vertices and the anti-cover is the graph shown on the right in Figure 4.9. Since the eigenvalues of a path on m vertices is $1-\cos(\pi j/(m-1))$ for $j = 0, 1, \ldots, m-1$, it follows that the Dirichlet eigenvalues for the set S shown in Figure 4.9 are the eigenvalues of the path on $2n+1$ vertices with the eigenvalues of the path on $n+1$ vertices removed. In particular, the Dirichlet eigenvalues are $1-\cos\big((2j-1)\pi/2n\big)$ for $j = 1, 2, \ldots, n$.

Acknowledgement

Sections 4.1 through 4.5 are based on "Eigenvalues of 2-edge-coverings" and Section 4.6 is based on "Cospectral graphs for both the adjacency and normalized Laplacian matrices", both of which are currently submitted.

5 Interlacing and the normalized Laplacian

5.1 Introduction

In the last chapter we saw an example of two coverings where the eigenvalues of a smaller graph also were eigenvalues of a larger graph. The principle at work was that the graphs were sharing a common structure. Now suppose that we have two graphs which differ only in a few edges. Then on the whole the structure of the two graphs are the same with a few local exceptions. We would then expect that the eigenvalues are also very similar and that the eigenfunctions have some small local perturbations. This turns us to the content of this chapter which is the interlacing of eigenvalues for the normalized Laplacian.

Suppose that A and B are real symmetric matrices, with eigenvalues $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_m$. Then the eigenvalues of A and B interlace if for some constants p and q and all $1 \leq i \leq n$ we have $\alpha_{i-p} \leq \beta_i \leq \alpha_{i+q}$. In many simple cases of interlacing we have $p, q \in \{0, 1\}$. Another way to think about interlacing is that the eigenvalues of B are bounded by the eigenvalues of A. In some cases when A has an eigenvalue which occurs with high multiplicity an eigenvalue of B can also be found.

A well known example of interlacing from spectral graph theory is to compare the adjacency matrix of G and $G \setminus \{v\}$ for any vertex v. In this case it is known by Cauchy's Interlacing Theorem that if the eigenvalues of $A(G)$ are α_i and the eigenvalues of $A(G \setminus \{v\})$ are β_i then the $\alpha_i \leq \beta_i \leq \alpha_{i+1}$ (so we have $p = 0$ and $q = 1$ in this case).

A recent development for interlacing and the normalized Laplacian was recently given by Chen et al. [10] (and even more recently a shortened proof by Li [28]).

Theorem 28. *Let G be a simple graph without loops or parallel edges, let $H = e$ be an*

edge of G, and $G - H$ the graph G with edge e removed. If $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ and $\theta_0 \leq \theta_1 \leq \cdots \leq \theta_{n-1}$ are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G - H)$ respectively, then

$$\lambda_{k-1} \leq \theta_k \leq \lambda_{k+1} \qquad \text{for each } k = 0, 1, \ldots, n-1,$$

where $\lambda_{-1} = 0$ and $\lambda_n = 2$. More generally if H is a subgraph of G with $|E(H)| = t$ then

$$\lambda_{k-t} \leq \theta_k \leq \lambda_{k+t} \qquad \text{for each } k = 0, 1, \ldots, n-1,$$

where $\lambda_{-t} = \cdots = \lambda_{-1} = 0$ and $\lambda_n = \cdots = \lambda_{n+t-1} = 2$.

We will establish an improved version of Theorem 28. Our first improvement will be to allow a broader range of graphs which include graphs with loops and multiple edges, and more generally any weighted graph. Our second improvement will be to show that when removing a graph from G we can use the number of vertices of the graph being removed rather than the number of edges to control the spread of the eigenvalues. So for instance when we are removing a dense graph such as K_{10} we will show that the eigenvalues spread by at most 10 (the number of vertices), while the above result states that the eigenvalues spread by at most 45 (the number of edges). Before we state the main result we need to introduce terminology needed for weighted graphs.

We will follow the conventions of Chapter 1 by letting a weighted graph be a graph (possibly with loops) with a nonnegative weight function $w : V \times V \to [0, \infty)$ with $w(u, v) = w(v, u)$ (i.e., undirected) and $w(u, v) > 0$ if and only if there is an edge joining u and v. Using the weight function we define the adjacency matrix by $A_{u,v} = w(u, v)$ and the diagonal degree matrix uses degrees defined by $d(u) = \sum_v w(u, v)$. With the adjacency matrix and the diagonal degree matrix we can then define the normalized Laplacian of such a graph as we have before, i.e., $\mathcal{L} = D^{-1/2}(D - A)D^{-1/2}$ (we will adopt the convention that if $d(v) = 0$ then $(d(v))^{-1/2} = 0$).

A simple graph is the special case when all weights are either 0 or 1 and $w(v, v) = 0$ for all v. However, by allowing the weights to vary we can model more graphs. For instance, multigraphs can be modeled by letting $w(u, v)$ be the number of edges connecting u to v.

Given a weighted graph G we say that H is a subgraph of G if $w_H(u, v) \leq w_G(u, v)$ for all u, v. When H is a subgraph of G we let $G - H$ be the graph which has weight function $w_{G-H}(u, v) = w_G(u, v) - w_H(u, v)$ for all u, v. From this definition it follows that $d_{G-H}(u) = d_G(u) - d_H(u)$. The graph $G + H$, where the two graphs are on the same set of vertices, is defined analogously. Our main result is as follows.

Theorem 29. *Let G be a weighted graph and H a subgraph of G with t nonisolated vertices. If $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ and $\theta_0 \leq \theta_1 \leq \cdots \leq \theta_{n-1}$ are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G-H)$ respectively, then for $k = 0, 1, \ldots, n-1$ we have*

$$\lambda_{k-t+1} \;\leq\; \theta_k \;\leq\; \begin{cases} \lambda_{k+t-1} & H \text{ is bipartite;} \\ \lambda_{k+t} & \text{otherwise;} \end{cases}$$

where $\lambda_{-t+1} = \cdots = \lambda_{-1} = 0$ and $\lambda_n = \cdots = \lambda_{n+t-1} = 2$.

In the proof for Theorem 29 we will see that when H is bipartite that we are (essentially) allowed one more degree of freedom than when H is not bipartite. This accounts for the difference in the bounds given by the theorem. The statement of the theorem is essentially the best possible; to see this consider the two graphs shown in Figure 5.1. The graph G has three loops of weight 1 on the three top vertices, three edges of weight 2 (as marked) and the remaining edges of weight 1. The graph H consists of a triangle of the three top vertices with edge weight 1 along with the loops (so $t = 3$). Calculating the eigenvalues we have that $\theta_1(G-H) = 5/4 > 8/7 = \lambda_3(G)$, showing that the bound for non-bipartite graphs cannot be improved in general.

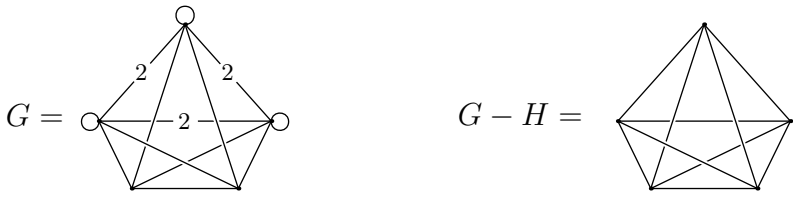

Figure 5.1: An example showing the result of Theorem 29 is tight.

As an example of an application of Theorem 29 we have the following.

Proposition. *If G is a simple graph on n vertices and more than $n/2$ of the vertices have degree $n-1$ then $n/(n-1)$ is an eigenvalue of G.*

This follows from the above theorem since $G = K_n - H$ where H has fewer than $n/2$ nonisolated vertices while $n/(n-1)$ is an eigenvalue of K_n with multiplicity $n-1$. This result is not the best possible (as a general rule finding eigenvalues by interlacing is one of the worst ways to find eigenvalues). For example, if a simple graph on n vertices has 2 vertices with degree $n-1$ then $n/(n-1)$ is an eigenvalue of G. This is a simple application of the following result (letting u and v be the vertices of degree $n-1$).

Lemma 30. *Let G be a simple graph and let* $\mathrm{nbd}(u) = \{w : u \sim w\}$ *(note that* $|\mathrm{nbd}(u)| = d_u$*). If* $\mathrm{nbd}(u) = \mathrm{nbd}(v)$ *and u is adjacent to v then $(d_u+1)/d_u$ is an eigenvalue of G for the normalized Laplacian. On the other hand if* $\mathrm{nbd}(u) = \mathrm{nbd}(v)$ *and u is not adjacent to v then 1 is an eigenvalue of G for the normalized Laplacian.*

The proof of the lemma is to let $\mathbf{x}$ be the vector which is 1 at u, -1 at v and 0 otherwise, and then check that this is an eigenvector for the indicated eigenvalue.

One interesting thing to note is that the result of Theorem 29 is independent of the amount of weight removed. This is because the proof relies on orthogonality conditions which are unaffected by changes in the weights. Instead of subtracting out a graph we could also add a graph. The following result immediately follows from Theorem 29 working with the graphs $G + H$ and $(G + H) - H = G$.

Corollary 31. *Let G be a weighted graph and H a graph on the vertices of G with t nonisolated vertices. If $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$ and $\theta_0 \le \theta_1 \le \cdots \le \theta_{n-1}$ are the eigenvalues of $\mathcal{L}(G)$ and $\mathcal{L}(G + H)$ respectively, then for $k = 0, 1, \ldots, n-1$ we have*

$$\lambda_{k+t-1} \;\ge\; \theta_k \;\ge\; \begin{cases} \lambda_{k-t+1} & H \text{ is bipartite;} \\ \lambda_{k-t} & \text{otherwise;} \end{cases}$$

where $\lambda_{-t} = \cdots = \lambda_{-1} = 0$ and $\lambda_n = \cdots = \lambda_{n+t-1} = 2$.

5.2 Proof of interlacing result

The proof of Theorem 29 will be adapted from the proof of Chen et al. [10] which follows by application of the Courant-Fischer Theorem (see [26]).

Theorem 32 (Courant-Fischer Theorem)**.** *Let M be a real symmetric matrix with eigenvalues $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$. Let $\mathcal{X}^k$ denote a k dimensional subspace of $\mathbf{R}^n$ and let $x \perp \mathcal{X}^k$ signify that $x \perp y$ for all $y \in \mathcal{X}^k$. Then*

$$\lambda_i = \min_{\mathcal{X}^{n-i-1}} \Big(\max_{x \perp \mathcal{X}^{n-i-1}, x \neq 0} \frac{x^T M x}{x^T x} \Big) = \max_{\mathcal{X}^i} \Big(\min_{x \perp \mathcal{X}^i, x \neq 0} \frac{x^T M x}{x^T x} \Big).$$

We will also use the notation $x \perp \mathcal{Z}$ for a set of vectors $\mathcal{Z}$ to indicate that $x \perp z$ for all $z \in \mathcal{Z}$. This is equivalent to saying that $x \perp \mathrm{span}(\mathcal{Z})$ (the span of the vectors of $\mathcal{Z}$).

In order to be able to use the Courant-Fischer Theorem we first note that

$$y^T(D-A)y = \sum_u y_u^2 d(u) - 2\sum_{u\sim v} y_u y_v w(u,v) = \sum_{u\sim v}(y_u - y_v)^2 w(u,v).$$

If we make the substitution $x = D^{1/2}y$ then we have

$$\frac{x^T\mathcal{L}x}{x^Tx} = \frac{(D^{1/2}y)^T\mathcal{L}(D^{1/2}y)}{(D^{1/2}y)^T(D^{1/2}y)} = \frac{y^TLy}{y^TDy} = \frac{\sum_{u\sim v}(y_u-y_v)^2w(u,v)}{\sum_u y_u^2 d(u)}. \tag{5.1}$$

Comment. *Here we run into small a problem when there are isolated vertices, in which case there might be no y for which $x = D^{1/2}y$. We will address this in the proof of Theorem 29 below, but for now will assume there are no isolated vertices in the graph.*

We now have

$$\begin{aligned}\lambda_i &= \min_{\mathcal{X}^{n-i-1}}\left(\max_{x\perp\mathcal{X}^{n-i-1},x\neq 0}\frac{x^T\mathcal{L}x}{x^Tx}\right) = \min_{\mathcal{X}^{n-i-1}}\left(\max_{D^{1/2}y\perp\mathcal{X}^{n-i-1},D^{1/2}y\neq 0}\frac{y^TLy}{y^TDy}\right)\\ &= \min_{\mathcal{Y}^{n-i-1}}\left(\max_{y\perp\mathcal{Y}^{n-i-1},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w(u,v)}{\sum_u y_u^2d(u)}\right),\end{aligned} \tag{5.2}$$

and similarly

$$\lambda_i = \max_{\mathcal{Y}^i}\left(\min_{y\perp\mathcal{Y}^i,y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w(u,v)}{\sum_u y_u^2d(u)}\right). \tag{5.3}$$

Proof of Theorem 29. Without loss of generality we may assume that the graph G has no isolated vertices. We first will consider the case when $G-H$ also has no isolated vertices. The result is trivial if $k \le t-1$ or $k \ge n-t-1$ ($k \ge n-t$ if H is bipartite), since the eigenvalues of the normalized Laplacian always lie in the interval between 0 and 2 inclusive (see [11]). So we may assume that $t-1 < k < n-t-1$ (or $n-t$ for H bipartite).

We now show that $\theta_k \ge \lambda_{k-t+1}$. Suppose that $\{u_1, u_2, \ldots, u_t\}$ are the nonisolated vertices of H, and let $\mathcal{Z} = \{e_{u_1}-e_{u_2}, e_{u_1}-e_{u_3}, \ldots, e_{u_1}-e_{u_t}\}$. Then using (5.2) we have

$$\begin{aligned}\theta_k &= \min_{\mathcal{Y}^{n-k-1}}\left(\max_{y\perp\mathcal{Y}^{n-k-1},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w_{G-H}(u,v)}{\sum_u y_u^2 d_{G-H}(u)}\right)\\
&= \min_{\mathcal{Y}^{n-k-1}}\left(\max_{y\perp\mathcal{Y}^{n-k-1},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w_G(u,v)-\sum_{u\sim v}(y_u-y_v)^2w_H(u,v)}{\sum_u y_u^2d_G(u)-\sum_u y_u^2 d_H(u)}\right)\\
&\ge \min_{\mathcal{Y}^{n-k-1}}\left(\max_{y\perp\mathcal{Y}^{n-k-1},y\perp\mathcal{Z},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w_G(u,v)}{\sum_u y_u^2d_G(u)-\sum_u y_u^2 d_H(u)}\right)\\
&\ge \min_{\mathcal{Y}^{n-k-1}}\left(\max_{y\perp\mathcal{Y}^{n-k-1},y\perp\mathcal{Z},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w_G(u,v)}{\sum_u y_u^2d_G(u)}\right)\\
&\ge \min_{\mathcal{Y}^{n-k+t-2}}\left(\max_{y\perp\mathcal{Y}^{n-k+t-2},y\neq 0}\frac{\sum_{u\sim v}(y_u-y_v)^2w_G(u,v)}{\sum_u y_u^2d_G(u)}\right) = \lambda_{k-t+1}.\end{aligned}$$

In going from the second to the third line we added the condition that y also be perpendicular to $\mathcal{Z}$ so that we are maximizing over a smaller set. With the condition that $y \perp \mathcal{Z}$ then $y_u = y_v$ for all u, v in H, and in particular the second term in the numerator drops out. While in going from the fourth to the fifth line we consider a broader optimization that would include the fourth line as a special case.

Next we show that $\theta_k \leq \lambda_{k+t}$ for general H. Suppose that $u_1, \ldots, u_t$ are the nonisolated vertices of H, and let $\mathcal{Z} = \{e_{u_1}, e_{u_2}, \ldots, e_{u_t}\}$. Then using (5.3) we have

$$\begin{aligned}
\theta_k &= \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v) - \sum_{u \sim v} (y_u - y_v)^2 w_H(u,v)}{\sum_u y_u^2 d_G(u) - \sum_u y_u^2 d_H(u)} \right) \\
&\leq \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \perp \mathcal{Z}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) \\
&\leq \max_{\mathcal{Y}^{k+t}} \left(\min_{y \perp \mathcal{Y}^{k+t}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) = \lambda_{k+t}. \qquad (5.4)
\end{aligned}$$

In going from the first to the second line we added the condition that y also be perpendicular to $\mathcal{Z}$ so that we are minimizing over a smaller set. With the condition that $y \perp \mathcal{Z}$ then $y_u = 0$ for all u in H. In particular the second terms in the numerator and denominator drop out. Finally, we consider a broader optimization that would include the second line as a special case.

For the case when H bipartite let $\{u_1, v_1\}, \{u_2, v_2\}, \ldots, \{u_{t-1}, v_{t-1}\}$ be edges of a spanning subgraph of H, and let $\mathcal{Z} = \{e_{u_1} + e_{v_1}, e_{u_2} + e_{v_2}, \ldots, e_{u_{t-1}} + e_{v_{t-1}}\}$. Note that if $y \perp \mathcal{Z}$ then for some γ, $y_u = \pm\gamma$, and in particular $(y_u - y_v)^2 = 4\gamma^2$ for all edges $\{u, v\}$ in H (here we are using that H is bipartite). So again using (5.3) we have

$$\begin{aligned}
\theta_k &= \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v) - \sum_{u \sim v} (y_u - y_v)^2 w_H(u,v)}{\sum_u y_u^2 d_G(u) - \sum_u y_u^2 d_H(u)} \right) \\
&\leq \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \perp \mathcal{Z}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v) - 2\gamma^2 \sum_u d_H(u)}{\sum_u y_u^2 d_G(u) - \gamma^2 \sum_u d_H(u)} \right) \\
&\leq \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \perp \mathcal{Z}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) \\
&\leq \max_{\mathcal{Y}^{k+t-1}} \left(\min_{y \perp \mathcal{Y}^{k+t-1}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) = \lambda_{k+t-1}.
\end{aligned}$$

We went from the first to the second line as before using the above comments on γ. In going from the second to the third line we used the following easily proved fact: let a, b, c be real with $2b \geq a \geq 2c \geq 0$ and $b > c \geq 0$, then $(a - 2c)/(b - c) \leq a/b$. That the assumptions on a, b, c are satisfied can be easily verified. Finally, we consider a broader optimization that would include the third line as a case.

We now turn to the case when $G - H$ has isolated vertices $u_1, u_2, \ldots, u_m$. We approach this by considering the graph G_ϵ which has loops of weight ϵ added to the vertices $u_1, u_2, \ldots, u_m$. A simple calculation shows that $\mathcal{L}(G_\epsilon - H) = \mathcal{L}(G - H)$, and in particular has the same eigenvalues, but now with the added loops has no isolated vertices. If we let λ_k^ϵ denote the kth eigenvalue of $\mathcal{L}(G_\epsilon)$ then the above derivation shows that

$$\lambda_{k-t+1}^\epsilon \;\leq\; \theta_k \;\leq\; \begin{cases} \lambda_{k+t-1}^\epsilon & H \text{ is bipartite,} \\ \lambda_{k+t}^\epsilon & \text{otherwise.} \end{cases}$$

We now let $\epsilon \to 0$, and since $\mathcal{L}(G_\epsilon) \to \mathcal{L}(G)$ then $\lambda_k^\epsilon \to \lambda_k$, and the result follows. □

5.3 Weak coverings and eigenvalues

The technique used to prove Theorem 29 can be used to prove other results. In this section we give an interlacing result for weak coverings. We say that G is a *weak cover* of H if there is some onto mapping $\pi : V(G) \to V(H)$ such that for all $u, v \in V(H)$,

$$w_H(u, v) = \sum_{\substack{x \in \pi^{-1}(u) \\ y \in \pi^{-1}(v)}} w_G(x, y).$$

From this definition it follows that $d_H(v) = \sum_{x \in \pi^{-1}(v)} d_G(x)$. Alternatively, for a weak covering we group the vertices of G in some manner then collapse the individual groups of vertices into single vertices of H. To find the edge weights of H we add the weights of any resulting parallel edges that are formed. An example of this is shown in Figure 5.2.

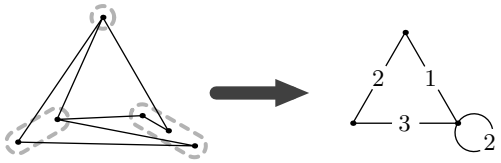

Figure 5.2: An example of a weak covering.

The idea of coverings for weighted graphs was previously considered by Chung and Yau [16]. In their definition of a cover they required additional structure which allows eigenvalues from H to be "lifted" up to G. Here we make fewer assumptions and so might no longer have eigenvalues lifting up (hence we give the name "weak cover").

Theorem 33. *Let G be a weak cover of H with $|V(G)| = n$ and $|V(H)| = m$, and further let $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$ and $\theta_0 \leq \theta_1 \leq \cdots \leq \theta_{m-1}$ be the eigenvalues of $\mathcal{L}(G)$*

and $\mathcal{L}(H)$ respectively. Then for $k = 0, 1, \ldots, m-1$ we have the following

$$\lambda_k \leq \theta_k \leq \lambda_{k+(n-m)}.$$

Proof. For $i = 1, \ldots, m$ let $V_i = \pi^{-1}(v_i)$, i.e., these are the groupings of the vertices of G, and let $\mathcal{Z}_i = \{e_{i_1} - e_{i_2}, e_{i_1} - e_{i_3}, \ldots, e_{i_1} - e_{i_j}\}$ where $V_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_j}\} \subseteq V(G)$. Further we will let $\mathcal{Z} = \bigcup_i \mathcal{Z}_i$. It is easy to check that the dimension of the span of $\mathcal{Z}$ is $n - m$. Now using (5.3) we have

$$\begin{aligned} \theta_k &= \max_{\mathcal{Y}^k \subseteq \mathbf{R}^m} \left(\min_{y \perp \mathcal{Y}^k, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_H(u,v)}{\sum_u y_u^2 d_H(u)} \right) \\ &= \max_{\mathcal{Y}^k \subseteq \mathbf{R}^n} \left(\min_{y \perp \mathcal{Y}^k, y \perp \mathcal{Z}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right). \end{aligned} \tag{5.5}$$

In the second step we used the defining property of weak covers to lift vectors from H to G so that we still satisfy the same Rayleigh quotient. Our only condition in lifting is that $y_i = y_{\pi(i)}$; in particular if $\pi(v_i) = \pi(v_j)$, then we need $y_{\pi(i)} = y_{\pi(j)}$. This last condition is easily achieved by requiring that the lifted vector be perpendicular to $\mathcal{Z}$.

We now bound (5.5) in two ways. First, we can drop the requirement that we remain perpendicular to $\mathcal{Z}$; thus we are minimizing over a larger set and so we have

$$\theta_k \geq \max_{\mathcal{Y}^k} \left(\min_{y \perp \mathcal{Y}^k, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) = \lambda_k.$$

The second approach is to maximize over some larger set that will also consider the case given in (5.5), i.e.,

$$\theta_k \leq \max_{\mathcal{Y}^{k+n-m}} \left(\min_{y \perp \mathcal{Y}^{k+n-m}, y \neq 0} \frac{\sum_{u \sim v} (y_u - y_v)^2 w_G(u,v)}{\sum_u y_u^2 d_G(u)} \right) = \lambda_{k+n-m}.$$

Combining the two inequalities above concludes the proof. □

Alternatively it is easy to show that Theorem 33 follows from a result of Haemers [23] on interlacing of eigenvalues of matrices. We have given the proof above to emphasize the nature of the approach.

5.4 Comments on interlacing for directed graphs

Recently Chung [12] defined a Laplacian for aperiodic strongly connected directed graphs and showed connections of its spectrum to mixing rates of random walks and isoperimetric properties, and in a subsequent paper gave a further connection to the

diameter of the graph [13]. A natural question is whether a similar interlacing result to Theorem 29 holds for the Laplacian of a directed graph. In this section we will partially answer this question in the negative.

5.4.1 Definition for the directed Laplacian

To define the Laplacian for a directed graph $\vec{G}$ we start with P the probability transition matrix with $P_{u,v}$ the probability of moving from u to v (for a weighted directed graph $P_{u,v} = w(u,v)/d_{out}(u)$ where $d_{out}(u) = \sum_t w(u,t)$ is the out-degree of u). If we let $\mathbf{1}$ denote the all 1s vector then $P\mathbf{1} = \mathbf{1}$. If we assume that the graph is strongly connected and aperiodic it follows from the Perron-Frobenius Theorem (see [26]) that there is a unique (row) vector ϕ for which $\phi P = \phi$ with $\phi(v) > 0$ for all v and $\sum_v \phi(v) = 1$. This vector ϕ is called the Perron vector of P.

If Φ is the diagonal matrix with $\Phi(v,v) = \phi(v)$, then the directed Laplacian is defined by Chung [12] as

$$\mathcal{L}(\vec{G}) = I - \frac{1}{2}\big(\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}\big),$$

where P^* denotes the transpose of P.

With this definition it is not too difficult to construct counterexamples to the corresponding statement of Theorem 28 for directed graphs. For example if we consider the simple directed graphs given in Figure 5.3 we have that $\lambda_2(\vec{G}) = \texttt{0.324609}\ldots < \texttt{0.362281}\ldots = \lambda_1(\vec{G} - \vec{H})$.

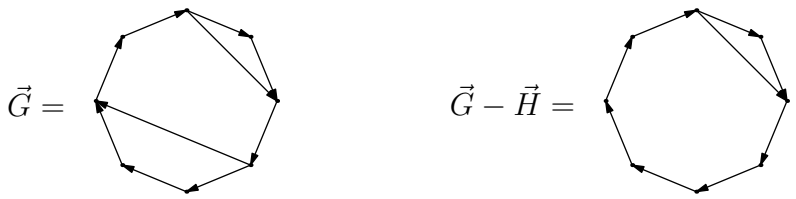

Figure 5.3: An example of a directed graph where the eigenvalues do not interlace.

5.4.2 Connecting directed Laplacians with undirected Laplacians

To understand why the corresponding statement of Theorem 28 could fail for directed graphs we connect the Laplacian for a directed graph with the Laplacian for a corresponding undirected graph.

Lemma 34. *Let $\vec{G}$ be an aperiodic strongly connected weighted directed graph and let H be a weighted undirected graph on the same vertex set with weights defined by*

$$w(u,v) = \phi(u)P(u,v) + \phi(v)P(v,u).$$

Then $\mathcal{L}(\vec{G}) = \mathcal{L}(H)$.

Proof. We note first that since ϕ is a left eigenvector of P and $\sum_u P(v,u) = 1$, it follows that $d_H(v) = \sum_u w(u,v) = 2\phi(v)$.

For terms on the diagonal we have

$$\mathcal{L}(\vec{G})_{v,v} \;=\; 1 - P(v,v) \;=\; 1 - \frac{w(v,v)}{2\phi(v)} \;=\; 1 - \frac{w(v,v)}{d_H(v)} = \mathcal{L}(H)_{v,v},$$

while for the off-diagonal terms

$$\begin{aligned}\mathcal{L}(\vec{G})_{u,v} &= -\frac{1}{2}\left(\sqrt{\frac{\phi(u)}{\phi(v)}}P(u,v) + \sqrt{\frac{\phi(v)}{\phi(u)}}P(v,u)\right) \\ &= -\frac{\big(\phi(u)P(u,v) + \phi(v)P(v,u)\big)}{\sqrt{(2\phi(u))(2\phi(v))}} \;=\; -\frac{w(u,v)}{\sqrt{d_H(u)d_H(v)}} \;=\; \mathcal{L}(H)_{u,v}. \qquad \square\end{aligned}$$

This connection between directed and undirected Laplacians can be used to establish several results. For instance it can be shown that the Cheeger inequality established for the directed graph $\vec{G}$ in Chung [12] is equivalent to the already known Cheeger inequality for the undirected graph H defined in Lemma 34.

The underlying principle of the directed Laplacian is based on circulations, i.e., a nonnegative function $F : V \times V \to [0,\infty)$ with the property that at each vertex u

$$\sum_v w(v,u) = \sum_t w(u,t).$$

If we think of the circulation F as a flow then the above equality can be interpreted as saying that at each vertex the in-flow equals the out-flow. Chung [12] showed that $F(u,v) = \phi(u)P(u,v)$ is a circulation and uses this to define the directed Laplacian. It is easy to adapt Lemma 34 to give new definitions for directed Laplacians using different types of circulations, and then establish some corresponding Cheeger inequalities and other similar results.

Applying Lemma 34 to the graphs in Figure 5.3 we get the weighted undirected graphs shown in Figure 5.4, where unspecified edges have weight 1. [We have scaled the weights in Figure 5.4 to more easily compare the two graphs; it is simple to see by (5.1) that scaling all the weights by some constant factor does not change the spectrum.]

Figure 5.4: The corresponding undirected graphs for graphs given in Figure 5.3.

Now we see that the removal of a single edge in the directed graph in Figure 5.3 had an effect on many edges in the underlying undirected graph in Figure 5.4. So by Theorem 29 the eigenvalues could spread by more than just 1, and in our case could spread by at most 3.

In general the removal of a single edge in a directed graph can have a tremendous impact on the underlying undirected graph. Using the results of Theorem 29 for the underlying undirected graph then we cannot in general guarantee a tight spread of the eigenvalues between the two graphs. Although for some special cases it can be shown that the removal of a single edge has a small impact on the underlying undirected graph in which case we can get a similar result.

This does not conclusively say that there is no corresponding statement such as Theorem 28, but only illustrates the difficulty of using the approach given by use of the Courant-Fischer Theorem. It would be interesting to see if there were some construction such that for each k there exists an aperiodic strongly connected graph directed graph $\vec{G}$ such that the removal of a single edge leaves an aperiodic strongly connected directed graph $\vec{G} - \vec{H}$ and either $\lambda_{m+k}(\vec{G}) < \lambda_m(\vec{G} - \vec{H})$ or $\lambda_{m-k}(\vec{G}) > \lambda_m(\vec{G} - \vec{H})$ for some m.

5.5 A heuristic reason why the middle of the spectrum is often ignored

The most commonly studied eigenvalues for the normalized Laplacian (and any matrix for that matter) have been λ_1 and λ_{n-1}. The further away from the extremes an eigenvalue lies, the less it has been studied. Theorem 29 gives a heuristic argument for why this should be the case.

The basic problem is that if a graph property can be changed by the addition or removal of a small number of edges (i.e., some graphs can go from connected to disconnected with the removal of a single edge), then the "interior" eigenvalues will not

change by much (since they tend to be clustered close together and so the eigenvalues will only have a small range that they can fall into), while the "fringe" eigenvalues can change by a nontrivial amount.

5.6 Some interlacing results using results of Haemers

In this section we will consider some results derived from interlacing that were known for the combinatorial Laplacian and show how to adopt them to the normalized Laplacian. Our starting point is the following classical result (see Haemers [23] for a short proof).

Theorem 35. *Let the matrix S of size $n \times m$ be such that $S^*S = I_m$ and let A be a Hermitian matrix of size n with eigenvalues $\mu_1 \leq \cdots \leq \mu_n$. Set $B = S^*AS$ and let $\eta_1 \leq \cdots \leq \eta_m$ be the eigenvalues of B. Then the eigenvalues of A and B interlace, that is $\mu_i \leq \eta_i \leq \mu_{n-m+i}$, for $i = 1, \ldots, m$.*

The proofs for the normalized Laplacian are very similar to those involving the combinatorial Laplacian and tend to involve a little more bookkeeping of the terms.

5.6.1 Interlacing sums of eigenvalues

We begin with the following which is a normalized version of a result given by Bollobás and Nikiforov [6].

Theorem 36. *Let G be a graph on n vertices with no isolated vertices, and let $\mathcal{L}$ denote the Laplacian of G with eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$. For every partitioning of the vertices $[n] = N_1 \cup \cdots \cup N_m$ we have*

$$\sum_{i=1}^{m-1} \lambda_i \leq m - \sum_{i=1}^{m} \frac{e(N_i, N_i)}{\operatorname{vol} N_i} = \sum_{1 \leq i < j \leq m} e(N_i, N_j)\left(\frac{1}{\operatorname{vol} N_i} + \frac{1}{\operatorname{vol} N_j}\right) \leq \sum_{i=0}^{m-1} \lambda_{n-1-i}.$$

Proof. Let $S = (s_{ij})$ be the $n \times m$ matrix defined by

$$s_{ij} = \begin{cases} \sqrt{d_i/\operatorname{vol} N_j} & \text{if } i \in N_j; \\ 0 & \text{otherwise.} \end{cases}$$

With this definition it is easy to check that $S^*S = I_m$. So by Theorem 35 we have that the eigenvalues of $B = S^*\mathcal{L}S$ are interlaced with the eigenvalues of $\mathcal{L}$. If we let $\eta_1 \leq \eta_2 \leq \cdots \leq \eta_m$ denote the eigenvalues of B, then we have

$$\lambda_{i-1} \leq \eta_i \leq \lambda_{n-1-m+i}.$$

From the interlacing inequalities it follows immediately that

$$\sum_{i=0}^{m-1} \lambda_i = \sum_{i=1}^{m} \lambda_{i-1} \leq \sum_{i=1}^{m} \eta_i \leq \sum_{i=1}^{m} \lambda_{n-1-m+i} = \sum_{i=0}^{m-1} \lambda_{n-1-i}.$$

Since $\lambda_0 = 0$ we can start the left hand sum at 1. On the other hand we can find the entries of $B = (b_{i,j})$, namely,

$$b_{i,j} = \begin{cases} 1 - \dfrac{e(N_i, N_i)}{\operatorname{vol} N_i} & \text{if } i = j; \\ -\dfrac{e(N_i, N_j)}{\sqrt{\operatorname{vol} N_i \operatorname{vol} N_j}} & \text{if } i \neq j. \end{cases}$$

Using this one has

$$\sum_{i=1}^{m} \eta_i = \operatorname{trace} B = \sum_{i=1}^{m} \left(1 - \frac{e(N_i, N_i)}{\operatorname{vol} N_i}\right) = m - \sum_{i=1}^{m} \frac{e(N_i, N_i)}{\operatorname{vol} N_i}.$$

Alternatively we note that

$$1 - \frac{e(N_i, N_i)}{\operatorname{vol} N_i} = \frac{\operatorname{vol} N_i - e(N_i, N_i)}{\operatorname{vol} N_i} = \frac{e(N_i, [n] \setminus N_i)}{\operatorname{vol} N_i} = \sum_{j \neq i} \frac{e(N_i, N_j)}{\operatorname{vol} N_i},$$

and so

$$\sum_{i=1}^{m} \mu_i = \sum_{i=1}^{m} \sum_{j \neq i} \frac{e(N_i, N_j)}{\operatorname{vol} N_i} = \sum_{1 \leq i < j \leq m} e(N_i, N_j) \left(\frac{1}{\operatorname{vol} N_i} + \frac{1}{\operatorname{vol} N_j}\right).$$

Combining everything gives the desired result. □

5.6.2 Large disjoint sets

A surprising application of Theorem 35 is the following which is a normalized version of a result of Haemers [23].

Theorem 37. *Let G be a connected graph on $n \geq 2$ vertices, and let $\mathcal{L}$ denote the Laplacian of G with eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$. If X and Y are disjoint subsets of the vertices of G with no edges between X and Y, then*

$$\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} \overline{X} \operatorname{vol} \overline{Y}} \leq \left(\frac{\lambda_{n-1} - \lambda_1}{\lambda_{n-1} + \lambda_1}\right)^2,$$

where $\overline{X} = V \setminus X$ and $\overline{Y} = V \setminus Y$.

Proof. We will let ψ_U denote the characteristic vector of a subset U, i.e.,

$$\psi_U(v) = \begin{cases} 1 & v \in U; \\ 0 & v \notin U. \end{cases}$$

Let $\lambda = -\frac{1}{2}(\lambda_{n-1} + \lambda_1)$, and consider the following matrix

$$\mathcal{A} = \begin{bmatrix} O & \mathcal{L} + \lambda I \\ \mathcal{L} + \lambda I & O \end{bmatrix}.$$

This is a $(2n)\times(2n)$ matrix with eigenvalues $\pm(\lambda_i + \lambda)$, and where $-\lambda$, $\frac{1}{2}(\lambda_{n-1} - \lambda_1)$, $-\frac{1}{2}(\lambda_{n-1} - \lambda_1)$, and λ are (respectively) the largest, second largest, second smallest, and smallest eigenvalues. Recall that D is the diagonal degree matrix of G and consider the following $2n \times 4$ matrix

$$S = \begin{bmatrix} D^{1/2}\psi_X & D^{1/2}\psi_{\overline{X}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & D^{1/2}\psi_{\overline{Y}} & D^{1/2}\psi_Y \end{bmatrix},$$

here $\mathbf{0}$ denotes the all 0s column vector. Finally, let

$$E = \begin{bmatrix} \operatorname{vol} X & 0 & 0 & 0 \\ 0 & \operatorname{vol} \overline{X} & 0 & 0 \\ 0 & 0 & \operatorname{vol} \overline{Y} & 0 \\ 0 & 0 & 0 & \operatorname{vol} Y \end{bmatrix}.$$

A straightforward calculation shows that $(SE^{-1/2})^*(SE^{-1/2}) = I_4$, so by Thereom 35 we have that the eigenvalues of $\mathcal{A}$ and $\mathcal{B} = (SE^{-1/2})^*\mathcal{A}(SE^{-1/2})$ interlace. More particularly, since $E^{-1/2}\mathcal{B}E^{1/2}$ is similar to $\mathcal{B}$ we have that the eigenvalues of $E^{-1}S^*\mathcal{A}S$ interlace with the eigenvalues of $\mathcal{A}$. Calculating we have

$$E^{-1}S^*\mathcal{A}S = \begin{bmatrix} O & \mathcal{B}_1 \\ \mathcal{B}_2 & O \end{bmatrix}, \text{ where,}$$

$$\mathcal{B}_1 = \begin{bmatrix} \dfrac{(\lambda+1)\operatorname{vol}(X \cap \overline{Y}) - e(X, \overline{Y})}{\operatorname{vol} X} & \dfrac{(\lambda+1)\operatorname{vol}(X \cap Y) - e(X, Y)}{\operatorname{vol} X} \\ \dfrac{(\lambda+1)\operatorname{vol}(\overline{X} \cap \overline{Y}) - e(\overline{X}, \overline{Y})}{\operatorname{vol} \overline{X}} & \dfrac{(\lambda+1)\operatorname{vol}(\overline{X} \cap Y) - e(\overline{X}, Y)}{\operatorname{vol} \overline{X}} \end{bmatrix},$$

$$\mathcal{B}_2 = \begin{bmatrix} \dfrac{(\lambda+1)\operatorname{vol}(X \cap \overline{Y}) - e(X, \overline{Y})}{\operatorname{vol} \overline{Y}} & \dfrac{(\lambda+1)\operatorname{vol}(\overline{X} \cap \overline{Y}) - e(\overline{X}, \overline{Y})}{\operatorname{vol} \overline{Y}} \\ \dfrac{(\lambda+1)\operatorname{vol}(X \cap Y) - e(X, Y)}{\operatorname{vol} Y} & \dfrac{(\lambda+1)\operatorname{vol}(\overline{X} \cap Y) - e(\overline{X}, Y)}{\operatorname{vol} Y} \end{bmatrix}.$$

By our assumptions we have that $X \cap Y = \emptyset$ (so $X \subseteq \overline{Y}$ and $Y \subseteq \overline{X}$), and $e(X, Y) = 0$. It follows that $\mathrm{vol}(X \cap \overline{Y}) = \mathrm{vol}\, X$ and $\mathrm{vol}(\overline{X} \cap Y) = \mathrm{vol}\, Y$, and a simple computation shows that

$$e(X, \overline{Y}) = e(X, Y) + e(X, \overline{Y}) = e(X, V) = \mathrm{vol}\, X,$$

so we have $e(X, \overline{Y}) = \mathrm{vol}(X \cap \overline{Y}) = \mathrm{vol}\, X$; similar computations show that $e(\overline{X}, Y) = \mathrm{vol}(\overline{X} \cap Y) = \mathrm{vol}\, Y$. Next note that

$$\mathrm{vol}\, \overline{X} = e(\overline{X}, V) = e(\overline{X}, Y) + e(\overline{X}, \overline{Y}) = \mathrm{vol}\, Y + e(\overline{X}, \overline{Y}),$$

showing that $e(\overline{X}, \overline{Y}) = \mathrm{vol}\, \overline{X} - \mathrm{vol}\, Y$, and a similar calculation shows that $e(\overline{X}, \overline{Y}) = \mathrm{vol}\, \overline{Y} - \mathrm{vol}\, X$.

Using the above we simplify to get

$$E^{-1} S^* \mathcal{A} S = \begin{bmatrix} 0 & 0 & \lambda & 0 \\ 0 & 0 & \left(1 - \dfrac{\mathrm{vol}\, Y}{\mathrm{vol}\, \overline{X}}\right)\lambda & \dfrac{\mathrm{vol}\, Y}{\mathrm{vol}\, \overline{X}}\lambda \\ \dfrac{\mathrm{vol}\, X}{\mathrm{vol}\, \overline{Y}}\lambda & \left(1 - \dfrac{\mathrm{vol}\, X}{\mathrm{vol}\, \overline{Y}}\right)\lambda & 0 & 0 \\ 0 & \lambda & 0 & 0 \end{bmatrix}.$$

If we let $\nu_1 \leq \nu_2 \leq \nu_3 \leq \nu_4$ denote the eigenvalues of $E^{-1} S^* \mathcal{A} S$ then it is easy to check that $\nu_1 = \lambda$, $\nu_4 = -\lambda$ and by interlacing that $-\nu_2 \leq \frac{1}{2}(\lambda_{n-1} - \lambda_1)$ and $\nu_3 \leq \frac{1}{2}(\lambda_{n-1} - \lambda_1)$. In particular, we have

$$\frac{\mathrm{vol}\, X\, \mathrm{vol}\, Y}{\mathrm{vol}\, \overline{X}\, \mathrm{vol}\, \overline{Y}} \lambda^4 = \det(E^{-1} S^* \mathcal{A} S) = \nu_1 \nu_2 \nu_3 \nu_4 \leq \frac{1}{4}(\lambda_{n-1} - \lambda_1)^2 \lambda^2,$$

which upon rearranging and substituting the definition of λ gives,

$$\frac{\mathrm{vol}\, X\, \mathrm{vol}\, Y}{\mathrm{vol}\, \overline{X}\, \mathrm{vol}\, \overline{Y}} \leq \left(\frac{\lambda_{n-1} - \lambda_1}{\lambda_{n-1} + \lambda_1}\right)^2. \qquad \square$$

5.6.3 An extension to bipartite graphs

We can mimic the proof of Theorem 37 to get a (slightly) better result for bipartite graphs.

Lemma 38. *Let G be a connected bipartite graph with $n \geq 2$ vertices, and vertex set $V_1 \cup V_2$ where edges in G go between V_1 and V_2. Let $\mathcal{L}$ denote the Laplacian of G with*

eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1} = 2$. *If* $X \subseteq V_1$ *and* $Y \subseteq V_2$ *with no edges between* X *and* Y, *then*

$$\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol}(V_1 \backslash X) \operatorname{vol}(V_2 \backslash Y)} \leq (1 - \lambda_1)^2.$$

Proof. We consider the matrix

$$I - \mathcal{L} = D^{-1/2} A D^{-1/2} = \begin{bmatrix} O & B \\ B^* & O \end{bmatrix},$$

where the initial rows are indexed by V_1 and the terminal rows are indexed by V_2. This matrix has eigenvalues $-1 \leq 1 - \lambda_{n-2} \leq \cdots \leq 1 - \lambda_1 \leq 1$. We interlace as in Theorem 37 where we now let S be the $n \times 4$ matrix

$$S = \begin{bmatrix} \psi_X & \psi_{V_1 \backslash X} & \psi_{V_2 \backslash Y} & \psi_Y \end{bmatrix},$$

and

$$E = \begin{bmatrix} \operatorname{vol} X & 0 & 0 & \\ 0 & \operatorname{vol} V_1 \backslash X & 0 & 0 \\ 0 & 0 & \operatorname{vol} V_2 \backslash Y & 0 \\ 0 & 0 & 0 & \operatorname{vol} Y \end{bmatrix}.$$

So the eigenvalues of $\mathcal{A} = I - \mathcal{L}$ and

$$E^{-1} S^* \mathcal{A} S = \begin{bmatrix} 0 & 0 & \dfrac{e(X, V_2 \backslash Y)}{\operatorname{vol} X} & \dfrac{e(X, Y)}{\operatorname{vol} X} \\ 0 & 0 & \dfrac{e(V_1 \backslash X, V_2 \backslash Y)}{\operatorname{vol} V_1 \backslash X} & \dfrac{e(V_1 \backslash X, Y)}{\operatorname{vol} V_1 \backslash X} \\ \dfrac{e(X, V_2 \backslash Y)}{\operatorname{vol} V_2 \backslash Y} & \dfrac{e(V_1 \backslash X, V_2 \backslash Y)}{\operatorname{vol} V_2 \backslash Y} & 0 & 0 \\ \dfrac{e(X, Y)}{\operatorname{vol} Y} & \dfrac{e(V_1 \backslash X, Y)}{\operatorname{vol} Y} & 0 & 0 \end{bmatrix}$$

interlace. Doing similar calculations as in Corollary 37 we have that

$$e(X, V_2 \backslash Y) = \operatorname{vol} X, \ e(V_1 \backslash X, Y) = \operatorname{vol} Y, \text{ and}$$
$$e(V_1 \backslash X, V_2 \backslash Y) = \operatorname{vol} V_1 \backslash X - \operatorname{vol} Y = \operatorname{vol} V_2 \backslash Y - \operatorname{vol} X.$$

Using these we have that

$$E^{-1} S^* \mathcal{A} S = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 - \dfrac{\operatorname{vol} Y}{\operatorname{vol} V_1 \backslash X} & \dfrac{\operatorname{vol} Y}{\operatorname{vol} V_1 \backslash X} \\ \dfrac{\operatorname{vol} X}{\operatorname{vol} V_2 \backslash Y} & 1 - \dfrac{\operatorname{vol} X}{\operatorname{vol} V_2 \backslash Y} & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

which has eigenvalues $-1 = \nu_1 \leq \nu_2 \leq \nu_3 \leq \nu_4 = 1$. By interlacing we have that $\nu_3 \leq 1 - \lambda_1$ and $-\nu_2 \leq \lambda_{n-2} - 1$.

So we now have that

$$\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol}(V_1 \backslash X) \operatorname{vol}(V_2 \backslash Y)} = \det E^{-1} S^* \mathcal{A} S = -\nu_2 \nu_3 \leq (1 - \lambda_1)(\lambda_{n-2} - 1).$$

Finally we note that by the Peron-Frobenius Theorem, the spectrum of $I - \mathcal{L}$ is symmetric and so $\lambda_{n-2} - 1 = 1 - \lambda_1$, substituting this in we have

$$\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol}(V_1 \backslash X) \operatorname{vol}(V_2 \backslash Y)} \leq (1 - \lambda_1)^2. \qquad \square$$

We can compare this result with a general result for discrepancy in a bipartite graph.

Lemma 39. *Let G be a connected bipartite graph with $n \geq 2$ vertices, and vertex set $V_1 \cup V_2$ where edges in G only go between V_1 and V_2. Then for all $X \subseteq V_1$ and $Y \subseteq V_2$ we have that*

$$\left| e(X, Y) - \frac{\operatorname{vol} X \ \operatorname{vol} Y}{\operatorname{vol} G} \right| \leq (1 - \lambda_1) \sqrt{\operatorname{vol} X \ \operatorname{vol} Y},$$

where λ_1 is the second smallest eigenvalue of $\mathcal{L}$, and we define $\operatorname{vol} G := \operatorname{vol} V_1 = \operatorname{vol} V_2$.

For the case when there are no edges between X and Y Lemma 39 implies

$$\frac{\operatorname{vol} X \ \operatorname{vol} Y}{\operatorname{vol} G \operatorname{vol} G} \leq (1 - \lambda_1)^2,$$

which is a weaker result than the one given in Lemma 38

Proof of Lemma 39. We recall that

$$I - \mathcal{L} = D^{-1/2} A D^{-1/2} = \begin{bmatrix} O & B \\ B^* & O \end{bmatrix},$$

has largest and smallest eigenvalues 1 and -1 respectively. In particular if we let D_{V_1} and D_{V_2} denote the restrictions of the degree matrices to V_1 and V_2 respectively we have that

$$S_1 = \frac{1}{\sqrt{2 \operatorname{vol} G}} \begin{bmatrix} D_{V_1}^{1/2} \mathbf{1} \\ D_{V_2}^{1/2} \mathbf{1} \end{bmatrix},$$

and,

$$S_2 = \frac{1}{\sqrt{2 \operatorname{vol} G}} \begin{bmatrix} D_{V_1}^{1/2} \mathbf{1} \\ -D_{V_2}^{1/2} \mathbf{1} \end{bmatrix},$$

are the normalized eigenvectors associated with 1 and -1 respectively. We consider the matrix $\mathcal{B}$ defined as

$$D^{-1/2}AD^{-1/2} + \frac{1}{2\operatorname{vol} G}D^{1/2}JD^{1/2} - \frac{1}{2\operatorname{vol} G}D^{1/2}\begin{bmatrix} J_{|V_1|} & -J_{|V_1|\times|V_2|} \\ -J_{|V_2|\times|V_1|} & J_{|V_2|} \end{bmatrix} D^{1/2}$$

which a straightforward calculation shows that

$$e(X,Y) - \frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} G} = \langle D^{1/2}\psi_X, \mathcal{B}D^{1/2}\psi_Y\rangle.$$

Finally, using that $|\langle x, My\rangle| \leq \|M\|\,\|x\|\,\|y\|$, where $\|M\|$ denote the largest singular value of M, we have that

$$\left|e(X,Y) - \frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} G}\right| \leq \|\mathcal{B}\|\,\|D^{1/2}\psi_X\|\,\|D^{1/2}\psi_Y\| = \|\mathcal{B}\|\sqrt{\operatorname{vol} X \operatorname{vol} Y}.$$

All that remains is to determine $\|\mathcal{B}\|$, this is done by noting that this is the largest (normed) eigenvalue of $\mathcal{B}$ which has the same eigenvalues as $I - \mathcal{L}$, except we have "subtracted" out ± 1. So the largest eigenvalue is $(1-\lambda_1)$ (as in Corollary 37 by symmetry this is equal to $(\lambda_{n-2} - 1)$). Substituting this in gives the desired result. □

Acknowledgement

Sections 5.1 through 5.4 are based on the paper "Interlacing for weighted graphs using the normalized Laplacian" in *Electronic Journal of Linear Algebra* (2007).

Bibliography

[1] Y. Bilu and N. Linial, *Lifts, discrepancy and nearly optimal spectral gap*, Combinatorica 26 (2006), pp. 495–519.

[2] Y. Bilu and N. Linial, *Constructing expander graphs by* 2*-lifts and discrepancy vs. spectral gap*, FOCS 2004 Proceedings, pp. 404–412.

[3] T. Bıyıkoğlu, J. Leydold and P.F. Stadler, Laplacian Eigenvectors of Graphs, Lecture Notes in Mathematics 1915, Springer, 2007.

[4] R.A. Brualdi and H.J. Ryser, Combinatorial Matrix Theory, Cambridge, 1991.

[5] B. Bollobás and V. Nikiforov, *Hermitian matrices and graphs: singular values and discrepancy*, Discrete Math. 285 (2004), pp.17–32.

[6] B. Bollobás and V. Nikiforov, *Graphs and Hermitian matrices: eigenvalue interlacing*, Discrete Math. 289 (2004), pp. 119–127.

[7] S. Butler, *Eigenvalues of* 2*-edge-coverings*, submitted.

[8] S. Butler, *Interlacing for weighted graphs using the normalized Laplacian*, Electronic Journal of Linear Algebra 16 (2007), pp. 90–98.

[9] S. Butler, "Relating singular values and discrepancy of weighted directed graphs", SODA 2006 Proceedings, pp. 1112-1116.

[10] G. Chen, G. Davis, F. Hall, Z. Li, K. Patel, and M. Stewart, *An interlacing result on normalized Laplacians*, SIAM J. Discrete Math. 18 (2004), pp. 353–361.

[11] F. Chung, Spectral Graph Theory, CBMS Lecture Notes, AMS, Providence, RI, 1997.

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

[13] F. Chung, *The diameter and Laplacian eigenvalues of directed graphs*, Electronic Journal of Combinatorics 13 (2006), N4, 6 pp.

[14] F. Chung and R.L. Graham, *Quasi-random graphs with given degree sequences*, Random Structures & Algorithms 32 (2008), pp. 1–19.

[15] F. Chung, R.L. Graham and R.W. Wilson, *Quasi-random graphs*, Combinatorica 9 (1989), pp. 345–362.

[16] F. Chung and S.-T. Yau, *Coverings, heat kernels and spanning trees*, Electronic Journal of Combinatorics 6 (1999), R12, 21 pp.

[17] S.S. D'Amato, *Eigenvalues of graphs with twofold symmetry*, Molecular Physics 37 (1979), pp. 1363–1369.

[18] S.S. D'Amato, *Eigenvalues of graphs with threefold symmetry*, Theoretica Chimica Acta 53 (1979), pp. 319–326.

[19] E.R. van Dam, and W.H. Haemers, *Which graphs are determined by their spectrum?*, Linear Algebra and its Applications **373** (2003), pp. 241–272.

[20] C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.

[21] C.D. Godsil and B.D. McKay, *Constructing cospectral graphs*, Aequationes Mathematicae **25** (1982), pp. 257–268.

[22] F. Goldberg, *Bounding the gap between extremal Laplacian eigenvalues of graphs*, Linear Algebra and its Applications 416 (2006), pp. 68–74.

[23] W.H. Haemers, Interlacing eigenvalues and graphs, *Linear Algebra and its Applications* **226-228** (1995), 593–616.

[24] W.H. Haemers and E. Spence, *Enumeration of cospectral graphs*, European Journal of Combinatorics **25** (2004), pp. 199–211.

[25] L. Hogben, *Spectral graph theory and the inverse eigenvalues problem of a graph*, Electronic Journal of Linear Algebra 14 (2005), pp. 12–31.

[26] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990.

[27] Y.P. Hou and J.S. Li, *On the Laplacian eigenvalues of signed graphs*, Linear and Multilinear Algebra 51 (2003), pp. 21–30.

[28] C.-K. Li, *A short proof of interlacing inequalities on normalized Laplacians*, Linear Algebra and its Applications 414 (2006), pp. 425–427.

[29] N. Ozeki, *On the estimation of the inequality by the maximum*, J. College Arts Chiba Univ. 5 (1968), pp. 199–203.

[30] G. Pólya and G. Szegö, Problems and Theorems in Analysis, Vol. I, Springer-Verlag, 1972.

[31] G.W. Stewart, *Matrix Algorithms*, SIAM, 1998.

[32] D. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, 1996.