



1-1-2013

Notes on Elementary Spectral Graph Theory Applications to Graph Clustering Using Normalized Cuts

Jean H. Gallier

University of Pennsylvania, jean@cis.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/cis_reports

 Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

Jean H. Gallier, "Notes on Elementary Spectral Graph Theory Applications to Graph Clustering Using Normalized Cuts", . January 2013.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-13-09.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/cis_reports/986
For more information, please contact repository@pobox.upenn.edu.

Notes on Elementary Spectral Graph Theory Applications to Graph Clustering Using Normalized Cuts

Abstract

These are notes on the method of normalized graph cuts and its applications to graph clustering. I provide a fairly thorough treatment of this deeply original method due to Shi and Malik, including complete proofs. I include the necessary background on graphs and graph Laplacians. I then explain in detail how the eigenvectors of the graph Laplacian can be used to draw a graph. This is an attractive application of graph Laplacians. The main thrust of this paper is the method of normalized cuts. I give a detailed account for $K = 2$ clusters, and also for $K > 2$ clusters, based on the work of Yu and Shi. Three points that do not appear to have been clearly articulated before are elaborated: 1. The solutions of the main optimization problem should be viewed as tuples in the K -fold cartesian product of projective space RP^{N-1} . 2. When $K > 2$, the solutions of the relaxed problem should be viewed as elements of the Grassmannian $G(K,N)$.

Disciplines

Computer Engineering | Computer Sciences

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-13-09.

Notes on Elementary Spectral Graph Theory Applications to Graph Clustering Using Normalized Cuts

Jean Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@cis.upenn.edu

© Jean Gallier

November 11, 2013

Abstract: These are notes on the method of normalized graph cuts and its applications to graph clustering. I provide a fairly thorough treatment of this deeply original method due to Shi and Malik, including complete proofs. I include the necessary background on graphs and graph Laplacians. I then explain in detail how the eigenvectors of the graph Laplacian can be used to draw a graph. This is an attractive application of graph Laplacians. The main thrust of this paper is the method of normalized cuts. I give a detailed account for $K = 2$ clusters, and also for $K > 2$ clusters, based on the work of Yu and Shi. Three points that do not appear to have been clearly articulated before are elaborated:

1. The solutions of the main optimization problem should be viewed as tuples in the K -fold cartesian product of projective space \mathbb{RP}^{N-1} .
2. When $K > 2$, the solutions of the relaxed problem should be viewed as elements of the Grassmannian $G(K, N)$.
3. Two possible Riemannian distances are available to compare the closeness of solutions: (a) The distance on $(\mathbb{RP}^{N-1})^K$. (b) The distance on the Grassmannian.

I also clarify what should be the necessary and sufficient conditions for a matrix to represent a partition of the vertices of a graph to be clustered.

Contents

1	Introduction	7
2	Graphs and Graph Laplacians; Basic Facts	15
2.1	Directed Graphs, Undirected Graphs, Weighted Graphs	15
2.2	Laplacian Matrices of Graphs	20
3	Spectral Graph Drawing	27
3.1	Graph Drawing and Energy Minimization	27
3.2	Examples of Graph Drawings	30
4	Graph Clustering	35
4.1	Graph Clustering Using Normalized Cuts	35
4.2	Special Case: 2-Way Clustering Using Normalized Cuts	36
4.3	K -Way Clustering Using Normalized Cuts	43
4.4	K -Way Clustering; Using The Dependencies Among X^1, \dots, X^K	55
4.5	Discrete Solution Close to a Continuous Approximation	58
A	The Rayleigh Ratio and the Courant-Fischer Theorem	63
B	Riemannian Metrics on Quotient Manifolds	69
	Bibliography	75

Chapter 1

Introduction

In the Fall of 2012, my friend Kurt Reillag suggested that I should be ashamed about knowing so little about graph Laplacians and normalized graph cuts. These notes are the result of my efforts to rectify this situation.

I begin with a review of basic notions of graph theory. Even though the graph Laplacian is fundamentally associated with an undirected graph, I review the definition of both directed and undirected graphs. For both directed and undirected graphs, I define the degree matrix D , the incidence matrix \tilde{D} , and the adjacency matrix A . I also define weighted graphs (with nonnegative weights), and the notions of *volume*, $\text{vol}(A)$ of a set of nodes A , of *links*, $\text{links}(A, B)$ between two sets of nodes A, B , and of *cut*, $\text{cut}(A) = \text{links}(A, \bar{A})$ of a set of nodes A . These concepts play a crucial role in the theory of normalized cuts. Then, I introduce the (unnormalized) *graph Laplacian* L of a directed graph G in an “old-fashion,” by showing that for any orientation of a graph G ,

$$\tilde{D}\tilde{D}^\top = D - A = L$$

is an invariant. I also define the (unnormalized) *graph Laplacian* L of a weighted graph (V, W) as $L = D - W$, and prove that

$$x^\top Lx = \frac{1}{2} \sum_{i,j=1}^m w_{ij}(x_i - x_j)^2 \quad \text{for all } x \in \mathbb{R}^m.$$

Consequently, $x^\top Lx$ does not depend on the diagonal entries in W , and if $w_{ij} \geq 0$ for all $i, j \in \{1, \dots, m\}$, then L is positive semidefinite. Then, if W consists of nonnegative entries, the eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ of L are real and nonnegative, and there is an orthonormal basis of eigenvectors of L . I show that the number of connected components of the graph $G = (V, W)$ is equal to the dimension of the kernel of L .

I also define the normalized graph Laplacians L_{sym} and L_{rw} , given by

$$\begin{aligned} L_{\text{sym}} &= D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2} \\ L_{\text{rw}} &= D^{-1} L = I - D^{-1} W, \end{aligned}$$

and prove some simple properties relating the eigenvalues and the eigenvectors of L , L_{sym} and L_{rw} . These normalized graph Laplacians show up when dealing with normalized cuts.

Next, I turn to *graph drawings* (Chapter 3). Graph drawing is a very attractive application of so-called spectral techniques, which is a fancy way of saying that eigenvalues and eigenvectors of the graph Laplacian are used. Furthermore, it turns out that graph clustering using normalized cuts can be cast as a certain type of graph drawing.

Given an undirected graph $G = (V, E)$, with $|V| = m$, we would like to draw G in \mathbb{R}^n for n (much) smaller than m . The idea is to assign a point $\rho(v_i)$ in \mathbb{R}^n to the vertex $v_i \in V$, for every $v_i \in V$, and to draw a line segment between the points $\rho(v_i)$ and $\rho(v_j)$. Thus, a *graph drawing* is a function $\rho: V \rightarrow \mathbb{R}^n$.

We define the *matrix of a graph drawing* ρ (in \mathbb{R}^n) as a $m \times n$ matrix R whose i th row consists of the row vector $\rho(v_i)$ corresponding to the point representing v_i in \mathbb{R}^n . Typically, we want $n < m$; in fact n should be much smaller than m .

Since there are infinitely many graph drawings, it is desirable to have some criterion to decide which graph is better than another. Inspired by a physical model in which the edges are springs, it is natural to consider a representation to be better if it requires the springs to be less extended. We can formalize this by defining the *energy* of a drawing R by

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} \|\rho(v_i) - \rho(v_j)\|^2,$$

where $\rho(v_i)$ is the i th row of R and $\|\rho(v_i) - \rho(v_j)\|^2$ is the square of the Euclidean length of the line segment joining $\rho(v_i)$ and $\rho(v_j)$.

Then, “good drawings” are drawings that minimize the energy function \mathcal{E} . Of course, the trivial representation corresponding to the zero matrix is optimum, so we need to impose extra constraints to rule out the trivial solution.

We can consider the more general situation where the springs are not necessarily identical. This can be modeled by a symmetric weight (or stiffness) matrix $W = (w_{ij})$, with $w_{ij} \geq 0$. In this case, our energy function becomes

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} w_{ij} \|\rho(v_i) - \rho(v_j)\|^2.$$

Following Godsil and Royle [8], we prove that

$$\mathcal{E}(R) = \text{tr}(R^\top L R),$$

where

$$L = D - W,$$

is the familiar unnormalized Laplacian matrix associated with W , and where D is the degree matrix associated with W .

It can be shown that there is no loss in generality in assuming that the columns of R are pairwise orthogonal and that they have unit length. Such a matrix satisfies the equation $R^\top R = I$ and the corresponding drawing is called an *orthogonal drawing*. This condition also rules out trivial drawings.

Then, I prove the main theorem about graph drawings (Theorem 3.2), which essentially says that the matrix R of the desired graph drawing is constituted by the n eigenvectors of L associated with the smallest nonzero n eigenvalues of L . We give a number examples of graph drawings, many of which are borrowed or adapted from Spielman [13].

The next chapter (Chapter 4) contains the “meat” of this document. This chapter is devoted to the method of normalized graph cuts for graph clustering. This beautiful and deeply original method first published in Shi and Malik [12], has now come to be a “textbook chapter” of computer vision and machine learning. It was invented by Jianbo Shi and Jitendra Malik, and was the main topic of Shi’s dissertation. This method was extended to $K \geq 3$ clusters by Stella Yu in her dissertation [15], and is also the subject of Yu and Shi [16].

Given a set of data, the goal of clustering is to partition the data into different groups according to their similarities. When the data is given in terms of a similarity graph G , where the weight w_{ij} between two nodes v_i and v_j is a measure of similarity of v_i and v_j , the problem can be stated as follows: Find a partition (A_1, \dots, A_K) of the set of nodes V into different groups such that the edges between different groups have very low weight (which indicates that the points in different clusters are dissimilar), and the edges within a group have high weight (which indicates that points within the same cluster are similar).

The above graph clustering problem can be formalized as an optimization problem, using the notion of cut mentioned earlier. If we want to partition V into K clusters, we can do so by finding a partition (A_1, \dots, A_K) that minimizes the quantity

$$\text{cut}(A_1, \dots, A_K) = \frac{1}{2} \sum_{i=1}^K \text{cut}(A_i).$$

For $K = 2$, the mincut problem is a classical problem that can be solved efficiently, but in practice, it does not yield satisfactory partitions. Indeed, in many cases, the mincut solution separates one vertex from the rest of the graph. What we need is to design our cost function in such a way that it keeps the subsets A_i “reasonably large” (reasonably balanced).

An example of a weighted graph and a partition of its nodes into two clusters is shown in Figure 1.1.

A way to get around this problem is to normalize the cuts by dividing by some measure of each subset A_i . A solution using the volume $\text{vol}(A_i)$ of A_i (for $K = 2$) was proposed and investigated in a seminal paper of Shi and Malik [12]. Subsequently, Yu (in her dissertation [15]) and Yu and Shi [16] extended the method to $K > 2$ clusters. The idea is to minimize

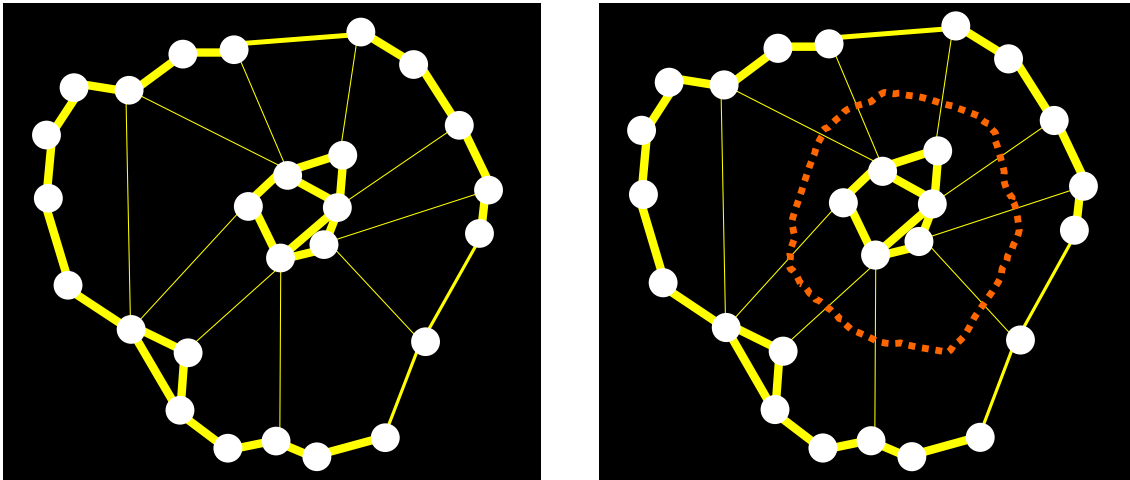


Figure 1.1: A weighted graph and its partition into two clusters.

the cost function

$$\text{Ncut}(A_1, \dots, A_K) = \sum_{i=1}^K \frac{\text{links}(A_i, \bar{A}_i)}{\text{vol}(A_i)} = \sum_{i=1}^K \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)}.$$

The first step is to express our optimization problem in matrix form. In the case of two clusters, a single vector X can be used to describe the partition $(A_1, A_2) = (A, \bar{A})$. We need to choose the structure of this vector in such a way that

$$\text{Ncut}(A, \bar{A}) = \frac{X^\top LX}{X^\top DX},$$

where the term on the right-hand side is a Rayleigh ratio.

After careful study of the original papers, I discovered various facts that were implicit in these works, but I feel are important to be pointed out explicitly.

First, I realized that it is important to pick a vector representation which is invariant under multiplication by a nonzero scalar, because the Rayleigh ratio is scale-invariant, and it is crucial to take advantage of this fact to make the denominator go away. This implies that *the solutions X are points in the projective space \mathbb{RP}^{N-1}* . This was my first revelation.

Let $N = |V|$ be the number of nodes in the graph G . In view of the desire for a scale-invariant representation, it is natural to assume that the vector X is of the form

$$X = (x_1, \dots, x_N),$$

where $x_i \in \{a, b\}$ for $i = 1, \dots, N$, for any two distinct real numbers a, b . This is an indicator vector in the sense that, for $i = 1, \dots, N$,

$$x_i = \begin{cases} a & \text{if } v_i \in A \\ b & \text{if } v_i \notin A. \end{cases}$$

The choice $a = +1, b = -1$ is natural, but premature. The correct interpretation is really to view X as a representative of a point in the real projective space \mathbb{RP}^{N-1} , namely the point $\mathbb{P}(X)$ of homogeneous coordinates $(x_1: \dots: x_N)$.

Let $d = \mathbf{1}^\top D \mathbf{1}$ and $\alpha = \text{vol}(A)$. I prove that

$$\text{Ncut}(A, \bar{A}) = \frac{X^\top L X}{X^\top D X}$$

holds iff the following condition holds:

$$a\alpha + b(d - \alpha) = 0. \quad (\dagger)$$

Note that condition (\dagger) applied to a vector X whose components are a or b is equivalent to the fact that X is orthogonal to $D\mathbf{1}$, since

$$X^\top D \mathbf{1} = \alpha a + (d - \alpha)b,$$

where $\alpha = \text{vol}(\{v_i \in V \mid x_i = a\})$.

If we let

$$\mathcal{X} = \{(x_1, \dots, x_N) \mid x_i \in \{a, b\}, a, b \in \mathbb{R}, a, b \neq 0\},$$

our solution set is

$$\mathcal{K} = \{X \in \mathcal{X} \mid X^\top D \mathbf{1} = 0\}.$$

Actually, to be perfectly rigorous, we are looking for solutions in \mathbb{RP}^{N-1} , so our solution set is really

$$\mathbb{P}(\mathcal{K}) = \{(x_1: \dots: x_N) \in \mathbb{RP}^{N-1} \mid (x_1, \dots, x_N) \in \mathcal{K}\}.$$

Consequently, our minimization problem can be stated as follows:

Problem PNC1

$$\begin{array}{ll} \text{minimize} & \frac{X^\top L X}{X^\top D X} \\ \text{subject to} & X^\top D \mathbf{1} = 0, \quad X \in \mathcal{X}. \end{array}$$

It is understood that the solutions are points $\mathbb{P}(X)$ in \mathbb{RP}^{N-1} .

Since the Rayleigh ratio and the constraints $X^\top D \mathbf{1} = 0$ and $X \in \mathcal{X}$ are scale-invariant, we are led to the following formulation of our problem:

Problem PNC2

$$\begin{array}{llll} \text{minimize} & X^\top L X & & \\ \text{subject to} & X^\top D X = 1, & X^\top D \mathbf{1} = 0, & X \in \mathcal{X}. \end{array}$$

Problem PNC2 is equivalent to problem PNC1 in the sense that they have the same set of minimal solutions as points $\mathbb{P}(X) \in \mathbb{RP}^{N-1}$ given by their homogenous coordinates X .

More precisely, if X is any minimal solution of PNC1, then $X/(X^\top DX)^{1/2}$ is a minimal solution of PNC2 (with the same minimal value for the objective functions), and if X is a minimal solution of PNC2, then λX is a minimal solution for PNC1 for all $\lambda \neq 0$ (with the same minimal value for the objective functions).

Now, as in the classical papers, we consider the relaxation of the above problem obtained by dropping the condition that $X \in \mathcal{X}$, and proceed as usual. However, having found a solution Z to the relaxed problem, we need to find a discrete solution X such that $d(X, Z)$ is minimum in $\mathbb{R}\mathbb{P}^{N-1}$. All this presented in Section 4.2.

If the number of clusters K is at least 3, then we need to choose a matrix representation for partitions on the set of vertices. It is important that such a representation be scale-invariant, and it is also necessary to state necessary and sufficient conditions for such matrices to represent a partition (to the best of our knowledge, these points are not clearly articulated in the literature).

We describe a partition (A_1, \dots, A_K) of the set of nodes V by an $N \times K$ matrix $X = [X^1 \dots X^K]$ whose columns X^1, \dots, X^K are indicator vectors of the partition (A_1, \dots, A_K) . Inspired by what we did when $K = 2$, we assume that the vector X^j is of the form

$$X^j = (x_1^j, \dots, x_N^j),$$

where $x_i^j \in \{a_j, b_j\}$ for $j = 1, \dots, K$ and $i = 1, \dots, N$, and where a_j, b_j are any two distinct real numbers. The vector X^j is an indicator vector for A_j in the sense that, for $i = 1, \dots, N$,

$$x_i^j = \begin{cases} a_j & \text{if } v_i \in A_j \\ b_j & \text{if } v_i \notin A_j. \end{cases}$$

The choice $\{a_j, b_j\} = \{0, 1\}$ for $j = 1, \dots, K$ is natural, but premature. I show that if we pick $b_i = 0$, then we have

$$\frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)} = \frac{(X^j)^\top LX^j}{(X^j)^\top DX^j} \quad j = 1, \dots, K,$$

which implies that

$$\text{Ncut}(A_1, \dots, A_K) = \sum_{j=1}^K \frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)} = \sum_{j=1}^K \frac{(X^j)^\top LX^j}{(X^j)^\top DX^j}.$$

Then, I give necessary and sufficient conditions for a matrix X to represent a partition.

If we let

$$\mathcal{X} = \left\{ [X^1 \dots X^K] \mid X^j = a_j(x_1^j, \dots, x_N^j), x_i^j \in \{1, 0\}, a_j \in \mathbb{R}, X^j \neq 0 \right\}$$

(note that the condition $X^j \neq 0$ implies that $a_j \neq 0$), then the set of matrices representing partitions of V into K blocks is

$$\mathcal{K} = \left\{ X = [X^1 \ \dots \ X^K] \mid X \in \mathcal{X}, \right. \\ \left. (X^i)^\top D X^j = 0, \quad 1 \leq i, j \leq K, i \neq j, \right. \\ \left. X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1} \right\}.$$

As in the case $K = 2$, to be rigorous, the *solution are really K -tuples of points in \mathbb{RP}^{N-1}* , so our solution set is really

$$\mathbb{P}(\mathcal{K}) = \left\{ (\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)) \mid [X^1 \ \dots \ X^K] \in \mathcal{K} \right\}.$$

In view of the above, we have our first formulation of K -way clustering of a graph using normalized cuts, called problem PNC1 (the notation PNCX is used in Yu [15], Section 2.1):

**K -way Clustering of a graph using Normalized Cut, Version 1:
Problem PNC1**

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^K \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j} \\ & \text{subject to} && (X^i)^\top D X^j = 0, \quad 1 \leq i, j \leq K, i \neq j, \\ & && X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}, \quad X \in \mathcal{X}. \end{aligned}$$

As in the case $K = 2$, the solutions that we are seeking are K -tuples $(\mathbb{P}(X^1), \dots, \mathbb{P}(X^K))$ of points in \mathbb{RP}^{N-1} determined by their homogeneous coordinates X^1, \dots, X^K .

Then, step by step, we transform problem PNC1 into an equivalent problem PNC2, which we eventually relax by dropping the condition that $X \in \mathcal{X}$.

Our second revelation is that the relaxation $(*_1)$ of version 2 of our minimization problem (PNC2), which is equivalent to version 1, reveals that that the solutions of the relaxed problem $(*_1)$ are members of the *Grassmannian* $G(K, N)$.

This leads us to our third revelation: *we have two choices of metrics to compare solutions*: (1) a metric on $(\mathbb{RP}^{N-1})^K$; (2) a metric on $G(K, N)$. We discuss the first choice, which is the choice implicitly adopted by Shi and Yu.

Some of the most technical material on the Rayleigh ratio, which is needed for some proofs in Chapter 3, is the object of Appendix A. Appendix B may seem a bit out of place. Its purpose is to explain how to define a metric on the projective space \mathbb{RP}^n . For this, we need to review a few notions of differential geometry.

I hope that these notes will be make it easier for people to become familiar with the wonderful theory of normalized graph cuts. As far as I know, except for a short section in one of Gilbert Strang's book, and von Luxburg [14] excellent survey on spectral clustering, there is no comprehensive writing on the topic of normalized cuts.

Chapter 2

Graphs and Graph Laplacians; Basic Facts

2.1 Directed Graphs, Undirected Graphs, Incidence Matrices, Adjacency Matrices, Weighted Graphs

Definition 2.1. A *directed graph* is a pair $G = (V, E)$, where $V = \{v_1, \dots, v_m\}$ is a set of *nodes* or *vertices*, and $E \subseteq V \times V$ is a set of ordered pairs of distinct nodes (that is, pairs $(u, v) \in V \times V$ with $u \neq v$), called *edges*. Given any edge $e = (u, v)$, we let $s(e) = u$ be the *source* of e and $t(e) = v$ be the *target* of e .

Remark: Since an edge is a pair (u, v) with $u \neq v$, self-loops are not allowed. Also, there is at most one edge from a node u to a node v . Such graphs are sometimes called *simple graphs*.

An example of a directed graph is shown in Figure 2.1.

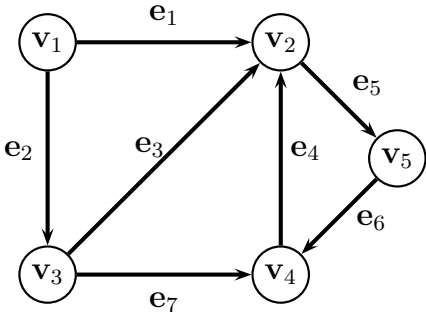


Figure 2.1: Graph G_1 .

For every node $v \in V$, the *degree* $d(v)$ of v is the number of edges leaving or entering v :

$$d(v) = |\{u \in V \mid (v, u) \in E \text{ or } (u, v) \in E\}|.$$

The *degree matrix* $D(G)$, is the diagonal matrix

$$D(G) = \text{diag}(d_1, \dots, d_m).$$

For example, for graph G_1 , we have

$$D(G_1) = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

Unless confusion arises, we write D instead of $D(G)$.

Definition 2.2. Given a directed graph $G = (V, E)$, with $V = \{v_1, \dots, v_m\}$, if $E = \{e_1, \dots, e_n\}$, then the *incidence matrix* $\tilde{D}(G)$ of G is the $m \times n$ matrix whose entries \tilde{d}_{ij} are given by

$$\tilde{d}_{ij} = \begin{cases} +1 & \text{if } e_j = (v_i, v_k) \text{ for some } k \\ -1 & \text{if } e_j = (v_k, v_i) \text{ for some } k \\ 0 & \text{otherwise.} \end{cases}$$

Here is the incidence matrix of the graph G_1 :

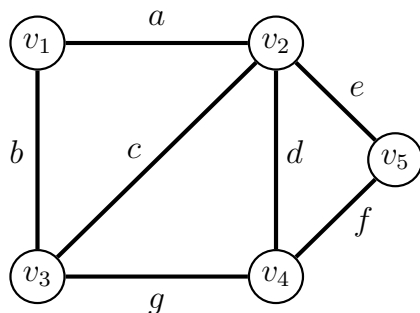
$$\tilde{D} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \end{pmatrix}.$$

Again, unless confusion arises, we write \tilde{D} instead of $\tilde{D}(G)$.

Undirected graphs are obtained from directed graphs by forgetting the orientation of the edges.

Definition 2.3. A *graph* (or *undirected graph*) is a pair $G = (V, E)$, where $V = \{v_1, \dots, v_m\}$ is a set of *nodes* or *vertices*, and E is a set of two-element subsets of V (that is, subsets $\{u, v\}$, with $u, v \in V$ and $u \neq v$), called *edges*.

Remark: Since an edge is a set $\{u, v\}$, we have $u \neq v$, so self-loops are not allowed. Also, for every set of nodes $\{u, v\}$, there is at most one edge between u and v . As in the case of directed graphs, such graphs are sometimes called *simple graphs*.

Figure 2.2: The undirected graph G_2 .

An example of a graph is shown in Figure 2.2.

For every node $v \in V$, the *degree* $d(v)$ of v is the number of edges adjacent to v :

$$d(v) = |\{u \in V \mid \{u, v\} \in E\}|.$$

The degree matrix D is defined as before. The notion of incidence matrix for an undirected graph is not as useful as the in the case of directed graphs

Definition 2.4. Given a graph $G = (V, E)$, with $V = \{v_1, \dots, v_m\}$, if $E = \{e_1, \dots, e_n\}$, then the *incidence matrix* $\tilde{D}(G)$ of G is the $m \times n$ matrix whose entries \tilde{d}_{ij} are given by

$$\tilde{d}_{ij} = \begin{cases} +1 & \text{if } e_j = \{v_i, v_k\} \text{ for some } k \\ 0 & \text{otherwise.} \end{cases}$$

Unlike the case of directed graphs, the entries in the incidence matrix of a graph (undirected) are nonnegative. We usually write \tilde{D} instead of $\tilde{D}(G)$.

The notion of adjacency matrix is basically the same for directed or undirected graphs.

Definition 2.5. Given a directed or undirected graph $G = (V, E)$, with $V = \{v_1, \dots, v_m\}$, the *adjacency matrix* $A(G)$ of G is the symmetric $m \times m$ matrix (a_{ij}) such that

(1) If G is directed, then

$$a_{ij} = \begin{cases} 1 & \text{if there is some edge } (v_i, v_j) \in E \text{ or some edge } (v_j, v_i) \in E \\ 0 & \text{otherwise.} \end{cases}$$

(2) Else if G is undirected, then

$$a_{ij} = \begin{cases} 1 & \text{if there is some edge } \{v_i, v_j\} \in E \\ 0 & \text{otherwise.} \end{cases}$$

As usual, unless confusion arises, we write A instead of $A(G)$. Here is the adjacency matrix of both graphs G_1 and G_2 :

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

In many applications, the notion of graph needs to be generalized to capture the intuitive idea that two nodes u and v are linked with a degree of certainty (or strength). Thus, we assign a nonnegative weights w_{ij} to an edge $\{v_i, v_j\}$; the smaller w_{ij} is, the weaker is the link (or similarity) between v_i and v_j , and the greater w_{ij} is, the stronger is the link (or similarity) between v_i and v_j .

Definition 2.6. A *weighted graph* is a pair $G = (V, W)$, where $V = \{v_1, \dots, v_m\}$ is a set of *nodes* or *vertices*, and W is a symmetric matrix called the *weight matrix*, such that $w_{ij} \geq 0$ for all $i, j \in \{1, \dots, m\}$, and $w_{ii} = 0$ for $i = 1, \dots, m$. We say that a set $\{v_i, v_j\}$ is an edge iff $w_{ij} > 0$. The corresponding (undirected) graph (V, E) with $E = \{\{e_i, e_j\} \mid w_{ij} > 0\}$, is called the *underlying graph* of G .

Remark: Since $w_{ii} = 0$, these graphs have no self-loops. We can think of the matrix W as a generalized adjacency matrix. The case where $w_{ij} \in \{0, 1\}$ is equivalent to the notion of a graph as in Definition 2.3.

We can think of the weight w_{ij} of an edge $\{v_i, v_j\}$ as a degree of similarity (or affinity) in an image, or a cost in a network. An example of a weighted graph is shown in Figure 2.3. The thickness of the edges corresponds to the magnitude of its weight.

For every node $v_i \in V$, the *degree* $d(v_i)$ of v_i is the sum of the weights of the edges adjacent to v_i :

$$d(v_i) = \sum_{j=1}^m w_{ij}.$$

Note that in the above sum, only nodes v_j such that there is an edge $\{v_i, v_j\}$ have a nonzero contribution. Such nodes are said to be *adjacent* to v_i . The degree matrix D is defined as before, namely by $D = \text{diag}(d(v_1), \dots, d(v_m))$.

Following common practice, we denote by $\mathbf{1}$ the (column) vector whose components are all equal to 1. Then, observe that $W\mathbf{1}$ is the (column) vector $(d(v_1), \dots, d(v_m))$ consisting of the degrees of the nodes of the graph.

Given any subset of nodes $A \subseteq V$, we define the *volume* $\text{vol}(A)$ of A as the sum of the weights of all edges adjacent to nodes in A :

$$\text{vol}(A) = \sum_{v_i \in A} d(v_i) = \sum_{v_i \in A} \sum_{j=1}^m w_{ij}.$$

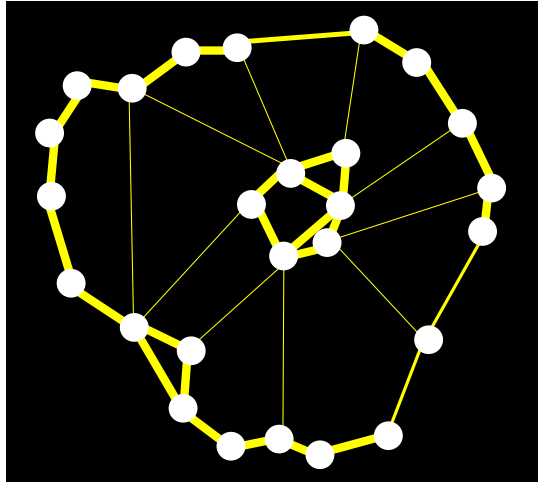


Figure 2.3: A weighted graph.

Remark: Yu and Shi [16] use the notation $\text{degree}(A)$ instead of $\text{vol}(A)$.

The notions of degree and volume are illustrated in Figure 2.4.

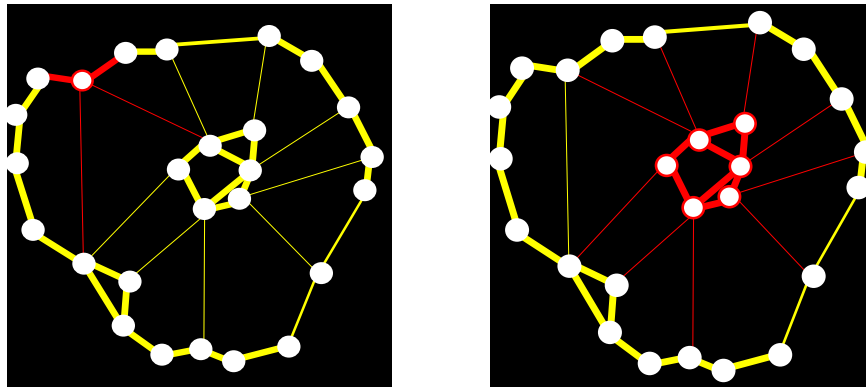


Figure 2.4: Degree and volume.

Observe that $\text{vol}(A) = 0$ if A consists of isolated vertices, that is, if $w_{ij} = 0$ for all $v_i \in A$. Thus, it is best to assume that G does not have isolated vertices.

Given any two subset $A, B \subseteq V$ (not necessarily distinct), we define $\text{links}(A, B)$ by

$$\text{links}(A, B) = \sum_{v_i \in A, v_j \in B} w_{ij}.$$

Since the matrix W is symmetric, we have

$$\text{links}(A, B) = \text{links}(B, A),$$

and observe that $\text{vol}(A) = \text{links}(A, V)$.

The quantity $\text{links}(A, \bar{A}) = \text{links}(\bar{A}, A)$, where $\bar{A} = V - A$ denotes the complement of A in V , measures how many links escape from A (and \bar{A}), and the quantity $\text{links}(A, A)$ measures how many links stay within A itself. The quantity

$$\text{cut}(A) = \text{links}(A, \bar{A})$$

is often called the *cut* of A , and the quantity

$$\text{assoc}(A) = \text{links}(A, A)$$

is often called the *association* of A . Clearly,

$$\text{cut}(A) + \text{assoc}(A) = \text{vol}(A).$$

The notions of cut is illustrated in Figure 2.5.

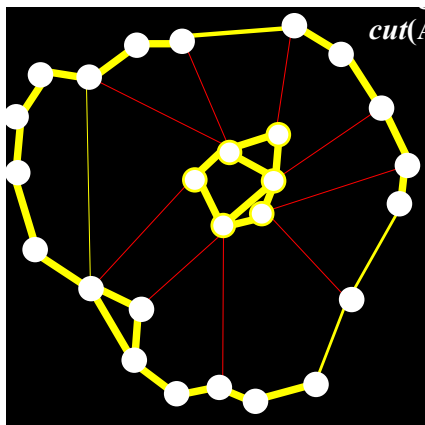


Figure 2.5: A Cut involving the set of nodes in the center and the nodes on the perimeter.

We now define the most important concept of these notes: The Laplacian matrix of a graph. Actually, as we will see, it comes in several flavors.

2.2 Laplacian Matrices of Graphs

Let us begin with directed graphs, although as we will see, graph Laplacians are fundamentally associated with undirected graph. The key proposition whose proof can be found in Gallier [5] and Godsil and Royle [8] is this:

Proposition 2.1. *Given any directed graph G if \tilde{D} is the incidence matrix of G , A is the adjacency matrix of G , and D is the degree matrix such that $D_{ii} = d(v_i)$, then*

$$\tilde{D}\tilde{D}^\top = D - A.$$

Consequently, $\tilde{D}\tilde{D}^\top$ is independent of the orientation of G and $D - A$ is symmetric, positive, semidefinite; that is, the eigenvalues of $D - A$ are real and nonnegative.

The matrix $L = \tilde{D}\tilde{D}^\top = D - A$ is called the (*unnormalized*) graph Laplacian of the graph G . For example, the graph Laplacian of graph G_1 is

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

The (*unnormalized*) graph Laplacian of an undirected graph $G = (V, E)$ is defined by

$$L = D - A.$$

Since L is equal to $\tilde{D}\tilde{D}^\top$ for any orientation of G , it is also positive semidefinite. Observe that each row of L sums to zero. Consequently, the vector $\mathbf{1}$ is in the nullspace of L .

Remark: With the unoriented version of the incidence matrix (see Definition 2.4), it can be shown that

$$\tilde{D}\tilde{D}^\top = D + A.$$

The natural generalization of the notion of graph Laplacian to weighted graphs is this:

Definition 2.7. Given any weighted directed graph $G = (V, W)$ with $V = \{v_1, \dots, v_m\}$, the (*unnormalized*) graph Laplacian $L(G)$ of G is defined by

$$L(G) = D(G) - W,$$

where $D(G) = \text{diag}(d_1, \dots, d_m)$ is the degree matrix of G (a diagonal matrix), with

$$d_i = \sum_{j=1}^m w_{ij}.$$

As usual, unless confusion arises, we write L instead of $L(G)$.

It is clear that each row of L sums to 0, so the vector $\mathbf{1}$ is the nullspace of L , but it is less obvious that L is positive semidefinite. An easy way to prove this is to evaluate the quadratic form $x^\top Lx$.

Proposition 2.2. For any $m \times m$ symmetric matrix W , if we let $L = D - W$ where D is the degree matrix of $W = (w_{ij})$, then we have

$$x^\top Lx = \frac{1}{2} \sum_{i,j=1}^m w_{ij} (x_i - x_j)^2 \quad \text{for all } x \in \mathbb{R}^m.$$

Consequently, $x^\top Lx$ does not depend on the diagonal entries in W , and if $w_{ij} \geq 0$ for all $i, j \in \{1, \dots, m\}$, then L is positive semidefinite.

Proof. We have

$$\begin{aligned}
x^\top Lx &= x^\top Dx - x^\top Wx \\
&= \sum_{i=1}^m d_i x_i^2 - \sum_{i,j=1}^m w_{ij} x_i x_j \\
&= \frac{1}{2} \left(\sum_{i=1}^m d_i x_i^2 - 2 \sum_{i,j=1}^m w_{ij} x_i x_j + \sum_{i=1}^m d_i x_i^2 \right) \\
&= \frac{1}{2} \sum_{i,j=1}^m w_{ij} (x_i - x_j)^2.
\end{aligned}$$

Obviously, the quantity on the right-hand side does not depend on the diagonal entries in W , and if $w_{ij} \geq 0$ for all i, j , then this quantity is nonnegative. \square

Proposition 2.2 immediately implies the following facts: For any weighted graph $G = (V, W)$,

1. The eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ of L are real and nonnegative, and there is an orthonormal basis of eigenvectors of L .
2. The smallest eigenvalue λ_1 of L is equal to 0, and $\mathbf{1}$ is a corresponding eigenvector.

It turns out that the dimension of the nullspace of L (the eigenspace of 0) is equal to the number of connected components of the underlying graph of G .

Proposition 2.3. *Let $G = (V, W)$ be a weighted graph. The number c of connected components K_1, \dots, K_c of the underlying graph of G is equal to the dimension of the nullspace of L , which is equal to the multiplicity of the eigenvalue 0. Furthermore, the nullspace of L has a basis consisting of indicator vectors of the connected components of G , that is, vectors (f_1, \dots, f_m) such that $f_j = 1$ iff $v_j \in K_i$ and $f_j = 0$ otherwise.*

Proof. A complete proof can be found in von Luxburg [14], and we only give a sketch of the proof.

First, assume that G is connected, so $c = 1$. A nonzero vector x is in the kernel of L iff $Lx = 0$, which implies that

$$x^\top Lx = \frac{1}{2} \sum_{i,j=1}^m w_{ij} (x_i - x_j)^2 = 0.$$

This implies that $x_i = x_j$ whenever $w_{ij} > 0$, and thus, $x_i = x_j$ whenever nodes v_i and v_j are linked by an edge. By induction, $x_i = x_j$ whenever there is a path from v_i to v_j . Since G is assumed to be connected, any two nodes are linked by a path, which implies that $x_i = x_j$

for all $i \neq j$. Therefore, the nullspace of L is spanned by $\mathbf{1}$, which is indeed the indicator vector of $K_1 = V$, and this nullspace has dimension 1.

Let us now assume that G has $c \geq 2$ connected components. If so, by renumbering the rows and columns of W , we may assume that W is a block matrix consisting of c blocks, and similarly L is a block matrix of the form

$$L = \begin{pmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_c \end{pmatrix},$$

where L_i is the graph Laplacian associated with the connected component K_i . By the induction hypothesis, 0 is an eigenvalue of multiplicity 1 for each L_i , and so the nullspace of L has dimension c . The rest is left as an exercise (or see von Luxburg [14]). \square

Proposition 2.3 implies that if the underlying graph of G is connected, then the second eigenvalue, λ_2 , of L is strictly positive.

Remarkably, the eigenvalue λ_2 contains a lot of information about the graph G (assuming that $G = (V, E)$ is an undirected graph). This was first discovered by Fiedler in 1973, and for this reason, λ_2 is often referred to as the *Fiedler number*. For more on the properties of the Fiedler number, see Godsil and Royle [8] (Chapter 13) and Chung [3]. More generally, the spectrum $(0, \lambda_2, \dots, \lambda_m)$ of L contains a lot of information about the combinatorial structure of the graph G . Leverage of this information is the object of *spectral graph theory*.

It turns out that normalized variants of the graph Laplacian are needed, especially in applications to graph clustering. These variants make sense only if G has no isolated vertices, which means that every row of W contains some strictly positive entry. In this case, the degree matrix D contains positive entries, so it is invertible and $D^{-1/2}$ makes sense; namely

$$D^{-1/2} = \text{diag}(d_1^{-1/2}, \dots, d_m^{-1/2}),$$

and similarly for any real exponent α .

Definition 2.8. Given any weighted directed graph $G = (V, W)$ with no isolated vertex and with $V = \{v_1, \dots, v_m\}$, the (*normalized*) graph Laplacians L_{sym} and L_{rw} of G are defined by

$$\begin{aligned} L_{\text{sym}} &= D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2} \\ L_{\text{rw}} &= D^{-1} L = I - D^{-1} W. \end{aligned}$$

Observe that the Laplacian $L_{\text{sym}} = D^{-1/2} L D^{-1/2}$ is a symmetric matrix (because L and $D^{-1/2}$ are symmetric) and that

$$L_{\text{rw}} = D^{-1/2} L_{\text{sym}} D^{1/2}.$$

The reason for the notation L_{rw} is that this matrix is closely related to a random walk on the graph G . There are simple relationships between the eigenvalues and the eigenvectors of L_{sym} , and L_{rw} . There is also a simple relationship with the generalized eigenvalue problem $Lx = \lambda Dx$.

Proposition 2.4. *Let $G = (V, W)$ be a weighted graph without isolated vertices. The graph Laplacians, L, L_{sym} , and L_{rw} satisfy the following properties:*

(1) *The matrix L_{sym} is symmetric, positive, semidefinite. In fact,*

$$x^\top L_{\text{sym}} x = \frac{1}{2} \sum_{i,j=1}^m w_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 \quad \text{for all } x \in \mathbb{R}^m.$$

(2) *The normalized graph Laplacians L_{sym} and L_{rw} have the same spectrum ($0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_m$), and a vector $u \neq 0$ is an eigenvector of L_{rw} for λ iff $D^{1/2}u$ is an eigenvector of L_{sym} for λ .*

(3) *The graph Laplacians, L, L_{sym} , and L_{rw} are symmetric, positive, semidefinite.*

(4) *A vector $u \neq 0$ is a solution of the generalized eigenvalue problem $Lu = \lambda Du$ iff $D^{1/2}u$ is an eigenvector of L_{sym} for the eigenvalue λ iff u is an eigenvector of L_{rw} for the eigenvalue λ .*

(5) *The graph Laplacians, L and L_{rw} have the same nullspace.*

(6) *The vector $\mathbf{1}$ is in the nullspace of L_{rw} , and $D^{1/2}\mathbf{1}$ is in the nullspace of L_{sym} .*

Proof. (1) We have $L_{\text{sym}} = D^{-1/2}LD^{-1/2}$, and $D^{-1/2}$ is a symmetric invertible matrix (since it is an invertible diagonal matrix). It is a well-known fact of linear algebra that if B is an invertible matrix, then a matrix S is symmetric, positive semidefinite iff BSB^\top is symmetric, positive semidefinite. Since L is symmetric, positive semidefinite, so is $L_{\text{sym}} = D^{-1/2}LD^{-1/2}$. The formula

$$x^\top L_{\text{sym}} x = \frac{1}{2} \sum_{i,j=1}^m w_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 \quad \text{for all } x \in \mathbb{R}^m$$

follows immediately from Proposition 2.2 by replacing x by $D^{-1/2}x$, and also shows that L_{sym} is positive semidefinite.

(2) Since

$$L_{\text{rw}} = D^{-1/2}L_{\text{sym}}D^{1/2},$$

the matrices L_{sym} and L_{rw} are similar, which implies that they have the same spectrum. In fact, since $D^{1/2}$ is invertible,

$$L_{\text{rw}}u = D^{-1}Lu = \lambda u$$

iff

$$D^{-1/2}Lu = \lambda D^{1/2}u$$

iff

$$D^{-1/2}LD^{-1/2}D^{1/2}u = L_{\text{sym}}D^{1/2}u = \lambda D^{1/2}u,$$

which shows that a vector $u \neq 0$ is an eigenvector of L_{rw} for λ iff $D^{1/2}u$ is an eigenvector of L_{sym} for λ .

(3) We already know that L and L_{sym} are positive semidefinite, and (2) shows that L_{rw} is also positive semidefinite.

(4) Since $D^{-1/2}$ is invertible, we have

$$Lu = \lambda Du$$

iff

$$D^{-1/2}Lu = \lambda D^{1/2}u$$

iff

$$D^{-1/2}LD^{-1/2}D^{1/2}u = L_{\text{sym}}D^{1/2}u = \lambda D^{1/2}u,$$

which shows that a vector $u \neq 0$ is a solution of the generalized eigenvalue problem $Lu = \lambda Du$ iff $D^{1/2}u$ is an eigenvector of L_{sym} for the eigenvalue λ . The second part of the statement follows from (2).

(5) Since D^{-1} is invertible, we have $Lu = 0$ iff $D^{-1}u = L_{\text{rw}}u = 0$.

(6) Since $L\mathbf{1} = 0$, we get $L_{\text{rw}}u = D^{-1}L\mathbf{1} = 0$. That $D^{1/2}\mathbf{1}$ is in the nullspace of L_{sym} follows from (2). \square

A version of Proposition 2.5 also holds for the graph Laplacians L_{sym} and L_{rw} . The proof is left as an exercise.

Proposition 2.5. *Let $G = (V, W)$ be a weighted graph. The number c of connected components K_1, \dots, K_c of the underlying graph of G is equal to the dimension of the nullspace of both L_{sym} and L_{rw} , which is equal to the multiplicity of the eigenvalue 0. Furthermore, the nullspace of L_{rw} has a basis consisting of indicator vectors of the connected components of G , that is, vectors (f_1, \dots, f_m) such that $f_j = 1$ iff $v_j \in K_i$ and $f_j = 0$ otherwise. For L_{sym} , a basis of the nullspace is obtained by multiplying the above basis of the nullspace of L_{rw} by $D^{1/2}$.*

Chapter 3

Spectral Graph Drawing

3.1 Graph Drawing and Energy Minimization

Let $G = (V, E)$ be some undirected graph. It is often desirable to draw a graph, usually in the plane but possibly in 3D, and it turns out that the graph Laplacian can be used to design surprisingly good methods. Say $|V| = m$. The idea is to assign a point $\rho(v_i)$ in \mathbb{R}^n to the vertex $v_i \in V$, for every $v_i \in V$, and to draw a line segment between the points $\rho(v_i)$ and $\rho(v_j)$. Thus, a *graph drawing* is a function $\rho: V \rightarrow \mathbb{R}^n$.

We define the *matrix of a graph drawing* ρ (in \mathbb{R}^n) as a $m \times n$ matrix R whose i th row consists of the row vector $\rho(v_i)$ corresponding to the point representing v_i in \mathbb{R}^n . Typically, we want $n < m$; in fact n should be much smaller than m . A representation is *balanced* iff the sum of the entries of every column is zero, that is,

$$\mathbf{1}^\top R = 0.$$

If a representation is not balanced, it can be made balanced by a suitable translation. We may also assume that the columns of R are linearly independent, since any basis of the column space also determines the drawing. Thus, from now on, we may assume that $n \leq m$.

Remark: A graph drawing $\rho: V \rightarrow \mathbb{R}^n$ is not required to be injective, which may result in degenerate drawings where distinct vertices are drawn as the same point. For this reason, we prefer not to use the terminology *graph embedding*, which is often used in the literature. This is because in differential geometry, an embedding always refers to an injective map. The term *graph immersion* would be more appropriate.

As explained in Godsil and Royle [8], we can imagine building a physical model of G by connecting adjacent vertices (in \mathbb{R}^n) by identical springs. Then, it is natural to consider a representation to be better if it requires the springs to be less extended. We can formalize this by defining the *energy* of a drawing R by

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} \|\rho(v_i) - \rho(v_j)\|^2,$$

where $\rho(v_i)$ is the i th row of R and $\|\rho(v_i) - \rho(v_j)\|^2$ is the square of the Euclidean length of the line segment joining $\rho(v_i)$ and $\rho(v_j)$.

Then, “good drawings” are drawings that minimize the energy function \mathcal{E} . Of course, the trivial representation corresponding to the zero matrix is optimum, so we need to impose extra constraints to rule out the trivial solution.

We can consider the more general situation where the springs are not necessarily identical. This can be modeled by a symmetric weight (or stiffness) matrix $W = (w_{ij})$, with $w_{ij} \geq 0$. Then our energy function becomes

$$\mathcal{E}(R) = \sum_{\{v_i, v_j\} \in E} w_{ij} \|\rho(v_i) - \rho(v_j)\|^2.$$

It turns out that this function can be expressed in terms of the matrix R and a diagonal matrix \widehat{W} obtained from W . Let $p = |E|$ be the number of edges in E and pick any enumeration of these edges, so that every edge $\{v_i, v_j\}$ is uniquely represented by some index e . Then, let \widehat{W} be the $p \times p$ diagonal matrix such that

$$\widehat{w}_{ee} = w_{ij}, \quad \text{where } e \text{ correspond to the edge } \{v_i, v_j\}.$$

We have the following proposition from Godsil and Royle [8].

Proposition 3.1. *Let $G = (V, E)$ be an undirected graph, with $|V| = m$, $|E| = p$, let W be a $m \times m$ weight matrix, and let R be the matrix of a graph drawing ρ of G in \mathbb{R}^n (a $m \times n$ matrix). If \widetilde{D} is the incidence matrix associated with any orientation of the graph G , and \widehat{W} is the $p \times p$ diagonal matrix associated with W , then*

$$\mathcal{E}(R) = \text{tr}(R^\top \widetilde{D} \widehat{W} \widetilde{D}^\top R).$$

Proof. Observe that the rows of $\widetilde{D}^\top R$ are indexed by the edges of G , and if $\{v_i, v_j\} \in E$, then the e th row of $\widetilde{D}^\top R$ is

$$\pm(\rho(v_i) - \rho(v_j)),$$

where e is the index corresponding to the edge $\{v_i, v_j\}$. As a consequence, the diagonal entries of $\widetilde{D}^\top R R^\top \widetilde{D}$ have the form $\|\rho(v_i) - \rho(v_j)\|^2$, where $\{v_i, v_j\}$ ranges over the edges in E . Hence,

$$\mathcal{E}(R) = \text{tr}(\widehat{W} \widetilde{D}^\top R R^\top \widetilde{D}) = \text{tr}(R^\top \widetilde{D} \widehat{W} \widetilde{D}^\top R),$$

since $\text{tr}(AB) = \text{tr}(BA)$ for any two matrices A and B . □

The matrix

$$L = \widetilde{D} \widehat{W} \widetilde{D}^\top$$

may be viewed as a weighted Laplacian of G . Observe that L is a $m \times m$ matrix, and that

$$L_{ij} = \begin{cases} -w_{ij} & \text{if } i \neq j \\ \sum_{\{v_i, v_k\} \in E} w_{ik} & \text{if } i = j. \end{cases}$$

Therefore,

$$L = D - W,$$

the familiar unnormalized Laplacian matrix associated with W , where D is the degree matrix associated with W , and so

$$\mathcal{E}(R) = \text{tr}(R^\top LR).$$

Note that

$$L\mathbf{1} = 0,$$

as we already observed.

Since the matrix $R^\top \widetilde{D}\widetilde{W}\widetilde{D}^\top R = R^\top LR$ is symmetric, it has real eigenvalues. Actually, since $L = \widetilde{D}\widetilde{W}\widetilde{D}^\top$ is positive semidefinite, so is $R^\top LR$. Then, the trace of $R^\top LR$ is equal to the sum of its positive eigenvalues, and this is the energy $\mathcal{E}(R)$ of the graph drawing.

If R is the matrix of a graph drawing in \mathbb{R}^n , then for any invertible matrix M , the map that assigns v_i to $\rho(v_i)M$ is another graph drawing of G , and these two drawings convey the same amount of information. From this point of view, a graph drawing is determined by the column space of R . Therefore, it is reasonable to assume that the columns of R are pairwise orthogonal and that they have unit length. Such a matrix satisfies the equation $R^\top R = I$, and the corresponding drawing is called an *orthogonal drawing*. This condition also rules out trivial drawings. The following result tells us how to find minimum energy graph drawings, provided the graph is connected.

Theorem 3.2. *Let $G = (V, W)$ be a weighted graph with $|V| = m$. If $L = D - W$ is the (unnormalized) Laplacian of G , and if the eigenvalues of L are $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_m$, then the minimal energy of any balanced orthogonal graph drawing of G in \mathbb{R}^n is equal to $\lambda_2 + \dots + \lambda_{n+1}$ (in particular, this implies that $n < m$). The $m \times n$ matrix R consisting of any unit eigenvectors u_2, \dots, u_{n+1} associated with $\lambda_2 \leq \dots \leq \lambda_{n+1}$ yields an orthogonal graph drawing of minimal energy; it satisfies the condition $R^\top R = I$.*

Proof. We present the proof given in Godsil and Royle [8] (Section 13.4, Theorem 13.4.1). The key point is that the sum of the n smallest eigenvalues of L is a lower bound for $\text{tr}(R^\top LR)$. This can be shown using an argument using the Rayleigh ratio; see Proposition A.3. Then, any n eigenvectors (u_1, \dots, u_n) associated with $\lambda_1, \dots, \lambda_n$ achieve this bound. Because the first eigenvalue of L is $\lambda_1 = 0$ and because we are assuming that $\lambda_2 > 0$, we have $u_1 = \mathbf{1}/\sqrt{m}$, and by deleting u_1 we obtain a balanced orthogonal graph drawing in \mathbb{R}^{n-1} with the same energy. The converse is true, so the minimum energy of an orthogonal graph drawing in \mathbb{R}^n is equal to the minimum energy of an orthogonal graph drawing in \mathbb{R}^{n+1} , and this minimum is $\lambda_2 + \dots + \lambda_{n+1}$. The rest is clear. \square

Observe that for any orthogonal $n \times n$ matrix Q , since

$$\text{tr}(R^\top LR) = \text{tr}(Q^\top R^\top LRQ),$$

the matrix RQ also yields a minimum orthogonal graph drawing.

In summary, if $\lambda_2 > 0$, an automatic method for drawing a graph in \mathbb{R}^2 is this:

1. Compute the two smallest nonzero eigenvalues $\lambda_2 \leq \lambda_3$ of the graph Laplacian L (it is possible that $\lambda_3 = \lambda_2$ if λ_2 is a multiple eigenvalue);
2. Compute two unit eigenvectors u_2, u_3 associated with λ_2 and λ_3 , and let $R = [u_2 \ u_3]$ be the $m \times 2$ matrix having u_2 and u_3 as columns.
3. Place vertex v_i at the point whose coordinates is the i th row of R , that is, (R_{i1}, R_{i2}) .

This method generally gives pleasing results, but beware that there is no guarantee that distinct nodes are assigned distinct images, because R can have identical rows. This does not seem to happen often in practice.

3.2 Examples of Graph Drawings

We now give a number of examples using `Matlab`. Some of these are borrowed or adapted from Spielman [13].

Example 1. Consider the graph with four nodes whose adjacency matrix is

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

We use the following program to compute u_2 and u_3 :

```
A = [0 1 1 0; 1 0 0 1; 1 0 0 1; 0 1 1 0];
D = diag(sum(A));
L = D - A;
[v, e] = eigs(L);
gplot(A, v(:, [3 2]))
hold on;
gplot(A, v(:, [3 2]), 'o')
```

The graph of Example 1 is shown in Figure 3.1. The function `eigs(L)` computes the six largest eigenvalues of L in decreasing order, and corresponding eigenvectors. It turns out that $\lambda_2 = \lambda_3 = 2$ is a double eigenvalue.

Example 2. Consider the graph G_2 shown in Figure 2.2 given by the adjacency matrix

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

We use the following program to compute u_2 and u_3 :

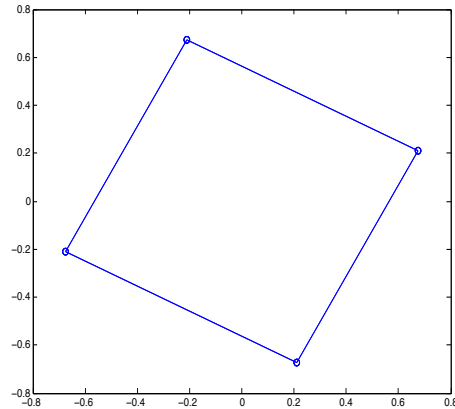


Figure 3.1: Drawing of the graph from Example 1.

```

A = [0 1 1 0 0; 1 0 1 1 1; 1 1 0 1 0; 0 1 1 0 1; 0 1 0 1 0];
D = diag(sum(A));
L = D - A;
[v, e] = eig(L);
gplot(A, v(:, [2 3]))
hold on
gplot(A, v(:, [2 3]), 'o')

```

The function `eig(L)` (with no `s` at the end) computes the eigenvalues of L in increasing order. The result of drawing the graph is shown in Figure 3.2. Note that node v_2 is assigned to the point $(0, 0)$, so the difference between this drawing and the drawing in Figure 2.2 is that the drawing of Figure 3.2 is not convex.

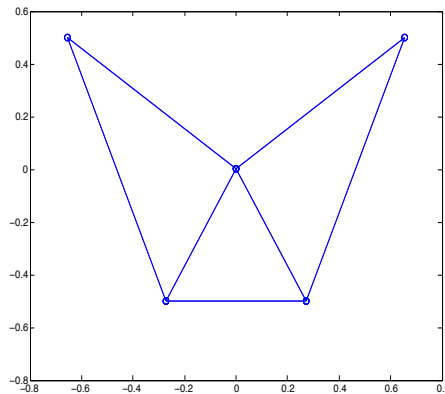


Figure 3.2: Drawing of the graph from Example 2.

Example 3. Consider the ring graph defined by the adjacency matrix A given in the Matlab program shown below:

```
A = diag(ones(1, 11),1);
A = A + A';
A(1, 12) = 1; A(12, 1) = 1;
D = diag(sum(A));
L = D - A;
[v, e] = eig(L);
gplot(A, v(:, [2 3]))
hold on
gplot(A, v(:, [2 3]),'o')
```

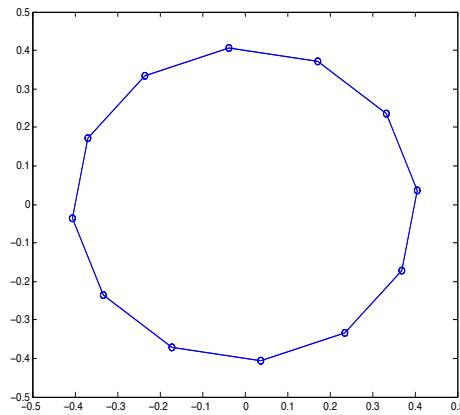


Figure 3.3: Drawing of the graph from Example 3.

Observe that we get a very nice ring; see Figure 3.3. Again $\lambda_2 = 0.2679$ is a double eigenvalue (and so are the next pairs of eigenvalues, except the last, $\lambda_{12} = 4$).

Example 4. In this example adapted from Spielman, we generate 20 randomly chosen points in the unit square, compute their Delaunay triangulation, then the adjacency matrix of the corresponding graph, and finally draw the graph using the second and third eigenvalues of the Laplacian.

```
A = zeros(20,20);
xy = rand(20, 2);
trigs = delaunay(xy(:,1), xy(:,2));
elemtrig = ones(3) - eye(3);
for i = 1:length(trigs),
    A(trigs(i,:),trigs(i,:)) = elemtrig;
end
```

```

A = double(A >0);
gplot(A,xy)
D = diag(sum(A));
L = D - A;
[v, e] = eigs(L, 3, 'sm');
figure(2)
gplot(A, v(:, [2 1]))
hold on
gplot(A, v(:, [2 1]), 'o')

```

The Delaunay triangulation of the set of 20 points and the drawing of the corresponding graph are shown in Figure 3.4. The graph drawing on the right looks nicer than the graph on the left but is no longer planar.

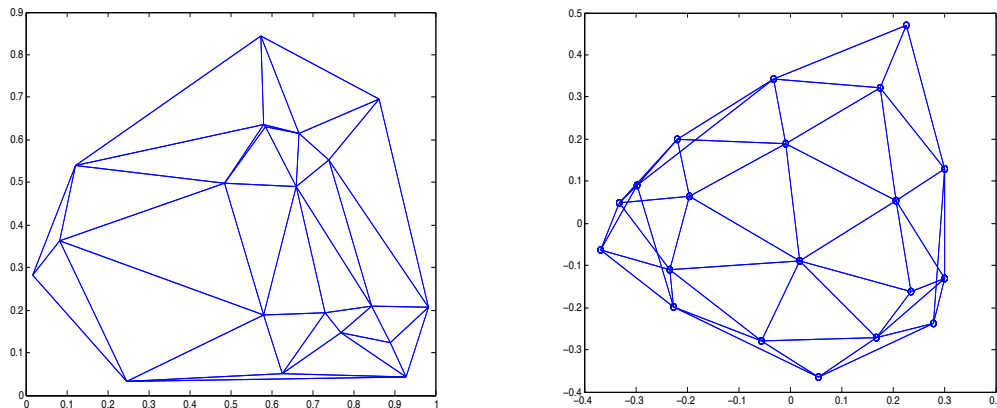


Figure 3.4: Delaunay triangulation (left) and drawing of the graph from Example 4 (right).

Example 5. Our last example, also borrowed from Spielman [13], corresponds to the skeleton of the “Buckyball,” a geodesic dome invented by the architect Richard Buckminster Fuller (1895–1983). The Montréal Biosphère is an example of a geodesic dome designed by Buckminster Fuller.

```

A = full(bucky);
D = diag(sum(A));
L = D - A;
[v, e] = eig(L);
gplot(A, v(:, [2 3]))
hold on;
gplot(A,v(:, [2 3]), 'o')

```

Figure 3.5 shows a graph drawing of the Buckyball. This picture seems a bit squashed for two reasons. First, it is really a 3-dimensional graph; second, $\lambda_2 = 0.2434$ is a triple

eigenvalue. (Actually, the Laplacian of L has many multiple eigenvalues.) What we should really do is to plot this graph in \mathbb{R}^3 using three orthonormal eigenvectors associated with λ_2 .

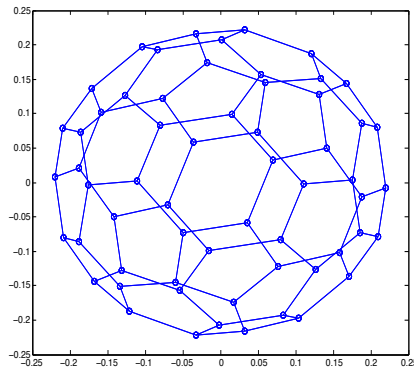


Figure 3.5: Drawing of the graph of the Buckyball.

A 3D picture of the graph of the Buckyball is produced by the following `Matlab` program, and its image is shown in Figure 3.6. It looks better!

```
[x, y] = gplot(A, v(:, [2 3]));
[x, z] = gplot(A, v(:, [2 4]));
plot3(x,y,z)
```

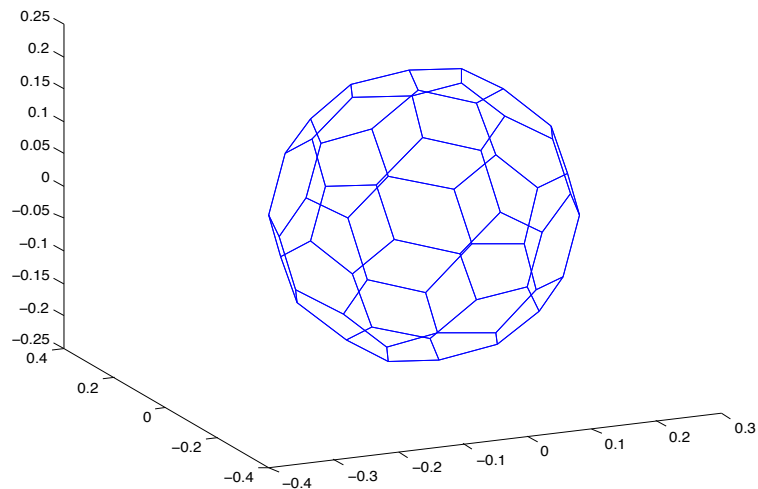


Figure 3.6: Drawing of the graph of the Buckyball in \mathbb{R}^3 .

Chapter 4

Graph Clustering

4.1 Graph Clustering Using Normalized Cuts

Given a set of data, the goal of clustering is to partition the data into different groups according to their similarities. When the data is given in terms of a similarity graph G , where the weight w_{ij} between two nodes v_i and v_j is a measure of similarity of v_i and v_j , the problem can be stated as follows: Find a partition (A_1, \dots, A_K) of the set of nodes V into different groups such that the edges between different groups have very low weight (which indicates that the points in different clusters are dissimilar), and the edges within a group have high weight (which indicates that points within the same cluster are similar).

The above graph clustering problem can be formalized as an optimization problem, using the notion of cut mentioned at the end of Section 2.1.

Given a subset A of the set of vertices V , recall that we define $\text{cut}(A)$ by

$$\text{cut}(A) = \text{links}(A, \bar{A}) = \sum_{v_i \in A, v_j \in \bar{A}} w_{ij},$$

and that

$$\text{cut}(A) = \text{links}(A, \bar{A}) = \text{links}(\bar{A}, A) = \text{cut}(\bar{A}).$$

If we want to partition V into K clusters, we can do so by finding a partition (A_1, \dots, A_K) that minimizes the quantity

$$\text{cut}(A_1, \dots, A_K) = \frac{1}{2} \sum_{i=1}^K \text{cut}(A_i).$$

The reason for introducing the factor $1/2$ is to avoid counting each edge twice. In particular,

$$\text{cut}(A, \bar{A}) = \text{links}(A, \bar{A}).$$

For $K = 2$, the mincut problem is a classical problem that can be solved efficiently, but in practice, it does not yield satisfactory partitions. Indeed, in many cases, the mincut solution

separates one vertex from the rest of the graph. What we need is to design our cost function in such a way that it keeps the subsets A_i “reasonably large” (reasonably balanced).

An example of a weighted graph and a partition of its nodes into two clusters is shown in Figure 4.1.

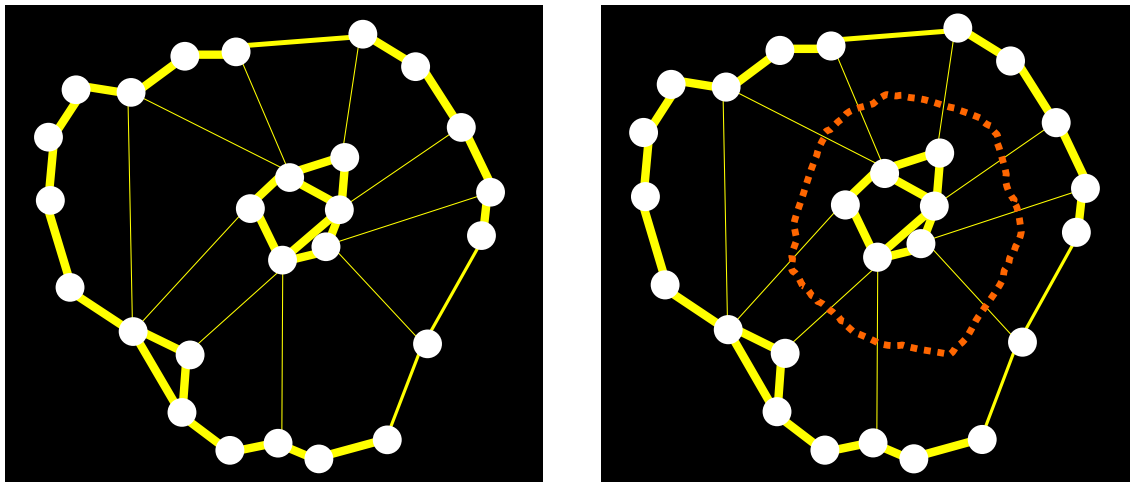


Figure 4.1: A weighted graph and its partition into two clusters.

A way to get around this problem is to normalize the cuts by dividing by some measure of each subset A_i . One possibility is to use the size (the number of elements) of A_i . Another is to use the volume $\text{vol}(A_i)$ of A_i . A solution using the second measure (the volume) (for $K = 2$) was proposed and investigated in a seminal paper of Shi and Malik [12]. Subsequently, Yu (in her dissertation [15]) and Yu and Shi [16] extended the method to $K > 2$ clusters. We will describe this method later. The idea is to minimize the cost function

$$\text{Ncut}(A_1, \dots, A_K) = \sum_{i=1}^K \frac{\text{links}(A_i, \bar{A}_i)}{\text{vol}(A_i)} = \sum_{i=1}^K \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)}.$$

We begin with the case $K = 2$, which is easier to handle.

4.2 Special Case: 2-Way Clustering Using Normalized Cuts

Our goal is to express our optimization problem in matrix form. In the case of two clusters, a single vector X can be used to describe the partition $(A_1, A_2) = (A, \bar{A})$. We need to choose the structure of this vector in such a way that $\text{Ncut}(A, \bar{A})$ is equal to the Rayleigh ratio

$$\frac{X^T L X}{X^T D X}.$$

It is also important to pick a vector representation which is invariant under multiplication by a nonzero scalar, because the Rayleigh ratio is scale-invariant, and it is crucial to take advantage of this fact to make the denominator go away.

Let $N = |V|$ be the number of nodes in the graph G . In view of the desire for a scale-invariant representation, it is natural to assume that the vector X is of the form

$$X = (x_1, \dots, x_N),$$

where $x_i \in \{a, b\}$ for $i = 1, \dots, N$, for any two distinct real numbers a, b . This is an indicator vector in the sense that, for $i = 1, \dots, N$,

$$x_i = \begin{cases} a & \text{if } v_i \in A \\ b & \text{if } v_i \notin A. \end{cases}$$

The correct interpretation is really to view X as a representative of a point in the real projective space \mathbb{RP}^{N-1} , namely the point $\mathbb{P}(X)$ of homogeneous coordinates $(x_1 : \dots : x_N)$. Therefore, from now on, we view X as a vector of homogeneous coordinates representing the point $\mathbb{P}(X) \in \mathbb{RP}^{N-1}$.

Let $d = \mathbf{1}^\top D \mathbf{1}$ and $\alpha = \text{vol}(A)$. Then, $\text{vol}(\bar{A}) = d - \alpha$. By Proposition 2.2, we have

$$X^\top L X = (a - b)^2 \text{cut}(A, \bar{A}),$$

and we easily check that

$$X^\top D X = \alpha a^2 + (d - \alpha) b^2.$$

Since $\text{cut}(A, \bar{A}) = \text{cut}(\bar{A}, A)$, we have

$$\text{Ncut}(A, \bar{A}) = \frac{\text{cut}(A, \bar{A})}{\text{vol}(A)} + \frac{\text{cut}(\bar{A}, A)}{\text{vol}(\bar{A})} = \left(\frac{1}{\text{vol}(A)} + \frac{1}{\text{vol}(\bar{A})} \right) \text{cut}(A, \bar{A}),$$

so we obtain

$$\text{Ncut}(A, \bar{A}) = \left(\frac{1}{\alpha} + \frac{1}{d - \alpha} \right) \text{cut}(A, \bar{A}) = \frac{d}{\alpha(d - \alpha)} \text{cut}(A, \bar{A}).$$

Since

$$\frac{X^\top L X}{X^\top D X} = \frac{(a - b)^2}{\alpha a^2 + (d - \alpha) b^2} \text{cut}(A, \bar{A}),$$

in order to have

$$\text{Ncut}(A, \bar{A}) = \frac{X^\top L X}{X^\top D X},$$

we need to find a and b so that

$$\frac{(a - b)^2}{\alpha a^2 + (d - \alpha) b^2} = \frac{d}{\alpha(d - \alpha)}.$$

The above is equivalent to

$$(a - b)^2\alpha(d - \alpha) = \alpha da^2 + (d - \alpha)db^2,$$

which can be rewritten as

$$a^2(\alpha d - \alpha(d - \alpha)) + b^2(d^2 - \alpha d - \alpha(d - \alpha)) + 2\alpha(d - \alpha)ab = 0.$$

The above yields

$$a^2\alpha^2 + b^2(d^2 - 2\alpha d + \alpha^2) + 2\alpha(d - \alpha)ab = 0,$$

that is,

$$a^2\alpha^2 + b^2(d - \alpha)^2 + 2\alpha(d - \alpha)ab = 0,$$

which reduces to

$$(a\alpha + b(d - \alpha))^2 = 0.$$

Therefore, we get the condition

$$a\alpha + b(d - \alpha) = 0. \quad (\dagger)$$

Note that condition (\dagger) applied to a vector X whose components are a or b is equivalent to the fact that X is orthogonal to $D\mathbf{1}$, since

$$X^\top D\mathbf{1} = \alpha a + (d - \alpha)b,$$

where $\alpha = \text{vol}(\{v_i \in V \mid x_i = a\})$.

We claim the following two facts. For any nonzero vector X whose components are a or b , if $X^\top D\mathbf{1} = \alpha a + (d - \alpha)b = 0$, then

- (1) $\alpha \neq 0$ and $\alpha \neq d$ iff $a \neq 0$ and $b \neq 0$.
- (2) if $a, b \neq 0$, then $a \neq b$.

(1) First assume that $a \neq 0$ and $b \neq 0$. If $\alpha = 0$, then $\alpha a + (d - \alpha)b = 0$ yields $db = 0$ with $d \neq 0$, which implies $b = 0$, a contradiction. If $d - \alpha = 0$, then we get $da = 0$ with $d \neq 0$, which implies $a = 0$, a contradiction.

Conversely, assume that $\alpha \neq 0$ and $\alpha \neq d$. If $a = 0$, then from $\alpha a + (d - \alpha)b = 0$ we get $(d - \alpha)b = 0$, which implies $b = 0$, contradicting the fact that $X \neq 0$. Similarly, if $b = 0$, then we get $\alpha a = 0$, which implies $a = 0$, contradicting the fact that $X \neq 0$.

(2) If $a, b \neq 0$, $a = b$ and $\alpha a + (d - \alpha)b = 0$, then $\alpha a + (d - \alpha)a = 0$, and since $a \neq 0$, we deduce that $d = 0$, a contradiction.

If $X^\top D\mathbf{1} = \alpha a + (d - \alpha)b = 0$ and $a, b \neq 0$, then

$$b = -\frac{\alpha}{(d - \alpha)} a,$$

so we get

$$\begin{aligned}\alpha a^2 + (d - \alpha)b^2 &= \alpha \frac{(d - \alpha)^2}{\alpha^2} b^2 + (d - \alpha)b^2 \\ &= (d - \alpha) \left(\frac{d - \alpha}{\alpha} + 1 \right) b^2 = \frac{(d - \alpha)db^2}{\alpha},\end{aligned}$$

and

$$\begin{aligned}(a - b)^2 &= \left(-\frac{(d - \alpha)}{\alpha} b - b \right)^2 \\ &= \left(\frac{d - \alpha}{\alpha} + 1 \right)^2 b^2 = \frac{d^2 b^2}{\alpha^2}.\end{aligned}$$

Since

$$\begin{aligned}X^\top DX &= \alpha a^2 + (d - \alpha)b^2 \\ X^\top LX &= (a - b)^2 \text{cut}(A, \bar{A}),\end{aligned}$$

we obtain

$$\begin{aligned}X^\top DX &= \frac{(d - \alpha)db^2}{\alpha} = \frac{\alpha da^2}{(d - \alpha)} \\ X^\top LX &= \frac{d^2 b^2}{\alpha^2} \text{cut}(A, \bar{A}) = \frac{d^2 a^2}{(d - \alpha)^2} \text{cut}(A, \bar{A}).\end{aligned}$$

If we wish to make α disappear, we pick

$$a = \sqrt{\frac{d - \alpha}{\alpha}}, \quad b = -\sqrt{\frac{\alpha}{d - \alpha}},$$

and then

$$\begin{aligned}X^\top DX &= d \\ X^\top LX &= \frac{d^2}{\alpha(d - \alpha)} \text{cut}(A, \bar{A}) = d \text{Ncut}(A, \bar{A}).\end{aligned}$$

In this case, we are considering indicator vectors of the form

$$\left\{ (x_1, \dots, x_N) \mid x_i \in \left\{ \sqrt{\frac{d - \alpha}{\alpha}}, -\sqrt{\frac{\alpha}{d - \alpha}} \right\}, \alpha = \text{vol}(A) \right\},$$

for any nonempty proper subset A of V . This is the choice adopted in von Luxburg [14]. Shi and Malik [12] use

$$a = 1, \quad b = -\frac{\alpha}{d - \alpha} = -\frac{k}{1 - k},$$

with

$$k = \frac{\alpha}{d}.$$

Another choice found in the literature (for example, in Belkin and Niyogi [1]) is

$$a = \frac{1}{\alpha}, \quad b = -\frac{1}{d - \alpha}.$$

However, there is no need to restrict solutions to be of either of these forms. So, let

$$\mathcal{X} = \{(x_1, \dots, x_N) \mid x_i \in \{a, b\}, a, b \in \mathbb{R}, a, b \neq 0\},$$

so that our solution set is

$$\mathcal{K} = \{X \in \mathcal{X} \mid X^\top D \mathbf{1} = 0\},$$

because by previous observations, since vectors $X \in \mathcal{X}$ have nonzero components, $X^\top D \mathbf{1} = 0$ implies that $\alpha \neq 0$, $\alpha \neq d$, and $a \neq b$, where $\alpha = \text{vol}(\{v_i \in V \mid x_i = a\})$. Actually, to be perfectly rigorous, we are looking for solutions in \mathbb{RP}^{N-1} , so our solution set is really

$$\mathbb{P}(\mathcal{K}) = \{(x_1 : \dots : x_N) \in \mathbb{RP}^{N-1} \mid (x_1, \dots, x_N) \in \mathcal{K}\}.$$

Consequently, our minimization problem can be stated as follows:

Problem PNC1

$$\begin{array}{ll} \text{minimize} & \frac{X^\top L X}{X^\top D X} \\ \text{subject to} & X^\top D \mathbf{1} = 0, \quad X \in \mathcal{X}. \end{array}$$

It is understood that the solutions are points $\mathbb{P}(X)$ in \mathbb{RP}^{N-1} .

Since the Rayleigh ratio and the constraints $X^\top D \mathbf{1} = 0$ and $X \in \mathcal{X}$ are scale-invariant (for any $\lambda \neq 0$, the Rayleigh ratio does not change if X is replaced by λX , $X \in \mathcal{X}$ iff $\lambda X \in \mathcal{X}$, and $(\lambda X)^\top D \mathbf{1} = \lambda X^\top D \mathbf{1} = 0$), we are led to the following formulation of our problem:

Problem PNC2

$$\begin{array}{llll} \text{minimize} & X^\top L X & & \\ \text{subject to} & X^\top D X = 1, & X^\top D \mathbf{1} = 0, & X \in \mathcal{X}. \end{array}$$

Problem PNC2 is equivalent to problem PNC1 in the sense that if X is any minimal solution of PNC1, then $X/(X^\top D X)^{1/2}$ is a minimal solution of PNC2 (with the same minimal value for the objective functions), and if X is a minimal solution of PNC2, then λX is a minimal solution for PNC1 for all $\lambda \neq 0$ (with the same minimal value for the objective functions). Equivalently, problems PNC1 and PNC2 have the same set of minimal solutions as points $\mathbb{P}(X) \in \mathbb{RP}^{N-1}$ given by their homogenous coordinates X .

Unfortunately, this is an NP-complete problem, as shown by Shi and Malik [12]. As often with hard combinatorial problems, we can look for a *relaxation* of our problem, which means looking for an optimum in a larger continuous domain. After doing this, the problem is to find a discrete solution which is close to a continuous optimum of the relaxed problem.

The natural relaxation of this problem is to allow X to be any nonzero vector in \mathbb{R}^N , and we get the problem:

$$\text{minimize } X^\top LX \quad \text{subject to } X^\top DX = 1, \quad X^\top D\mathbf{1} = 0.$$

As usual, let $Y = D^{1/2}X$, so that $X = D^{-1/2}Y$. Then, the condition $X^\top DX = 1$ becomes

$$Y^\top Y = 1,$$

the condition

$$X^\top D\mathbf{1} = 0$$

becomes

$$Y^\top D^{1/2}\mathbf{1} = 0,$$

and

$$X^\top LX = Y^\top D^{-1/2}LD^{-1/2}Y.$$

We obtain the problem:

$$\text{minimize } Y^\top D^{-1/2}LD^{-1/2}Y \quad \text{subject to } Y^\top Y = 1, \quad Y^\top D^{1/2}\mathbf{1} = 0.$$

Because $L\mathbf{1} = 0$, the vector $D^{1/2}\mathbf{1}$ belongs to the nullspace of the symmetric Laplacian $L_{\text{sym}} = D^{-1/2}LD^{-1/2}$. By Proposition A.2, minima are achieved by any unit eigenvector Y of the second eigenvalue ν_2 of L_{sym} . Then, $Z = D^{-1/2}Y$ is a solution of our original relaxed problem. Note that because Z is nonzero and orthogonal to $D\mathbf{1}$, a vector with positive entries, it must have negative and positive entries.

The next question is to figure how close is Z to an exact solution in \mathcal{X} . Actually, because solutions are points in \mathbb{RP}^{N-1} , the correct statement of the question is: Find an exact solution $\mathbb{P}(X) \in \mathbb{P}(\mathcal{X})$ which is the closest (in a suitable sense) to the approximate solution $\mathbb{P}(Z) \in \mathbb{RP}^{N-1}$. However, because \mathcal{X} is closed under the antipodal map, as explained in Appendix B, minimizing the distance $d(\mathbb{P}(X), \mathbb{P}(Z))$ on \mathbb{RP}^{N-1} is equivalent to minimizing the Euclidean distance $\|X - Z\|_2$ (if we use the Riemannian metric on \mathbb{RP}^{N-1} induced by the Euclidean metric on \mathbb{R}^N).

We may assume $b < 0$, in which case $a > 0$. If all entries in Z are nonzero, due to the projective nature of the solution set, it seems reasonable to say that the partition of V is defined by the signs of the entries in Z . Thus, A will consist of nodes those v_i for which $x_i > 0$. Elements corresponding to zero entries can be assigned to either A or \bar{A} , unless additional information is available.

Now, using the fact that

$$b = -\frac{\alpha a}{d - \alpha},$$

a better solution is to look for a vector $X \in \mathbb{R}^N$ with $X_i \in \{a, b\}$ which is closest to a minimum Z of the relaxed problem. Here is a proposal for an algorithm.

For any solution Z of the relaxed problem, let $I_Z^+ = \{i \mid Z_i > 0\}$ be the set of indices of positive entries in Z , $I_Z^- = \{i \mid Z_i < 0\}$ the set of indices of negative entries in Z , $I_Z^0 = \{i \mid Z_i = 0\}$ the set of indices of zero entries in Z , and let Z^+ and Z^- be the vectors given by

$$Z_i^+ = \begin{cases} Z_i & \text{if } i \in I_Z^+ \\ 0 & \text{if } i \notin I_Z^+ \end{cases} \quad Z_i^- = \begin{cases} Z_i & \text{if } i \in I_Z^- \\ 0 & \text{if } i \notin I_Z^- \end{cases}.$$

Also let $n_a = |I_Z^+|$, $n_b = |I_Z^-|$, let \bar{a} and \bar{b} be the average of the positive and negative entries in Z respectively, that is,

$$\bar{a} = \frac{\sum_{i \in I_Z^+} Z_i}{n_a} \quad \bar{b} = \frac{\sum_{i \in I_Z^-} Z_i}{n_b},$$

and let \overline{Z}^+ and \overline{Z}^- be the vectors given by

$$(\overline{Z}^+)_i = \begin{cases} \bar{a} & \text{if } i \in I_Z^+ \\ 0 & \text{if } i \notin I_Z^+ \end{cases} \quad (\overline{Z}^-)_i = \begin{cases} \bar{b} & \text{if } i \in I_Z^- \\ 0 & \text{if } i \notin I_Z^- \end{cases}.$$

If $\|\overline{Z}^+ - Z^+\| > \|\overline{Z}^- - Z^-\|$, then replace Z by $-Z$. Then, perform the following steps:

(1) Let

$$n_a = |I_Z^+|, \quad \alpha = \text{vol}(\{v_i \mid i \in I_Z^+\}), \quad \beta = \frac{\alpha}{d - \alpha},$$

and form the vector \overline{X} with

$$\overline{X}_i = \begin{cases} a & \text{if } i \in I_Z^+ \\ -\beta a & \text{otherwise,} \end{cases}$$

such that $\|\overline{X} - Z\|$ is minimal; the scalar a is determined by finding the solution of the equation

$$\mathcal{Z}a = Z,$$

in the least squares sense, where

$$\mathcal{Z} = \begin{cases} 1 & \text{if } i \in I_Z^+ \\ -\beta & \text{otherwise,} \end{cases}$$

and is given by

$$a = (\mathcal{Z}^\top \mathcal{Z})^{-1} \mathcal{Z}^\top Z;$$

that is,

$$a = \frac{\sum_{i \in I_Z^+} Z_i}{n_a + \beta^2(N - n_a)} - \frac{\sum_{i \in I_Z^-} \beta Z_i}{n_a + \beta^2(N - n_a)}.$$

(2) While $I_Z^0 \neq \emptyset$, pick the smallest index $i \in I_Z^0$, compute

$$\begin{aligned} \tilde{I}_Z^+ &= I_Z^+ \cup \{i\} \\ \tilde{n}_a &= n_a + 1 \\ \tilde{\alpha} &= \alpha + d(v_i) \\ \tilde{\beta} &= \frac{\tilde{\alpha}}{d - \tilde{\alpha}}, \end{aligned}$$

and then \tilde{X} with

$$\tilde{X}_j = \begin{cases} \tilde{a} & \text{if } j \in \tilde{I}_Z^+ \\ -\tilde{\beta}\tilde{a} & \text{otherwise,} \end{cases}$$

and

$$\tilde{a} = \frac{\sum_{j \in \tilde{I}_Z^+} Z_j}{\tilde{n}_a + \tilde{\beta}^2(N - \tilde{n}_a)} - \frac{\sum_{j \in I_Z^-} \tilde{\beta} Z_j}{\tilde{n}_a + \tilde{\beta}^2(N - \tilde{n}_a)}.$$

If $\|\tilde{X} - Z\| < \|\bar{X} - Z\|$, then let $\bar{X} = \tilde{X}$, $I_Z^+ = \tilde{I}_Z^+$, $n_a = \tilde{n}_a$, $\alpha = \tilde{\alpha}$, and $I_Z^0 = I_Z^0 - \{i\}$; go back to (2).

(3) The final answer is \bar{X} .

4.3 *K*-Way Clustering Using Normalized Cuts

We now consider the general case in which $K \geq 3$. Two crucial issues need to be addressed (to the best of our knowledge, these points are not clearly articulated in the literature).

1. The choice of a matrix representation for partitions on the set of vertices. It is important that such a representation be scale-invariant. It is also necessary to state necessary and sufficient conditions for such matrices to represent a partition.
2. The choice of a metric to compare solutions. It turns out that the space of discrete solutions can be viewed as a subset of the K -fold product $(\mathbb{RP}^{N-1})^K$ of the projective space \mathbb{RP}^{N-1} . Version 1 of the formulation of our minimization problem (PNC1) makes this point clear. However, the relaxation $(*_1)$ of version 2 of our minimization problem (PNC2), which is equivalent to version 1, reveals that the solutions of the relaxed problem $(*_1)$ are members of the *Grassmannian* $G(K, N)$. Thus, we have two choices of metrics: (1) a metric on $(\mathbb{RP}^{N-1})^K$; (2) a metric on $G(K, N)$. We discuss the first choice, which is the choice implicitly adopted by Shi and Yu.

We describe a partition (A_1, \dots, A_K) of the set of nodes V by an $N \times K$ matrix $X = [X^1 \dots X^K]$ whose columns X^1, \dots, X^K are indicator vectors of the partition (A_1, \dots, A_K) . Inspired by what we did in Section 4.2, we assume that the vector X^j is of the form

$$X^j = (x_1^j, \dots, x_N^j),$$

where $x_i^j \in \{a_j, b_j\}$ for $j = 1, \dots, K$ and $i = 1, \dots, N$, and where a_j, b_j are any two distinct real numbers. The vector X^j is an indicator vector for A_j in the sense that, for $i = 1, \dots, N$,

$$x_i^j = \begin{cases} a_j & \text{if } v_i \in A_j \\ b_j & \text{if } v_i \notin A_j. \end{cases}$$

When $\{a_j, b_j\} = \{0, 1\}$ for $j = 1, \dots, K$, such a matrix is called a *partition matrix* by Yu and Shi. However, such a choice is premature, since it is better to have a scale-invariant representation to make the denominators of the Rayleigh ratios go away.

Since the partition (A_1, \dots, A_K) consists of nonempty pairwise disjoint blocks whose union is V , some conditions on X are required to reflect these properties, but we will worry about this later.

Let $d = \mathbf{1}^\top D \mathbf{1}$ and $\alpha_j = \text{vol}(A_j)$, so that $\alpha_1 + \dots + \alpha_K = d$. Then, $\text{vol}(\overline{A_j}) = d - \alpha_j$, and as in Section 4.2, we have

$$\begin{aligned} (X^j)^\top L X^j &= (a_j - b_j)^2 \text{cut}(A_j, \overline{A_j}), \\ (X^j)^\top D X^j &= \alpha_j a_j^2 + (d - \alpha_j) b_j^2. \end{aligned}$$

When $K \geq 3$, unlike the case $K = 2$, in general we have $\text{cut}(A_j, \overline{A_j}) \neq \text{cut}(A_k, \overline{A_k})$ if $j \neq k$, and since

$$\text{Ncut}(A_1, \dots, A_K) = \sum_{j=1}^K \frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)},$$

we would like to choose a_j, b_j so that

$$\frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)} = \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j} \quad j = 1, \dots, K,$$

because this implies that

$$\mu(X) = \text{Ncut}(A_1, \dots, A_K) = \sum_{j=1}^K \frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)} = \sum_{j=1}^K \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j}.$$

Since

$$\frac{(X^j)^\top L X^j}{(X^j)^\top D X^j} = \frac{(a_j - b_j)^2 \text{cut}(A_j, \overline{A_j})}{\alpha_j a_j^2 + (d - \alpha_j) b_j^2}$$

and $\text{vol}(A_j) = \alpha_j$, in order to have

$$\frac{\text{cut}(A_j, \overline{A_j})}{\text{vol}(A_j)} = \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j} \quad j = 1, \dots, K,$$

we need to have

$$\frac{(a_j - b_j)^2}{\alpha_j a_j^2 + (d - \alpha_j) b_j^2} = \frac{1}{\alpha_j} \quad j = 1, \dots, K.$$

Thus, we must have

$$(a_j^2 - 2a_j b_j + b_j^2) \alpha_j = \alpha_j a_j^2 + (d - \alpha_j) b_j^2,$$

which yields

$$2\alpha_j b_j (b_j - a_j) = d b_j^2.$$

The above equation is trivially satisfied if $b_j = 0$. If $b_j \neq 0$, then

$$2\alpha_j (b_j - a_j) = d b_j,$$

which yields

$$a_j = \frac{2\alpha_j - d}{2\alpha_j} b_j.$$

This choice seems more complicated than the choice $b_j = 0$, so we will opt for the choice $b_j = 0, j = 1, \dots, K$. With this choice, we get

$$(X^j)^\top D X^j = \alpha_j a_j^2.$$

Thus, it makes sense to pick

$$a_j = \frac{1}{\sqrt{\alpha_j}} = \frac{1}{\sqrt{\text{vol}(A_j)}}, \quad j = 1, \dots, K,$$

which is the solution presented in von Luxburg [14]. This choice also corresponds to the scaled partition matrix used in Yu [15] and Yu and Shi [16].

When $N = 10$ and $K = 4$, an example of a matrix X representing the partition of $V = \{v_1, v_2, \dots, v_{10}\}$ into the four blocks

$$\{A_1, A_2, A_3, A_4\} = \{\{v_2, v_4, v_6\}, \{v_1, v_5\}, \{v_3, v_8, v_{10}\}, \{v_7, v_9\}\},$$

is shown below:

$$X = \begin{pmatrix} 0 & a_2 & 0 & 0 \\ a_1 & 0 & 0 & 0 \\ 0 & 0 & a_3 & 0 \\ a_1 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \\ a_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_4 \\ 0 & 0 & a_3 & 0 \\ 0 & 0 & 0 & a_4 \\ 0 & 0 & a_3 & 0 \end{pmatrix}.$$

Let us now consider the problem of finding necessary and sufficient conditions for a matrix X to represent a partition of V .

When $b_j = 0$, the pairwise disjointness of the A_i is captured by the orthogonality of the X^i :

$$(X^i)^\top X^j = 0, \quad 1 \leq i, j \leq K, i \neq j. \quad (*)$$

This is because, for any matrix X where the nonzero entries in each column have the same sign, for any $i \neq j$, the condition

$$(X^i)^\top X^j = 0$$

says that for every $k = 1, \dots, N$, if $x_k^i \neq 0$ then $x_k^j = 0$.

When we formulate our minimization problem in terms of Rayleigh ratios, conditions on the quantities $(X^i)^\top DX^i$ show up, and it is more convenient to express the orthogonality conditions using the quantities $(X^i)^\top DX^j$ instead of the $(X^i)^\top X^j$, because these various conditions can be combined into a single condition involving the matrix $X^\top DX$. Now, because D is a diagonal matrix with positive entries and because the nonzero entries in each column of X have the same sign, for any $i \neq j$, the condition

$$(X^i)^\top X^j = 0$$

is equivalent to

$$(X^i)^\top DX^j = 0, \quad (**)$$

since, as above, it means that for $k = 1, \dots, N$, if $x_k^i \neq 0$ then $x_k^j = 0$. Observe that the orthogonality conditions (*) (and (**)) are equivalent to the fact that every row of X has at most one nonzero entry.

Remark: The disjointness condition

$$X\mathbf{1}_K = \mathbf{1}_N$$

is used in Yu [15]. However, this condition does not guarantee the disjointness of the blocks. For example, it is satisfied by the matrix X whose first column is $\mathbf{1}_N$, with 0 everywhere else.

Each A_j is nonempty iff $X^j \neq 0$, and the fact that the union of the A_j is V is captured by the fact that each row of X must have some nonzero entry (every vertex appears in some block). It is not obvious how to state conveniently this condition in matrix form.

Observe that the diagonal entries of the matrix XX^\top are the square Euclidean norms of the rows of X . Therefore, we can assert that these entries are all nonzero. Let DIAG be the function which returns the diagonal matrix (containing the diagonal of A),

$$\text{DIAG}(A) = \text{diag}(a_{11}, \dots, a_{nn}),$$

for any square matrix $A = (a_{ij})$. Then, the condition for the rows of X to be nonzero can be stated as

$$\det(\text{DIAG}(XX^\top)) \neq 0.$$

Observe that the matrix

$$\text{DIAG}(XX^\top)^{-1/2}X$$

is the result of normalizing the rows of X so that they have Euclidean norm 1. This normalization step is used by Yu [15] in the search for a discrete solution closest to a solution of a relaxation of our original problem. For our special matrices representing partitions, normalizing the rows will have the effect of rescaling the columns (if row i has a_j in column j , then all nonzero entries in column j are equal to a_j), but for a more general matrix, this is false. Since our solution matrices are invariant under rescaling the columns, but not the rows, rescaling the rows does not appear to be a good idea.

A better idea which leads to a scale-invariant condition stems from the observation that since every row of any matrix X representing a partition has a single nonzero entry a_j , we have

$$X^\top \mathbf{1} = \begin{pmatrix} n_1 a_1 \\ \vdots \\ n_K a_K \end{pmatrix}, \quad X^\top X = \text{diag}(n_1 a_1^2, \dots, n_K a_K^2),$$

where n_j is the number of elements in A_j , the j th block of the partition, which implies that

$$(X^\top X)^{-1} X^\top \mathbf{1} = \begin{pmatrix} \frac{1}{a_1} \\ \vdots \\ \frac{1}{a_K} \end{pmatrix},$$

and thus,

$$X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}. \quad (\dagger)$$

When $a_j = 1$ for $j = 1, \dots, K$, we have $(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}$, and condition (\dagger) reduces to

$$X \mathbf{1}_K = \mathbf{1}_N.$$

Note that because the columns of X are linearly independent, $(X^\top X)^{-1} X^\top$ is the pseudo-inverse of X . Consequently, condition (\dagger) , can also be written as

$$XX^+ = \mathbf{1},$$

where $X^+ = (X^\top X)^{-1} X^\top$ is the pseudo-inverse of X . However, it is well known that XX^+ is the orthogonal projection of \mathbb{R}^K onto the range of X (see Gallier [6], Section 14.1), so the condition $XX^+ = \mathbf{1}$ is equivalent to the fact that $\mathbf{1}$ belongs to the range of X . In retrospect, this should have been obvious since the columns of a solution X satisfy the equation

$$a_1^{-1} X^1 + \dots + a_K^{-1} X^K = \mathbf{1}.$$

We emphasize that it is important to use conditions that are invariant under multiplication by a nonzero scalar, because the Rayleigh ratio is scale-invariant, and it is crucial to take advantage of this fact to make the denominators go away.

If we let

$$\mathcal{X} = \left\{ [X^1 \ \dots \ X^K] \mid X^j = a_j(x_1^j, \dots, x_N^j), x_i^j \in \{1, 0\}, a_j \in \mathbb{R}, X^j \neq 0 \right\}$$

(note that the condition $X^j \neq 0$ implies that $a_j \neq 0$), then the set of matrices representing partitions of V into K blocks is

$$\mathcal{K} = \left\{ X = [X^1 \ \dots \ X^K] \mid X \in \mathcal{X}, \right. \\ \left. (X^i)^\top D X^j = 0, \quad 1 \leq i, j \leq K, i \neq j, \right. \\ \left. X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1} \right\}.$$

As in the case $K = 2$, to be rigorous, the solution are really K -tuples of points in \mathbb{RP}^{N-1} , so our solution set is really

$$\mathbb{P}(\mathcal{K}) = \left\{ (\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)) \mid [X^1 \ \dots \ X^K] \in \mathcal{K} \right\}.$$

In view of the above, we have our first formulation of K -way clustering of a graph using normalized cuts, called problem PNC1 (the notation PNCX is used in Yu [15], Section 2.1):

**K -way Clustering of a graph using Normalized Cut, Version 1:
Problem PNC1**

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^K \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j} \\ & \text{subject to} && (X^i)^\top D X^j = 0, \quad 1 \leq i, j \leq K, i \neq j, \\ & && X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}, \quad X \in \mathcal{X}. \end{aligned}$$

As in the case $K = 2$, the solutions that we are seeking are K -tuples $(\mathbb{P}(X^1), \dots, \mathbb{P}(X^K))$ of points in \mathbb{RP}^{N-1} determined by their homogeneous coordinates X^1, \dots, X^K .

Remark: Because

$$(X^j)^\top L X^j = (X^j)^\top D X^j - (X^j)^\top W X^j = \text{vol}(A_j) - (X^j)^\top W X^j,$$

Instead of minimizing

$$\mu(X^1, \dots, X^K) = \sum_{j=1}^K \frac{(X^j)^\top L X^j}{(X^j)^\top D X^j},$$

we can maximize

$$\epsilon(X^1, \dots, X^K) = \sum_{j=1}^K \frac{(X^j)^\top W X^j}{(X^j)^\top D X^j},$$

since

$$\epsilon(X^1, \dots, X^K) = K - \mu(X^1, \dots, X^K).$$

This second option is the one chosen by Yu [15] and Yu and Shi [16] (actually, they work with $\frac{1}{K}(K - \mu(X^1, \dots, X^K))$, but this doesn't make any difference).

Let us now show how our original formulation (PNC1) can be converted to a more convenient form, by chasing the denominators in the Rayleigh ratios, and by expressing the objective function in terms of the *trace* of a certain matrix.

For any $N \times N$ matrix A , because

$$\begin{aligned} X^\top AX &= \begin{bmatrix} (X^1)^\top \\ \vdots \\ (X^K)^\top \end{bmatrix} A[X^1 \dots X^K] \\ &= \begin{pmatrix} (X^1)^\top AX^1 & (X^1)^\top AX^2 & \dots & (X^1)^\top AX^K \\ (X^2)^\top AX^1 & (X^2)^\top AX^2 & \dots & (X^2)^\top AX^K \\ \vdots & \vdots & \ddots & \vdots \\ (X^K)^\top AX^1 & (X^K)^\top AX^2 & \dots & (X^K)^\top AX^K \end{pmatrix}, \end{aligned}$$

we have

$$\text{tr}(X^\top AX) = \sum_{j=1}^K (X^j)^\top AX^j,$$

and the conditions

$$(X^i)^\top AX^j = 0, \quad 1 \leq i, j \leq K, \quad i \neq j,$$

are equivalent to

$$X^\top AX = \text{diag}((X^1)^\top AX^1, \dots, (X^K)^\top AX^K).$$

As a consequence, if we assume that

$$(X^1)^\top AX^1 = \dots = (X^K)^\top AX^K = \alpha^2,$$

then we have

$$X^\top AX = \alpha^2 I,$$

and if R is any orthogonal $K \times K$ matrix, then by multiplying on the left by R^\top and on the right by R , we get

$$R^\top X^\top AX R = R^\top \alpha^2 I R = \alpha^2 R^\top R = \alpha^2 I.$$

Therefore, if

$$X^\top AX = \alpha^2 I,$$

then

$$(XR)^\top A(XR) = \alpha^2 I,$$

for any orthogonal $K \times K$ matrix R . Furthermore, because $\text{tr}(AB) = \text{tr}(BA)$ for all matrices A, B , we have

$$\text{tr}(R^\top X^\top AX R) = \text{tr}(X^\top AX).$$

Since the Rayleigh ratios

$$\frac{(X^j)^\top LX^j}{(X^j)^\top DX^j}$$

are invariant under rescaling by a nonzero number, we have

$$\begin{aligned} \mu(X) &= \mu(X^1, \dots, X^K) = \sum_{j=1}^K \frac{(X^j)^\top LX^j}{(X^j)^\top DX^j} \\ &= \mu(((X^1)^\top DX^1)^{-1/2} X^1, \dots, ((X^K)^\top DX^K)^{-1/2} X^K) \\ &= \sum_{j=1}^K ((X^j)^\top DX^j)^{-1/2} (X^j)^\top L ((X^j)^\top DX^j)^{-1/2} X^j \\ &= \text{tr}(\Lambda^{-1/2} X^\top LX \Lambda^{-1/2}), \end{aligned}$$

where

$$\Lambda = \text{diag}((X^1)^\top DX^1, \dots, (X^K)^\top DX^K).$$

If $(X^1)^\top DX^1 = \dots = (X^K)^\top DX^K = \alpha^2$, then $\Lambda = \alpha^2 I_K$, so Λ commutes with any $K \times K$ matrix which implies that

$$\text{tr}(\Lambda^{-1/2} R^\top X^\top LX R \Lambda^{-1/2}) = \text{tr}(R^\top \Lambda^{-1/2} R^\top X^\top LX \Lambda^{-1/2} R) = \text{tr}(\Lambda^{-1/2} X^\top LX \Lambda^{-1/2}),$$

and thus,

$$\mu(X) = \mu(XR),$$

for any orthogonal $K \times K$ matrix R .

The condition

$$X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}$$

is also invariant if we replace X by XR , where R is any invertible matrix, because

$$\begin{aligned} XR((XR)^\top (XR))^{-1} (XR)^\top \mathbf{1} &= XR(R^\top X^\top XR)^{-1} R^\top X^\top \mathbf{1} \\ &= XRR^{-1} (X^\top X)^{-1} (R^\top)^{-1} R^\top X^\top \mathbf{1} \\ &= X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}. \end{aligned}$$

In summary we proved the following proposition:

Proposition 4.1. *For any orthogonal $K \times K$ matrix R , any symmetric $N \times N$ matrix A , and any $N \times K$ matrix $X = [X^1 \dots X^K]$, the following properties hold:*

(1) $\mu(X) = \text{tr}(\Lambda^{-1/2} X^\top LX \Lambda^{-1/2})$, where

$$\Lambda = \text{diag}((X^1)^\top DX^1, \dots, (X^K)^\top DX^K).$$

- (2) If $(X^1)^\top DX^1 = \dots = (X^K)^\top DX^K = \alpha^2$, then $\mu(X) = \mu(XR)$.
- (3) The condition $X^\top AX = \alpha^2 I$ is preserved if X is replaced by XR .
- (4) The condition $X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}$ is preserved if X is replaced by XR .

Now, by Proposition 4.1(1) and the fact that the conditions in PNC1 are scale-invariant, we are led to the following formulation of our problem:

$$\begin{array}{ll}
\text{minimize} & \text{tr}(X^\top LX) \\
\text{subject to} & (X^i)^\top DX^j = 0, \quad 1 \leq i, j \leq K, i \neq j, \\
& (X^j)^\top DX^j = 1, \quad 1 \leq j \leq K, \\
& X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}, \quad X \in \mathcal{X}.
\end{array}$$

Conditions on lines 2 and 3 can be combined in the equation

$$X^\top DX = I,$$

and, we obtain the following formulation of our minimization problem:

***K*-way Clustering of a graph using Normalized Cut, Version 2:
Problem PNC2**

$$\begin{array}{ll}
\text{minimize} & \text{tr}(X^\top LX) \\
\text{subject to} & X^\top DX = I, \\
& X(X^\top X)^{-1} X^\top \mathbf{1} = \mathbf{1}, \quad X \in \mathcal{X}.
\end{array}$$

Problem PNC2 is equivalent to problem PNC1 in the sense that for every minimal solution (X^1, \dots, X^K) of PNC1, $((X^1)^\top DX^1)^{-1/2} X^1, \dots, ((X^K)^\top DX^K)^{-1/2} X^K$ is a minimal solution of PNC2 (with the same minimum for the objective functions), and that for every minimal solution (Z^1, \dots, Z^K) of PNC2, $(\lambda_1 Z^1, \dots, \lambda_K Z^K)$ is a minimal solution of PNC1, for all $\lambda_i \neq 0$, $i = 1, \dots, K$ (with the same minimum for the objective functions). In other words, problems PNC1 and PNC2 have the same set of minimal solutions as K -tuples of points $(\mathbb{P}(X^1), \dots, \mathbb{P}(X^K))$ in $\mathbb{R}\mathbb{P}^{N-1}$ determined by their homogeneous coordinates X^1, \dots, X^K .

Formulation PNC2 reveals that finding a minimum normalized cut has a geometric interpretation in terms of the graph drawings discussed in Section 3.1. Indeed, PNC2 has the following equivalent formulation: Find a minimal energy graph drawing X in \mathbb{R}^K of the weighted graph $G = (V, W)$ such that:

1. The matrix X is orthogonal with respect to the inner product $\langle -, - \rangle_D$ in \mathbb{R}^N induced by D , with

$$\langle x, y \rangle_D = x^\top Dy, \quad x, y \in \mathbb{R}^N.$$

2. The rows of X are nonzero; this means that no vertex $v_i \in V$ is assigned to the origin of \mathbb{R}^K (the zero vector 0_K).
3. Every vertex v_i is assigned a point of the form $(0, \dots, 0, a_j, 0, \dots, 0)$ on some axis (in \mathbb{R}^K).
4. Every axis in \mathbb{R}^K is assigned at least some vertex.

Condition 1 can be reduced to the standard condition for graph drawings ($R^\top R = I$) by making the change of variable $Y = D^{1/2}X$ or equivalently $X = D^{-1/2}Y$. Indeed,

$$\text{tr}(X^\top LX) = \text{tr}(Y^\top D^{-1/2}LD^{-1/2}Y),$$

so we use the normalized Laplacian $L_{\text{sym}} = D^{-1/2}LD^{-1/2}$ instead of L ,

$$X^\top DX = Y^\top Y = I,$$

and conditions (2), (3), (4) are preserved under the change of variable $Y = D^{1/2}X$, since $D^{1/2}$ is invertible. However, conditions (2), (3), (4) are “hard” constraints, especially condition (3). In fact, condition (3) implies that the columns of X are orthogonal with respect to both the Euclidean inner product and the inner product $\langle -, - \rangle_D$, so condition (1) is redundant, except for the fact that it prescribes the norm of the columns, but this is not essential due to the projective nature of the solutions.

The main problem in finding a good relaxation of problem PNC2 is that it is very difficult to enforce the condition $X \in \mathcal{X}$. Also, the solutions X are not preserved under arbitrary rotations, but only by very special rotations which leave \mathcal{X} invariant (they exchange the axes). The first natural relaxation of problem PNC2 is to drop the condition that $X \in \mathcal{X}$, and we obtain the

Problem $(*_1)$

$$\begin{array}{ll} \text{minimize} & \text{tr}(X^\top LX) \\ \text{subject to} & X^\top DX = I, \\ & X(X^\top X)^{-1}X^\top \mathbf{1} = \mathbf{1}. \end{array}$$

By Proposition 4.1, for every orthogonal matrix $R \in \mathbf{O}(K)$ and for every X minimizing $(*_1)$, the matrix XR also minimizes $(*_1)$. As a consequence, as explained below, we can view the solutions of problem $(*_1)$ as elements of the *Grassmannian* $G(N, K)$.

Recall that the *Stiefel manifold* $St(k, n)$ consists of the set of orthogonal k -frames in \mathbb{R}^n , that is, the k -tuples of orthonormal vectors (u_1, \dots, u_k) with $u_i \in \mathbb{R}^n$. For $k = n$, the manifold $St(n, n)$ is identical to the orthogonal group $\mathbf{O}(n)$. For $1 \leq k \leq n - 1$, the group $\mathbf{SO}(n)$ acts transitively on $St(k, n)$, and $St(k, n)$ is isomorphic to the coset manifold

$\mathbf{SO}(n)/\mathbf{SO}(n-k)$. The *Grassmann manifold* $G(k, n)$ consists of all (linear) k -dimensional subspaces of \mathbb{R}^n . Again, the group $\mathbf{SO}(n)$ acts transitively on $G(k, n)$, and $G(k, n)$ is isomorphic to the coset manifold $\mathbf{SO}(n)/S(\mathbf{SO}(k) \times \mathbf{SO}(n-k))$. The group $\mathbf{O}(k)$ acts on the right on the Stiefel manifold $St(k, n)$ (by multiplication), and the orbit manifold $St(k, n)/\mathbf{O}(k)$ is isomorphic to the Grassmann manifold $G(k, n)$. Furthermore, both $St(k, n)$ and $G(k, n)$ are *naturally reductive homogeneous manifolds* (for the Stiefel manifold, when $n \geq 3$), and $G(k, n)$ is even a *symmetric space* (see O’Neill [11]). The upshot of all this is that to a large extent, the differential geometry of these manifolds is completely determined by some subspace \mathfrak{m} of the Lie algebra $\mathfrak{so}(n)$, such that we have a direct sum

$$\mathfrak{so}(n) = \mathfrak{m} \oplus \mathfrak{h},$$

where $\mathfrak{h} = \mathfrak{so}(n-k)$ in the case of the Stiefel manifold, and $\mathfrak{h} = \mathfrak{so}(k) \times \mathfrak{so}(n-k)$ in the case of the Grassmannian manifold (some additional condition on \mathfrak{m} is required). In particular, the geodesics in both manifolds can be determined quite explicitly, and thus we obtain closed form formulae for distances, *etc.*

The Stiefel manifold $St(k, n)$ can be viewed as the set of all $n \times k$ matrices X such that

$$X^\top X = I_k.$$

In our situation, we are considering $N \times K$ matrices X such that

$$X^\top DX = I.$$

This is not quite the Stiefel manifold, but if we write $Y = D^{1/2}X$, then we have

$$Y^\top Y = I,$$

so the space of matrices X satisfying the condition $X^\top DX = I$ is the image $\mathcal{D}(St(K, N))$ of the Stiefel manifold $St(K, N)$ under the linear map \mathcal{D} given by

$$\mathcal{D}(X) = D^{1/2}X.$$

Now, the right action of $\mathbf{O}(K)$ on $\mathcal{D}(St(K, N))$ yields a coset manifold $\mathcal{D}(St(K, N))/\mathbf{O}(K)$ which is obviously isomorphic to the Grassmann manifold $G(K, N)$.

Therefore, *the solutions of problem $(*_1)$ can be viewed as elements of the Grassmannian $G(N, K)$* . We can take advantage of this fact to find a discrete solution of our original optimization problem PNC2 approximated by a continuous solution of $(*_1)$.

Recall that condition $X(X^\top X)^{-1}X^\top \mathbf{1} = \mathbf{1}$ is equivalent to $XX^\top \mathbf{1} = \mathbf{1}$, which is also equivalent to the fact that $\mathbf{1}$ is in the range of X . If we make the change of variable $Y = D^{1/2}X$ or equivalently $X = D^{-1/2}Y$, the condition that $\mathbf{1}$ is in the range of X becomes the condition that $D^{1/2}\mathbf{1}$ is in the range of Y , which is equivalent to

$$YY^\top D^{1/2}\mathbf{1} = D^{1/2}\mathbf{1}.$$

However, since $Y^\top Y = I$, we have

$$Y^+ = Y^\top,$$

so we get the equivalent problem

Problem (₁)**

$$\begin{aligned} & \text{minimize} && \text{tr}(Y^\top D^{-1/2} L D^{-1/2} Y) \\ & \text{subject to} && Y^\top Y = I, \\ & && Y Y^\top D^{1/2} \mathbf{1} = D^{1/2} \mathbf{1}. \end{aligned}$$

This time, the matrices Y satisfying condition $Y^\top Y = I$ do belong to the Stiefel manifold $St(K, N)$, and again, we view the solutions of problem (**₁) as elements of the Grassmannian $G(K, N)$. We pass from a solution Y of problem (**₁) in $G(K, N)$ to a solution Z of problem (*₁) in $G(K, N)$ by the linear map \mathcal{D}^{-1} ; namely, $Z = \mathcal{D}(Y) = D^{-1/2} Y$.

The Rayleigh–Ritz Theorem (see Proposition A.2) tells us that if we temporarily ignore the second constraint, minima of problem (**₁) are obtained by picking any K unit eigenvectors (u_1, \dots, u_k) associated with the smallest eigenvalues

$$0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_K$$

of $L_{\text{sym}} = D^{-1/2} L D^{-1/2}$. We may assume that $\nu_2 > 0$, namely that the underlying graph is connected (otherwise, we work with each connected component), in which case $Y^1 = D^{1/2} \mathbf{1} / \|D^{1/2} \mathbf{1}\|_2$, because $\mathbf{1}$ is in the nullspace of L . Since $Y^1 = D^{1/2} \mathbf{1} / \|D^{1/2} \mathbf{1}\|_2$, the vector $D^{1/2} \mathbf{1}$ is in the range of Y , so the condition

$$Y Y^\top D^{1/2} \mathbf{1} = D^{1/2} \mathbf{1}$$

is also satisfied. Then, $Z = D^{-1/2} Y$ with $Y = [u_1 \dots u_K]$ yields a minimum of our relaxed problem (*₁) (the second constraint is satisfied because $\mathbf{1}$ is in the range of Z).

By Proposition 2.4, the vectors Z^j are eigenvectors of L_{rw} associated with the eigenvalues $0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_K$. Recall that $\mathbf{1}$ is an eigenvector for the eigenvalue $\nu_1 = 0$, and $Z^1 = \mathbf{1} / \|D^{1/2} \mathbf{1}\|_2$. Because, $(Y^i)^\top Y^j = 0$ whenever $i \neq j$, we have

$$(Z^i)^\top D Z^j = 0, \quad \text{whenever } i \neq j.$$

This implies that Z^2, \dots, Z^K are all orthogonal to $D\mathbf{1}$, and thus, that each Z^j has both some positive and some negative coordinate, for $j = 2, \dots, K$.

The conditions $(Z^i)^\top D Z^j = 0$ do not necessarily imply that Z^i and Z^j are orthogonal (w.r.t. the Euclidean inner product), but we can obtain a solution of Problem (*₁) achieving the same minimum for which distinct columns Z^i and Z^j are simultaneously orthogonal and D -orthogonal, by multiplying Z by some $K \times K$ orthogonal matrix R on the right. Indeed,

the $K \times K$ symmetric matrix $Z^\top Z$ can be diagonalized by some orthogonal $K \times K$ matrix R as

$$Z^\top Z = R\Sigma R^\top,$$

where Σ is a diagonal matrix, and thus,

$$R^\top Z^\top ZR = (ZR)^\top ZR = \Sigma,$$

which shows that the columns of ZR are orthogonal. By Proposition 4.1, ZR also satisfies the constraints of $(*_1)$, and $\text{tr}((ZR)^\top L(ZR)) = \text{tr}(Z^\top LZ)$.

Remark: Since Y has linearly independent columns (in fact, orthogonal) and since $Z = D^{-1/2}Y$, the matrix Z also has linearly independent columns, so $Z^\top Z$ is positive definite and the entries in Σ are all positive.

In summary, we should look for a solution X that corresponds to an element of the Grassmannian $G(K, N)$, and hope that for some suitable orthogonal matrix R , the vectors in XR are close to a true solution of the original problem.

4.4 K -Way Clustering; Using The Dependencies Among X^1, \dots, X^K

At this stage, it is interesting to reconsider the case $K = 2$ in the light of what we just did when $K \geq 3$. When $K = 2$, X^1 and X^2 are not independent, and it is convenient to assume that the nonzero entries in X^1 and X^2 are both equal to some positive real $c \in \mathbb{R}$, so that

$$X^1 + X^2 = c\mathbf{1}.$$

To avoid subscripts, write (A, \bar{A}) for the partition of V that we are seeking, and as before let $d = \mathbf{1}^\top D\mathbf{1}$ and $\alpha = \text{vol}(A)$. We know from Section 4.2 that

$$\begin{aligned} (X^1)^\top DX^1 &= \alpha c^2 \\ (X^2)^\top DX^2 &= (d - \alpha)c^2, \end{aligned}$$

so we normalize X^1 and X^2 so that $(X^1)^\top DX^1 = (X^2)^\top DX^2 = c^2$, and we consider

$$X = \begin{bmatrix} X^1 & X^2 \\ \sqrt{\alpha} & \sqrt{d - \alpha} \end{bmatrix}.$$

Now, we claim that there is an orthogonal matrix R so that if X as above is a solution to our discrete problem, then XR contains a multiple of $\mathbf{1}$ as a first column. A similar observation is made in Yu [15] and Yu and Shi [16] (but beware that in these works $\alpha = \text{vol}(A)/\sqrt{d}$). In fact,

$$R = \frac{1}{\sqrt{d}} \begin{pmatrix} \sqrt{\alpha} & \sqrt{d - \alpha} \\ \sqrt{d - \alpha} & -\sqrt{\alpha} \end{pmatrix}.$$

Indeed, we have

$$\begin{aligned} XR &= \begin{bmatrix} X^1 & c\mathbf{1} - X^1 \\ \sqrt{\alpha} & \sqrt{d-\alpha} \end{bmatrix} R \\ &= \begin{bmatrix} X^1 & c\mathbf{1} - X^1 \\ \sqrt{\alpha} & \sqrt{d-\alpha} \end{bmatrix} \frac{1}{\sqrt{d}} \begin{pmatrix} \sqrt{\alpha} & \sqrt{d-\alpha} \\ \sqrt{d-\alpha} & -\sqrt{\alpha} \end{pmatrix} \\ &= \frac{1}{\sqrt{d}} \left[c\mathbf{1} \sqrt{\frac{d-\alpha}{\alpha}} X^1 - \sqrt{\frac{\alpha}{d-\alpha}} (c\mathbf{1} - X^1) \right]. \end{aligned}$$

If we let

$$a = c\sqrt{\frac{d-\alpha}{\alpha}}, \quad b = -c\sqrt{\frac{\alpha}{d-\alpha}},$$

then we check that

$$\alpha a + b(d-\alpha) = 0,$$

which shows that the vector

$$Z = \sqrt{\frac{d-\alpha}{\alpha}} X^1 - \sqrt{\frac{\alpha}{d-\alpha}} (c\mathbf{1} - X^1)$$

is a potential solution of our discrete problem in the sense of Section 4.2. Furthermore, because $L\mathbf{1} = 0$,

$$\text{tr}(X^\top LX) = \text{tr}((XR)^\top L(XR)) = Z^\top LZ,$$

the vector Z is indeed a solution of our discrete problem. Thus, we reconfirm the fact that the second eigenvector of $L_{\text{rw}} = D^{-1}L$ is indeed a continuous approximation to the clustering problem when $K = 2$. This can be generalized for any $K \geq 2$.

Again, we may assume that the nonzero entries in X^1, \dots, X^K are some positive real $c \in \mathbb{R}$, so that

$$X^1 + \dots + X^K = c\mathbf{1},$$

and if (A_1, \dots, A_K) is the partition of V that we are seeking, write $\alpha_j = \text{vol}(A_j)$. We have $\alpha_1 + \dots + \alpha_K = d = \mathbf{1}^\top D\mathbf{1}$. Since

$$(X^j)^\top DX^j = \alpha_j c^2,$$

we normalize the X^j so that $(X^j)^\top DX^j = \dots = (X^K)^\top DX^K = c^2$, and we consider

$$X = \begin{bmatrix} X^1 & X^2 & \dots & X^K \\ \sqrt{\alpha_1} & \sqrt{\alpha_2} & \dots & \sqrt{\alpha_K} \end{bmatrix}.$$

Then, we have the following result.

Proposition 4.2. *If $X = \left[\frac{X^1}{\sqrt{\alpha_1}} \frac{X^2}{\sqrt{\alpha_2}} \cdots \frac{X^K}{\sqrt{\alpha_K}} \right]$ is a solution of our discrete problem, then there is an orthogonal matrix R such that its first column R^1 is*

$$R^1 = \frac{1}{\sqrt{d}} \begin{pmatrix} \sqrt{\alpha_1} \\ \sqrt{\alpha_2} \\ \vdots \\ \sqrt{\alpha_K} \end{pmatrix}$$

and

$$XR = \left[\frac{c}{\sqrt{d}} \mathbf{1} \ Z^2 \ \cdots \ Z^K \right].$$

Furthermore,

$$(XR)^\top D(XR) = c^2 I$$

and

$$\text{tr}((XR)^\top L(XR)) = \text{tr}(Z^\top LZ),$$

with $Z = [Z^2 \ \cdots \ Z^K]$.

Proof. Apply Gram–Schmidt to (R^1, e_2, \dots, e_K) (where (e_1, \dots, e_K) is the canonical basis of \mathbb{R}^K) to form an orthonormal basis. The rest follows from Proposition 4.1. \square

Proposition 4.2 suggests that if $Z = [Z^2 \ \cdots \ Z^K]$ is a solution of the relaxed problem $(*_1)$, then there should be an orthogonal matrix R such that ZR^\top is an approximation of a solution of the discrete problem PNC1.

The next step is to find an exact solution $(\mathbb{P}(X^1), \dots, \mathbb{P}(X^K)) \in \mathbb{P}(\mathcal{K})$ which is the closest (in a suitable sense) to our approximate solution $(Z^1, \dots, Z^K) \in G(K, N)$. The set \mathcal{K} is not necessarily closed under all orthogonal transformations in $\mathbf{O}(K)$, so we can't view \mathcal{K} as a subset of the Grassmannian $G(K, N)$. However, we can think of \mathcal{K} as a subset of $G(K, N)$ by considering the subspace spanned by (X^1, \dots, X^K) for every $[X^1 \ \cdots \ X^K] \in \mathcal{K}$. Then, we have two choices of distances.

1. We view \mathcal{K} as a subset of $(\mathbb{R}\mathbb{P}^{N-1})^K$. Because \mathcal{K} is closed under the antipodal map, as explained in Appendix B, minimizing the distance $d(\mathbb{P}(X^j), \mathbb{P}(Z^j))$ on $\mathbb{R}\mathbb{P}^{N-1}$ is equivalent to minimizing the Euclidean distance $\|X^j - Z^j\|_2$, for $j = 1, \dots, K$ (if we use the Riemannian metric on $\mathbb{R}\mathbb{P}^{N-1}$ induced by the Euclidean metric on \mathbb{R}^N). Then, minimizing the distance $d(X, Z)$ in $(\mathbb{R}\mathbb{P}^{N-1})^K$ is equivalent to minimizing $\|X - Z\|_F$, where

$$\|X - Z\|_F^2 = \sum_{j=1}^K \|X^j - Z^j\|_2^2$$

is the Frobenius norm. This is implicitly the choice made by Yu.

2. We view \mathcal{K} as a subset of the Grassmannian $G(K, N)$. In this case, we need to pick a metric on the Grassmannian, and we minimize the corresponding Riemannian distance $d(X, Z)$. A natural choice is the metric on $\mathfrak{sc}(n)$ given by

$$\langle X, Y \rangle = \text{tr}(X^\top Y).$$

This choice remains to be explored, and will be the subject of a forthcoming report.

4.5 Finding a Discrete Solution Close to a Continuous Approximation

Inspired by Yu [15] and the previous section, given a solution Z_0 of problem $(*_1)$, we look for pairs $(X, R) \in \mathcal{K} \times \mathbf{O}(K)$ (where R is a $K \times K$ orthogonal matrix), with $\|X^j\| = \|Z_0^j\|$ for $j = 1, \dots, K$, that minimize

$$\varphi(X, R) = \|X - Z_0 R\|_F.$$

Here, $\|A\|_F$ is the Frobenius norm of A , with $\|A\|_F^2 = \text{tr}(A^\top A)$.

It may seem desirable to look for discrete solutions $X \in \mathcal{K}$ whose entries are 0 or 1, in which case

$$X \mathbf{1}_K = \mathbf{1}_N.$$

Therefore, we begin by finding a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ such that

$$\|Z_0 \Lambda \mathbf{1}_K - \mathbf{1}_N\|_2$$

is minimal in the least-square sense. As we remarked earlier, since the columns of Z_0 are orthogonal with respect to the inner product $\langle u, v \rangle_D = x^\top D y$, they are linearly independent, thus the pseudo-inverse of Z_0 is $(Z_0^\top Z_0)^{-1} Z_0^\top$, and the best solution $(\lambda_1, \dots, \lambda_K)$ of least Euclidean norm is given by

$$(Z_0^\top Z_0)^{-1} Z_0^\top \mathbf{1}_N.$$

Therefore, we form the (column-rescaled) matrix

$$Z = Z_0 \text{diag}((Z_0^\top Z_0)^{-1} Z_0^\top \mathbf{1}_N).$$

Remark: In Yu [15] and Yu and Shi [16], the rows of Z_0 are normalized by forming the matrix

$$\text{DIAG}(Z_0 Z_0^\top)^{-1/2} Z_0.$$

However, this does not yield a matrix whose columns are obtained from those of Z_0 by rescaling, so the resulting matrix is no longer a rescale of a correct solution of problem $(*_1)$, which seems undesirable.

Actually, even though the columns of Z are D -orthogonal, the matrix Z generally does not satisfy the condition $Z^\top DZ = I$, so ZR may not have D -orthogonal columns (with $R \in \mathbf{O}(K)$), yet $\text{tr}(Z^\top LZ) = \text{tr}((ZR)^\top L(ZR))$ holds! The problem is that the conditions $Z^\top DZ = I$ and $Z\mathbf{1} = \mathbf{1}$ are antagonistic. If we try to force condition $Z\mathbf{1} = \mathbf{1}$, we modify the D -norm of the columns of Z , and then ZR may no longer have D -orthogonal columns. Unless these methods are implemented and tested, it seems almost impossible to tell which option yields the best result. We will proceed under the assumption that Z_0 has been rescaled as explained above, but the method described next also applies if we pick $Z = Z_0$. In this latter case, by Proposition 4.1, Z satisfies the condition $Z^\top DZ = I$, and so does ZR .

The key to minimizing $\|X - ZR\|_F$ rests on the following computation:

$$\begin{aligned} \|X - ZR\|_F^2 &= \text{tr}((X - ZR)^\top (X - ZR)) \\ &= \text{tr}((X^\top - R^\top Z^\top)(X - ZR)) \\ &= \text{tr}(X^\top X - X^\top ZR - R^\top Z^\top X + R^\top Z^\top ZR) \\ &= \text{tr}(X^\top X) - \text{tr}(X^\top ZR) - \text{tr}(R^\top Z^\top X) + \text{tr}(R^\top Z^\top ZR) \\ &= \text{tr}(X^\top X) - \text{tr}((R^\top Z^\top X)^\top) - \text{tr}(R^\top Z^\top X) + \text{tr}(Z^\top ZRR^\top) \\ &= \text{tr}(X^\top X) - 2\text{tr}(R^\top Z^\top X) + \text{tr}(Z^\top Z). \end{aligned}$$

Therefore, minimizing $\|X - ZR\|_F^2$ is equivalent to maximizing $\text{tr}(R^\top Z^\top X)$. This will be done by alternating steps during which we minimize $\varphi(X, R) = \|X - ZR\|_F$ with respect to X holding R fixed, and steps during which we minimize $\varphi(X, R) = \|X - ZR\|_F$ with respect to R holding X fixed. For this second step, we need the following proposition.

Proposition 4.3. *For any $n \times n$ matrix A and any orthogonal matrix Q , we have*

$$\max\{\text{tr}(QA) \mid Q \in \mathbf{O}(n)\} = \sigma_1 + \cdots + \sigma_n,$$

where $\sigma_1 \geq \cdots \geq \sigma_n$ are the singular values of A . Furthermore, this maximum is achieved by $Q = VU^\top$, where $A = U\Sigma V^\top$ is any SVD for A .

Proof. Let $A = U\Sigma V^\top$ be any SVD for A . Then we have

$$\begin{aligned} \text{tr}(QA) &= \text{tr}(QU\Sigma V^\top) \\ &= \text{tr}(V^\top QU\Sigma). \end{aligned}$$

The matrix $Z = V^\top QU$ is an orthogonal matrix so $|z_{ij}| \leq 1$ for $1 \leq i, j \leq n$, and Σ is a diagonal matrix, so we have

$$\text{tr}(Z\Sigma) = z_{11}\sigma_1 + \cdots + z_{nn}\sigma_n \leq \sigma_1 + \cdots + \sigma_n,$$

which proves the first statement of the proposition. For $Q = VU^\top$, we get

$$\begin{aligned} \text{tr}(QA) &= \text{tr}(QU\Sigma V^\top) \\ &= \text{tr}(VU^\top U\Sigma V^\top) \\ &= \text{tr}(V\Sigma V^\top) = \sigma_1 + \cdots + \sigma_n, \end{aligned}$$

which proves the second part of the proposition. □

As a corollary of Proposition 4.3 (with $A = Z^\top X$ and $Q = R^\top$), we get the following result (see Golub and Van Loan [9], Section 12.4.1):

Proposition 4.4. *For any two fixed $N \times K$ matrices X and Z , the minimum of the set*

$$\{\|X - ZR\|_F \mid R \in \mathbf{O}(K)\}$$

is achieved by $R = UV^\top$, for any SVD decomposition $U\Sigma V^\top = Z^\top X$ of $Z^\top X$.

We now deal with step 1. The solutions Z of the relaxed problem $(*_1)$ have columns Z^j of norm ρ_j . Then, for fixed Z and R , we would like to find some $X \in \mathcal{K}$ with $\|X^j\| = \|Z^j\| = \rho_j$ for $j = 1, \dots, K$, so that $\|X - ZR\|_F$ is minimal. Without loss of generality, we may assume that the entries a_1, \dots, a_K occurring in the matrix X are positive. To find $X \in \mathcal{K}$, first we find the shape \widehat{X} of X , which is the matrix obtained from X by rescaling the columns of X so that \widehat{X} has entries $+1, 0$. Then, we rescale the columns of \widehat{X} so that $\|X^j\| = \rho_j$ for $j = 1, \dots, K$.

Since

$$\|X - ZR\|_F^2 = \|X\|_F^2 + \|Z\|_F^2 - \text{tr}(R^\top Z^\top X) = 2 \sum_{j=1}^K \rho_j^2 - \text{tr}(R^\top Z^\top X),$$

minimizing $\|X - ZR\|_F$ is equivalent to maximizing

$$\text{tr}(R^\top Z^\top X) = \text{tr}((ZR)^\top X) = \text{tr}(X(ZR)^\top),$$

and since the i th row of X contains a single nonzero entry, say a_{j_i} (in column j_i , $1 \leq j_i \leq K$), if we write $Y = ZR$, then

$$\text{tr}(XY^\top) = \sum_{i=1}^N a_{j_i} y_{i j_i}. \quad (*)$$

By $(*)$, $\text{tr}(XY^\top)$ is maximized iff $a_{j_i} y_{i j_i}$ is maximized for $i = 1, \dots, N$. Since the a_k are positive, this is achieved if, for the i th row of X , we pick a column index ℓ such that $y_{i\ell}$ is maximum.

Observe that if we change the ρ_j s, minimal solutions for these new values of the ρ_j are obtained by rescaling the a_ℓ 's. Thus, to find the shape \widehat{X} of X , we may assume that $a_\ell = 1$.

Actually, to find the shape \widehat{X} of X , we first find a matrix \overline{X} according to the following method. If we let

$$\begin{aligned} \mu_i &= \max_{1 \leq j \leq K} y_{ij} \\ J_i &= \{j \in \{1, \dots, K\} \mid y_{ij} = \mu_i\}, \end{aligned}$$

for $i = 1, \dots, N$, then

$$\bar{x}_{ij} = \begin{cases} +1 & \text{for some chosen } j \in J_i, \\ 0 & \text{otherwise.} \end{cases}$$

Of course, a single column index is chosen for each row. Unfortunately, the matrix \bar{X} may not be a correct solution, because the above prescription does not guarantee that every column of \bar{X} is nonzero. Therefore, we may have to reassign certain nonzero entries in columns having “many” nonzero entries to zero columns, so that we get a matrix in \mathcal{K} . When we do so, we set the nonzero entry in the column from which it is moved to zero. This new matrix is \hat{X} , and finally we normalize each column of \hat{X} to obtain X , so that $\|X^j\| = \rho_j$, for $j = 1, \dots, K$. This last step may not be necessary since Z was chosen so that $\|Z\mathbf{1}_K - \mathbf{1}_N\|_2$ is minimal. A practical way to deal with zero columns in \bar{X} is to simply decrease K . Clearly, further work is needed to justify the soundness of such a method.

The above method is essentially the method described in Yu [15] and Yu and Shi [16], except that in these works (in which X, Z and Y are denoted by X^*, \tilde{X}^* , and \tilde{X} , respectively) the entries in X belong to $\{0, 1\}$; as described above, for row i , the index ℓ corresponding to the entry $+1$ is given by

$$\arg \max_{1 \leq j \leq K} \tilde{X}(i, j).$$

The fact that \bar{X} may have zero columns is not addressed by Yu. Furthermore, it is important to make sure that each column of X has the same norm as the corresponding column of ZR , but this normalization step is not performed in the above works. On the hand, the rows of Z are normalized, but the resulting matrix may no longer be a correct solution of the relaxed problem. Only a comparison of tests obtained by implementating both methods will reveal which method works best in practice.

The method due to Yu and Shi (see Yu [15] and Yu and Shi [16]) to find $X \in \mathcal{K}$ and $R \in \mathbf{O}(K)$ that minimize $\varphi(X, R) = \|X - ZR\|_F$ is to alternate steps during which either R is held fixed (step PODX) or X is held fixed (step PODR).

- (1) In step PODX, the next discrete solution X^* is obtained fom the previous pair (R^*, Z) by computing \bar{X} and then $X^* = \hat{X}$ from $Y = ZR^*$, as just explained above.
- (2) In step PODR, the next matrix R^* is obtained from the previous pair (X^*, Z) by

$$R^* = UV^\top,$$

for any SVD decomposition $U\Sigma V^\top$ of $Z^\top X^*$.

It remains to initialize R^* to start the process, and then steps (1) and (2) are iterated, starting with step (1). The method advocated by Yu [15] is to pick K rows of Z that are as orthogonal to each other as possible. This corresponds to a K -means clustering strategy with K nearly orthogonal data points as centers. Here is the algorithm given in Yu [15].

Given the $N \times K$ matrix Z (whose columns all have the same norm), we compute a matrix R whose columns are certain rows of Z . We use a vector $c \in \mathbb{R}^N$ to keep track of the inner products of all rows of Z with the columns R^1, \dots, R^{k-1} that have been constructed so far, and initially when $k = 1$, we set $c = 0$.

The first column R^1 of R is any chosen row of Z .

Next, for $k = 2, \dots, K$, we compute all the inner products of R^{k-1} with all rows in Z , which are recorded in the vector ZR^{k-1} , and we update c as follows:

$$c = c + \mathbf{abs}(ZR^{k-1}).$$

We take the absolute values of the entries in ZR^{k-1} so that the i th entry in c is a score of how orthogonal is the i th row of Z to R^1, \dots, R^{k-1} . Then, we choose R^k as any row Z_i of Z for which c_i is minimal (the customary (and ambiguous) $i = \arg \min c$).

Appendix A

The Rayleigh Ratio and the Courant-Fischer Theorem

The most important property of symmetric matrices is that they have real eigenvalues and that they can be diagonalized with respect to an orthogonal matrix. Thus, if A is an $n \times n$ symmetric matrix, then it has n real eigenvalues $\lambda_1, \dots, \lambda_n$ (not necessarily distinct), and there is an orthonormal basis of eigenvectors (u_1, \dots, u_n) (for a proof, see Gallier [6]). Another fact that is used frequently in optimization problem is that the eigenvalues of a symmetric matrix are characterized in terms of what is known as the *Rayleigh ratio*, defined by

$$R(A)(x) = \frac{x^\top Ax}{x^\top x}, \quad x \in \mathbb{R}^n, x \neq 0.$$

The following proposition is often used to prove various optimization or approximation problems (for example PCA).

Proposition A.1. (*Rayleigh–Ritz*) *If A is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and if (u_1, \dots, u_n) is any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i , then*

$$\max_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \lambda_1$$

(with the maximum attained for $x = u_1$), and

$$\max_{x \neq 0, x \in \{u_1, \dots, u_k\}^\perp} \frac{x^\top Ax}{x^\top x} = \lambda_{k+1}$$

(with the maximum attained for $x = u_{k+1}$), where $1 \leq k \leq n - 1$. Equivalently, if V_k is the subspace spanned by (u_k, \dots, u_n) , then

$$\lambda_k = \max_{x \neq 0, x \in V_k} \frac{x^\top Ax}{x^\top x}, \quad k = 1, \dots, n.$$

Proof. First, observe that

$$\max_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \max_x \{x^\top Ax \mid x^\top x = 1\},$$

and similarly,

$$\max_{x \neq 0, x \in \{u_1, \dots, u_k\}^\perp} \frac{x^\top Ax}{x^\top x} = \max_x \{x^\top Ax \mid (x \in \{u_1, \dots, u_k\}^\perp) \wedge (x^\top x = 1)\}.$$

Since A is a symmetric matrix, its eigenvalues are real and it can be diagonalized with respect to an orthonormal basis of eigenvectors, so let (u_1, \dots, u_n) be such a basis. If we write

$$x = \sum_{i=1}^n x_i u_i,$$

a simple computation shows that

$$x^\top Ax = \sum_{i=1}^n \lambda_i x_i^2.$$

If $x^\top x = 1$, then $\sum_{i=1}^n x_i^2 = 1$, and since we assumed that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, we get

$$x^\top Ax = \sum_{i=1}^n \lambda_i x_i^2 \leq \lambda_1 \left(\sum_{i=1}^n x_i^2 \right) = \lambda_1.$$

Thus,

$$\max_x \{x^\top Ax \mid x^\top x = 1\} \leq \lambda_1,$$

and since this maximum is achieved for $e_1 = (1, 0, \dots, 0)$, we conclude that

$$\max_x \{x^\top Ax \mid x^\top x = 1\} = \lambda_1.$$

Next, observe that $x \in \{u_1, \dots, u_k\}^\perp$ and $x^\top x = 1$ iff $x_1 = \dots = x_k = 0$ and $\sum_{i=1}^n x_i^2 = 1$. Consequently, for such an x , we have

$$x^\top Ax = \sum_{i=k+1}^n \lambda_i x_i^2 \leq \lambda_{k+1} \left(\sum_{i=k+1}^n x_i^2 \right) = \lambda_{k+1}.$$

Thus,

$$\max_x \{x^\top Ax \mid (x \in \{u_1, \dots, u_k\}^\perp) \wedge (x^\top x = 1)\} \leq \lambda_{k+1},$$

and since this maximum is achieved for $e_{k+1} = (0, \dots, 0, 1, 0, \dots, 0)$ with a 1 in position $k+1$, we conclude that

$$\max_x \{x^\top Ax \mid (x \in \{u_1, \dots, u_k\}^\perp) \wedge (x^\top x = 1)\} = \lambda_{k+1},$$

as claimed. □

Proposition A.2 is often known as part of the *Rayleigh–Ritz theorem*. For our purposes, we need the version of Proposition A.1 applying to min instead of max, whose proof is obtained by a trivial modification of the proof of Proposition A.1.

Proposition A.2. (*Rayleigh–Ritz*) *If A is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and if (u_1, \dots, u_n) is any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i , then*

$$\min_{x \neq 0} \frac{x^\top A x}{x^\top x} = \lambda_n$$

(with the minimum attained for $x = u_n$), and

$$\min_{x \neq 0, x \in \{u_{i+1}, \dots, u_n\}^\perp} \frac{x^\top A x}{x^\top x} = \lambda_i$$

(with the minimum attained for $x = u_i$), where $1 \leq i \leq n - 1$. Equivalently, if $W_k = V_{k+1}^\perp$ denotes the subspace spanned by (u_1, \dots, u_k) (with $V_{n+1} = (0)$), then

$$\lambda_k = \min_{x \neq 0, x \in W_k} \frac{x^\top A x}{x^\top x} = \min_{x \neq 0, x \in V_{k+1}^\perp} \frac{x^\top A x}{x^\top x}, \quad k = 1, \dots, n.$$

As an application of Propositions A.1 and A.2, we give a proof of a proposition which is the key to the proof of Theorem 3.2. First, we need a definition. Given an $n \times n$ symmetric matrix A and an $m \times m$ symmetric B , with $m \leq n$, if $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of A and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_m$ are the eigenvalues of B , then we say that the eigenvalues of B *interlace* the eigenvalues of A if

$$\lambda_{n-m+i} \leq \mu_i \leq \lambda_i, \quad i = 1, \dots, m.$$

Proposition A.3. *Let A be an $n \times n$ symmetric matrix, R be an $n \times m$ matrix such that $R^\top R = I$ (with $m \leq n$), and let $B = R^\top A R$ (an $m \times m$ matrix). The following properties hold:*

- (a) *The eigenvalues of B interlace the eigenvalues of A .*
- (b) *If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of A and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_m$ are the eigenvalues of B , and if $\lambda_i = \mu_i$, then there is an eigenvector v of B with eigenvalue μ_i such that Rv is an eigenvector of A with eigenvalue λ_i .*

Proof. (a) Let (u_1, \dots, u_n) be an orthonormal basis of eigenvectors for A , and let (v_1, \dots, v_m) be an orthonormal basis of eigenvectors for B . Let U_j be the subspace spanned by (u_1, \dots, u_j) and let V_j be the subspace spanned by (v_1, \dots, v_j) . For any i , the subspace V_i has dimension i and the subspace $R^\top U_{i-1}$ has dimension at most $i - 1$. Therefore, there is some nonzero vector $v \in V_i \cap (R^\top U_{i-1})^\perp$, and since

$$v^\top R^\top u_j = (Rv)^\top u_j = 0, \quad j = 1, \dots, i - 1,$$

we have $Rv \in (U_{i-1})^\perp$. By Proposition A.1 and using the fact that $R^\top R = I$, we have

$$\lambda_i \geq \frac{(Rv)^\top ARv}{(Rv)^\top Rv} = \frac{v^\top Bv}{v^\top v}.$$

On the other hand, by Proposition A.2,

$$\mu_i = \min_{x \neq 0, x \in \{v_{i+1}, \dots, v_n\}^\perp} \frac{x^\top Bx}{x^\top x} = \min_{x \neq 0, x \in \{v_1, \dots, v_i\}} \frac{x^\top Bx}{x^\top x},$$

so

$$\mu_i \leq \frac{w^\top Bw}{w^\top w} \quad \text{for all } w \in V_i,$$

and since $v \in V_i$, we have

$$\mu_i \leq \frac{v^\top Bv}{v^\top v} \leq \lambda_i, \quad i = 1, \dots, m.$$

We can apply the same argument to the symmetric matrices $-A$ and $-B$, to conclude that

$$-\mu_i \leq -\lambda_{n-m+i},$$

that is,

$$\lambda_{n-m+i} \leq \mu_i, \quad i = 1, \dots, m.$$

Therefore,

$$\lambda_{n-m+i} \leq \mu_i \leq \lambda_i, \quad i = 1, \dots, m,$$

as desired.

(b) If $\lambda_i = \mu_i$, then

$$\lambda_i = \frac{(Rv)^\top ARv}{(Rv)^\top Rv} = \frac{v^\top Bv}{v^\top v} = \mu_i,$$

so v must be an eigenvector for B and Rv must be an eigenvector for A , both for the eigenvalue $\lambda_i = \mu_i$. \square

Observe that Proposition A.3 implies that

$$\lambda_n + \dots + \lambda_{n-m+1} \leq \text{tr}(R^\top AR) \leq \lambda_1 + \dots + \lambda_m.$$

The left inequality is used to prove Theorem 3.2.

For the sake of completeness, we also prove the Courant–Fischer characterization of the eigenvalues of a symmetric matrix.

Theorem A.4. (*Courant–Fischer*) Let A be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and let (u_1, \dots, u_n) be any orthonormal basis of eigenvectors of A , where u_i is a unit eigenvector associated with λ_i . If \mathcal{V}_k denotes the set of subspaces of \mathbb{R}^n of dimension k , then

$$\lambda_k = \max_{W \in \mathcal{V}_k} \min_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}$$

$$\lambda_k = \min_{W \in \mathcal{V}_{n-k+1}} \max_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}.$$

Proof. Let us consider the second equality, the proof of the first equality being similar. Observe that the space V_k spanned by (u_k, \dots, u_n) has dimension $n-k+1$, and by Proposition A.1, we have

$$\lambda_k = \max_{x \neq 0, x \in V_k} \frac{x^\top A x}{x^\top x} \geq \min_{W \in \mathcal{V}_{n-k+1}} \max_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}.$$

Therefore, we need to prove the reverse inequality; that is, we have to show that

$$\lambda_k \leq \max_{x \neq 0, x \in W} \frac{x^\top A x}{x^\top x}, \quad \text{for all } W \in \mathcal{V}_{n-k+1}.$$

Now, for any $W \in \mathcal{V}_{n-k+1}$, if we can prove that $W \cap V_{k+1}^\perp \neq (0)$, then for any nonzero $v \in W \cap V_{k+1}^\perp$, by Proposition A.2, we have

$$\lambda_k = \min_{x \neq 0, x \in V_{k+1}^\perp} \frac{x^\top A x}{x^\top x} \leq \frac{v^\top A v}{v^\top v} \leq \max_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}.$$

It remains to prove that $\dim(W \cap V_{k+1}^\perp) \geq 1$. However, $\dim(V_{k+1}) = n-k$, so $\dim(V_{k+1}^\perp) = k$, and by hypothesis $\dim(W) = n-k+1$. By the Grassmann relation,

$$\dim(W) + \dim(V_{k+1}^\perp) = \dim(W \cap V_{k+1}^\perp) + \dim(W + V_{k+1}^\perp),$$

and since $\dim(W + V_{k+1}^\perp) \leq \dim(\mathbb{R}^n) = n$, we get

$$n - k + 1 + k \leq \dim(W \cap V_{k+1}^\perp) + n;$$

that is, $1 \leq \dim(W \cap V_{k+1}^\perp)$, as claimed. \square

Appendix B

Riemannian Metrics on Quotient Manifolds

In order to define a metric on the projective space $\mathbb{R}P^n$, we need to review a few notions of differential geometry. First, we need to define the quotient M/G of a manifold by a group acting on M . This section relies heavily on Gallot, Hulin, Lafontaine [7] and Lee [10], which contain thorough expositions and should be consulted for details.

Definition B.1. Recall that an *action* of a group G (with identity element 1) on a set X is a map $\gamma: G \times X \rightarrow X$ satisfying the following properties:

- (1) $\gamma(1, x) = x$, for all $x \in X$.
- (2) $\gamma(g_1, \gamma(g_2, x)) = \gamma(g_1 g_2, x)$, for all $g_1, g_2 \in G$, and all $x \in X$.

We usually abbreviate $\gamma(g, x)$ by $g \cdot x$.

If X is a topological space and G is a topological group, we say that the action is *continuous* iff the map γ is continuous. In this case, for every $g \in G$, the map $x \mapsto g \cdot x$ is a homeomorphism. If X is a smooth manifold and G is a Lie group, we say that the action is *smooth* iff the map γ is smooth. In this case, for every $g \in G$, the map $x \mapsto g \cdot x$ is a diffeomorphism.

Remark: To be more precise, what we have defined in Definition B.1 is a *left action* of the group G on the set X . There is also a notion of a *right action*, but we won't need it.

The *quotient of X by G* , denoted X/G , is the set of orbits of G ; that is, the set of equivalence classes of the equivalence relation \simeq defined such that, for any $x, y \in X$,

$$x \simeq y \quad \text{iff} \quad (\exists g \in G)(y = g \cdot x).$$

The *orbit* of $x \in X$ (the equivalence class of x) is the set

$$O_x = \{g \cdot x \mid g \in G\},$$

also denoted by $G \cdot x$. If X is a topological space, we give X/G the quotient topology.

For any subset V of X and for any $g \in G$, we denote by gV the set

$$gV = \{g \cdot x \mid x \in V\}.$$

One problem is that even if X is Hausdorff, X/G may not be. Thus, we need to find conditions to ensure that X/G is Hausdorff.

By a *discrete group*, we mean a group equipped with the discrete topology (every subset is open). In other words, we don't care about the topology of G ! The following conditions prove to be useful.

Definition B.2. Let $\cdot : G \times X \rightarrow X$ be the action of a group G on a set X . We say that G acts *freely* (or that the action is *free*) iff for all $x \in X$ and all $g \in G$, if $g \neq 1$, then $g \cdot x \neq x$.

If X is a locally compact space and G is a discrete group acting continuously on X , we say that G acts *properly* (or that the action is *proper*) iff

- (i) For every $x \in X$, there is some open subset V with $x \in V$ such that $gV \cap V \neq \emptyset$ for only finitely many $g \in G$.
- (ii) For all $x, y \in X$, if $y \notin G \cdot x$ (y is not in the orbit of x), then there exist some open sets V, W with $x \in V$ and $y \in W$ such that $gV \cap W = \emptyset$ for all $g \in G$.

The following proposition gives necessary and sufficient conditions for a discrete group to act freely and properly often found in the literature (for instance, O'Neill [11], Berger and Gostiaux [2], and do Carmo [4], but beware that in this last reference Hausdorff separation is not required!).

Proposition B.1. *If X is a locally compact space and G is a discrete group, then a smooth action of G on M is free and proper iff the following conditions hold:*

- (i) *For every $x \in X$, there is some open subset V with $x \in V$ such that $gV \cap V = \emptyset$ for all $g \in G$ such that $g \neq 1$.*
- (ii) *For all $x, y \in X$, if $y \notin G \cdot x$ (y is not in the orbit of x), then there exist some open sets V, W with $x \in V$ and $y \in W$ such that $gV \cap W = \emptyset$ for all $g \in G$.*

Proof. Condition (i) of Proposition B.1 implies condition (i) of Definition B.2, and condition (ii) is the same in Proposition B.1 and Definition B.2. If (i) holds, then the action must be free since if $g \cdot x = x$, then $gV \cap V \neq \emptyset$, which implies that $g = 1$.

Conversely, we just have to prove that the conditions of Definition B.2 imply condition (i) of Proposition B.1. By (i) of Definition B.2, there is some open subset U containing x and a finite number of elements of G , say g_1, \dots, g_m , with $g_i \neq 1$, such that

$$g_i U \cap U \neq \emptyset, \quad i = 1, \dots, m.$$

Since our action is free and $g_i \neq 1$, we have $g_i \cdot x \neq x$, so by Hausdorff separation, there exist some open subsets W_i, W'_i , with $x \in W_i$ and $g_i \cdot x \in W'_i$, such that $W_i \cap W'_i = \emptyset$, $i = 1, \dots, m$. Then, if we let

$$V = W \cap \left(\bigcap_{i=1}^m (W_i \cap g_i^{-1} W'_i) \right),$$

we see that $V \cap g_i V = \emptyset$, and since $V \subseteq W$, we also have $V \cap gV = \emptyset$ for all other $g \in G$. \square

Remark: The action of a discrete group satisfying the properties of Proposition B.1 is often called “properly discontinuous.” However, as pointed out by Lee ([10], just before Proposition 9.18), this term is self-contradictory since such actions are smooth, and thus continuous!

We also need covering maps.

Definition B.3. Let X and Y be two topological spaces. A map $\pi: X \rightarrow Y$ is a *covering map* iff the following conditions hold:

- (1) The map π is continuous and surjective.
- (2) For every $y \in Y$, there is some open subset $W \subseteq Y$ with $y \in W$, such that

$$\pi^{-1}(W) = \bigcup_{i \in I} U_i,$$

where the $U_i \subseteq X$ are pairwise disjoint open subsets such that the restriction of π to U_i is a homeomorphism for every $i \in I$.

If X and Y are smooth manifolds, we assume that π is smooth and that the restriction of π to each U_i is a diffeomorphism.

Then, we have the following useful result.

Theorem B.2. *Let M be a smooth manifold and let G be discrete group acting smoothly, freely and properly on M . Then there is a unique structure of smooth manifold on M/G such that the projection map $\pi: M \rightarrow M/G$ is a covering map.*

For a proof, see Gallot, Hulin, Lafontaine [7] (Theorem 1.88) or Lee [10] (Theorem 9.19).

Real projective spaces are illustrations of Theorem B.2. Indeed, if M is the unit n -sphere $S^n \subseteq \mathbb{R}^{n+1}$ and $G = \{I, -I\}$, where $-I$ is the antipodal map, then the conditions of Proposition B.1 are easily checked (since S^n is compact), and consequently the quotient

$$\mathbb{R}\mathbb{P}^n = S^n/G$$

is a smooth manifold and the projection map $\pi: S^n \rightarrow \mathbb{R}\mathbb{P}^n$ is a covering map. The fiber $\pi^{-1}([x])$ of every point $[x] \in \mathbb{R}\mathbb{P}^n$ consists of two antipodal points: $x, -x \in S^n$.

The next step is see how a Riemannian metric on M induces a Riemannian metric on the quotient manifold M/G .

Definition B.4. Given any two Riemannian manifolds (M, g) and (N, h) a smooth map $f: M \rightarrow N$ is a *local isometry* iff for all $p \in M$, the tangent map $df_p: T_p M \rightarrow T_{f(p)} N$ is an orthogonal transformation of the Euclidean spaces $(T_p M, g_p)$ and $(T_{f(p)} N, h_{f(p)})$. Furthermore, if f is a diffeomorphism, we say that f is an *isometry*.

The Riemannian version of a covering map is the following:

Definition B.5. Let (M, g) and (N, h) be two Riemannian manifolds. A map $\pi: M \rightarrow N$ is a *Riemannian covering map* iff the following conditions hold:

- (1) The map π is a smooth covering.
- (2) The map π is a local isometry.

The following theorem is the Riemannian version of Theorem B.2.

Theorem B.3. Let (M, h) be a Riemannian manifold and let G be discrete group acting smoothly, freely and properly on M , and such that the map $x \mapsto \sigma \cdot x$ is an isometry for all $\sigma \in G$. Then there is a unique structure of Riemannian manifold on $N = M/G$ such that the projection map $\pi: M \rightarrow M/G$ is a Riemannian covering map.

Proof sketch. For a complete proof see Gallot, Hulin, Lafontaine [7] (Proposition 2.20). To define a Riemannian metric g on $N = M/G$ we need to define an inner product g_p on the tangent space $T_p N$ for every $p \in N$. Pick any $q_1 \in \pi^{-1}(p)$ in the fibre of p . Because π is a Riemannian covering map, it is a local diffeomorphism, and thus $d\pi_{q_1}: T_{q_1} M \rightarrow T_p M$ is an isometry. Then, given any two tangent vectors $u, v \in T_p N$, we define their inner product $g_p(u, v)$ by

$$g_p(u, v) = h_{q_1}(d\pi_{q_1}^{-1}(u), d\pi_{q_1}^{-1}(v)).$$

Now, we need to show that g_p does not depend on the choice of $q_1 \in \pi^{-1}(p)$. So, let $q_2 \in \pi^{-1}(p)$ be any other point in the fibre of p . By definition of $N = M/G$, we have $q_2 = g \cdot q_1$ for some $g \in G$, and we know that the map $f: q \mapsto g \cdot q$ is an isometry of M . Now, since $\pi = \pi \circ f$ we have

$$d\pi_{q_1} = d\pi_{q_2} \circ df_{q_1},$$

and since $d\pi_{q_1}: T_{q_1} M \rightarrow T_p M$ and $d\pi_{q_2}: T_{q_2} M \rightarrow T_p M$ are isometries, we get

$$d\pi_{q_2}^{-1} = df_{q_1} \circ d\pi_{q_1}^{-1}.$$

But $df_{q_1}: T_{q_1} M \rightarrow T_{q_2} M$ is also an isometry, so

$$h_{q_2}(d\pi_{q_2}^{-1}(u), d\pi_{q_2}^{-1}(v)) = h_{q_2}(df_{q_1}(d\pi_{q_1}^{-1}(u)), df_{q_1}(d\pi_{q_1}^{-1}(v))) = h_{q_1}(d\pi_{q_1}^{-1}(u), d\pi_{q_1}^{-1}(v)).$$

Therefore, the inner product g_p is well defined on $T_p N$. □

Theorem B.3 implies that every Riemannian metric g on the sphere S^n induces a Riemannian metric \widehat{g} on the projective space $\mathbb{R}\mathbb{P}^n$, in such a way that the projection $\pi: S^n \rightarrow \mathbb{R}\mathbb{P}^n$ is a Riemannian covering. In particular, if U is an open hemisphere obtained by removing its boundary S^{n-1} from a closed hemisphere, then π is an isometry between U and its image $\mathbb{R}\mathbb{P}^n - \pi(S^{n-1}) \approx \mathbb{R}\mathbb{P}^n - \mathbb{R}\mathbb{P}^{n-1}$.

We also observe that for any two points $p = [x]$ and $q = [y]$ in $\mathbb{R}\mathbb{P}^n$, where $x, y \in S^n$, if $x \cdot y = \cos \theta$, with $0 \leq \theta \leq \pi$, then there are two possibilities:

1. $x \cdot y \geq 0$, which means that $0 \leq \theta \leq \pi/2$, or
2. $x \cdot y < 0$, which means that $\pi/2 < \theta \leq \pi$.

In the second case, since $[-y] = [y]$ and $x \cdot (-y) = -x \cdot y$, we can replace the representative y of q by $-y$, and we have $x \cdot (-y) = \cos(\pi - \theta)$, with $0 \leq \pi - \theta < \pi/2$. Therefore, in all cases, for any two points $p, q \in \mathbb{R}\mathbb{P}^n$, we can find an open hemisphere U such that $p = [x], q = [y], x, y \in U$, and $x \cdot y \geq 0$; that is, the angle $\theta \geq 0$ between x and y is at most $\pi/2$. This fact together with the following simple proposition will allow us to figure out the distance (in the sense of Riemannian geometry) between two points in $\mathbb{R}\mathbb{P}^n$.

Proposition B.4. *Let $\pi: M \rightarrow N$ be a Riemannian covering map between two Riemannian manifolds (M, g) and (N, h) . Then, the geodesics of (N, h) are the projections of geodesics in (M, g) (i.e., curves $\pi \circ \gamma$ in (N, h) , where γ is a geodesic in (M, g)), and the geodesics of (M, g) are the liftings of geodesics in (N, h) (i.e., curves γ of (M, g) , such that $\pi \circ \gamma$ is a geodesic in (N, h)).*

The proof of Proposition B.4 can be found in Gallot, Hulin, Lafontaine [7] (Proposition 2.81).

Now, if (M, g) is a connected Riemannian manifold, recall that we define the distance $d(p, q)$ between two points $p, q \in M$ as

$$d(p, q) = \inf\{L(\gamma) \mid \gamma: [0, 1] \rightarrow M\},$$

where γ is any piecewise C^1 -curve from p to q , and

$$L(\gamma) = \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))} dt$$

is the length of γ . It is well known that d is a metric on M . The Hopf-Rinow Theorem (see Gallot, Hulin, Lafontaine [7], Theorem 2.103) says among other things that (M, g) is geodesically complete (which means that every geodesic γ of M can be extended to a geodesic $\tilde{\gamma}$ defined on all of \mathbb{R}) iff any two points of M can be joined by a minimal geodesic iff (M, d) is a complete metric space. Therefore, in a complete (connected) manifold

$$d(p, q) = \inf\{L(\gamma) \mid \gamma: [0, 1] \rightarrow M \text{ is a geodesic}\}.$$

In particular, compact manifolds are complete, so the distance between two points is the infimum of the length of minimal geodesics joining these points.

Applying this to \mathbb{RP}^n and the canonical Euclidean metric induced by \mathbb{R}^{n+1} , since geodesics of S^n are great circles, by the discussion above, for any two points $p = [x]$ and $q = [y]$ in \mathbb{RP}^n , with $x, y \in S^n$, the distance between them is given by

$$d(p, q) = d([x], [y]) = \begin{cases} \cos^{-1}(x \cdot y) & \text{if } x \cdot y \geq 0 \\ \cos^{-1}(-x \cdot y) & \text{if } x \cdot y < 0. \end{cases}$$

Here $\cos^{-1}(z) = \arccos(z)$ is the unique angle $\theta \in [0, \pi]$ such that $\cos(\theta) = z$. Equivalently,

$$d([x], [y]) = \cos^{-1}(|x \cdot y|),$$

and

$$d([x], [y]) = \min\{\cos^{-1}(x \cdot y), \pi - \cos^{-1}(x \cdot y)\}.$$

If the representatives $x, y \in \mathbb{R}^{n+1}$ of $p = [x]$ and $q = [y]$ are not unit vectors, then

$$d([x], [y]) = \cos^{-1}\left(\frac{|x \cdot y|}{\|x\| \|y\|}\right).$$

Note that $0 \leq d(p, q) \leq \pi/2$.

Now, the Euclidean distance between x and y on S^n is given by

$$\|x - y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2x \cdot y = 2 - 2\cos\theta = 4\sin^2(\theta/2).$$

Thus,

$$\|x - y\|_2 = 2\sin(\theta/2), \quad 0 \leq \theta \leq \pi.$$

It follows that for any $x \in S^n$, and for any subset $A \subseteq S^n$, a point $a \in A$ minimizes the distance $d_{S^n}(x, a) = \cos^{-1}(x \cdot a) = \theta$ on S^n iff it minimizes the Euclidean distance $\|x - a\|_2 = 2\sin(\theta/2)$ (since $0 \leq \theta \leq \pi$). Then, on \mathbb{RP}^n , for any point $p = [x] \in \mathbb{RP}^n$ and any $A \subseteq \mathbb{RP}^n$, a point $[a] \in A$ minimizes the distance $d([x], [a])$ on \mathbb{RP}^n iff it minimizes $\min\{\|x - a\|_2, \|x + a\|_2\}$. So, we are looking for $[b] \in A$ such that

$$\begin{aligned} \min\{\|x - b\|_2, \|x + b\|_2\} &= \min_{[a] \in A} \min\{\|x - a\|_2, \|x + a\|_2\} \\ &= \min\left\{\min_{[a] \in A} \|x - a\|_2, \min_{[a] \in A} \|x + a\|_2\right\}. \end{aligned}$$

If the subset $A \subseteq S^n$ is closed under the antipodal map (which means that if $x \in A$, then $-x \in A$), then finding $\min_{a \in A} d([x], [a])$ on \mathbb{RP}^n is equivalent to finding $\min_{a \in A} \|x - a\|_2$, the minimum of the Euclidean distance. This is the case for the set \mathcal{X} in Section 4.2 and the set \mathcal{K} in Section 4.3.

Acknowledgments: First, it must be said that the seminal and highly original work of Jianbo Shi and Stella Yu on normalized cuts, was the source of inspiration for this document. I also

wish to thank Katerina Fragkiadaki for pointing out a number of mistakes in an earlier version of this paper. Roberto Tron also made several suggestions that contributed to improving this report. Katerina, Ryan Kennedy, Andrew Yeager, and Weiyu Zhang made many useful comments and suggestions. Thanks to Dan Spielman for making available his lovely survey on spectral graph theory, and to Kostas for giving me the opportunity to hold hostage a number of people for three Friday afternoons in a row.

Bibliography

- [1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*, pages 585–591. MIT Press, 2001.
- [2] Marcel Berger and Bernard Gostiaux. *Géométrie différentielle: variétés, courbes et surfaces*. Collection Mathématiques. Puf, second edition, 1992. English edition: Differential geometry, manifolds, curves, and surfaces, GTM No. 115, Springer Verlag.
- [3] Fan R. K. Chung. *Spectral Graph Theory*, volume 92 of *Regional Conference Series in Mathematics*. AMS, first edition, 1997.
- [4] Manfredo P. do Carmo. *Riemannian Geometry*. Birkhäuser, second edition, 1992.
- [5] Jean H. Gallier. *Discrete Mathematics*. Universitext. Springer Verlag, first edition, 2011.
- [6] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, second edition, 2011.
- [7] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Universitext. Springer Verlag, second edition, 1993.
- [8] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. GTM No. 207. Springer Verlag, first edition, 2001.
- [9] H. Golub, Gene and F. Van Loan, Charles. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.
- [10] John M. Lee. *Introduction to Smooth Manifolds*. GTM No. 218. Springer Verlag, first edition, 2006.
- [11] Barrett O’Neill. *Semi-Riemannian Geometry With Applications to Relativity*. Pure and Applied Math., Vol 103. Academic Press, first edition, 1983.
- [12] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

- [13] Daniel Spielman. Spectral graph theory. In Uwe Naumann and Olaf Schenk, editors, *Combinatorial Scientific Computing*. CRC Press, 2012.
- [14] von Luxburg Ulrike. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.
- [15] Stella X. Yu. *Computational Models of Perceptual Organization*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA 15213, USA, 2003. Dissertation.
- [16] Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *9th International Conference on Computer Vision, Nice, France, October 13-16*. IEEE, 2003.