

Spectral graph theory: Cheeger constants and discrepancy*

Steve Butler

September 2006

Abstract

In this third talk we will discuss properties related to edge expansion. In particular, we will define the Cheeger constant (which measures how easy it is to cut off a large piece of the graph) and state the Cheeger inequalities. We also will define and discuss discrepancy for undirected and directed graphs. We also state the Perron-Frobenius Theorem which is a useful tool in spectral graph theory, particularly for directed graphs.

1 Introduction

In the first talk we introduced the common matrices whose spectrum is used for spectral graph theory, and in the second talk we defined the Rayleigh quotient and used the Courant-Fischer Theorem to give a way of finding the eigenvalues of the normalized Laplacian. In this talk we will examine properties of the edges.

Our first question that we will be looking at with edges is how easy it is to trim a few edges to break off a large portion of the graph. This question has applications to communication networks in that it can measure how stable a network is, i.e., how many communication failures must occur before a large portion of the network is isolated.

The second question that we will be looking at is how well the edges are placed in a graph, that is given two subsets X and Y we compare the actual and expected edges between the two sets. The closer the actual and expected edges agree the more “random-like” the graph is in the way that sets “expand”. This has applications in proving structure theorems about a graph, for instance when discrepancy is sufficiently small for almost-regular graphs then the graph has a Hamiltonian cycle.

*Third of three talks given at the *Center for Combinatorics*, Nankai University, Tianjin

2 The Cheeger constant

In this section we will consider the problem of separating the graph into two large components by making a small cut. A cut typically involves the removal of edges and/or vertices of the graph. For our problem we will consider edge-cuts (which can be easily related to machinery that we have already set up). The goal is to find an optimal cut, or at least an estimation of what an optimal cut should be,

Given a subset U of the vertices the number of edges needed to separate U from the remaining vertices (which we denote \bar{U}) will be $e(U, \bar{U})$. That is, $e(U, \bar{U})$ will equal the number of edges with its first vertex in U and second vertex in \bar{U} . We also want a measurement of the size of U and for this we will use the volume. We define the volume of U by $\text{vol}(U) = \sum_{v \in U} d_v$.

Comment. The choice of how to measure the size of a subset U can make very large differences in the approach needed to take. For instance if we use $|U|$, i.e., the number of vertices of U , then the combinatorial Laplacian is the better tool to use. For the normalized Laplacian the normalization can be thought of as a redistribution of the weights moving from equally weighted vertices to weighing vertices by their degree. This can have important consequences.

Comment. The term $e(U, \bar{U})$ can also be looked at as the size of the “edge boundary”. The edge boundary of a set U is the set of all edges with one endpoint in U and one endpoint not in U . Alternatively we could use the vertex boundary of a set U , which is the set of all vertices adjacent to U but not in U . Different types of boundaries will give you different type of cuts and lead to different results. We are working with edge boundary in this case because of the nice result that we can state.

Given a subset U one measurement of the cut to remove U from the graph is

$$\frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}.$$

Note that if we need fewer edges or remove a larger portion of the graph then this ratio gets smaller. So intuitively a good cut will be associated with a smaller value. We will define the Cheeger constant of the graph G , denoted h_G , as the minimum ratio achievable, i.e.,

$$h_G = \min_{\emptyset \subset U \subset V} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}. \quad (1)$$

Note by the definition if we start with a small subset U (small in the sense that $\text{vol}(U) \leq \frac{1}{2} \text{vol}(G)$ ($= \frac{1}{2} \text{vol}(V)$)) then the number of edges needed to cut U from the rest of the graph is at least $h_G \text{vol}(U)$, while there exists some small subset which will take exactly $h_G \text{vol}(U)$ edges to remove U from the remainder of the graph.

The Cheeger inequalities

Looking at the definition of the Cheeger inequality in (1) we see that in order to compute the constant we would have to consider approximately $2^{|G|-1}$ subsets of the vertices. In particular, finding the Cheeger constant for a general graph is nontrivial, especially as the number of vertices gets large.

However, even though we may not be able to compute the Cheeger constant directly we can still get a good estimate for it by using the eigenvalues of the normalized Laplacian. In particular, we have the following result.

Theorem 1. *Let $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ be the eigenvalues of the normalized Laplacian and h_G be the Cheeger constant of G . Then*

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

In particular, if we can have control of the first nontrivial eigenvalue then we have control of the Cheeger constant. On the other hand, for some graphs we can get an estimate for the size of the Cheeger constant and in such a setting we also can get control of the size of the first nontrivial eigenvalue.

We will prove $2h_G \geq \lambda_1$ and give a brief outline of $\lambda_1 \geq h_G^2/2$. So suppose that we have found an optimum U that will give the minimum in the definition of the Cheeger constant. Then define a vector \hat{y} by

$$\hat{y}_i = \begin{cases} \frac{1}{\text{vol}(U)} & \text{if } i \in U; \\ -\frac{1}{\text{vol}(\bar{U})} & \text{if } i \in \bar{U}. \end{cases}$$

A simple calculation will show that $\hat{y} \perp D\mathbf{1}$ so that

$$\begin{aligned} \lambda_1 &= \min_{y \perp D\mathbf{1}, y \neq 0} \frac{\sum_{i \sim j} (y_i - y_j)^2}{\sum_i y_i^2 d_i} \\ &\leq \frac{\sum_{i \sim j} (\hat{y}_i - \hat{y}_j)^2}{\sum_i \hat{y}_i^2 d_i} \\ &= e(U, \bar{U}) \left(\frac{1}{\text{vol}(U)} + \frac{1}{\text{vol}(\bar{U})} \right) \\ &\leq 2 \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}} = 2h_G. \end{aligned}$$

For the other inequality the idea is to use the eigenvector y for λ_1 to create some special cuts. This is done by taking the eigenvector for λ_1 and arranging the vertices

as v_1, v_2, \dots, v_n so that the corresponding entries in y are nondecreasing. We can then define cuts C_i by letting $U_i = \{v_1, v_2, \dots, v_i\}$, see Figure 1.

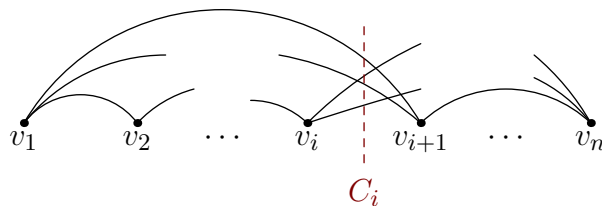


Figure 1: Using an eigenvector to find a good cut.

Taking the best among these cuts gives an upper bound for the Cheeger constant, at the same time because we used the eigenvector to define the cuts it can also be related back to λ_1 . [This last step is nontrivial, but can be found in the literature.]

Quasirandom graphs

Above we saw that we could use something which is easy to compute (i.e., the first eigenvalue which can be computed in polynomial time) to bound the behavior of something which is difficult to compute (i.e., the Cheeger constant which can take exponential time to compute). This is a major theme of graph theory and, in particular, spectral graph theory.

This is the power of quasirandom graphs which was first developed twenty years ago but has gotten increasing attention in the last five years. The basic idea is that there is a list of graph properties which we call quasirandom (the idea behind the name being that these properties are often ones we would associate with a random graph), such that if a graph (or more generally, a family of graphs) satisfies any single property then it will satisfy every property.

One example of a quasirandom property is that the graph has roughly $n^2/4$ edges and roughly $n^4/16$ four-cycles. Another property is that *each* possible subgraph occurs as an induced subgraph approximately the right number of times. The surprising result is that if the first statement is true then the second statement must also be true (the reverse direction also holds but is not as surprising).

3 Edges between sets (Discrepancy)

One important property of graphs is how the edges expand. That is, starting with an initial vertex we look at its neighbors, then the neighbors of the neighbors and so on. In a random graph model we have a fairly good control on expansion, but what about a specific graph?

One way to approach this problem is to look at how the edges are distributed between sets, and in particular is this distribution somehow “randomish”. So let X and Y be two subsets of the vertices of the graph. The number of edges that start in X and end in Y will be denoted $e(X, Y)$ (by convention any edge in $X \cap Y$ will be counted twice).

To estimate the number of edges that should be between X and Y we will use $\text{vol}(X)\text{vol}(Y)/\text{vol}(G)$. Underlying this is a random graph model where the probability of joining vertices u and v is $d_u d_v / \text{vol} G$. This random graph model works well for graphs with uneven degree distributions. In particular, note that the expected degree of vertex v is d_v .

We are now ready to define the discrepancy of a graph, denoted $\text{disc}(G)$, as the minimal α so that

$$\left| \underbrace{e(X, Y)}_{\text{actual edges}} - \underbrace{\frac{\text{vol}(X)\text{vol}(Y)}{\text{vol}(G)}}_{\text{expected edges}} \right| \leq \alpha \underbrace{\sqrt{\text{vol}(X)\text{vol}(Y)}}_{\text{error normalizing term}}.$$

The smaller the value of α is the more random-like our edges have been placed. Not surprisingly, there is a tie between the discrepancy of a graph and the spectrum. We have the following result.

Theorem 2. *Let G be a graph with $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ the eigenvalues of the normalized Laplacian. Then*

$$\text{disc}(G) \leq \max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\} \leq 150 \text{disc}(G)(1 - 8 \log(\text{disc}(G))).$$

So if the nontrivial eigenvalues cluster close to 1 (as random graphs do) then the graph has good expansion property. Also note that if we have control on the expansion of the graph then we also have control of the eigenvalues. Again we will prove half of the above result and give a brief outline of the other half.

To start we define indicator functions for subsets X by

$$(\psi_X)_i = \begin{cases} 1 & \text{if } i \in X; \\ 0 & \text{if } i \notin X. \end{cases}$$

With this convention it is easy to check that $e(X, Y) = \psi_X^T A \psi_Y$ (where A is again the adjacency matrix), while $\text{vol}(X)\text{vol}(Y) = \psi_X^T D \mathbf{1} \mathbf{1}^T D \psi_Y = \psi_X^T D J D \psi_Y$ (where $\mathbf{1}$ is the all 1s vector and J is the all 1s matrix).

We now have the following

$$\begin{aligned} \left| e(X, Y) - \frac{\text{vol}(X)\text{vol}(Y)}{\text{vol}(G)} \right| &= \left| \psi_X^T A \psi_Y - \frac{\psi_X^T D \mathbf{1} \mathbf{1}^T D \psi_Y}{\text{vol}(G)} \right| \\ &= \left| \psi_X^T D^{1/2} \left(D^{-1/2} A D^{-1/2} - \frac{D^{1/2} \mathbf{1} \mathbf{1}^T D^{1/2}}{\text{vol}(G)} \right) D^{1/2} \psi_Y \right|. \end{aligned}$$

The matrix in the middle may look familiar, it is similar to the matrix that we saw in the first lecture as it relates to random walks. In particular $D^{-1/2}AD^{-1/2}$ has an eigenvalue of 1 that corresponds to the eigenvector $D^{1/2}\mathbf{1}/\sqrt{\text{vol}(G)}$, so the matrix above corresponds to the matrix $D^{-1/2}AD^{-1/2}$ with the eigenvalue of 1 replaced by 0, the remaining eigenvalues are $1 - \lambda_{n-1} \leq \dots \leq 1 - \lambda_1$.

Our final tool is that for M a real symmetric matrix with eigenvalues $\theta_1, \dots, \theta_n$ and x, y vectors $|x^T My| \leq \max_i |\theta_i| \|x\| \|y\|$, where $\|\cdot\|$ represents the 2-norm of a vector (as seen in the first lecture). More about this will be given below.

Combining all of the above with some simple calculations, we now have

$$\begin{aligned} \left| e(X, Y) - \frac{\text{vol}(X) \text{vol}(Y)}{\text{vol}(G)} \right| &\leq (\max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}) \|D^{1/2}\psi_X\| \|D^{1/2}\psi_Y\| \\ &= (\max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}) \sqrt{\text{vol}(X) \text{vol}(Y)}. \end{aligned}$$

Thus we have shown that $\text{disc}(G) \leq \max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}$.

For the other direction the idea is to start with the eigenvector x corresponding to the eigenvalue $\max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}$ and then estimate it by $\sum_i D^{1/2}y^{(i)}$ where each $y^{(i)}$ has a simple form. The key is that the vectors in this estimation are closely related to discrepancy for which we have control. Thus we are able to get control of $\max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}$ by using discrepancy.

Here symmetry was not an important tool, and this is a case where we can generalize the result to directed graphs. But before we do that we should first talk about singular values and the Perron-Frobenius Theorem.

Singular values and the Perron-Frobenius Theorem

As mentioned in the first lecture we started with real symmetric matrices since such matrices are nice, i.e., have real eigenvalues and a complete set of orthonormal eigenvectors. If we drop symmetry (which is a natural condition when dealing with directed graphs), then our eigenvalues may no longer be real, and moreover we may not even have a complete set of eigenvectors (let alone being orthonormal!). If we drop the assumption of being square the situation becomes even worse in that we no longer have any eigenvalues or eigenvectors.

In such settings the singular values of the matrix become more reasonable to work with. For example, *any* matrix will always have a complete set of singular values and they will always be nonnegative. Further, there are orthonormal vectors on the left and the right associated with the singular values.

Intuitively, a matrix is a linear transformation from one space to another. Because it is linear it will map a unit ball to an ellipsoid. Further, the orthonormal directions that run in the direction of the axes of the ellipsoid come from an orthonormal set of

vectors in the unit ball. The singular values are the lengths of the axes of the ellipsoid (given in order from largest to smallest, see Figure 2).

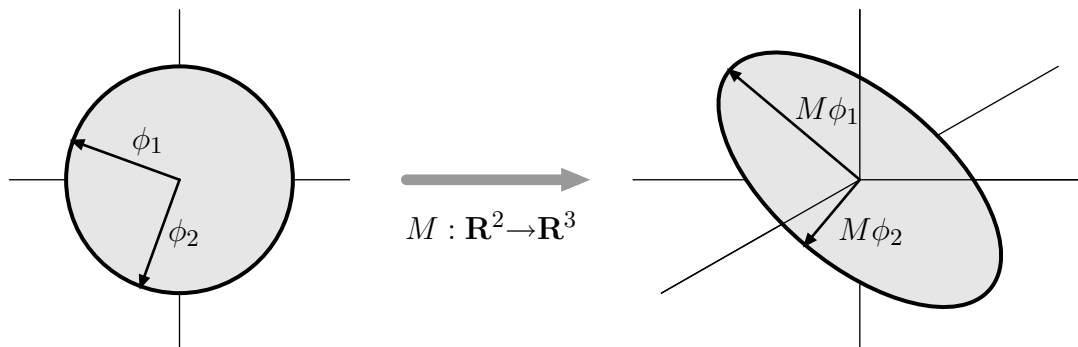


Figure 2: A geometric intuition for singular values.

By way of example, in the figure $\|M\phi_1\|/\|\phi_1\|$ and $\|M\phi_2\|/\|\phi_2\|$ would be the first and second singular values of M . In particular, the first singular value measures the maximum amount of “stretching” that M can do to a vector under the L^2 -norm (here stretching need not preserve direction, in general the ϕ_i are not eigenvectors). Singular values are traditionally denoted by σ_i and the largest singular value by $\sigma_1(M)$ and then arranging the rest in decreasing order, i.e., $\sigma_1(M) \geq \sigma_2(M) \geq \dots$.

From the above discussion we have the following definition for $\sigma_1(M)$:

$$\sigma_1(M) = \max_{x \neq 0} \frac{\|Mx\|}{\|x\|},$$

in particular, the singular values of M are the square root of the eigenvalues of $M^T M$. Alternatively we could define it by

$$\sigma_1(M) = \max_{x, y \neq 0} \frac{|x^T Ay|}{\|x\| \|y\|}$$

(in the picture above $y = \phi_1$ and $x = M\phi_1$).

This last definition gives a very useful formula, namely $|x^T My| \leq \sigma_1(M) \|x\| \|y\|$. We used this formula above in the special case when M is symmetric. For such matrices it is simple to check that the singular values are the absolute eigenvalues of the matrix arranged in decreasing order. So we have that $\sigma_1(D^{-1/2}AD^{-1/2}) = 1$ while $\sigma_2(D^{-1/2}AD^{-1/2}) = \max\{|1 - \lambda_1|, |\lambda_{n-1} - 1|\}$.

An important tool for working with singular values for nonnegative matrices, and so by extension directed graphs, is the Perron-Frobenius Theorem.

Theorem 3 (Perron-Frobenius Theorem). *Let M be a square nonnegative irreducible matrix. Then there is a $\rho > 0$ so that*

- (i) ρ is an eigenvalue of M and all other eigenvalues of M satisfy $|\lambda| \leq \rho$.
- (ii) there is an eigenvector for the eigenvalue ρ with all positive entries.
- (iii) if there are k eigenvalues with $|\lambda_i| = \rho$ then the eigenvalues are fixed under rotation by an angle of $2\pi/k$ in the complex plane.

In the theorem above ρ is sometimes referred to as the spectral radius.

A square matrix M is said to be irreducible if there is no way to permute the rows and columns of the matrix so that it has the form

$$\left(\begin{array}{c|c} D & E \\ \hline O & F \end{array} \right),$$

i.e., no nontrivial $k \times (n - k)$ block of 0s. For undirected graphs this is equivalent to being connected while for directed graphs this will be equivalent to being strongly connected (i.e., there is a directed path between any pair of vertices).

If we have a nonnegative irreducible matrix M , it follows that if we can find *any* nonnegative eigenvector for $M^T M$, then the corresponding square root of that eigenvalue is $\sigma_1(M)$.

Directed graphs and discrepancy

Finally, we will consider discrepancy for directed graphs. For a directed graph we can define the adjacency matrix A by $A_{i,j} = 1$ if there is a directed edge starting at i and ending at j , and 0 otherwise. There will be no combinatorial Laplacian in general because now there are two different degrees and there is no clear choice of what should go on the diagonal. However, we will still have a normalized adjacency matrix (which would be very similar to a normalized Laplacian).

To normalize we first note that the row-sum of the adjacency matrix corresponds to the out-degree and the column-sum corresponds to the in-degree. We can define D_{in} and D_{out} as the two diagonal degree matrices similar to what we had before. Then the normalized adjacency matrix will be $D_{out}^{-1/2} A D_{in}^{-1/2}$. With this definition it can be checked that $D_{in}^{1/2} \mathbf{1}$ is an eigenvector with eigenvalue 1 for the matrix $(D_{out}^{-1/2} A D_{in}^{-1/2})^T (D_{out}^{-1/2} A D_{in}^{-1/2})$. In particular, $\sigma_1(D_{out}^{-1/2} A D_{in}^{-1/2}) = 1$, and moreover the right and left vectors associated with the singular value are $D_{in}^{1/2} \mathbf{1}$ and $D_{out}^{1/2} \mathbf{1}$ respectively.

Half of the trick to generalizing results is figuring out how to generalize the definitions. Instead of $e(X, Y)$ we will count $e(X \rightarrow Y)$, which counts the number of edges which start in X and end in Y . Also since there are now two types of degrees associated with each vertex, namely the in-degree and the out-degree, we will also

consider two types of volume, namely the in-volume and the out-volume. That is $\text{vol}_{in}(X) = \sum_{v \in X} d_{in}(v)$ and $\text{vol}_{out}(X) = \sum_{v \in X} d_{out}(v)$.

The discrepancy for a directed graph G , again denoted $\text{disc}(G)$, can now be defined as the minimal α so that

$$\left| e(X \rightarrow Y) - \frac{\text{vol}_{out}(X) \text{vol}_{in}(Y)}{\text{vol}(G)} \right| \leq \alpha \sqrt{\text{vol}_{out}(X) \text{vol}_{in}(Y)}.$$

Following the same line of reasoning for discrepancy as before with properties of singular values,

$$\begin{aligned} \left| e(X \rightarrow Y) - \frac{\text{vol}_{out}(X) \text{vol}_{in}(Y)}{\text{vol}(G)} \right| &= \left| \psi_X^T D_{out}^{1/2} \left(D_{out}^{-1/2} A D_{in}^{-1/2} - \frac{D_{out}^{1/2} \mathbf{1} \mathbf{1}^T D_{in}^{1/2}}{\text{vol}(G)} \right) D_{in}^{1/2} \psi_Y \right| \\ &= \sigma_1 \left(D_{out}^{-1/2} A D_{in}^{-1/2} - \frac{D_{out}^{1/2} \mathbf{1} \mathbf{1}^T D_{in}^{1/2}}{\text{vol}(G)} \right) \| D_{out}^{1/2} \psi_X \| \| D_{in}^{1/2} \psi_Y \| \\ &= \sigma_2(D_{out}^{-1/2} A D_{in}^{-1/2}) \sqrt{\text{vol}_{out}(X) \text{vol}_{in}(Y)}. \end{aligned}$$

In the last step we noticed that the matrix under consideration was $D_{out}^{-1/2} A D_{in}^{-1/2}$ with the largest singular value “subtracted out”, and so we can move to considering σ_2 of the matrix. In particular, this establishes

$$\text{disc}(G) \leq \sigma_2(D_{out}^{-1/2} A D_{in}^{-1/2}).$$

By a similar technique as outlined before (with a little more bookkeeping) it can also be shown that for directed graphs

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

So in some sense the second largest singular value and the discrepancy for a directed graph are equivalent. (This hints at a class of quasirandom properties for directed graphs, an area not yet developed.)

Comment. Several results that hold for the eigenvalues of undirected graphs tend to generalize to singular values for directed graphs. The discrepancy result stated above is one such example. In some sense, many of the results for undirected graphs were using singular values to begin with (it was just hard to tell since singular values and eigenvalues essentially agree for the symmetric case).