Lecture 3

In which we analyze the power method to approximate eigenvalues and eigenvectors, and we describe some more algorithmic applications of spectral graph theory.[1]

1 The power method

Last week, we showed that, if $G = (V, E)$ is a d-regular graph, and L is its normalized Laplacian matrix with eigenvalues $0 = \lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$, given an eigenvector of λ_2, the algorithm SpectralPartition finds, in nearly-linear time $O(|E| + |V| \log |V|)$, a cut $(S, V - S)$ such that $\phi(S) \leq 2\sqrt{\phi(G)}$.

More generally, if, instead of being given an eigenvector $\mathbf{x}$ such that $L\mathbf{x} = \lambda_2 \mathbf{x}$, we are given a vector $\mathbf{x} \perp \mathbf{1}$ such that $\mathbf{x}^T L \mathbf{x} \leq (\lambda_2 + \epsilon)\mathbf{x}^T \mathbf{x}$, then the algorithm finds a cut such that $\phi(S) \leq \sqrt{4\phi(G) + 2\epsilon}$. In this lecture we describe and analyze an algorithm that computes such a vector using $O((|V| + |E|) \cdot \frac{1}{\epsilon} \cdot \log \frac{|V|}{\epsilon})$ arithmetic operations.

A symmetric matrix is *positive semi-definite* (abbreviated PSD) if all its eigenvalues are nonnegative. We begin by describing an algorithm that approximates the largest eigenvalue of a given symmetric PSD matrix. This might not seem to help very much because because we want to compute the second smallest, not the largest, eigenvalue. We will see, however, that the algorithm is easily modified to accomplish what we want.

1.1 The Power Method to Approximate the Largest Eigenvalue

The algorithm works as follows

Algorithm *Power*

Input: PSD matrix M, parameter k

- Pick uniformly at random $\mathbf{x}_0 \sim \{-1,1\}^n$
- for $i := 1$ to k
  $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return $\mathbf{x}_k$

That is, the algorithm simply picks uniformly at random a vector $\mathbf{x}$ with ± 1 coordinates, and outputs $M^k \mathbf{x}$.

Note that the algorithm performs $O(k \cdot (n+m))$ arithmetic operations, where m is the number of non-zero entries of the matrix M.

Theorem 1 *For every PSD matrix M, positive integer k and parameter $\epsilon > 0$, with probability $\geq 3/16$ over the choice of $\mathbf{x}_0$, the algorithm Power outputs a vector $\mathbf{x}_k$ such that*

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1 + 4n(1-\epsilon)^{2k}}$$

where λ_1 is the largest eigenvalue of M.

Note that, in particular, we can have $k = O(\log n/\epsilon)$ and $\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq (1 - O(\epsilon)) \cdot \lambda_1$.

Let $\lambda_1 \geq \cdots \lambda_n$ be the eigenvalues of M, with multiplicities, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a system of orthonormal eigenvectors such that $M\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Theorem 1 is implied by the following two lemmas

Lemma 2 *Let $\mathbf{v} \in \mathbb{R}^n$ be a vector such that $||\mathbf{v}|| = 1$. Sample uniformly $\mathbf{x} \sim \{-1,1\}^n$. Then*

$$\mathbb{P}\left[|\langle \mathbf{x}, \mathbf{v} \rangle| \geq \frac{1}{2} \right] \geq \frac{3}{16}$$

Lemma 3 *Let $\mathbf{x} \in \mathbb{R}^n$ be a vector such that $|\langle \mathbf{x}, \mathbf{v}_1 \rangle| \geq \frac{1}{2}$. Then, for every positive integer t and positive $\epsilon > 0$, if we define $\mathbf{y} := M^k \mathbf{x}$, we have*

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1 + 4||\mathbf{x}||^2 (1-\epsilon)^{2k}}$$

It remains to prove the two lemmas.

Proof: (Of Lemma 2) Let $\mathbf{v} = (v_1, \ldots, v_n)$. The inner product $\langle \mathbf{x}, \mathbf{v} \rangle$ is the random variable

$$S := \sum_i x_i v_i$$

Let us compute the first, second, and fourth moment of S.

$$\mathbb{E}\, S = 0$$

$$\mathbb{E}\, S^2 = \sum_i v_i^2 = 1$$

$$\mathbb{E}\, S^4 = 3\left(\sum_i v_i^2\right) - 2\sum_i v_i^4 \le 3$$

Recall that the Paley-Zygmund inequality states that if Z is a non-negative random variable with finite variance, then, for every $0 \le \delta \le 1$, we have

$$\mathbb{P}[Z \ge \delta\, \mathbb{E}\, Z] \ge (1-\delta)^2 \cdot \frac{(\mathbb{E}\, Z)^2}{\mathbb{E}\, Z^2} \tag{1}$$

which follows by noting that

$$\mathbb{E}\, Z = \mathbb{E}[Z \cdot 1_{Z < \delta\, \mathbb{E}\, Z}] + \mathbb{E}[Z \cdot 1_{Z \ge \delta\, \mathbb{E}\, Z}] \ ,$$

that

$$\mathbb{E}[Z \cdot 1_{Z < \delta\, \mathbb{E}\, Z}] \le \delta\, \mathbb{E}\, Z \ ,$$

and that

$$\begin{aligned} \mathbb{E}[Z \cdot 1_{Z \ge \delta\, \mathbb{E}\, Z}] &\le \sqrt{\mathbb{E}\, Z^2} \cdot \sqrt{\mathbb{E}\, 1_{Z \ge \delta\, \mathbb{E}\, Z}} \\ &= \sqrt{\mathbb{E}\, Z^2}\sqrt{\mathbb{P}[Z \ge \delta\, \mathbb{E}\, Z]} \end{aligned}$$

We apply the Paley-Zygmund inequality to the case $Z = S^2$ and $\delta = 1/4$, and we derive

$$\mathbb{P}\left[S^2 \ge \frac{1}{4}\right] \ge \left(\frac{3}{4}\right)^2 \cdot \frac{1}{3} = \frac{3}{16}$$

□

Remark 4 *The proof of Lemma 2 works even if* $\mathbf{x} \sim \{-1,1\}^n$ *is selected according to a 4-wise independent distribution. This means that the algorithm can be derandomized in polynomial time.*

Proof: (Of Lemma 3) Let us write $\mathbf{x}$ as a linear combination of the eigenvectors

$$\mathbf{x} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n$$

where the coefficients can be computed as $a_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$. Note that, by assumption, $|a_1| \geq .5$, and that, by orthonormality of the eigenvectors, $||\mathbf{x}||^2 = \sum_i a_i^2$.

We have

$$\mathbf{y} = a_1 \lambda_1^k \mathbf{v}_1 + \cdots + a_n \lambda_n^k \mathbf{v}_n$$

and so

$$\mathbf{y}^T M \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k+1}$$

and

$$\mathbf{y}^T \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k}$$

We need to prove a lower bound to the ratio of the above two quantities. We will compute a lower bound to the numerator and an upper bound to the denominator in terms of the same parameter.

Let ℓ be the number of eigenvalues larger than $\lambda_1 \cdot (1-\epsilon)$. Then, recalling that the eigenvalues are sorted in non-increasing order, we have

$$\mathbf{y}^T M \mathbf{y} \geq \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k+1} \geq \lambda_1 (1-\epsilon) \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k}$$

We also see that

$$\sum_{i=\ell+1}^{n} a_i^2 \lambda_i^{2k}$$

$$\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \sum_{i=\ell+1}^{n} a_i^2$$

$$\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \cdot ||\mathbf{x}||^2$$

$$\leq 4a_1^2\lambda_1^{2k}(1-\epsilon)^{2t}||\mathbf{x}||^2$$

$$\leq 4||\mathbf{x}||^2(1-\epsilon)^{2k}\sum_{i=1}^{\ell} a_i^2\lambda_i^{2k}$$

So we have

$$\mathbf{y}^T\mathbf{y} \leq (1+4||\mathbf{x}||^2(1-\epsilon)^{2k}) \cdot \sum_{i=1}^{\ell} a_i^2$$

giving

$$\frac{\mathbf{y}^T M\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1+4||\mathbf{x}||^2(1-\epsilon)^{2k}}$$

$\square$

Remark 5 *Where did we use the assumption that M is positive semidefinite? What happens if we apply this algorithm to the adjacency matrix of a bipartite graph?*

1.2 Approximating the Second Largest Eigenvalue

Suppose now that we are interested in finding the second largest eigenvalue of a given PSD matrix M. If M has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$, and we know the eigenvector $\mathbf{v}_1$ of λ_2, then M is a PSD linear map from the orthogonal space to $\mathbf{v}_1$ to itself, and λ_2 is the largest eigenvalue of this linear map. We can then run the previous algorithm on this linear map.

Algorithm *Power2*

Input: PSD matrix M, vector $\mathbf{v}_1$ parameter k

- Pick uniformly at random $\mathbf{x} \sim \{-1,1\}^n$
- $\mathbf{x}_0 := \mathbf{x} - \mathbf{v}_1 \cdot \langle \mathbf{x}, \mathbf{v}_1 \rangle$
- for $i := 1$ to k
  $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return $\mathbf{x}_k$

If $\mathbf{v}_1,\ldots,\mathbf{v}_n$ is an orthonormal basis of eigenvectors for the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of M, then, at the beginning, we pick a random vector

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$$

that, with probability at least 3/16, satisfies $|a_2| \geq 1/2$. (Cf. Lemma 2.) Then we compute $\mathbf{x}_0$, which is the projection of $\mathbf{x}$ on the subspace orthogonal to $\mathbf{v}_1$, that is

$$\mathbf{x}_0 = a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$$

Note that $||\mathbf{x}||^2 = n$ and that $||\mathbf{x}_0||^2 \leq n$.

The output is the vector $\mathbf{x}_k$

$$\mathbf{x}_k = a_2\lambda_2^k\mathbf{v}_2 + \cdots a_n\lambda_n^k\mathbf{v}_n$$

If we apply Lemma 3 to subspace orthogonal to $\mathbf{v}_1$, we see that when $|a_2| \geq 1/2$ we have that, for every $0 < \epsilon < 1$,

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1-\epsilon) \cdot \frac{1}{4n(1-\epsilon)^{2k}}$$

We have thus established the following analysis.

Theorem 6 *For every PSD matrix M, positive integer k and parameter $\epsilon > 0$, if $\mathbf{v}_1$ is a length-1 eigenvector of the largest eigenvalue of M, then with probability $\geq 3/16$ over the choice of $\mathbf{x}_0$, the algorithm Power2 outputs a vector $\mathbf{x}_k \perp \mathbf{v}_1$ such that*

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1-\epsilon) \cdot \frac{1}{1 + 4n(1-\epsilon)^{2k}}$$

where λ_2 is the second largest eigenvalue of M, counting multiplicities.

1.3 The Second Smallest Eigenvalue of the Laplacian

Finally, we come to the case in which we want to compute the second smallest eigenvalue of the normalized Laplacian matrix $L = I - \frac{1}{d}A$ of a d-regular graph $G = (V, E)$, where A is the adjacency matrix of G.

Consider the matrix $M := 2I - L = I + \frac{1}{d}A$. Then if $0 = \lambda_1 \leq \ldots \leq \lambda_n \leq 2$ are the eigenvalues of L, we have that

$$2 = 2 - \lambda_1 \geq 2 - \lambda_2 \geq \cdots \geq 2 - \lambda_n \geq 0$$

are the eigenvalues of M, and that M is PSD. M and L have the same eigenvectors, and so $\mathbf{v}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ is a length-1 eigenvector of the largest eigenvalue of M.

By running algorithm Power2, we can find a vector $\mathbf{x}$ such that

$$\mathbf{x}^T M \mathbf{x}^T \geq (1-\epsilon)\cdot(2-\lambda_2)\cdot \mathbf{x}^T\mathbf{x}$$

and

$$\mathbf{x}^T M \mathbf{x}^T = 2\mathbf{x}^T\mathbf{x} - \mathbf{x}^T L\mathbf{x}$$

so, rearranging, we have

$$\frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \leq \lambda_2 + 2\epsilon$$

If we want to compute a vector whose Rayleigh quotient is, say, at most $2\lambda_2$, then the running time will be $\tilde{O}((|V|+|E|)/\lambda_2)$, because we need to set $\epsilon = \lambda_2/2$, which is not nearly linear in the size of the graph if λ_2 is, say $O(1/|V|)$.

For a running time that is nearly linear in n for all values of λ_2, one can, instead, apply the power method to the *pseudoinverse* L^+ of L. (Assuming that the graph is connected, $L^+\mathbf{x}$ is the unique vector $\mathbf{y}$ such that $L\mathbf{y} = \mathbf{x}$, if $\mathbf{x} \perp (1,\ldots,1)$, and $L^+\mathbf{x} = 0$ if $\mathbf{x}$ is parallel to $(1,\ldots,1)$.) This is because L^+ has eigenvalues $0, 1/\lambda_2, \ldots, 1/\lambda_n$, and so L^+ is PSD and $1/\lambda_2$ is its largest eigenvalue.

Although computing L^+ is not known to be doable in nearly linear time, there are nearly linear time algorithms that, given $\mathbf{x}$, solve in $\mathbf{y}$ the linear system $L\mathbf{y} = \mathbf{x}$, and this is the same as computing the product $L^+\mathbf{x}$, which is enough to implement algorithm *Power* applied to L^+.

In time $O((V+|E|)\cdot(\log|V|/\epsilon)^{O(1)})$, we can find a vector $\mathbf{y}$ such that $\mathbf{y} = (L^+)^k\mathbf{x}$, where $\mathbf{x}$ is a random vector in $\{-1,1\}^n$, shifted to be orthogonal to $(1,\ldots,1)$ and $k = O(\log|V|/\epsilon)$. What is the Rayleigh quotient of such a vector with respect to L?

Let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be a basis of orthonormal eigenvectors for L and L^+. If $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of L, then we have

$$L\mathbf{v}_1 = L^+\mathbf{v}_1 = \mathbf{0}$$

and, for $i = 1,\ldots,n$, we have

$$L\mathbf{v}_i = \lambda_i \quad L^+\mathbf{v}_i = \frac{1}{\lambda_i}$$

Write $\mathbf{x} = a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$, where $\sum_i a_i^2 \leq n$, and ssume that, as happens with probability at least $3/16$, we have $a_2^2 \geq \frac{1}{4}$. Then

$$\mathbf{y} = \sum_{i=2}^{n} a_i \frac{1}{\lambda_i^k}$$

and the Rayleigh quotient of $\mathbf{y}$ with respect to L is

$$\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}}}{\sum_i a_i^2 \frac{1}{\lambda_i^{2k}}}$$

and the analysis proceeds similarly to the analysis of the previous section. If we let ℓ be the index such that $\lambda_\ell \leq (1+\epsilon) \cdot \lambda_2 \leq \lambda_{\ell+1}$ then we can upper bound the numerator as

$$\begin{aligned}
\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}} &\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \sum_{i > \ell} a_i^2 \\
&\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot n \\
&\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot 4 n a_2^2 \\
&\leq \left(1 + \frac{4n}{(1+\epsilon)^{2k-1}}\right) \cdot \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}
\end{aligned}$$

and we can lower bound the denominator as

$$\begin{aligned}
\sum_i a_i^2 \frac{1}{\lambda_i^{2k}} &\geq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k}} \\
&\geq \frac{1}{(1+\epsilon)\lambda_2} \cdot \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}
\end{aligned}$$

and the Rayleigh quotient is

$$\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \leq \lambda_2 \cdot (1+\epsilon) \cdot \left(1 + \frac{4n}{(1+\epsilon)^{2k-1}}\right) \leq (1+2\epsilon) \cdot \lambda_2$$

when $k = O\left(\frac{1}{\epsilon} \log \frac{n}{\epsilon}\right)$.

An $O((|V|+|E|) \cdot (\log |V|)^{O(1)})$ algorithm to solve in $\mathbf{y}$ the linear system $L\mathbf{y} = \mathbf{x}$ was first developed by Spielman and Teng. Faster algorithms (with a lower exponent in the $(\log |V|)^{O(1)}$ part of the running time, and smaller constants) were later developed by Koutis, Miller and Peng, and, very recently, by Kelner, Orecchia, Sidford, and Zhu.

2 Other applications of spectral graph theory

2.1 Spectral graph theory in irregular graphs

Let $G = (V, E)$ be an undirected graph in which every vertex has positive degree and A be the adjacency matrix of G. We want to define a Laplacian matrix L and a Rayleigh quotient such that the k-th eigenvalue of L is the minimum over all k-dimensional spaces of the maximum Rayleigh quotient in the space, and we want the conductance of a set to be the same as the Rayleigh quotient of the indicator vector of the set. All the facts that we have proved in the regular case essentially reduce to these two properties of the Laplacian and the Rayleigh quotient.

Let d_v be the degree of vertex v in G. We define the Rayleigh quotient of a vector $\mathbf{x} \in \mathbb{R}^V$ as

$$R_G(\mathbf{x}) := \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\sum_v d_v x_v^2}$$

Let D be the diagonal matrix of degrees such that $D_{u,v} = 0$ if $u \neq v$ and $D_{v,v} = d_v$. Then define the Laplacian of G as

$$L_G := I - D^{-1/2} A D^{-1/2}$$

Note that in a d-regular graph we have $D = dI$ and $L_G = I - \frac{1}{d}A$, which is the standard definition.

Since $L = L_G$ is a real symmetric matrix, the k-th smallest eigenvalue of L is

$$\lambda_k = \min_{k-\text{dimensional } S} \max_{\mathbf{x} \in S} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

Now let us do the change of variable $\mathbf{y} \leftarrow D^{-1/2}\mathbf{x}$. We have

$$\lambda_k = \min_{k-\text{dimensional } S'} \max_{\mathbf{y} \in S'} \frac{\mathbf{y}^T D^{1/2} L D^{1/2} \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}$$

In the numerator, $\mathbf{y}^T D \mathbf{y} = \sum_v d_v y_v^2$, and in the denominator a simple calculation shows

$$\mathbf{y}^T D^{1/2} L D^{1/2} \mathbf{y} = \mathbf{y}^T (D - A) \mathbf{y} = \sum_{\{u,v\}} |y_v - y_u|^2$$

so indeed

$$\lambda_k = \min_{k-\text{dimensional } S} \; \max_{\mathbf{y} \in S} \; R_G(\mathbf{y})$$

For two vectors $\mathbf{y}, \mathbf{z}$, define the inner product

$$\langle \mathbf{y}, \mathbf{z} \rangle_G := \sum_v d_v \overline{y_v} z_v$$

Then we can prove that

$$\lambda_2 = \min_{\mathbf{y}:\langle \mathbf{y},(1,\ldots,1)\rangle_G = 0} R_G(\mathbf{y})$$

With these definitions and observations in place, it is now possible to repeat the proof of Cheeger's inequality step by step (replacing the condition $\sum_v x_v = 0$ with $\sum_i d_v x_v = 0$, adjusting the definition of Rayleigh quotient, etc.) and prove that if λ_2 is the second smallest eigenvalue of the Laplacian of an irregular graph G, and $\phi(G)$ is the conductance of G, then

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

2.2 Higher-order Cheeger inequality

The Cheeger inequality gives a "robust" version of the fact that $\lambda_2 = 0$ if and only if G is disconnected. It is possible to also give a robust version of the fact that $\lambda_k = 0$ if and only if G has at least k connected components. We will restrict the discussion to regular graphs.

For a size parameter $s \leq |V|/2$, denote the size-s *small-set expansion* of a graph

$$SSE_s(G) := \min_{S \subseteq V: \; |S| \leq s} \phi(S)$$

So that $SSE_{\frac{n}{2}}(G) = \phi(G)$. This is an interesting optimization problem, because in many settings in which non-expanding sets correspond to clusters, it is more interesting to find small non-expanding sets (and, possibly, remove them and iterate to find more) than to find large ones. It has been studied very intensely in the past five years because of its connection with the *Unique Games Conjecture*, which is in turn one of the key open problems in complexity theory.

If $\lambda_k = 0$, then we know that are at least k connected components, and, in particular, there is a set $S \subseteq V$ such that $\phi(S) = 0$ and $|S| \leq \frac{n}{k}$, meaning that $SSE_{\frac{n}{k}} = 0$. By analogy with the Cheeger inequality, we may look for a robust version of this fact, of the form $SSE_{\frac{n}{k}} \leq O(\sqrt{\lambda_k})$. Unfortunately there are counterexamples, but Arora, Barak and Steurer have proved that, for every δ,

$$SSE_{\frac{n^{1+\delta}}{k}} \leq O\left(\sqrt{\frac{\lambda_k}{\delta}}\right)$$

To formulate a "higher-order" version of the Cheeger inequality, we need to define a quantity that generalize expansion in a different way. For an integer parameter $k \geq 2$, define "order k expansion" as

$$\phi_k(G) = \min_{S_1,\ldots S_k \subseteq V \text{ disjoint}} \max_{i=1,\ldots,k} \phi(S_i)$$

Note that $\phi_2(G) = \phi(G)$. Then Lee, Oveis-Gharan and Trevisan prove that

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^2) \cdot \sqrt{\lambda_k}$$

and

$$\phi_{.9 \cdot k}(G) \leq O(\sqrt{\lambda_k \cdot \log k})$$

(which was also proved by Louis, Raghavendra, Tetali and Vempala). The upper bounds are algorithmic, and given k orthogonal vectors all of Rayleigh quotient at most λ, there are efficient algorithms that find at least k disjoint sets each of expansion at most $O(k^2\sqrt{\lambda})$ and at least $.9 \cdot k$ disjoint sets each of expansion at most $O(\sqrt{\lambda \log k})$.

2.3 A Cheeger-type inequality for λ_n

We proved that $\lambda_n = 2$ if and only if G has a bipartite connected component. What happens when λ_n is, say, 1.999?

We can define a "bipartite" version of expansion as follows:

$$\beta(G) := \min_{\mathbf{x} \in \{-1,0,1\}^V} \frac{\sum_{\{u,v\} \in E} |x_u + x_v|}{\sum_v d_v |x_v|}$$

The above quantity has the following combinatorial interpretation: take a set S of vertices, and a partition of S into two disjoint sets A, B. Then define

$$\beta(S, A, B) := \frac{2E(A) + 2E(B) + 2E(S, V - S)}{vol(S)}$$

where $E(A)$ is the number of edges entirely contained in A, and $E(S, V-S)$ is the number of edges with one endpoint in S and one endpoint in $V-S$. We can think of $\beta(S, A, B)$ as measuring what fraction of the edges incident on S we need to delete in order to make S disconnected from the rest of the graph and A, B be a bipartition of the subgraph induced by S. In other words, it measure how close S is to being a bipartite connected component. Then we see that

$$\beta(G) = \min_{S \subseteq V,\ A,B \text{ partition of } S} \beta(S, A, B)$$

Trevisan proves that

$$\frac{1}{2} \cdot (2 - \lambda_n) \leq \beta(G) \leq \sqrt{2 \cdot (2 - \lambda_n)}$$