

# Spectral clustering

## 1 Graph conductance

As in the last lecture, consider an undirected graph  $G = (V, E)$  with  $n = |V|$  vertices. One way to capture the idea that a subset  $S \subseteq V$  of vertices is a good “cluster” is through the notion of *conductance*. Define the *volume* of  $S$  as the quantity

$$\text{vol}(S) = \sum_{u \in S} \deg(u),$$

i.e., the sum of vertex degrees in  $S$ . Then we define the *conductance* of the set  $S$  by

$$\phi_G(S) := \frac{|E(S, \bar{S})|}{\min(\text{vol}(S), \text{vol}(\bar{S}))},$$

where  $E(S, \bar{S})$  denotes the set of edges that run from  $S$  to its complement.

Let us define  $\rho_2(G)$  to be the minimum of the conductance as it runs over all non-empty subsets of  $G$ :

$$\rho_2(G) := \min_{\emptyset \neq S \subseteq V} \phi_G(S).$$

## 2 Spectral clusters

It turns out that the eigenvectors of the Laplacian of  $G$  encode information about sets of small conductance. It will be more useful to work with the *normalized Laplacian*. For simplicity, let's assume that our graph is  $d$ -regular, which means that  $\deg(v) = d$  for every  $v \in V$ . Then the normalized Laplacian is just the usual (“combinatorial”) Laplacian scaled down by  $d$ :

$$\mathcal{L}_G := \frac{1}{d} L_G = I - \frac{1}{d} A,$$

where  $A$  is the adjacency matrix of  $G$ . Recall that if we sort the eigenvalues of  $\mathcal{L}_G$  as  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ , then  $\lambda_1 = 0$  corresponds to the constant eigenvector  $\mathbf{x}_1 = (1/\sqrt{n}, \dots, 1/\sqrt{n})$ . The next claim gives the first connection between the 2nd smallest eigenvalue  $\lambda_2(G)$  of  $\mathcal{L}_G$  and the sets of minimal conductance in  $G$ .

**Theorem 2.1.** *It holds that  $\lambda_2(G) \leq 2\rho_2(G)$  for any graph  $G$ .*

To prove this claim, it will help to use the variational characterization of  $\lambda_2(G)$  from the last lecture:

$$\begin{aligned} \lambda_2(G) &= \min \left\{ \langle \mathbf{x}, \mathcal{L}_G \mathbf{x} \rangle : \mathbf{x} \perp \mathbf{x}_1, \|\mathbf{x}\|^2 = 1 \right\} \\ &= \min \left\{ \frac{\langle \mathbf{x}, \mathcal{L}_G \mathbf{x} \rangle}{\|\mathbf{x}\|^2} : \mathbf{x} \perp \mathbf{x}_1 \right\} \\ &= \min \left\{ \frac{\langle \mathbf{x}, \mathcal{L}_G \mathbf{x} \rangle}{\|\mathbf{x} - \bar{\mathbf{x}}\|^2} : \mathbf{x} \neq 0 \right\}, \end{aligned}$$

where  $\bar{\mathbf{x}} := (\frac{1}{n} \sum_{u \in V} x_u) (1, 1, \dots, 1)$  is the vector containing the average value of the coordinates of  $\mathbf{x}$  in every coordinate. Note that we have removed the condition  $\mathbf{x} \perp \mathbf{x}_1$  by removing the part of  $\mathbf{x}$  that is not orthogonal to  $\mathbf{x}_1$ :

$$\mathbf{x} - \langle \mathbf{x}, \mathbf{x}_1 \rangle \mathbf{x}_1 = \mathbf{x} - \bar{\mathbf{x}}.$$

Using the expression we derived last time for the quadratic form  $\langle \mathbf{x}, \mathcal{L}_G \mathbf{x} \rangle$ , we can rewrite the last line above as

$$\lambda_2(G) = \min \left\{ \frac{\frac{1}{d} \sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_{u \in V} (x_u - \bar{x})^2} : \mathbf{x} \neq 0 \right\}, \quad (1)$$

where  $\bar{x} = \frac{1}{n} \sum_{u \in V} x_u$  is the average value of the components of  $x$ .

*Proof.* Note that for a  $d$ -regular graph, we have  $\text{vol}(S) = d|S|$ . Consider any subset  $S \subseteq V$ . We will show that  $\lambda_2(G) \leq 2\phi_G(S)$ . Clearly we may assume that  $|S| \leq n/2$ . Define the indicator vector  $\mathbf{v}_S$  so that  $\mathbf{v}_S(u) = 1$  if  $u \in S$  and  $\mathbf{v}_S(u) = 0$  otherwise.

Let us now plug  $\mathbf{v}_S$  into (1) to get an upper bound on  $\lambda_2(G)$ :

$$\begin{aligned} \lambda_2(G) &\leq \frac{\frac{1}{d} \sum_{\{u,v\} \in E} (\mathbf{v}_S(u) - \mathbf{v}_S(v))^2}{\sum_{u \in V} (\mathbf{v}_S(u) - \frac{|S|}{n})^2} = \frac{\frac{1}{d} |E(S, \bar{S})|}{|S| \left(1 - \frac{|S|}{n}\right)^2 + (n - |S|) \left(\frac{|S|}{n}\right)^2} \\ &= \frac{\frac{1}{d} |E(S, \bar{S})|}{\frac{|S|}{n} (n - |S|)}. \end{aligned}$$

Since we have assumed that  $|S| \leq n/2$ , we know that  $\frac{n-|S|}{n} = 1 - \frac{|S|}{n} \geq 1/2$ , and therefore

$$\lambda_2(G) \leq 2 \frac{\frac{1}{d} |E(S, \bar{S})|}{|S|} = 2 \frac{|E(S, \bar{S})|}{\text{vol}(S)} = 2\phi_G(S).$$

□

### 3 Cheeger's inequality and sweep cuts

In the previous section, we saw that sets of small conductance force  $\lambda_2(G)$  to be small. It turns out that the other direction is true as well.

**Theorem 3.1** (Discrete Cheeger inequality). *It holds that  $\rho_2(G) \leq \sqrt{2\lambda_2(G)}$  for any graph  $G$ .*

While we will not prove this inequality, we will now discuss the simple algorithm behind it:

**Algorithm SWEEP( $\mathbf{x}$ ):**

Sort the vertices  $V = \{v_1, v_2, \dots, v_n\}$  according to the value of the corresponding coordinate in  $\mathbf{x}$ :

$$\mathbf{x}(v_1) \leq \mathbf{x}(v_2) \leq \dots \leq \mathbf{x}(v_n).$$

This sorted order gives us  $n - 1$  different cuts:

$$(\{v_1\}, \{v_2, \dots, v_n\}), (\{v_1, v_2\}, \{v_3, \dots, v_n\}), \dots, (\{v_1, \dots, v_{n-1}\}, v_n)$$

Now return the cut  $S$  among these  $n - 1$  that has the smallest  $\phi_G(S)$  value.

**Theorem 3.2.** *The algorithm is well-defined for any vector  $\mathbf{x} \in \mathbb{R}^V$ . If we additionally assume that  $\mathbf{x} \perp \mathbf{x}_1$ , then this algorithm is guaranteed to output a cut  $S$  that satisfies*

$$\phi_G(S) \leq \sqrt{2\mathcal{R}_G(\mathbf{x})},$$

where  $\mathcal{R}_G(\mathbf{x})$ —called the “Rayleigh quotient of  $\mathbf{x}$ ”—is defined by

$$\mathcal{R}_G(\mathbf{x}) = \frac{\sum_{\{u,v\} \in E} (\mathbf{x}(u) - \mathbf{x}(v))^2}{d\|\mathbf{x}\|^2}$$

## 4 Multi-way spectral clustering

If we are interested in finding a clustering into more pieces, we can use higher eigenvectors. Define the *k-way conductance of  $G$*  as the quantity

$$\rho_k(G) := \min_{S_1, S_2, \dots, S_k \subseteq V} \max \{ \phi_G(S_i) : i = 1, \dots, k \},$$

where the minimum is over all non-empty, *pairwise disjoint* subsets of  $V$ . Then one has the following higher-order Cheeger inequality [1]:

**Theorem 4.1.**

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

It is not known whether one can reduce the factor  $O(k^2)$ , but this is possible if we know there are at least  $1.1k$  small eigenvalues:

**Theorem 4.2.**

$$\rho_k(G) \leq O\left(\sqrt{\lambda_{1.1k} \log k}\right).$$

One general way to recover a  $k$ -clustering is to first construct the *spectral embedding* along the first  $k$  eigenvectors: This is the map  $F : V \rightarrow \mathbb{R}^k$  defined by

$$F(u) = (\mathbf{x}_1(u), \mathbf{x}_2(u), \dots, \mathbf{x}_k(u)),$$

where  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are the first  $k$  eigenvectors of  $\mathcal{L}_G$ .

Now one can look for geometric clusters, and using the right procedure, it is possible to give an algorithm that is guaranteed to produce  $k$  disjoint clusters with small conductance. In practice, this second geometric partitioning step often is done using some variant of the  $k$ -means algorithm [2].

## References

- [1] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. arXiv:1111.1055, 2011. <https://arxiv.org/pdf/1111.1055.pdf>.
- [2] Wikipedia contributors.  $k$ -means clustering. *Wikipedia*. [https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering).