

Math 690: Topics in Data Analysis and Computation

Lecture notes for October 5, 2017

Scribed by Dev Dabke and Andrew Cho

1 Introduction

The lecture covered the following:

- Continue spectral clustering
- Graph cut interpretation of spectral clustering

2 Setup Laplacian for spectral clustering

**Insert joke about creating a subroutine for this setup since we start every class with this*

Recall $W_{n \times n}$ where $W = W^\top$ (i.e. symmetric) and $w_{ij} > 0$. Additionally, let $d = \text{diag } d_1, \dots, d_n$ where $d_i = \sum_{j=0}^n w_{ij}$. Next, denote

$$\begin{aligned}\mathcal{L}_{un} &= D - W \\ \mathcal{L}_{rw} &= I - D^{-1}W\end{aligned}$$

and $P = D^{-1}W$.

Recall that if a graph has k connected components, then the eigenspace of \mathcal{L}_{un} associated with eigenvalue 0 is the span of $\{\mathbf{1}_{\mathbf{A}_1}, \dots, \mathbf{1}_{\mathbf{A}_k}\}$.

3 Exploring \mathcal{L}_{rw}

3.1 Redefining Matrices

What if \mathcal{L}_{rw} instead of \mathcal{L}_{un} ? We see that $\mathcal{L}_{rw} = \Psi(1 - \Lambda)\Psi^\top$ and that $P = \Psi\Lambda\Psi^\top$. Next, we note that $\Psi^\top D \Psi = I$. Now, denoting $\Psi = [\varphi_1, \dots, \varphi_n]$, we can see that $\varphi_k^\top D \varphi_l = 0$ if $k \neq l$. Finally, let $P\phi_k = \lambda_k \phi_l$.

3.2 Eigenvectors

If f is an eigenvector of \mathcal{L}_{un} with $\lambda = 0$, $\mathcal{L}_{un}f = 0f$ and $(D - W)f = 0$. If f is an eigenvector of \mathcal{L}_{rw} with $\lambda = 0$, then $D^{-1}(D - W)f = 0$ and $(D - W)f = 0$.

Next, we can also write the eigenvectors $\varphi_1 = c_1 \mathbf{1}_{\mathbf{A}_1} + \dots + c_k \mathbf{1}_{\mathbf{A}_k}$.

What are the constants?

$$c_1 = \sqrt{\frac{1}{\sum_{i \in C_1} d_i}}$$

so $c_1 = \{1, \dots, n_1\}$ and $c_2 = \{n_1 + 1, \dots, n_1 + n_2\}$.

3.3 The Special Case of $k = 2$

We can write $\widetilde{\varphi}_2 = \alpha\varphi_1 + \beta\varphi_2$. From **section 3.1**, we know $\varphi_k^\top D\varphi_l = 0$ if $k \neq l$ and $\varphi_k^\top D\varphi_l = 1$ if $k = l$. So for $k = 2$ we have,

$$\begin{aligned}\widetilde{\varphi}_2^\top D\widetilde{\varphi}_2 &= 1 \implies \alpha^2 + \beta^2 = 1 \\ \widetilde{\varphi}_2^\top D\widetilde{\varphi}_1 &= 0 \implies \widetilde{\varphi}_2^\top d = 0\end{aligned}$$

And since $d = [d_1, \dots, d_n]^\top$, we have $\alpha(\varphi_1^\top d) + \beta(\varphi_2^\top d) = 0$ and $\alpha\sqrt{\sum_{i \in C_1} d_i} + \beta\sqrt{\sum_{i \in C_2} d_i} = 0$.

Now, α and β should have different signs because $\widetilde{\varphi}_2 = \alpha c_1 \mathbf{1}_{A_1} + \beta c_2 \mathbf{1}_{A_2}$. In other words,

$$\widetilde{\varphi}_2(i) = \begin{cases} \alpha c_1 & \text{if } i \in c_1 \\ \beta c_2 & \text{if } i \in c_2 \end{cases}$$

Therefore, we can use the signs of α and β to get the cluster or in general, we don't have to necessarily cut at zero per say but can tune to any cut-off other than zero.

3.4 Deciding k for spectral clustering

We can use the spectral gap heuristic, which tells us that the first k eigenvalues will be close to one another, and then there will be a "gap" after k . We could use the log values of the ordered eigenvalues can provide some estimate as to where this gap could be.

4 Graph Cut

Remark. Here we take a graph cut point of view in understanding spectral clustering and why it works. The following is the introduction given by the 2007 Luxburg paper. This wasn't necessarily part of the lecture discussion, but we thought it helps understand the section better.

The intuition of clustering is to separate points in different groups according to their similarities. For data given in form of a similarity graph, this problem can be restated as follows: we want to find a partition of the graph such that the edges between different groups have a very low weight (which means that points in different clusters are dissimilar from each other) and the edges within a group have high weight (which means that points within the same cluster are similar to each other). In this section we will see how spectral clustering can be derived as an approximation to such graph partitioning problems (Luxburg '07 pg 8).

4.1 Getting Started

Definition (Cut). Let $A \cap B = \emptyset$ where $A, B \subset V = \{1, \dots, n\}$. Then a cut $W(A, B)$ is defined as

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

Definition (Volume Set). $\text{Vol}(A) = \sum_{i \in A} d_i$ is the sum of degrees.

Definition (Normalized Cut Partition). Let $C = \{C_1, \dots, C_k\}$ and denote

$$NCut(C) = \frac{\sum_{l=1}^k W(C_l, C_l^c)}{\text{Vol } C_l}$$

Proposition.

$$\begin{aligned} NCut(C) &= \sum_{l=1}^k \varphi_l^\top L \varphi_l \\ \varphi_l &= c_l \mathbf{1}_{C_l} & l = 1, \dots, k \\ c_l &= \frac{1}{\sqrt{\text{Vol } C_l}} \end{aligned}$$

Proof.

$$\begin{aligned} \varphi_l^\top L \varphi_l &= \frac{1}{2} \sum_{ij} w_{ij} (\varphi_l(i) - \varphi_l(j))^2 \\ &= \sum_{i \in C_l, j \in C_l^c} w_{ij} \left(\frac{1}{\text{Vol}(C_l)} - 0 \right)^2 \\ &= \sum_{i \in C_l, j \in C_l^c} w_{ij} \frac{1}{\text{Vol}(C_l)} \\ &= \frac{W(C_l, C_l^c)}{\text{Vol}(C_l)} \end{aligned}$$

□

4.2 Problem of Min NCut

Let H be the matrix such that $H = [\varphi_1, \dots, \varphi_k]$. Then, the minimum $NCut$ over the C s is the trace of $H^\top L H$.

The spectral minimum $NCut$ is defined to be the above trace, but where $H^\top D H = I_k$.

Proposition. *The solution of the spectral minimum $NCut$ is the first k eigenvectors of \mathcal{L}_{rw} normalized.*

4.3 Next class

We will discuss which Laplacian to use between \mathcal{L}_{un} , \mathcal{L}_{rw} , and \mathcal{L}_{sym} next class.