Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-8

Lecture 27: Basics of Graph Theory-2

So, what is graph Laplacian? So, it is usually defined as. d minus a ok. So, is graph Laplacian symmetric for undirected graph is it symmetric right it is symmetric ok. So, what is each element L ij that is going to be d ij minus a ij right. and if I sum this to n, this is nothing but summation j 1 through n d ij minus summation j 1 through n a ij. What is this quantity? d sub i and what is this quantity? So, every because it is a diagonal D, capital D is a diagonal matrix.

So, it is going to pick up a number at D ii, right, which is basically the degree. So, this is also going to be D i. So, this is essentially going to be 0. So, the row sum of the Laplacian is going to be 0 and likewise the column sum, ok.

The other way to write this is, if I multiply this Laplacian by a vector of ones. which is essentially what we are trying to do right. We are multiplying this by a vector of 1's, transposing it with a vector of 1's. So, what do you get? 0 right. So, that means 0 is a, 0 is an eigenvalue of, eigenvalue of L with vector of 1's corresponding eigenvector right.

* Laplacian: $\left. L := D - A \right\}$ symmetric

$$L_{ij} = d_{ij} - a_{ij} \qquad \sum_{j=1}^{N} l_{ij} = \underbrace{\sum_{j=1}^{N} d_{ij}}_{d_i} - \underbrace{\sum_{j=1}^{N} a_{ij}}_{d_i} = 0$$

$$L \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 0$$

$\rightarrow$ 0 is an eigenvalue of L with $\mathbb{1}_N$ being the corresponding eigenvector.

So, sometimes we use this kind of Laplacian is called unnormalized Laplacian and sometimes people use the normalized version of Laplacian. So, normalized Laplacian and there are two ways to normalize this Laplacian, one is a So, this is the normalized

Laplacian essentially you pre and post multiply your original Laplacian with D D inverse half D inverse half right. So, that is your or you have a random walk kind of Laplacian which is again a normalized Laplacian which basically you pre multiply it by 1 minus D inverse A. And if you are aware of spectral clustering algorithms say for the Shi and Malik had this normalized cut or the spectral clustering algorithm. So, that algorithm by Shi and Malik or the normalized cut it uses this Laplacian and ratio cut algorithm if you are aware of it actually uses this Laplacian, this Laplacian.

So, these Laplacians are used in spectral clustering algorithm I mean if you are aware of it. So, that this side thing but. So, it is related to connectivity of the graph. So, let us I mean at least it will give you certain idea about the graph.

Normalized Laplacian $\begin{cases} L_{sym} = I - D^{-1/2} A D^{-1/2} \rightarrow \text{nCut} \\ L_{rw} = I - D^{-1} A \rightarrow \text{Ratio Cut} \end{cases}$

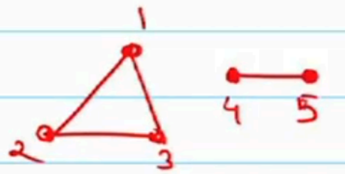
Spectral Clustering.

So, let us take a look at this particular example where we have this kind of graph. So, let me redraw this graph or a version of it. So, the algebraic multiplicity is just 1. So, if I have this kind of graph is the graph connected no right. So, this is a disconnected graph and in fact, number of connected components is 2 ok.

Now, what does the adjacency matrix look like? So, you have a block diagonal kind of structure here right. And degree matrix anyway it is a diagonal matrix. So, degree matrix here is 2, 2, 2 and 1 and 1 in all zeros right. So, degree matrix anyway a diagonal matrix. So, if I look at the Laplacian which is d minus a, it is going to have this block diagonal So, now if I multiply this Laplacian with the vector of let us say 1, 1, 1 and I do not care what are the entries here, 0.

So, right or if I multiply this with the this and this again you will get a 0 right. So, how many 0 eigenvalues are there? 2 So, 0 is not a simple eigenvalue anymore and there are 2 corresponding eigenvectors. So, in fact number of. So, algebraic multiplicity of 2 of the of 0 eigenvalue that is equal to the number of connected components algebraic multiplicity of 0 eigenvalue of L is equal to number of connected components. So, you would have to yeah I mean it would be something else I am saying that.

$L \mathbb{1} = 0 \longrightarrow$ Is 0 a simple eigenvalue?



$$A = \left[\begin{array}{c|c} A_1 & 0 \\ \hline 0 & A_2 \end{array}\right]$$

Disconnected graph

connected components is 2.

$$D = \begin{bmatrix} 2 & & & & 0 \\ & 2 & & & \\ & & 2 & & \\ & & & 1 & \\ 0 & & & & 1 \end{bmatrix}$$

$$L = D - A = \left[\begin{array}{c|c} L_1 & 0 \\ \hline 0 & L_2 \end{array}\right]$$

or what you can do is you can you can use 0 0 you just use 0 0 you can not start you just use 0 0 here ok. So, these are two different eigenvectors for the same eigenvalue 0. So, the algebraic multiplicity is 2 and the number of connected components is also 2 and that is what Laplacian kind of tells you right. So, in fact you can show that So, all the eigenvalues, so first of all Laplacian is a positive semi definite matrix, all eigenvalues are. So, in fact, let me define 0 to be the first eigenvalue, because it is anyway the smallest eigenvalue, this is less than equal to lambda 2 and so on.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Algebraic multiplicity of 0 eigenvalues of L = # of connected components.

Eigenvalues of Laplacian:

$$0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$$

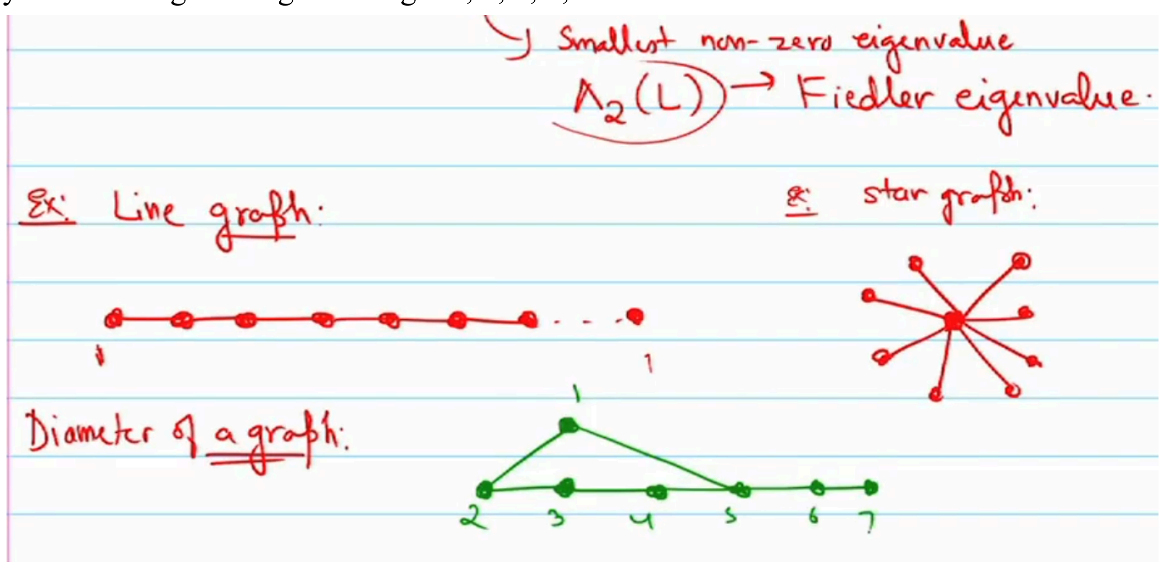
λ_2: Smallest non-zero eigenvalue

So, smallest non-zero eigenvalue. So, this is called. So, we are going to usually denote this by lambda 2 because it is a second small eigenvalue lambda 2 of L and this is also called Fiedler eigenvalue. and this directly tells you about the connectivity of the graph. So, is the graph well connected not well connected.

So, what do we mean by the connectivity of the graph. So, even if let us say the graph is fully connected right or graph is connected ok. So, there are multiple scenarios right. So, if you if I consider examples let us say a line graph which looks something like this.

and so on. So, this is a line graph. So, is this a connected graph? But what do you think like if I were to propagate information from one node to another node? Is it going to take a lot of time? Because for the information to travel from here to here, it is going to take a lot of steps. Whereas, if I consider another graph, let us say star graph. something like this. Now, in two steps you can reach from one node to any other node in just two steps.

So, this graph you are going to say that this graph is better connected than the previous one, the line graph is well connected because you can reach from you can propagate information from one node to any other node in fewer steps and this gets captured through this Fiedler Eigenvalue. So, what is diameter of a graph? So, what is the graph diameter? So, let us take an example I think it will be much clearer through an example. Suppose this is your graph ok and let me. So, in how many steps can you reach from let us say node 1 to node 2? So, 1 to 2 you can reach in one step 1, 2, let's say 3, 2 steps. But you can also go through here right 1, 2, 5, 4, 3 and so on.



But then we consider the minimum value because at we are guaranteed to reach in 2 steps that way right. And likewise so you basically enumerate all pairs all i, j pairs right. And then you look at the maximum entry and that is the diameter of the graph. So, for instance from here like if I were to go from 3 to 7. So, either I mean it will take 1, 2, 3, 4 and that is the diameter of the graph because from 2 let us say from 2 to 7 if I were to reach I can go 2 to 1, 1 to 5 again 4 steps right.

So, between any 2 pair of nodes or between any pair of nodes it will take at least 4 steps right for the information to propagate. So, diameter of the graph is. So, graph diameter is equal to 4. in this case diameter is equal to 2 because you can reach from one node to any. So, diameter here is equal to 2 and likewise this diameter is not in I mean let us say I mean there are n nodes n minus 1 right.

So, diameter is also a notion of it also captures the notion of connectivity in the graph. So, how many steps it is going to take to propagate one piece of information from like between any two pair of nodes, but as you can see that we have to compare all n square possible combinations right. Is there a better way through which we can capture and it turns out yes. So, there is a result by Brandon McKay which says that that the Fiedler value and it is not an equality kind of thing, but it gives you an idea quote this is for connected graphs. So, the diameter is large then you generally expect I mean even though it is greater than equal to, but it is it is I mean you can assume that this is fairly satisfies with equality.

Brendan Mckay Fiedler value $\geq \dfrac{4}{n\,(\text{diameter})}$ $\Big\}$ → for connected graphs

So, if the diameter is large then the Fiedler Eigen value is going to be small and vice versa. So, for graphs which are well connected that means the diameter is small the Fiedler value is going to be large ok. Underrated yes n is the number of edges number of nodes. The other thing that we said about L is that Laplacian is positive semi definite. So, what does it mean lambda of any eigenvalue of L that should be greater than equal to So, let us try and prove this.

So, if we want to show that this is positive semi definite then for any vector x. So, we have to show that x transpose L x this is greater than equal to 0 right for positive semi definiteness ok. this is x transpose L x right. So, how can we show positive semi-differences? So, what property do we know about Laplacian? It is symmetric and so, let me write it this way right. So, I can say and if you were to write L because somehow we want to use the fact that summation of any row like sum of any row should be equal to 0.

So, what is this, what is summation j 1 through n d ij? Sorry, d ij xj. So, d ij xj that is fine and d ij is again going to be summation a ij right. So, I think what we can show from here is by the way d i j's are non-zero only if i equal to j right. So, this term is equal to x i square d i right and this I can again write this is this I can write as half x i square d i, I can just use it like because i is just a dummy variable or dummy notation like dummy index, I can write this as x j square d j minus. So, this turns out to be half because summation di is I mean di is nothing but summation a ij like summation j a ij.

* L is a positive semi-definite matrix $\Rightarrow \lambda(L) \geq 0$.

$$x^T L x \geq 0$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i l_{ij} x_j$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i (D_{ij} - A_{ij}) x_j$$

$$= \sum_{i,j=1}^{N} x_i D_{ij} x_j - \sum_{i,j=1}^{N} x_i A_{ij} x_j$$

$$= \sum_{i=1}^{N} x_i^2 d_i - \sum_{i,j=1}^{N} x_i x_j A_{ij}$$

So, you can show that this is nothing but through n a ij x i minus x j whole square which is greater than equal to 0 and 0 on if and only if x i is equal to x j. So, that means like this. So, first of all it is positive semi definite matrix, 0 is one of the eigenvalues with the vector mean with when x i is equal to x j for all i j. So, that means vector of 1 is one of the eigenvectors as well right if the graph is connected. So, that is that shows that this particular order of eigenvalues and the smallest nonzero eigenvalue is your field of value.

$$= \frac{1}{2} \sum_{i=1}^{N} x_i^2 d_i + \frac{1}{2} \sum_{j=1}^{N} x_j^2 d_j - \sum_{i,j=1}^{N} x_i x_j A_{ij}$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} A_{ij} (x_i - x_j)^2 \geq 0$$

In fact, you will see later that when we prove the convergence of different distributed optimization algorithms, it actually depends on the Fiedler value. So, for instance, if I try to run a consensus here, you can expect that it is going to be much faster than running a consensus here. because then it the information would have to travel at least n minus 1 steps for you to be able to run any consensus. Here you can achieve consensus in two steps. So, your algorithms they are going to be dependent this convergence rate of the algorithms are going to be dependent on the Fiedler value and that is something we are going to be looking at in the subsequent lectures. Thank you.