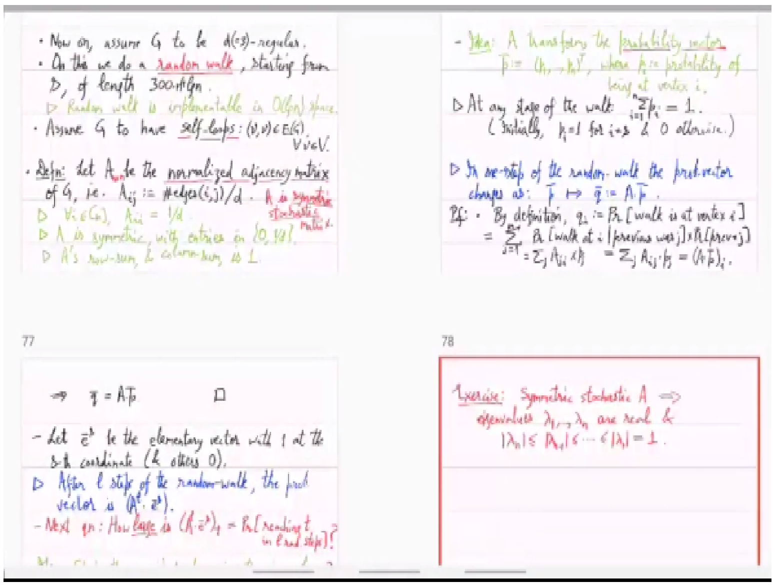
Randomized Methods in Complexity
Prof. Nitin Saxena
Department of Computer Science & Engineering
Indian Institute of Technology-Kanpur

Lecture - 10
Graph Expansion Properties

(Refer Slide Time: 00:18)



Last time we started this new topic called expanders or the concept is called expansion. So the theorem that we are proving and it will take some time to finish the proof is by AKLLR 1979. So in this theorem we will show that if you are given an undirected graph and vertices s and t, s for source t for target, you want to check whether there is a path from s to t okay.

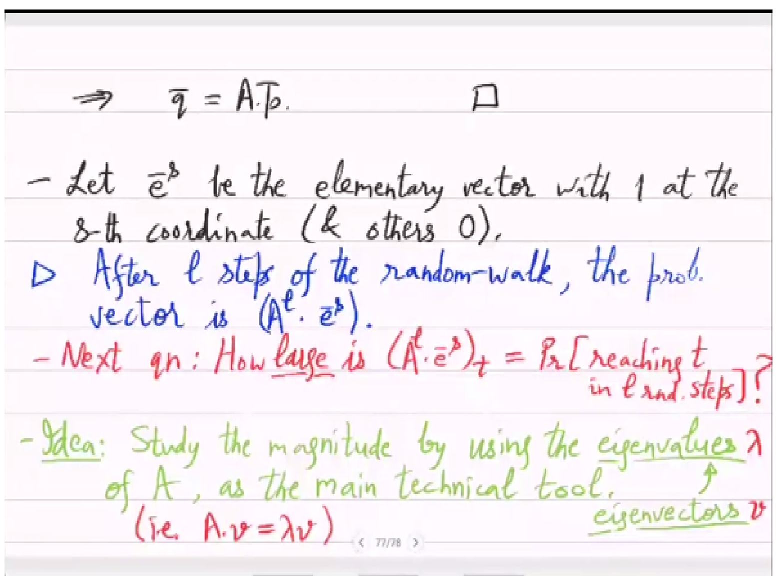
And in fact the, I mean in the end you will see that you can also output the path and everything can be done using only logarithmic space in the work tape, okay. So input size is n, output size is also n, but the workspace is only log n or order log n. And the idea is very simple. The idea is just that you start your random walk from vertex s, which means that whatever neighbors you see you pick a random one and then proceed okay.

So obviously, the main point here will be calculating the success probability, okay. So if s and t are in the same connected component, what is the success probability of this random walk reaching t from s in l steps where l is something like this 300 times n the 4 log n. So what we have done till now is we defined this normalized adjacency matrix.

So normalized adjacency matrix is this 0,1 matrix but you will divide by d where d is the degree of each vertex in the graph. Remember, we have already made our graph regular by a simple transformation. So A is a symmetric stochastic matrix. So it has only real eigenvalues, right? This I left as an exercise and the maximum eigenvalue is in magnitude 1. Next important thing we defined was the probability vector.

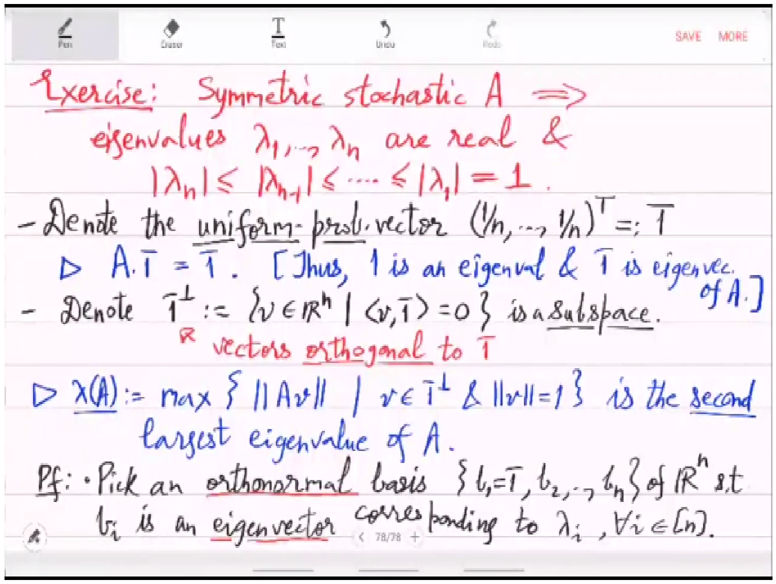
So initially the probability vector is all concentrated on the vertex s. So the probability of being there is 1, everywhere else it is 0. And as you take step, so every step is multiplying A with the probability vectors, okay. This stochastic matrix is the kind of the probability density matrix and this acts on the probability vector to get a new probability vector, which will be where you will be with what probability after one step and two step and so on.

(Refer Slide Time: 03:08)



So we were then looking at this question in red, how large is $(A^{l}.\bar{e})_t$ place, okay. So we are interested in the probability of reaching t in l steps. So that is exactly this question $(A^{l}.\bar{e})_t$ coordinate and we will study it by eigenvalues and eigenvectors of A.

(Refer Slide Time: 03:36)



So let us start that study. So first you denote the uniform probability vector which is $(\frac{1}{n}, \ldots, \frac{1}{n})^T$ So this is we call it the uniform probability vector because this is just equal probability of being anywhere in the graph, which is 1 / n. And this is the ideal. So you want to reach this state, right? So after l steps you want to reach a state wherein probability of being at any vertex is equal, which is 1 /n.

This ideal vector we will call $\bar{1}$ and observe that A. $\bar{1}=\bar{1}$ okay. So what this means is thus 1 is an eigenvalue And $\bar{1}$ is eigenvector of A. Okay, this is the first property you learn. And we will also be interested in the vectors that are orthogonal to $\bar{1}$. So orthogonal means that $\bar{1}^{\perp} := \{v \in \mathbb{R}^{n} | \langle v, \bar{1} \rangle = 0\}$ So notice that this means, in particular that v has negative coordinates and positive coordinates. So this is not really a probability vector. It is just a subspace, okay. This is a subspace, vector space. So this is called vectors orthogonal to $\bar{1}$, okay orthogonal vectors. And one nice property that we will show is the action of A on this subspace $\bar{1}$ orthogonal, okay. That is very nice.

And if you look at the maximum length of Av for every v in this space that will give you the second eigenvalue which is λ_2. So define $\lambda(A) := max\{ ||Av|| \ | \ v \in \ \bar{1}\perp$ & $||v|| \ = \ 1\}$So if you look at the unit vectors in this space and apply A on that look at the length pick the maximum length, call it $\lambda(A)$, this is the second largest eigenvalue.

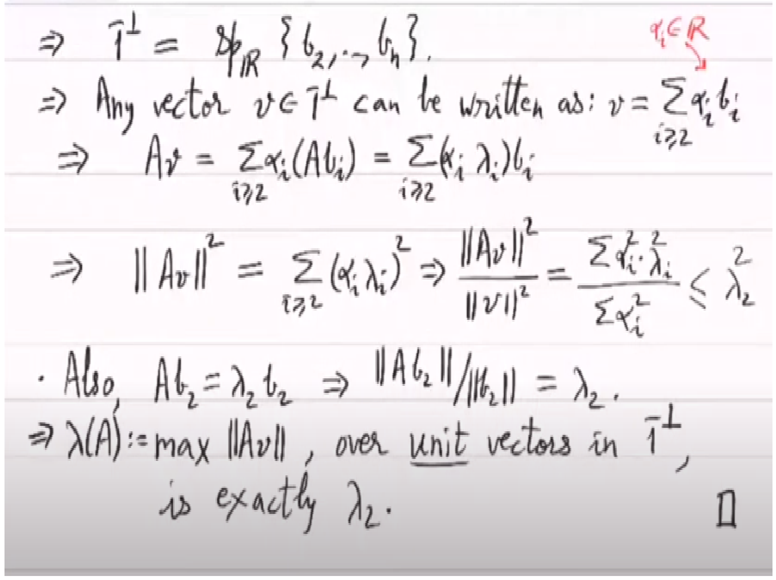
So why is that why is this thing related toλ_2, in fact equal to λ_2. What is the reason? So this will lead, this will need a proof. So first you identify a basis of this subspace. So pick an orthonormal basis. Let us take b_1 to be $\bar{1}$ because that is an eigenvector of A. And then you take b_2 and b_n. So orthonormal basis $b_1$to b_n such that b_i is, in fact this is a basis of $\mathbb{R}^n$, okay.$\{b_1 = \bar{1}, b_2, \ldots, b_n\}$ of $\mathbb{R}^n$

Basically we are picking n linearly independent vectors, real vectors The first one is $\bar{1}$ and each of these b_i is a is an eigenvector such that b_i is an eigenvector of corresponding toλ_i $\forall\, i \in [n]$ so we have picked this orthonormal basis. So these are mutually orthogonal. Moreover, they are unit vectors they are eigenvectors of A, okay.

This can be found this exists and it also can be found it exists mainly because of this above exercise A is symmetric stochastic. So the eigenvalues are real and then when you look at the linear system $Av = \lambda Av$, v will be a real vector okay and then you can also pick them to be mutually orthogonal and unit vector.

So this, we fix this v_1 to b_n and what happens is then by definition other than $\bar{1}$, so b_2 to b_n, if you look at b_2 ,b_3 ……. b_n, they form a orthonormal basis of $\bar{1}$orthogonal, okay. That is now by construction.

(Refer Slide Time: 11:09)



So now $\bar{1}\perp = sp_{\mathbb{R}}\{b_2, \dots, b_n\}$. So which means that any vector $v \in \bar{1}\perp$ in this space can be written as $v = \sum_{i\geq 2} \alpha_i b_i$ and now you can study the action of A on this $\bar{1}$ orthogonal space. So $Av = \sum_{i\geq 2} \alpha_i (Ab_i) = \sum_{i\geq 2} (\alpha_i \lambda_i) b_i$

So $\alpha_i \lambda_i$ is a constant, b_i is a vector. So this is the combination of Av which means what?

Now remember that b_i's are orthogonal, right? So this is actually an orthogonal, it is a linear combination of orthogonal vectors. So it behaves very well with the Euclidean norm. So the, if you look at the length of this vector square, this is exactly equal to $(\alpha_i \lambda_i)^2$, okay because b_i's are orthogonal and unit vectors.

So the length can be expressed as just this sum of squares, which means that so we did not pick v to be a unit vector but we can make it unit by just dividing by $||v||^2$ and that will be $\frac{\Sigma \alpha_i^2 \lambda_i^2}{\Sigma \alpha_i^2}$.

$$||Av||^2 = \sum_{i\geq 2} (\alpha_i \lambda_i)^2 \Rightarrow \frac{||Av||^2}{||v||^2} = \frac{\Sigma \alpha_i^2 \lambda_i^2}{\Sigma \alpha_i^2}$$

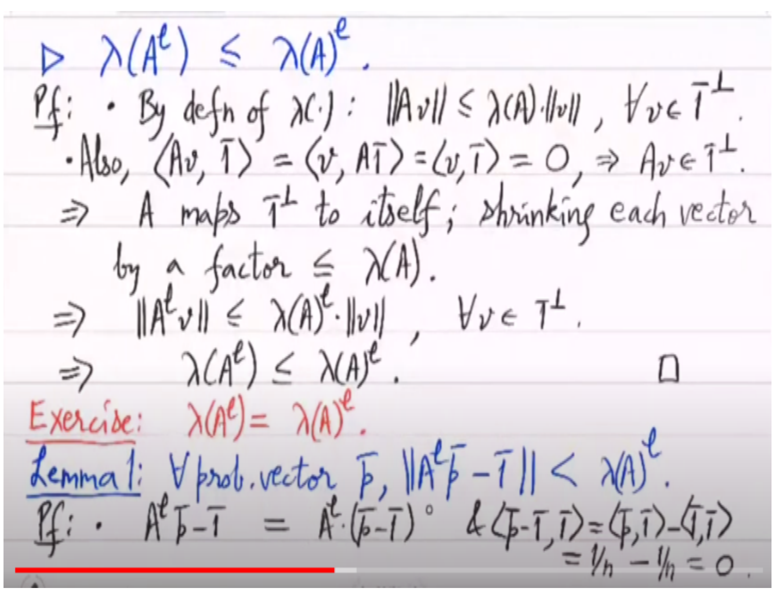
So we are just taking a convex combination of λ_i^2, okay.

Now when you take this convex combination or when you take this kind of weighted average you cannot exceed the maximum. So a maximum is λ_2, okay. So this A's action on $\overline{1}$ orthogonal unit vectors right this cannot give a scaling more thanλ_2. Also $Ab_2 = \lambda_2 b_2 \Rightarrow ||Ab_2||/||b_2|| = \lambda_2$, right?

So what we have is that in $\overline{1}$ orthogonal space there is a unit vector whose scaling is exactly λ_2 by A and no other vector can exceed scaling of λ_2. So this overall means that $max\ ||Av||$ over unit vectors in $\overline{1}\perp$ orthogonal is exactly λ_2 okay. So this proves this theorem that $\lambda(A)$which was the max scaling, this is exactly λ_2. And this we are calling $\lambda(A)$.

So lambda A is exactly λ_2, okay. This is a very useful property that we have just shown. So we are still just studying eigenvalues of A right. So next property that we will prove is remember we wanted to study the action of A^l. So how do the eigenvalues and especially the second largest eigenvalue, how does this change as you power A. That is the next property to study.

(Refer Slide Time: 16:58)



So what we will show is that $\lambda(A^l) \leq \lambda(A)^l$. This is again a cute property. So by definition of λ: $||Av|| \leq \lambda(A)||v||$, $v \in \overline{1}\perp$. That we just showed this, the max cannot exceed $\lambda(A)$. So every action Av divided by v is this$\lambda(A)$ in magnitude.

Also, so what we are doing here is that from this inequality we will now study say the action of A^2. What is A^{2v} So that we can see as action of A on Av. But then what is Av? So we actually saw in the previous proof that Avremains in $\overline{1}\perp$, okay. So in that sense we can apply this inequality a second time. So in so formally speaking this Av is orthogonal to $\overline{1}$.

Why because $< Av, \overline{1} > \; = \; < v, A\overline{1} > \; = \; < v, \overline{1} > \; = 0 \Rightarrow Av \in \overline{1}\perp$ okay. This is a shorter proof than the previous one. So what we have learned is that A maps $\overline{1}\perp$ to itself, okay. So A is a map from $\overline{1^{\perp}}$ to itself. And the good thing about this is then A^2 is also a map from $\overline{1^{\perp}}$ to itself and A^las well.

And the shrinking that it does, so shrinking each vector by a factor $\leq \lambda(A)$. So if you apply two times it will be $\lambda(A)^2$. Three times or$l$ times then it will be $\lambda(A)l$ okay. That is the proof; for every v in $\overline{1^{\perp}}$. This means that$\lambda(A)l$ is at most this factor that you have shown above, okay.

So this kind of $\lambda(A^l)$ does not exceed $\lambda(A)^l$. That is what we wanted to show. In fact, as an exercise, you can also show that they are equal, but I would not need that. You can show this because basically, you look at the eigenvalues of A^l in terms of λ_1 to λ_n. And you can show that the eigenvalues are exactly l-th powers of those, okay.

So the order of the eigenvalues that you started with will remain the same. So it is actually a very strong relationship. Powering of a matrix is very nicely related to the original matrix in terms of eigenvectors and eigenvalues. So these were the basic properties of eigenvalues. And now we will go back to questions that are more relevant to the theorem that we want to prove, which is what is the action of A^l on probability vector.

So let us prove it in two lemmas. First lemma is for all probability vector, for every probability vector$\overline{p}$,$||A^{l}\overline{p} - \overline{1}|| < \lambda(A)^{l}$ okay. And since $\lambda(A)$is a fraction as you walk more and more in a random fashion in the graph, you approach the uniform probability.

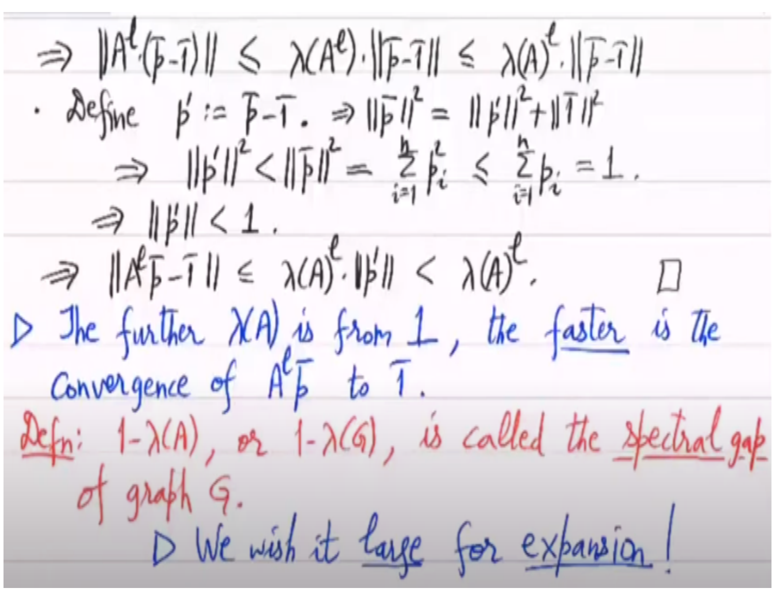
So the way we will show this is again we will use this above property of λ. So we have to see $A^{l}\overline{p} - \overline{1}$as an action of A$l$ on some vector. So for that simply observe that

$||A^{l}\overline{p} - \overline{1}||$=$||A^{l}\overline{(p} - \overline{1})||$. Because, these two vectors are exactly equal, $\overline{1}$ =$A^{l}\overline{1}$

then you repeat this l times. And now $\overline{(p} - \overline{1})$ is orthogonal to $\overline{1}$, you can show that. And because of that okay, let us just claim it and $<\overline{p} - \overline{1}, \overline{1}> = <\overline{p}, \overline{1}> - <\overline{1}, \overline{1}>$ So remember that $\overline{p}$ is a probability vector. So the inner product with 1 by n will give you overall 1 /n.

And the second thing for the same reason 1 / n. So they are equal okay, it is 0. So which means that A^{l} is actually acting on a vector which is orthogonal to $\overline{1}$. So you can use $\lambda(A^{l})$ scaling factor.

(Refer Slide Time: 25:21)



So this means that $||A^l(\overline{p} - \overline{1})|| \leq \lambda(A^l) . ||\overline{p} - \overline{1}|| \leq \lambda(A)^l . ||\overline{p} - \overline{1}||$ So we will actually show that this is less than 1. So we can drop it.

So let us do that. So define$p' := .\overline{p} - \overline{1} \Rightarrow \ ||\overline{p}||^2$ So you can write it as sum of squares $p' := .\overline{p} - \overline{1} \Rightarrow \ ||\overline{p}||^2 = \ ||p'||^2 + ||\overline{1}||^2$. So which means that $||p'|| < ||\overline{p}||$ right. Because this $\overline{1}$s length is positive.

So $||p'||$ is then strictly smaller than $||\overline{p}||$ which is or let me continue using squares. So $||\overline{p}||^2 = \sum_{i=1}^{n} p_i^2 \leq \sum_{i=1}^{n} p_i = 1$. So what we have learned is that $||p'|| < 1$.

So which means that this $||A^l\overline{p} - \overline{1}|| \leq \lambda(A)^l . ||p'|| < \lambda(A)^l$. So after l steps of the random walk the difference between your probability distribution with the uniform probability distribution is λ^l. λ is a fraction so it keeps on decreasing. So you are reaching the uniform distribution, okay.

So this is a very natural result, but we have obtained it with a great number of calculations and matrix analysis. But there is an easy way to read this, there is an intuitive way. So what you learn from this the qualitative reading of this equation is the further $\lambda(A)$ is from 1, okay. So in other words the smaller $\lambda(A)$ is, it is a fraction, but the smaller it is.

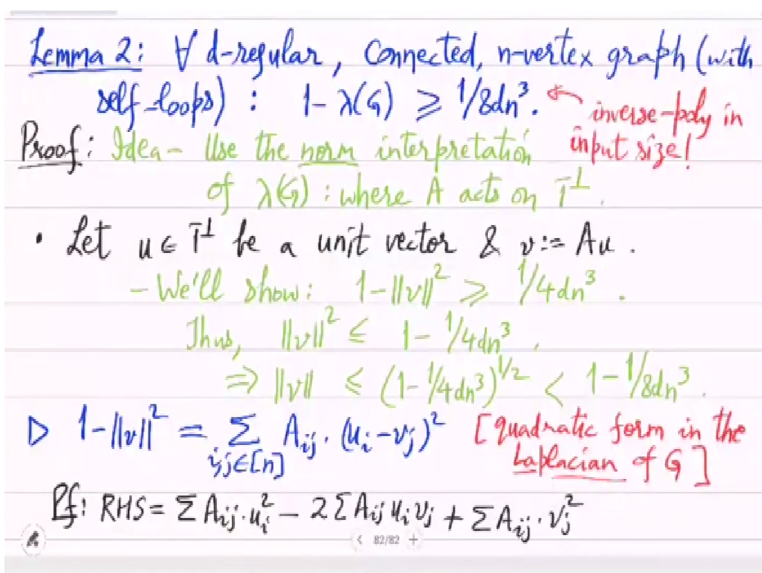
So the further away it is from 1, the faster is the convergence $A^l\overline{p}$to$\overline{1}$ to the uniform distribution, okay. And this is what motivates spectral gap, the notion of spectral gap. So for a graph like this given by adjacency matrix, normalized adjacency matrix A, the spectral gap is
1-$\lambda(A)$ or 1- $\lambda(G)$ and we want it to be large. So 1 -$\lambda(A)$ or we can equivalently define it for the graph as well in the same way.

This is called the spectral gap of the graph G and we wish it large, okay. So two things we have defined here, two major concepts in graph, in spectral graph theory. First is the spectral gap and the larger it is, we want it to be large because the larger it is expansion in the graph is rapid, okay. And the reason is this inequality we showed that after l steps the error term from uniform distribution is λ^{l}.

So if λ is very small l is large then this converges very rapidly to 0 okay making your reachability everywhere with equal chance, okay. So this is a major point in our understanding, but we are still not done because remember we wanted to fix the value of l in the algorithm, right? So how long should we walk or how far should we walk? How many times should we pick these random neighbors?

So for that actually we have to prove a bound on the spectral gap, okay. So no matter what graph you are given, because your input is an arbitrary graph, what can you say about the spectral gap? How large is it? So we will show that it is sufficiently large for every graph.

(Refer Slide Time: 33:40)



That is lemma 2. And then we will be done. So for every d-regular connected graph G(with self-loops). In the proof you will see how we will use connectedness and these self-loops, okay. It will appear in some equations estimates. So for any such graph, spectral gap $1-\lambda(G) \geq 1/8dn^3$. So somewhere we should say n is the number of vertices connected n-vertex graph with self-loops okay.

So every vertex has degree the same which is d and n vertices. It is connected and there are self-loops. Then the spectral gap is $1/n^3$. The reason why we are calling it large is because it is inverse polynomially away from 0. So in contrast to this, this spectral gap could have been extremely close to 0, but it is not, okay. It is inverse polynomially away from 0, so we call it large.

It is inverse poly in the input size. That is why we call it large. It is not inverse exponential, it is not $1/2^n$. So this is a good thing and the proof will be tricky. So first thing we will do is we will replace this $\lambda(G)$ by as a norm on $\overline{1}^{\perp}$ space, okay. So remember that $\lambda(A)$ is the norm can be thought of as the maximum scaling that or maximum, yeah maximum scaling that you get when A acts on $\overline{1}^{\perp}$, right.

So we will use that interpretation that is the idea. So use the norm interpretation of $\lambda(G)$when A acts on $\overline{1}^{\perp}$. So that is the basic idea, but then you have to do a number of calculations to actually get to this $1/n^3$. Okay, so let us start that. So you can write, so let $u \epsilon \overline{1}^{\perp}$ be a unit vector and $v := Au$. And then we will look at the stretch or the scaling that has happened when you went from u to v.

So we will show that $1 - ||v||^2$ is large.$1 - ||v||^2 \geq 1/4dn^3$ And once you have shown this you will get that, in other words we are showing that$1 - ||v||^2$ is small. So $1 - ||v||^2 \leq 1/4dn^3$ and $||v|| \leq (1/4dn^3)^{1/2} < 1/8dn^3$. So this is the sequence of some of the basic kind of landmarks we will have in the proof. We will start with $1 - ||v||^2$. We will show that this is small.

And because this$||v||$ will be smaller than 1 $-1/8dn^3$. Now this will mean that 1 -$||v||$ is large, okay. And 1 -$||v||$ is actually 1 -$||Au||$, which will exactly tell you about$\lambda(G)$ So you will have the statement in lemma 2, fine? So we will now focus on$1 - ||v||^2$ for this reason. So the first claim is that$1 - ||v||^2$ has a nice expression, okay.

It has a nice expression, which is why we actually chose it. So the expression will be $1 - ||v||^2 = \sum_{i,j\in[n]} A_{i,j}.(u_i - v_j)^2$. So you can write this 1 minus square as sum of squares with the coefficients being $A_{i,j}$. So this is called, in the literature it is called quadratic form in the Laplacian of G, okay. So this equation has a fancy name, it is actually related to the Laplacian of the graph. It is a quadratic form in that, okay.

So remember$v$ is Au. So u has n coordinates v has n coordinates and what this quadratic form is capturing is it is actually sum of squares for all possible differences across, right? So $u_i - v_j$. So this is what we want to show. So once we show this, then the strategy or the plan would be to show that this $u_i - v_j$ for some i, j is large, okay. We will show that for some i, j it is large and hence, the left hand side is large.

That would be the eventual plan. So let us first prove this. So this is just a calculation on the right hand side. So $R.H.S = \sum A_{i,j}u_i^2 - 2\sum A_{i,j}u_iv_j + \sum A_{i,j}v_j^2$ Then we will calculate these three terms separately. So this is equal to first term you sum up over i.

(Refer Slide Time: 43:40)

$$= \sum_i \Big(\sum_j A_{ij}\Big)\cdot u_i^2 - 2\cdot\langle Au, v\rangle + \sum_j \Big(\sum_i A_{ij}\Big)\cdot v_j^2$$

$$= \sum u_i^2 - 2\langle Au, v\rangle + \sum v_j^2 = ||u||^2 - 2\langle v, v\rangle + ||v||^2$$

$$= 1 - ||v||^2 = LHS. \quad \square$$

- Thus, it suffices to show: $\exists i, j,\ A_{ij}\cdot(u_i - v_j)^2 \geq 1/4dn^3$.
- If $\exists i,\ (u_i - v_i)^2 \geq 1/4n^3$ then we are done.
  - So, assume: $\forall i,\ |u_i - v_i| < 1/2n^{1.5}$.
- Sort the coordinates of u: $\underset{\geq 0}{u_1} \geq u_2 \geq \cdots \geq \overset{\leq 0}{u_n}$.
  - ▷ $\sum u_i = 0$ & $\sum u_i^2 = 1$.
    - $\Rightarrow$ either $u_1 \geq 1/\sqrt{n}$ or $u_n \leq -1/\sqrt{n}$.
- ▷ $u_1 - u_n \geq 1/\sqrt{n}$.

So you get this$\sum_i(\sum_j A_{i,j})u_i^2$. It is basically the i-th row sum or sum of the i-th row in A, right the second was this . This is the cross term. This is actually related to the inner product <Au,v>plus a symmetric thing like the first one$\sum_i(\sum_j A_{i,j})v_j^2$. And as you can guess, this will simplify because A is stochastic. So this $\sum_j A_{i,j}$ is 1 and then you get $\sum u_i^2 - 2 < Au, v > \ + \sum v_j^2 = ||u||^2 - 2 < v, v >+ ||v||^2$.

But u we have assumed to be unit vector, right? So this is$||u||^2$ is 1. So 1-$||v||^2$as claimed, okay. So this is equal to the left hand side; that is all. So now the strategy as I said is thus it suffices to $\exists i, j, A_{i,j}$... $(u_i - v_j)^2 \geq 1/4dn^3$, okay.

So we will show it next. Here we will use the connectedness and self-loops, okay. We will get this and once we have$1/4dn^3$ for some time note that all these terms are actually non-negative, okay. So the sum cannot be smaller than that. So that will give you the lower bound on 1-$||v||^2$, right. So remember that we have added, we assume that there are self-loops, so we can also focus on $u_i - v_i$ right? What is happening in those terms?

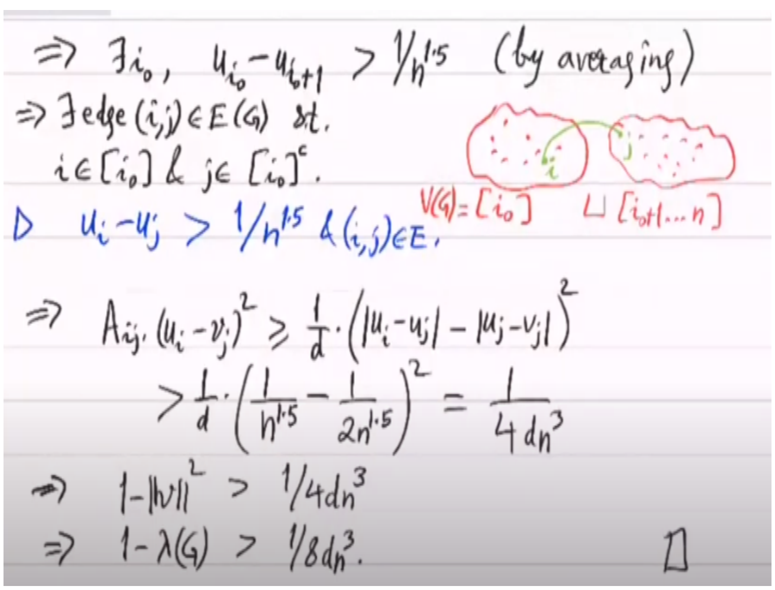
So if $\exists$ i, the self-loop term, so$(u_i - v_i)^2 \geq 1/4n^3$. So if there is an i such an i, then you are done already. So the bad case is when for all the i's this $(u_i - v_i)^2 < 1/4n^3$. This is $|u_i - v_i| < 1/2n^{1.5}$ for every i. Then we cannot pick these, the self-loop contributions.

Then we have to look at edges that are not self-loops, right? So there actually we will use connectedness. So what you do is remember $\bar{u}$was a unit vector. So let us first sort the coordinates, sought the coordinates of u. So what you can say is $u\colon u_1 \geq u_2 .... \geq u_n$. And we have this information about u_i's that first is that since it is orthogonal to $\overline{1}$, $\sum u_i = 0$.

And since it is a unit vector, $\sum u_i = 1$. So u 1, if you look at the magnitude, either u_1 is the biggest or u_n is the biggest, right? So either u_1^2 is the largest or u_n^2 is the largest. So one of them has to be at least 1 / n just by averaging, because their sum is 1. So one of them has to be at least 1 /n. So what you learn is that either $u_1 \geq 1/\sqrt{n}$ *or* $u_n \leq -1/\sqrt{n}$.

And because the sum has to be zero, so for positives, there will also be negatives. And you know that u_1 is non-negative. And if u_1 is positive then u_n will be negative. And with the sum of square being 1, you get either u_1 bigger than $1/\sqrt{n}$ or u_n smaller than ($-1/\sqrt{n}$). So these are the properties you know. So you know that there is a gap between u 1 and u n. And the gap $(u_1 - u_n)$is at least $1/\sqrt{n}$, right? We have gotten a gap, let me deduce the following.

(Refer Slide Time: 51:28)



$\exists i_0, u_{i_0} - u_{i_0+1} > 1/n^{1.5}$ because between the extremes the difference is $1/\sqrt{n}$. So there will be some position where the consecutive differences, difference of the consecutive u_i's or u_{i_0} is $1/n^{1.5}$ or more. Again by averaging argument.

So now look at vertices 1 to i_0. So we wanted this consecutive because we will have some information about u_1 to u_{i_0}, it is in sorted order. And we will have information from u_{i_0+1} to u_n, which is again sorted. So look at these two parts in the graph. So 1 to i_0 and $i_0 + 1 \dots n$, okay. Look at these vertices. This is the whole graph, I mean the vertex set.

And since the graph is a connected graph, there will be some way to go from this part to that part, right? It is not disconnected. So there is some edge i to j. There is some bridge between these two connected components. So pick this bridge and look at this cross term in the sum. So there exists an edge(i, j)$\epsilon E(G)$such that $i\epsilon[i_0]$ & $j[i_0]^c$. So which means that u_i is more than u_{i_0}, right?

Andu_j<u_{i_0+1}. So which means in particular that if you look at the difference$u_i - u_j$ this cannot decrease, okay. So this is a very good thing. We have identified that the difference is large, and it is an edge. We have identified an edge where the difference is large, okay. So now just focus on this term. So this means, if you look at the term $A_{i,j}(u_i - u_j)^2 \geq 1/d\ (u_i - u_j) - |u_j - v_j|)^2$. This you can say because this operator's property, look at this. So the norm of $|u_i - v_j|$ it is at least the difference of the $|u_i - v_j|$ and $|u_j - v_j|$. And then you use the above property in blue. So $|u_i - u_j|$ is large.

what do you know about $u_j - v_j$? So the nice thing is that we know you already assumed something about $u_j - v_j$which was the self-loop thing we did, right? We had assumed that all these$u_i - v_i$'s are small. So let us invoke that. That is a useful property and then square. So which comes out to be $1/4dn^3$,

$A_{i,j}(u_i - u_j)^2 \geq 1/d\ (u_i - u_j) - |u_j - v_j|)^2$>1/d($1/n^{1.5} - 1/2n^{1.5}$ $)^2 = 1/4dn^3$

which means that we are almost there, right?

So we have shown that in fact$1 - ||v||^2 \geq 1/4dn^3$, in fact strictly because we have shown and we showed this only by just by using one term, right? So sum will possibly be bigger but cannot be smaller than this and from which we will deduce by the stretching that A is doing. By that will deduce something about λ(A). It is same as λ(G), right?

1-λ(G) >$1/8dn^3$

That was the original claim. That is what we wanted to show. So 1-λ(G) is greater than that. So these are all strict. Okay, no so we have greater than equal to here. So this will probably, this creates a problem. Oh it is too much. Okay anyway, so the proof is done.