Chapter - 02
Lecture - 09
Lecture - 04

So far in this chapter on Network Measure, so we have discussed degree you know degree distribution, we have also looked at different measures which are roughly divided into 3 buckets, right; microscopic, macroscopic and mesoscoping. We have seen metrics like clustering coefficient, local clustering coefficient, global clustering coefficient, and connected components, right; strongly connected components, weakly connected components.

So, here we will discuss another very interesting aspect of network measure, it is called centrality. Now, this is widely used, right.

(Refer Slide Time: 01:03)

Centrality is basically a measure of how central a particular node is with respect to the network, right. So, you can think of it, I mean if I say that you know to check whether a node is central to a particular network, so the first thing that comes to our mind is whether the node is node has equal distance to all the other nodes, right.

Say, if a node has equal distance to all the other nodes, more or less equal distance to all the other nodes, you can say that you know that node is basically a central node. But we are not

talking about that kind of centrality here, we are talking about. Of course, this is one aspect of centrality, but there are other notions of centralities.

And remember the notion of centrality depends on the particular application. Say, if the application is say outlier detection, you have different notion of centrality. If the application is information spreading, you have another type of centrality. If the application is you know vaccinate nodes, you have another type of centrality and so on and so forth.

So, in general when we talk about centrality I generally look at it you know using this 3 these 4 Ps, right. So, the first P is the prestige. We want to quantify prestige in the formulation of centrality. We will try to quantify prominence in the formulation of centrality. We quantify importance of a node in the formulation of centrality and then power, ok. So, these 4 Ps are something that we will try to incorporate in the definition of a centrality, ok.
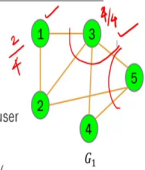
(Refer Slide Time: 02:46)

So, let us look at the first centrality, the basic one, the first centrality measure which is called the degree centrality. So, we know what is degree? Degree is basically the number of edges adjacent, number of edges which are incident on a particular node, right.

So, what is degree centrality? Degree centrality is essentially the degree of a node, but it is normalized, right. So, degree centrality of a node v is the degree of the node v and divided by the maximum degree of a node present in the graph, right. You see the max degree and that is the denominator.

So, for example, in this particular graph G 1, if you look at say node 1, the degree centrality of node 1 would be 2 by the max degree is 3, right. So, the max degree is 3 or 4, right, the max degree is 4. So, this would be 2 by 4, ok. Degree centrality of node 3 is 4 by 4 and so on and so forth.

So, as you can understand it basically ranges between 0 to 1, and the more the degree centrality the higher the case that, I mean the higher the likelihood that the node has the node has maximum degree. And therefore, you know in certain applications for example, say you want to identify prominent node or you want to identify celebrities for example, right.

And so, I mean in any application where you use degree as a measure, right, you can use degree centrality. That is a very simple notion of a centrality.

(Refer Slide Time: 04:33)

The second one is called closeness centrality. Remember, all these central centrality measures that we are talking about here, these are all node centric properties, but of course, you can map it to its edge centric as well, right.

So, the next one is closeness centrality. So, as the name suggest we will look at how close a particular node is with respect to other nodes in the graph and depending on that we define this closeness centrality.

So, closeness centrality of a node v is basically, I mean you can basically say that let us look at the shortest path distance, right of say distance of v from the remaining nodes u, where u

belongs to the vortex set V minus the v itself, right because we will not we will not measure the distance of v with respect to itself.

So, with all the other nodes, what is the distance of v and you know you just normalize it in some ways, right. If you do it, that is also ok. But the problem is that we also want that higher the value of centrality better would be the node, right. Higher the value of closeness centrality better would be the quality of the particular node.

Say, think of a node, right which is closer to all the other nodes. So, the sum of the distance would be lower, compared to another node which is farther you know which is farther from the other nodes, right. So, say let us say you have u and v, u is closer to all the remaining nodes, so the distance would be the sum of the distance would be lower whereas, v is farther from the other nodes, so the distance would be higher. Therefore, the closeness centrality value, forget about the denominator, you can use any denominator, right. I will discuss.

So, irrespective of the denominator the closeness centrality value would be higher for node v, because for node v the numerator value is higher, right. But that is not the desired output, right. We basically want that node; the node which is closer to the other nodes should have higher centrality values, right.

So, how can we then you know make to do some sort of tweaks here in this particular formula? We actually reverse it, ok. So, closeness centrality is basically defined by the fraction where the denominator would be the sum of distance, ok and the numerator would be the normalizing factor.

So, in this case the normalizing factor is number of nodes minus 1, mod V minus 1. So, if you do this, in that case, lower the distance lower the sum of distance higher the centrality value. So, the node which is at the center at the center position of a network, which has lower distance with the other with the remaining nodes that will have higher closeness centrality value, ok. So, using the closeness centrality, we can essentially identify nodes which are central to a particular network right.

What is the uses of this particular centrality measure? Now, think about it. So, I am basically trying to identify a node which has small distance with the other nodes, right.

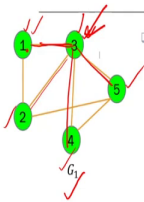So, it means that you know if I want to spread, say, for example, I want to spread up information a particular information to the entire network, ok, I would basically identify a node, right which has higher closeness centrality, right. If I identify that node and if I convince that node to spread the information, it is highly likely that the information will spread across the network faster than the case where the node has lower closeness centrality values, ok.

So, in case of say viral marketing or say spread of information, so spread of fake news misinformation, right in fact, spread of say any sort of epidemics, right. So, you basically I mean from an adversarial point of view, think of it as an attacker you are an attacker, and you want to attack the network by spreading the fake news piece information, you basically choose nodes whose central whose closeness centrality value is maximum, and you basically convince that nodes to spread misinformation, ok.

(Refer Slide Time: 09:35)

So, now let us move to the; so, I have already you know discussed I mean how to formulate closeness centrality. Let us take an example. So, let us say we have G 1 as a graph, right and let us look at the closeness centrality of node 1. So, node 1 if you look at the distance from 1 to 2 the distance is 1, from 1 to 3 the distance is 1, from remember we are interested in shortest path distance, ok. We discussed what is shortest path distance in the last lecture.
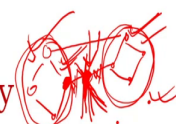
From 1 to 4 we have distance to this and this, right and from 1 to 5 we have distance 2, right. So, the sum of the distance would be 6. Total number of nodes is 5. So, 5 minus 4 is 5 minus 1 would be 4, so 4 by 6 you have 0.67, right.

But what about 3? So, 3 has distance one to all the other nodes from 1, from 2, from 4, from 5, right. So, you see the denominator is 4 and the numerator is also 4, therefore, the closeness centrality of node 3 is 1, right. So, you can also look at it manually that 3 is basically at the center of the network, right, therefore, its closeness centrality is maximum.

So, if you want to spread certain information you basically need to convince node 3, and if node 3 you know spreads some information or to its something, right immediately it is one of neighbours will receive that information, ok.

(Refer Slide Time: 11:15)

Now, let us look at this third notion of centrality which is betweenness centrality. So, this is very interesting. So, as the name suggest we will look at nodes for a particular node, we will basically look at you know possible paths and see whether this that particular node is a part of the paths, right.

So, let us try to quantify it. So, what we will do, we will take all pairs of nodes. There are in-situ pairs we take all pairs of nodes and for every pair for every pair x, y for example, right for every pair x, y we look at the shortest path. Now, remember there can be multiple shortest

paths, right the value of the shortest path would be unique of course, but you will have you can have multiple shortest paths.

So, for every pair x, y I will figure out the shortest paths and I will see whether the particular node v for which I am measuring the between a centrality, whether node v is present in how many of the shortest paths, right. Say, between x and y there are 5 shortest paths, and out of these 5 shortest paths, there are 3 shortest paths where the node v exists. These 3 shortest paths actually move through v, right and that would give you the quantity.

Now, we will do this for all pairs of shortest paths, right. So, let us assume that the sigma xy v is the; so, let us first look at sigma xy. So, sigma xy is the number of shortest paths between x and y, number of shortest paths between x and y. And I am interested in computing the betweenness centrality for node v. So, out of all the shortest paths between x and y, right how many shortest paths, in how many shortest paths this node v exists?

So, this sigma v xy indicates that out of all the shortest paths between x and y, how many shortest in the how many shortest paths between x and y the node v exists, ok. And I basically do this for all pairs, x comma y which is a part of V cross V for all pairs, ok.

Now, if now think about it. If a node has higher betweenness centrality, what does it mean? It means that the node is present in almost all the shortest paths between pairs in the network, right. So, whatever path you want to choose that particular node will always come, that particular node will always you know be encountered. What does it mean? And where this betweenness centrality would be useful?

So, think of, you know think of a vaccination drive, right. So, say an epidemic is going on and you want to you want to stop the spread of epidemic, so and you want to vaccinate certain nodes, ok. Now, think of a think of a network like this, right like this, ok.

Now, think of this node, and let us assume that the epidemic has been detected in this part of the network, in this part of the network, right, and you and the epidemic has already been detected. And you want that the epidemic will should not flow from this part of the network to this part of the network.

So, what would you do? You basically vaccinate this node because you know that this is the junction point, and if you vaccinate this node, right it will the epidemic will not move from this part of the network to this part of the network, ok.

So, if you compute the betweenness centrality of all the nodes in this particular graph, you will see that the betweenness centrality is maximum for this junction node, which is very obvious because when you take all pairs, right when you take one node from this component and one node from this component, you will have to move through this particular node, right.

Now, this kind of node is also useful for identifying clusters, ok. Say for example, if you identify nodes which has higher betweenness centrality or the highest betweenness centrality, and if you remove that node, what would happen? You may see that the network would get disconnected.

Say, for example, in this case if you remove this node the adjacent edges will also be removed, right. So, you will see there are two you know disconnected components emerging. So, this betweenness centrality concept is also useful for the purpose of attacking a network.

Say, you want to decentralize a network, now think of a terrorist network. So, let us think of a terrorist network, where nodes are terrorist and links are relations between terrorists and you identify a node which basically acts a bridge between two terrorist organizations, right.

If you break the node, if you say if you kill the terrorist or if you arrest the terrorist, what would happen? The information that is flowing from one group to another group that will be broken, that will be stopped, right. So, these are some of the some of the applications of betweenness centrality.

Now, this node whose betweenness centrality is maximum, this is also called articulation point. Now, this terminology is very very useful, right. So, articulation point is a node, whose removal will disconnect a particular network, ok. It is not necessary that if you remove a node it would disconnect the network, right. But let us say you have another edge like this, ok, so if you disconnect this node nothing would happen, the network will still remain connected.

But at least what would happen is that your cluster would become prominent, ok meaning that your inter cluster edges, so these are inter cluster edges, the inter cluster edges will be
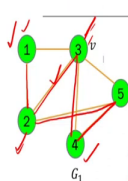
reduced and intra cluster edges will remains same. Therefore, clusters the clustering structure will become prominent, ok.

So, and this is also useful for and this you know articulation points or the nodes with higher betweenness centrality. These nodes are also very dangerous. Why? Because say for example, you want to you want that some of the information should not you know should not move from one component to another component, right.

So, and this these nodes would act as a spy, ok because this is the only node which connects two different components in a network. So, you should be very careful you know from this nodes. You should not you know, you may not want to share critical information with this node. These are also called gatekeepers. Gatekeepers, articulation points, these are these terminologies are used in you know different applications, right.

(Refer Slide Time: 19:07)

So, let us take an example. So, say this is G 1, and what we are doing here we are creating a matrix, this is a matrix 5 cross 5 matrix, you have 5 nodes here 5 nodes here, right. And a cell say x bar y, this is basically used to indicate that say.

So, what is y? y is the let us say let us say let us look at this component, ok, 0 slash 1. Means that there are two shortest paths, between the there are one shortest path, between the, I mean there is one shortest path between 1 and 2, and there is no shortest path between 1 and 2 where the particular node v exists.

So, I want to compute the betweenness centrality for node v which is 3, in this case. There is no shortest path between 1 and 2 where the node 3 exists. Let us look at it. So, between 1 and 2, you see that there is only one shortest path and 3 does not exist, right.

Let us look at this one, say between 2 and 4, between 2 and 4 you have this shortest path and you have this shortest path, ok. So, among these two shortest paths. So, therefore, this is 2, then the denominator is 2, and among these two shortest paths in this shortest path which is 2, 3, 4 the node 3 exists. So, therefore, the numerator is 1, right. So, the fraction is 1 by 2, right.

Similarly, we do the calculation for all pairs between 1 and 1, 1 and 2, 1 and 3 and so on and so forth. And then what we do? We basically take the sum. So, in this case, 0 by 1 plus 0 by 1 plus 1 by 1 plus 1 by 2 plus 0 by 1, 0 by 1, sum of all the cells. And for node 3, it would be 4, ok. So, this is betweenness centrality.

So, sometimes we mix up the concept of betweenness centrality and closeness centrality. Think about it very carefully. These two metrics capture two different notions.

Closeness centrality captures how close you are with respect to other nodes whereas, betweenness centrality indicates you know what is the what is the likelihood that whenever I basically want to move from one node to another node you will be encountered, ok. So, you cannot avoid the particular node in every shortest path distance calculation, ok. There are two different notions. I hope you understand what I am trying to say, alright.

(Refer Slide Time: 22:01)

Betweenness Centrality: Variants

❏ The edge betweenness centrality refers to the fraction of all pairs of shortest paths of the network that pass through a given edge.

❏ Computation is more-or-less similar to that of betweenness centrality

❏ The flow betweenness centrality the fraction of all paths (not necessarily the shortest paths) of the network that pass through a given edge.

❏ Clearly, flow betweenness centrality measure is computationally expensive than betweenness or edge betweenness centrality measures.

So, now this betweenness centrality can also be computed for edges, right. So far we have computed betweenness centrality for a node and you can easily extend it for edges because straight forward, right. So, now, again you look at all shortest paths, all pairs of shortest paths and see for a particular edge, how many times that edge exists, right in all pairs of shortest paths and the calculation will remains same.
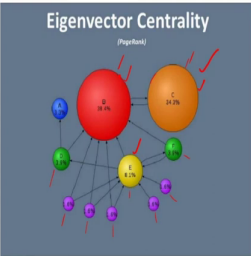
There is another notion called flow betweenness centrality, right. So, in betweenness centrality, we are only looking at the shortest path shortest paths between two nodes. Sometimes you may not be able to move through shortest paths because it may happen that shortage paths are you know are congested in some ways, you may not be able to use the shortest paths. You may want to use some other paths which may not be shortest path, but you know slightly longer, but that would still you know fulfil your purpose.

So, in flow between a centrality instead of looking at the shortest path, I would actually look at all possible paths between a node, right between a pair of nodes. So, all possible paths between a pair of nodes is computationally expensive. Therefore, we generally do not calculate flow betweenness centrality, but for certain applications, right we may want to calculate flow betweenness centrality, ok.

(Refer Slide Time: 23:35)

Now, we are moving to a bit complicated centrality measure. So far we have looked at you know simple measures like distance, shortest path distance and so on and so forth degree. Now, we are moving to the to another notion which is called eigenvector centrality.

And believe me or not this eigenvector centrality is something based on which a series of centrality measures have been proposed, ok. So, what is eigenvector centrality? So, eigenvector centrality says that you know when we measure the centrality of a node, you would also look at the other nodes which are linked to the given node.

Say for example, so this is a network, right you see that node E is linked to many other nodes in the network, ok. But node B is only linked to, ok node B is also linked to many other nodes, but node C is only linked to B, ok. And the size of this of every node indicates the prestige or centrality or eigenvector centrality.

You see that node B has maximum size followed by C, followed by E and so on and so forth, right. So, what it is saying is that when we measure the centrality of a node, you also look at other nodes which are actually pointing to this given node and you and you derive the centrality of the given node based on the centralities of the other nodes.

And the idea is that if I am pointed by some of the highly prestigious nodes, my prestige will automatically increase, ok. In other words, say on twitter, if I am followed by Shah Rukh Khan, for example, Amitabh Bachchan automatically my follower count will increase, Does not matter whether the other you know other users are following me or not. So, if I am followed by some you know social media celebrities my prestige will automatically increase, irrespective of whether others are following me or not.

On the other hand, if you think of another user who is followed by many users, but their prestige is not that high, my prestige would not be that higher compared to the case where I am followed by a celebrity for example, ok. And this is the this is a very important notion and based on that a series of metrics I have been proposed. I will discuss a few in the later slides, right.

So, essentially you see that you know the same the eigenvector centrality measure is a kind of a recursive approach, right. So, I am measuring the, I am measuring the centrality of a node based on the centrality of other nodes. So, I will measure the centrality of the other nodes based on the centrality of their followers, alright, their neighbours and so on and so forth. So, it is kind of a recursive definition, ok.

So, now let us move to the definition, right. What it is saying is that a centrality of a node, the eigenvector centrality of a node v is the sum of the eigenvector centrality of all its neighbours. So, N v is the neighbours of v, set of neighbours of v and t is one of such neighbour, so one of such neighbours. So, so x t is the eigenvector centrality of t. So, I am taking the sum of eigenvector centrality of all the neighbours, right.

And I am actually using a some sort of normalization, say some sort of constant say this is lambda 1, ok. So, this is the formula. Now, you can write it in a different manner you can say that, ok 1 by lambda 1, sum of all the nodes present in V, how do I know that the t is a neighbour of this V, right.

So, I can use the entry in the adjacency matrix. So, a v t, now think of it. So, a is the adjacency matrix, ok and one such entry is small a v t, small a such sub v t. So, say this is v th row and t th column. So, this would be either 1 or 0. If v and t are connected then it would be 1, otherwise 0, ok. So, this times x t. If this would be 0, then this will not be considered. If this would be 1, then x t will be considered, right.

So, now, you can see that I can write it in the matrix form. I can say that this is. So, let me you know adjust this equation in some other ways. I let me move this lambda 1 here, so it would be x v lambda 1 equals to sum of, right a v t x t, right.

Now, let us make it compact. Let us make it compact; let us. So, x v is the eigenvector centrality of node v; x t is the eigenvector centrality of node t. Let us assume that I have a I have a 1D matrix x, where we have elements x 0, x 1, x 2, dot dot dot x v, x t everything.

So, all these entries indicate eigenvector centralities of different nodes. And I have this adjacency matrix A, right. So, can I write it in this manner? Right. Say lambda 1 X equals to A X. So, this is basically a and this is X. So, you see this is the equation A X equals to lambda 1 X.

And if you know matrix operation, right linear algebra you should have heard about something called eigenvalues and eigenvectors, right. So, this is exactly eigenvalue eigenvector formula, right equation. So, we know that this is the equation, right A X equals to lambda X; this is the eigenvalue eigenvector equation, right.

So, if you do not know what is eigenvalue eigenvector, you go back and check, but roughly what it is saying is that you know eigen; so, let me you know briefly talk about matrix multiplication. I know this is not a linear algebra course, but let us briefly talk about it, right.

So, A is a matrix and X is a vector, ok. So, now, in a 2D space or a 3D space, right X is basically an equation of say let us take a 2D, right something like this. So, you can think of X as a as an equation of a of I mean as a data point here, as a vector here, right a 2D point. So, when we multiply a vector with a matrix, what actually happens? It basically makes a linear projection or linear transformation, ok.

So, it. So, when you multiply this by a matrix you get another vector say Y. So, Y is the linear transformation of X, right and this transformation happens due to A, ok. Now, this linear transformation. So, what actually this you know matrix indicate. So, if you think of this columns of this matrix A, right these are basically different you know basis vectors in your transform space. If you do not know what is basis vector please go back and check.

You basically you are transforming now, in a normal Euclidean space or coordinate geometry we have two basis vectors i hat, right and j hat, right which is 1 0 and 0 1, these are two basis vectors, right. Any vector can be generated based on this basis vectors.

So, when we multiply X with A, as if A matrix is transforming X to another space, right where you have two you have two different basis vectors and these basis vectors are basically the columns of A, ok.

So, during the transformation you will, I mean after the transformation you will get a transformed vector Y. Now, what is special about this one? This is also transformation A X equals to lambda X. So, you are multiplying X with A and the transformed vector itself is basically X, the same vector. So, this is a special type of transformation, ok.

And what is lambda? Lambda is a constant, ok. So, lambda indicates whether you, whether the transform vector is squeezed or expanded, right. If it is greater than 1 then you basically expand it. If you if it is a fraction, right, so 0.1 or 0.3 whatever you basically squeeze it. It can also be negative by the way, negative means you are changing the direction of the resultant vector, ok. So, this is a special kind of matrix I mean linear transformation where the resultant vector actually resides on the span of a of the vector itself.

Now, what is span? Again a span, again go back and check the term span, right, basis vector these are very important concepts, right. So, I am not going into the details of eigenvector, eigenvalue you know concepts. But what it basically does in this particular case you see that essentially, X which is the eigenvector I mean the eigenvector, right of A, so this actually captures all the eigenvector centralities, ok.

Now, you may have multiple eigenvectors, eigenvalues. There can be multiple such transformation. There can be multiple such vectors which you which when you multiply with A, the vector will remain same with some constants, constant lambda 1, lambda 2. So, which eigenvector which eigenvector eigenvalue I should consider? So, this is called eigenvector, this is eigenvalue, right.

So, it turned out that I should consider the principal eigenvector, I should consider the principal eigenvector. What is principal eigenvector? Principal eigenvector is eigenvector corresponding to the largest eigenvalue. So, you will have multiple such eigenvalues, lambda 1, lambda 2, lambda 3 dot dot, right. So, you choose the maximum eigenvalue and you take the corresponding eigenvector, right.

Why so? Your question would be why we suddenly need to consider the principal eigenvector. Because what I wanted to do is that I wanted to come up with a vector an

eigenvector which is whose values are non-negative, right. I want to come up with a eigenvector whose values are non-negative.

So, how do I guarantee that which eigenvalue I should consider, so that my eigenvector the all the entries in the eigenvector should be non-negative? Right. There is a very nice you know theorem called Perron Frobenius theorem, right which basically says that if you want to identify an eigenvector whose all elements are non-negative, you should choose eigenvalue which is the maximum, right.

So, therefore, I want to choose the principal eigenvector, right of A. What is A? A is the A is the adjacency matrix. So, it is a very simple thing. Although, the idea is little bit complicated, but I mean the formulation is very simple. I have the adjacency matrix. I will get the principal eigenvector of the adjacency matrix. And each element in the principle eigenvector indicates the eigenvector centrality of the corresponding node, ok.

So, again if you do not know what is how to calculate eigenvector, eigenvalues, go back and look at some you know Fundamentals on Matrix Operation, ok.

So, I stop here. The next part of the lecture again that would be a continuation of this chapter. I will discuss about other variations of eigenvector centrality. I will discuss about PageRank, card centrality, and so on and so forth.

Thanks.