Complex Network: Theory and Application

Prof. Animesh Mukherji
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 05
Network Analysis – IV

Welcome back to this session on Network Analysis. We will continue with some of the ideas that we have discussed in the last lecture. To start off, I would just briefly recap the concept of eigenvector centrality that we developed on the last lecture.

(Refer Slide Time: 00:39)

So if you look at the slides, I have written exactly whatever we done last day on the slides.

(Refer Slide Time: 00:51)

The basic thing that we saw was that, if $X(t)$ is the popularity value at some time t then we can express $X(t)$ in terms of $A^t X(0)$ this is what we have already seen. This we call equation 1, say for simplicity. Now what we have further said is that let us express $X(0)$ as a linear combination of some eigenvectors of the matrix A, the adjacency matrix A. Then one can write$X(0)$ as nothing but the sum of some linear combination of some of the vectors of the adjacency matrix, where v_i is an arbitrary eigenvector of A, this is what we wrote last day already. Further from the eigenvector equation we all know that $Av = \lambda v$ where lambda is the eigen value corresponding to the eigenvector v.

So, now if we multiply both sides of this equation by A once again, we have $A.A.v = \lambda.A.v$, that should be equal to $\lambda.A.v$ once again that is nothing but $\lambda.\lambda.v = \lambda^2 v$. So, continuing like this 40 times this would imply $A^t v = \lambda^t v$. So this is what is we call equation 3, and let us say that we call this equation 2. The equation $X(0) = \sum_i c_i v_i$ has equation 2.

(Refer Slide Time: 03:51)

So from 1, 2 and 3 we can write the following expression; $X(t)$ should be nothing but equal to lambda $\sum_i \lambda_i^t c_i v_i$. Now we will do a small trick out here, what we will do we will divide both sides by λ_1^t, where λ_1 is the principal eigen value and the corresponding principal vector is say v_1. Then you can write this as nothing but $\lambda_1^t c_i v_i$. Now if we have gone sufficiently many numbers of steps.

So, we are talking in terms of thermo dynamical limit that is when t is very, very large. The point after which there is no significant change in the popularity, so at that point we will have limit on t it really goes to large numbers we can put on both sides this limit. So this, what does this tell you now for i is equal to 1 this value the value here changes to 1, so you have $c_1 v_1$ plus numbers which have a fractional pre factor power to some infinite power. All those fractional pre factors that are power to some infinite power go to 0 only $c_1 v_1$ remains. So, this tells you that the popularity converges to the principal eigenvector of the adjacency matrix A.

So, that like kind of completes the idea of eigenvector centrality. Basically, in a nutshell what you need to remember is that if you have to compute the eigenvector centrality of a network.

(Refer Slide Time: 07:13)

Then you for that network you have the adjacency matrix say A and you compute the principal eigenvector of A, say that is v_1 which corresponds to the eigenvector centrality of the nodes. So, all this time we were discussing mostly in the context of undirected networks, where edges are not directed. This entire exercise that we did in for computation of eigenvectors is for an undirected graph. The adjacency matrix A is assumed to be symmetric.

So, all this explanation that i have given you so far is under the assumption that the adjacency matrix that we are considering is symmetric, that the graph is an undirected graph.

(Refer Slide Time: 08:36)

Now, the immediate next question if you look in to this slide is how to convert this definition in the context of a directed network.

(Refer Slide Time: 08:36)

As soon as you try to do that there are certain problems that crop up and one of the important problems is what I have shown in the figure in the small network that I have described here. So what happens is think of the node A, now this node does not have any in degree it's in degree is 0, so that means the centrality value that this node will have is 0. Then the node B, consider the node B this has only one in degree from A.

This node will also have a centrality value of 0, because A has a centrality value of 0 which is actually borrowed by B and that centrality is also 0, and in this way it continues and propagates over the entire network. So, the entire acyclic parts of the network actually for all the nodes that are part of that acycle of that network have a 0 centrality. This is a very big problem, when you try to translate the concept of eigenvector centralities for directed networks.

So, I repeat the problem is very simple. You have this node A here which do not have any in degree, so that means since it does not have any in degree its centrality is going to be 0. So if its centrality is going to be 0 then the centrality of B which is solely completely dependent on A is also going to be 0. And in this way anybody whose centrality is just dependent on B is its centrality the node the centrality of that node is also going to be 0, and this will continue until and unless there is some cycle in the network. This is a problem when you try to convert the definition of eigenvector centrality for directed networks. So, what could be a solution?

(Refer Slide Time: 10:49)

So, what we try to do is we try to sprinkle some initial popularity to each individual node. Since this is a problem, now if we sprinkle a very small equal centrality to all the nodes. We do not disseminate any node, so we try to sprinkle a very, very small tiny centrality value to all the nodes equivalently. Now we start from that configuration. Then in that configuration you can immediately see that A’s centrality will no longer be equal to 0. Therefore, if you do such a thing then you have to readjust the formula that we introduced last.

So now again we can write $X(t)$. Or say for a particular node x_i should be, now we are rebalancing things. So here in this formula what we have brought in it looks very similar to the previous formula that we have introduced for the eigenvector centralities, but then we have brought in two important changes if you look carefully; one is this parameter β here and the other is this parameter a α here. So, both β and α are constants this is the first premise. The second thing is beta is the small initial value of centrality that we give to each node. So, β is the initial small centrality that we actually give to each node equivalently, and α actually is the readjustment constant so that the entire formula remains balanced.

Since you are introducing this beta component into the formula we have to accordingly rescale the other part of the popularity, so one popularity is inherent that is given to me that is sprinkled on all of us and there is another popularity that is coming from the neighborhood. So, these two has to be rebalanced again in order to keep the formula equivalent to the previous case. So, now given this we can immediately write the notations in terms of X(t)'s as functions or in terms of vectors and matrices.

(Refer Slide Time: 14:12)

So now, let us say X is the vector popularity vector so this is like X_1 is the popularity for node 1, X_2 is the popularity for node 2 and so on and so forth. This is my vector of popularities. From here you can given the formula that we have written just now we can immediately write the following expression as I have already written also on this slides; $\alpha A\ x + \beta\ 1$, where 1 the darkened 1 that I am writing, the bold 1 is the vector of all one's. Therefore, express X in terms of alpha A again the vector $Ax + \beta$ factor which actually is nothing but the initial popularity that you have sprinkle to on each and every individual node.

Now this if we try to readjust and write express x as the subject of the formula, so expressing x as the subject of the formula we can write $x = \beta(I\text{-}\alpha)A^{-1}$ and there is this vector of all one's. Let us denote the vectors by the vector sign it would be easier for us to note. Now see from that original formula we have come up with an expression to find out the exact value of popularity given the constants alpha and beta and the initial popularity β, the other readjustment constant α, even these two and the adjacency matrix of the network you can immediately compute the popularity value x.

But then just a point of note here that computing inverse, this is the identity matrix inverse of matrices is difficult. Thus it is better to compute using the iterative equation, which in this case is again very simple you can write $X(t)$ is equal to nothing but $\alpha Ax(t\text{-}1) + \beta$. And then there will be a point after which $X(t)$ will change no longer and you will have an expression for the final value of $X(t)$. All of these ideas are borrowed from the first definition of eigenvectors then we are gradually incrementing that same idea, we are building up on that same idea and developing better and better matrix for application in different network settings. For instance, this particular development is in the settings of are directed network. And this particular centrality is termed as by the name of the inventor Katz centrality.

So now, since we had talking about eigenvector centralities, Katz centrality, directed graphs, there is one thing that actually is indispensible and needs to be discussed in this context and that is like the worldwide web graph. And in the context of the World Wide Web graph how are the web pages ranked. So, all these initial settings that we have been developing so far in the last lecture and today's lecture is in the direction of building the idea of how to rank the web pages in a World Wide Web graph.

(Refer Slide Time: 19:18)

Basically, in a World Wide Web graph you will have nodes which are pages, say page P_1, page P_2, page P_3, and so on. So in this way you have a World Wide Web graph, and if there is a hyperlink from page P_1 you draw a directed edge like this. There could be another hyper link say from page P_2 to P_1 you draw a directed edge like this. This is a typical representation of a World Wide Web graph.

Now, the question is like all of us have been using Google search engine almost every day and multiple times a day and probably many a times a question have come across is like when you search using a query term there Google returns you back a bunch of web pages. Now, these web pages actually are ranked in some way and this ranking actually is done in terms of the importance of a particular page. And this importance is measured using some formula which is very similar to what we have just now discussed, which is very similar to Katz centrality. And this formula was developed in the frame work of the very famous algorithm called the Page Rank.

Page rank basically tries to determine the rank of a web page being displayed in return to a query on the search engine. This algorithm actually tries to assign importance to each individual page so that whenever there is a search query the pages that are return could be actually sorted in terms of their importance. The basic idea is again same as we have discussed for the case of the eigenvector centrality.

(Refer Slide Time: 22:11)

If you look at the slide you I give a very simple example of a small snap shot of the world wide web. So, what you see here are a few nodes marked like A, B, C, D, etcetera and the size of the node is actually expresses the extent of popularity of that particular node. You see that the node C has only 1 in degree, but then you see that the popularity of node C is quite high, this is because node C is actually pointed by node B who himself is very, very popular. This is the basic idea of the eigenvector centrality that we have already discussed. Since node C is being pointed already highly popular node B that is why node C's popularity automatically increases.

Whereas, take for the example the case of node E. The node e here has a lot of in degree, but each of these nodes that point node E are not themselves very popular. So that is why node E is not actually very popular. Even if it has a very high in degree it is not actually very, very popular. Whereas, in contrast nodes C which has just 1 in degree is much more popular by virtue of having one very popular person pointing to node C. This is the idea that we have already discussed, I have just illustrated the same using this example on the slide.

So then, the question is how was page rank defined. It is again very, very simple. Even before we go to that definition the quantity definition one question that can immediately come to our mind is how to make our own web page. Suppose, you have (Refer Time: 24:11) a webpage what could be a criteria to actually make your webpage important or popular. This is a very important question and various companies have been working in order to help the promotion of certain web pages.

And the crux is the following; if very highly popular websites say for instance Google, Yahoo, MSN, etcetera are pointing towards your website. If there is a hyperlink from all these websites to your website, of course this is a dream for every person to your website then immediately your page rank your importance actually raises. And whatever search term is fed to the engine there is a high probability that your pages retreat back and shown to the user.

(Refer Slide Time: 25:12)

So, this is the basic idea.

(Refer Slide Time: 25:14)

So, now with this basic idea we will (Refer Time: 25:17) to finally defining a quantitative measure for the page rank.

(Refer Slide Time: 25:27)

Again, you have say the popularity of a particular node being expressed by X_i. Then you can write here $X_i = \alpha$, the same type of expression that we have already written. $A_{ij} X_j$ + β beta, but there is only one small difference from the Katz centrality in that, what you do is you now divide the popularity of the node X_j by the total out degree of X_j .

Suppose, there is this node j it has an out degree of k^{out}_j and say the popularity of X_j of the node j is X_j , then each of the individual nodes that j is pointing to one of this is actually the node i with popularity X_i . So each of them is receiving a fraction X_j / k^{out}_j from the node j. So, this fraction of popularity from node j actually is given to node i. So you basically appropriately normalize the popularity of X_j and distribute it among all its neighbor. That is the only difference from the Katz centrality; Katz centrality did not have this normalization.

Now, one problem that immediately comes up is that if $k^{out}_j = 0$, that is the out degree of j is 0. That means, this would immediately say that j will not be able to contribute to the centrality of other nodes. This will immediately say that node j will not be able to contribute to the centralities of any other node. In such cases we assume that k^{out}_j is preset to 1 in order to make our computations easy. In all such cases where k^{out}_j is 0. So, X_j will be the popularity of j, so this component will actually go to 0, because it will contribute to 0 popularity to all other neighbors. We set k^{out}_j to equal to 1 so that we avoid division by 0.

In that particular situation we can write again the vector form of the equation as αA, now since this k^{out}_j is here we have this matrix $D^{-1}X + \beta 1$, where these are all vectors. So now we can again express the vector x as the subject of the formula and this will give us D^{-1}. Where D^{-1}, where D is the diagonal matrix containing all the out degree values. Where D is the diagonal matrix containing all the out degree values and D_{ii} is nothing but $\max(k^{out}_j, 1)$. If k^{out}_j is 0 then it is set to 1.

Thank you very much.