Complex Network: Theory and Application
Prof. Animesh Mukherjee
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 08
Social Network Principles – I

In the last few lectures we have been talking about the Basic Statically Metrics for analyzing complex large, complex networks. And we have got introduced to different centrality measures, page rank etcetera.

In this set of lectures from now on wards we will mostly talk about Social Network Principles, and one of the first social network principles that we will discuss is called Assortativity or Homophily.

(Refer Slide Time: 00:53)

The idea is somewhat like this, that given a social network rich people always tend to make friendship with other rich people. So this is the idea of Homophily or Assortativity. Also in other words you can say that the like goes with the like, so rich goes with the rich and possibly the poor goes with the poor.

(Refer Slide Time: 01:36)

So if you look in to the slides the first example that we have here, is a friendship network from the one of the US high schools and what you see here there are three types of nodes in this network. The black ones correspond to black people in the school, the white ones corresponds to white people in the school and the grey ones are the others which could not be people who cannot be classified into either of this groups. And an edge in this network indicates a friendship relationship.

So, what you observe here immediately is that there is this existence of homophily. That there are more blacks are more friends with other blacks, where as whites are more friends with other whites, and there are hardly any connections between blacks and white. This is the idea of homophily that we will build up on from now. So this is one of the very interesting examples.

(Refer Slide Time: 02:33)

Another example was this experiment that was conducted in the San Francisco where there were 1958 couples who are interviewed. Now, these couples are like they classify themselves into four basic classes; the blacks, the whites, the hispanic or the people from Spanish Portuguese origin and others, who could not be classified into any of these three. And people from all this origins were interviewed and the question they were asked was about their sexual partnership. So, given a chance what type of sexual partner they would prefer. And this particular matrix in the slide shows you like what is their preferences, in general what is the preferences.

So, one of the immediate observations from this particular slide or specifically this particular table is that the cells that are on the diagonal are the heaviest. Which again indicates that people who are of the same type are interested to have partner from their same own class like; blacks want to have more partners from the black class itself, hispanics want to have partners from mostly from the hispanic class itself, white tend to choose partners mostly from the white class and the others from the other class. You see that this is one very typical example in majority of social networks mostly which are built on this idea of friendship this particular phenomena is very, very, very prevalent.

So, the idea is that again to iterate is that if there are people from the same class then partnerships or friendships between them is more probable than people from two different classes. Also this idea could be thought of as like people tend to go with other like people, so rich people tend to go with rich people like, so you can interpret it in various different forms. But the basic idea is this.

So, some more examples; if you now look into this slide you see two typical examples. The left hand side network as it shows is much more assortativity than the right hand side network, the right hand side network on the other side is less hemophilic. And in general this type of networks are termed as Disassortativity Networks, that is rich do not go with rich; rich usually tend to go with poor. As we have seen long back in one of our introductory lectures in biological networks you see such disassortativity networks. Even in technological networks like routed networks you see this sort of disassortativity networks where like, many small computers, many mini computers connect to a large router. So it is mostly a disassortativity network.

Where, social networks or friendship networks are mostly assortativity in nature. That is popular people tend to go with other popular people, tend to make friendship with other popular people rich people tend to make friendship with other rich people, that is the basic idea. Now given this observation from various social networks what immediate question is like, how can we have a quantitative measure of these particular phenomena?

(Refer Slide Time: 06:24)

Now we will see how to Quantify Assortativity. The quantification goes like this, let us say that consider a node of degree k. Now the assortativity can be expressed by a factor called k_{nn} that is nearest neighbor degree. And this is defined as the following; $k' p(k'/k)$

, where p_k’ given k is nothing but the conditional probability that a node of degree k ends up in connecting with another node of degree k’. So this is the conditional probability that a node with degree k will connect at its other end with the node of degree k’.

So, this conditional probability multiplied by the node degree at the other end the k’ sum of this over all nodes or all such k’s defines the nearest neighbor degree. The idea is very, very simple. So what you do is, let us say that we have a node x now we look at the degree of the node x, we also look at the degree of each of neighbors of the x. Let us draw it like this.

(Refer Slide Time: 08:52)

Suppose you have a node x here, now say x as k neighbors N_1, N_2, N_3 up on till such k neighbors. Then what we do is we see what is the degree of each of the individual neighbors; we check the degree of each of the individual neighbors. We find an average of the degree of the neighbors that is the nearest degree neighbors. We find an average of the degree of all the neighbors, so you have the degree of the node x and the average degree of the neighbors. You have these two things, on the x axis you have the degree of the node x and on the y axis you have the average degree of the neighbors of x.

Now, if this plot is a scatter diagram which mostly concentrates on the y equals x line then you have a high probability that nodes with similar degree or nodes of similar degree at friends in a social network. So what you see is that, my degree which is k is highly related with the average degree of my neighbors, so that is the idea. If my degree is highly correlated with the degree of my neighbors then it is an assortativity network.

And such co-relation is reflected by the scatter diagram which is concentrated close to the y equals x line on this particular plot. So this is how you basically identify by plotting the degree and the degree of a node and the average degree of the neighbors of that node by plotting them on the x and the y axis and looking at how well they concentrate around the y equals x axis you identify whether a particular graph is assortativity or not.

For instance, if you have a similar plot where you have the k and the average degree of the neighbors of x, k is basically the degree of x. And if you have a scatter plot which is just opposite like this then you have a high chance to believe that this particular network is disassortativity in nature. So, one side when it is highly correlated it is assortativity in nature, on the other side if it is negatively correlated then the network is thought to be disassortatvity.

Just to make things more clear look at this diagram in each of this plot what we have plotted on the x axis is the degree values of all the nodes. So, every node x in the network we have plotted the degree of every node x in the network and on the y axis we have plotted the average degree of the neighbors of each such node x in the network that generates this plot.

Now looking at this plot and having this fit having, this co relation analysis you can immediately say whether this is an assortativity network or disassortativity network.

(Refer Slide Time: 13:09)

Now in order to further nicely quantify this idea there was this concept of Mixing introduced. Now in order to understand what exactly we mean by mixing in a social network we will look into the same example that I should you last time. The example of the partnership choices of these 4 categories of inhabitants of San Francisco: Black, Hispanic, White and the Others. Now, from this particular table that we see here we will translate this table into a more normalized version.

(Refer Slide Time: 13:50)

So what we will do in this normalized version, if you look at this slides each cell of this table is normalized by the sum of all the entries across all this cells of the table. Basically, you normalize each cell by sum of all the entries in all the cells of this table. That means, now the sum of all the individual cells will adapt to 1. If you look at the slides that is way we write here $\Sigma_{ij}e_{ij}$ is equal to 1. Now again even by looking at this table you can very nicely observe that the diagonalies heavy.

Now, if we have a matrix where the diagonal contains all the values there is no other values in no other cells, then that would mean that the network is perfectly assortative, that is there is no other value in any other cell except the diagonal. So, blacks only go with black, hispanics only go with Hispanics, others only goes with others, and white only goes with white. Then in such case only the diagonal will have all the concentration of the values while the other cells will be empty or 0.

In order to quantify this particular notion we will define the assortative mixing coefficient r. On one extreme you have e_{ii}, which is the diagonal element this is the sum of all the diagonal elements so you are counting the total density of the diagonal elements by sum of e_{ii}. Now you are subtracting from there the chance that a black chooses a hispanic or a black chooses some other group with some random chance independently, so that is quantified by this $\Sigma\ a_i\ b_i$. As you see here, as we have shown in the table a i is the sum of the elements on the rows, where as b_i or b_j is the some of the elements on the columns.

Basically, this is independently if there is a chance those two nodes from two different groups' pair up for sexual partnership so that you discount from the total volume. Basically, you see what is the actual partnership that, you are getting from the data minus the part that you could have observed just by random chance. This is similar to the idea of defining correlation coefficient in statistics. Basic idea is again if I iterate that looking at the data you have the probability, you can estimate the probability of pair of people grouping for sexual partnership. This is say black going with black, white going with white, these value is counted or this fraction is counted in some of e_{ii}. And from there we remove the part which could be just absorbed by random chance which is sum of $a_i\ b_i$.

Now, this is normalized by, as I say perfect assortativity would be when some of e_{ii} will be 1 everything else is 0 that is perfect assortativity. So that extreme is 1, that is the extreme value of e_{ii} - Σ a_i b_i. So that is the extreme value of e_{ii} - Σ a_i b_i. This fraction is what we call the mixing coefficient.

Basically, what you see is you find out what is the probability or what is the chance that blacks goes with blacks, white go with whites, and you sum up all this counts minus what is the probability that you see by chance that two people pair up that is what you discount from this value and then you normalize this whole metric with 1 - Σ a_i b_i. Where 1 is the extreme value of e_{ii} that is the maximum that you can achieve. So if it is a perfectly assortative network then what will happen is this mixing coefficient again will be 1.

(Refer Slide Time: 18:51)

Because, in such case you have r is equal to Σ e_{ii} - Σ a_i b_i/ 1- Σ a_i b_i. Now for perfectly assortative networks sum of e_{ii} will be equal to 1 as we said, that implies r will be equal to 1- Σ a_i b_i/ 1- Σ a_i b_i which is equal to 1. So, for perfectly assortative graphs we will have a mixing coefficient equal to 1. However, if it is a disassortative network then e_{ii} will be 0 and we will have a negative mixing coefficient value.

(Refer Slide Time: 19:58)

Then after this the after we have got a little bit of idea about homophily or assortativity we will now look into another very interesting concept called Signed Graphs.

(Refer Slide Time: 20:10)

Basically, this is a formal structure of graphs through which you can express, for instance in a social network or in a friendship network you can express both friendship as well as enmity. A network by which one can express both - friendship and enmity, some of the examples are one that we have given here in the slides, so look at this graphs. So, a plus sign on an age of this network would indicate friendship, whereas a minus sign would indicate enmity. If two nodes are connected by an edge which as a plus sign then it is a friendship relationship between these two nodes.

However, if two nodes are connected by a negative edge, then this relationship is enmity relationship. And I have this interesting question given our online class it would be a nice exercise to measure how it will look in terms of this sign graph. Do you really have enemies here?

(Refer Slide Time: 21:53)

Once we have this concept of sign graphs the first thing that people where interested in studying was this idea of balancing. Basically, these idea barrows from the traditional balancing theory; if you look at these graphs are given here. For instance the first graph, the graph marked as a. You see there are three nodes u v and w, it is a triangle basically. Now all the edges are marked as plus. So everybody is a friend of everybody else in this network. This is very stable configuration.

Now let us take the second example. The second example is a bit tricky. So what you have here that, there are two nodes who are friend among each other and both of them actually share an enmity relationship with the third node. This is again a possible configuration because two friends might have a common enemy in general that is also a stable configuration. The third one is where you have at least two edges which are positive. Whereas, the third edge between these two is negative. This is a rare case. And the forth case is impossible. That there are three enemies in a triangle is a completely impossible case.

(Refer Slide Time: 23:40)

Now given this examples of triangles we can also imagine cases of 4 cycles. Now like how should be the sign graphs taking 4 nodes together look like. Some examples are here. So, some of the stable configuration are shown here. These are the 2 friends each of each are enemies or these are the two friends and then there are 2 enemies on the other side. So these are some of the stable configurations that you observe here.

In general the idea is that you should have even number of negative signs in the graphs, unless you have an even number of negative signs in the graph the configuration is not stable. Only if you have an even number of negative signs on edges in a graph then only your configuration is a stable configuration.

For instance, in this particular example you see c and d are having uneven number of negative edges, and that is why these are unstable configurations. Whereas, in this particular case the 4 cycles you have only even number of negative edges that is why both of them are stable configurations.

(Refer Slide Time: 25:07)

So, the next idea that we will talk about is Structural Holes. This is also again a very interesting idea and we have already looked into some sort of a quantification of this idea in one of our previous lectures when we discussed about betweenness centrality. Basically, structural holes are nothing but nodes or social actors in a network who are like brokers, like they actually transmit relevant information from one part of the network to the other part; they actually behave like information brokers.

For instance, let us take these examples here. So, structural holes, as it reads out actually will separate non-redundant sources of information, sources that are additive and not over lapping.

(Refer Slide Time: 26:32)

If you have two parts of the network say, one here and the other here. Basically, this green node here is denoted as a structural hole, because we are imagining that the information that is there within this particular group of members in the social network is very different from the information that is stored here in this group of networks, so that is why we call this particular node a Structural Hole. We have a word of caution here; there are two things that one needs to be careful about.

A cohesive group cannot have a structural hole, for instance if you have a network like this, so this very cohesive network. And since this is a very cohesive network everybody has similar piece of information that is why nobody in this network actually qualifies as a structural hole. Similarly, if there is another similar concept of equivalence. For instance, suppose you have a node here and on two sides of it you have nodes that have equivalent information, and then also this is not an example of a structural hole.

For instance say, this node or this node or this node or this node none of them are structural holes. Here also this particular black node is not a structural hole, because it does not enjoy any extra information more than, either of this green node. However, if you have a case where you have a node same black node here, but then the nodes on the left hand side have a very different set of information from the nodes on the right hand side. Then this particular node actually qualifies as a structural hole.

So, we will stop here.