

**Social Network Analysis**  
**Prof. Tanmoy Chakraborty**  
**Department of Computer Science and Engineering**  
**Indraprastha Institute of Information Technology, Delhi**

**Chapter - 02**  
**Lecture - 11**  
**Lecture - 06**

So, far we have discussed many important concepts right. We have discussed centrality, we have discussed PageRank eigenvector centrality right card centrality and so on. So, we are almost there we will discuss a few other measures which are also very important right. So, let us look at the new one which is called the assortative mixing. So, what is assortativeness?

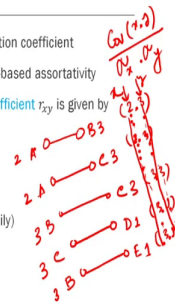
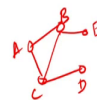
(Refer Slide Time: 00:50)

### Assortative Mixing

- A common practice to find similarity between nodes is to use a correlation coefficient
- The **Pearson correlation coefficient** is a good choice if we want degree-based assortativity
- For two data (degree) distribution  $x$  and  $y$ , the **Pearson correlation coefficient**  $r_{xy}$  is given by

$$r_{xy} = \frac{N \sum xy - \sum x \sum y}{\sqrt{(N \sum x^2 - (\sum x)^2)(N \sum y^2 - (\sum y)^2)}}$$

- If  $r_{xy} = 1$  then nodes  $x$  and  $y$  are perfectly assortative (homophily)
- ✓ If  $r_{xy} = -1$  then nodes  $x$  and  $y$  are perfectly disassortative (heterophily)
- ✓ If  $r_{xy} = 0$  then nodes  $x$  and  $y$  are non-assortative



So, assortativeness is a very important concept in network science. So, it basically says that whether same types of users are linked to each other or not ok. Let us say you have a network of users right and you also know the say political alignment of users right. So, say pro-BJP anti BJP pro-Congress anti-Congress and so on and so forth. So, in the in this network you basically want to understand whether a pro-BJP supporter a pro-BJP user is linked to another pro-BJP user or a pro-BJP user is linked to a anti-Congress user and so on and so forth.

Now, this you know this political alignment is a feature, it can be any feature, it can be a degree a simple degree. So, you want to understand whether degree 2 nodes are connected degree 3 nodes are connected to each other or not degree 4 nodes are connected to each other

or not and so on and so forth right. So, if you think of degree as a property right. So, a network with high assortativity would basically say that nodes with similar degree are connected.

Meaning that degree 1 nodes are connected to each other, but degree 1 and degree 5 nodes are not connected. Degree 2 nodes are connected to each other, but degree 2 and degree 6 nodes are not connected. So, nodes with similar degrees they are connected ok and how do you quantify? It is very simple you can use simple Pearson correlation ok. So, what you can do? Let us assume that you know this is a network and say A B C D E right.

So, let us first arrange all the edges ok right. So, you have edge A B, you have edge A C, you have edge B C, then C D, then B E ok. The degree of node A is 2, degree of node B is 3 right degree of node A is 2, 3, 3, 3 right this is 3, 1, 3 and 1 ok. So, now, you have got pairs. So, 2, 3 2, 3 right 3, 3 3, 1 3, 1 ok.


So, now you can easily calculate Pearson correlation what is Pearson correlation? Pearson correlation is basically covariance of x comma y by standard deviation of x times standard deviation of y right and you have these two variables this and this. So, this is 1 say you can think of x. So, x has this columns, y is this column. So, you can easily compute covariance between x and y.


You can also compute standard deviation of this column standard deviation of this column and that is all right. So, if 2 if nodes have similar degree and they are connected therefore, this would be; this would be this would tend to be 1 if you know dissimilar nodes with different degrees they are connected this would be I mean you can think of either dis assortative or you know random.

If the Pearson correlation it tends to be 1 then it is perfectly assortative if it tends to be 0 then its basically random right and if this is minus 1 then it is negatively assortative meaning that you know meaning that node 1 is connected to node 5 say for example, right. So, it is completely disassociative network right and this kind of information is very important for example in case of link prediction right you want to understand whether nodes with similar features they are connected or not because that something which is which you want to analyze at the beginning right.

If nodes with similar features are connected then it is easy to predict links right because you will say that ok let us look at x and y and x and y have similar features therefore, but they are not connected but it is highly likely that they will be connected in in the future ok. So, we always study assortativity of a network before analysing it further.


(Refer Slide Time: 05:56)





---

- A metric to determine the linkage between a pair of nodes
- Very important in social networks, and to a lesser degree in other networks
- In abstract mathematics, if entity  $x$  is related to entity  $y$ , and also entity  $y$  is related to entity  $z$ , then the transitivity of the relation ensures that entity  $x$  is related to entity  $z$ .
- In social networks, a complete transitivity may yield: "Friends of my friends are my friends"
  - Utterly Absurd in real networks!
- In fact, a complete transitivity would imply that each component of a network is a clique!!
- However, partial transitivity is useful: "Friends of my friend are more likely my friend than some randomly chosen member from the population"



Let us look at the you know the second metric which is called the transitivity. What is transitivity? Transitivity property says that if A and B are connected and B and C are connected what is the probability? What is the propensity that A and C will also be connected? A and B are connected B and C are connected what is the likelihood then A and C will also be connected?

If you recall this was discussed earlier in the earlier lecture is something called clustering coefficient right. So, what is clustering coefficient? We have discussed local and global clustering coefficient. Local clustering coefficient will basically in case of global clustering coefficient what you what we have measured? We have measured the number of triplets total triplets and number of closed triplets.

So, if the number of closed triplets is same as number of triplets, then transitivity would be higher. So, this is one triplet right and if this edges this edge is also there then it is a closed triplet. So, if the clustering coefficient is high meaning the that indicates that network will have high transitivity value ok.

(Refer Slide Time: 07:15)

## Transitivity

- A complete graph is surely transitive
- A measure of transitivity intends to capture how close a network is to a complete graph
- A network with higher transitivity are likely to form dense clusters
- Two ways to capture this tendency
  - Local clustering coefficient
  - Global clustering coefficient



(Refer Slide Time: 07:18)

## Reciprocity



- Relevant for directed networks
- A measure of the likelihood of vertices in a directed network to be mutually linked.
- Networks that transport information or material, mutual links facilitate the transportation process
- An important phenomenon for such applications
- Informally, reciprocity refers to: "If you would follow me, most likely I shall follow you back"
- May be considered a simplified version of transitivity



Let us move to the next concept which is called reciprocity. What is reciprocity? So, reciprocity says that if edges are reciprocal or not. If there is a connection from A to B whether B is also connected to A or not. Now this is applicable of course, for a directed graph and if a network has high reciprocity, it basically indicates that it looks like a it looks like an undirected graph because an reciprocal edge I mean 2 reciprocal such edges would equivalent to an undirected edge right.

(Refer Slide Time: 08:02)



### Reciprocity $\binom{n}{2}$

$G$

- Reciprocity counts the closed loops of length 2
- The reciprocity  $R$  of a network  $G$  is defined as
$$R = \frac{\text{Total number of reciprocal pairs in } G}{\text{Total number of pairs (reciprocal \& nonreciprocal) in } G}$$
- For graph  $G$ , the reciprocity is  $\frac{1}{3}$



How do we quantify this? How do we quantify reciprocity? So, reciprocity is total number of reciprocal pairs right  $A$  to  $B$   $B$  to  $A$  and it is divided by total number of pairs reciprocal or non-reciprocal right total number pairs is  $n \times n$  ok, but remember  $n \times n$  would not be enough why? Because in  $n \times n$  will also capture such pairs which are not connected. So, I am interested in only those pairs which are connected either by 1 either by unidirectional edge or bidirectional edge right.

So, what I am trying to say is that say  $A$   $B$   $C$  and  $D$  ok. So, how many pairs are there? How many reciprocal non-reciprocal pairs are there? You have  $A$   $B$ , you have  $B$   $D$ , you have  $A$   $C$  right, but we will not consider  $B$   $C$  because  $B$   $C$  is not connected. So, I am trying to understand that if there is a connection from  $B$  to  $C$  whether there is a reverse connection from  $C$  to  $B$ . So, the forward connection should exist right.

So, I will not consider this pair I will also not consider  $A$   $D$  ok you think of this right here in this example how many pairs are there? How many valid pairs are there? We have a we have  $1$   $2$ ,  $2$   $3$  and  $1$   $3$ . So, the denominator would be  $3$  and the numerator would be  $1$  because there is one such pair which you know which has a reciprocal edge right.

(Refer Slide Time: 09:53)

NPTEL

## Reciprocity

□ The reciprocity  $R$  for a graph  $G(V, E)$  having adjacency matrix  $A = (a_{ij})$  is given by

$$R = \frac{2}{|E|} \sum_{i < j} (a_{ij} \cdot a_{ji})$$

□ On simplification,

$$R = \frac{2}{|E|} \times \frac{\text{Trace}(A^2)}{2} = \frac{\text{Trace}(A^2)}{|E|}$$

□ In the above expression,  $\text{Trace}(\cdot)$  function denotes the sum of the diagonal elements of its argument square matrix

AND

$\sum_{(i,j)} a_{ij} a_{ji}$

$\frac{\sum_{(i,j)} a_{ij} a_{ji}}{2}$

$A^2$

$(i,j) = 1$



We can also systematically you know formulate reciprocity. So, in this manner. So, essentially we are looking at all pairs  $i$  comma  $j$  right and if the adjacent matrix  $a$  if  $i$  a  $ij$  is 1 and a  $ji$  is also 1 a  $ji$  then this would contribute to 1. So, for all such pairs I will multiply a  $ij$  and a  $ji$  if both of them are 1 then it would be; then it would be 1 right it is basically and operation ok, but we need to we also need to normalize it, it would be mod  $E$  by 2. Why it is mod  $E$  by 2? Because mod  $E$  is the total number of edges.

And since one reciprocal edge is formed by 2 edges. So, mod  $E$  by 2 would be the maximum number of reciprocal edges that is possible right. You can also; you can also you know quantify it in a more sophisticated way, you look at the trace what is the trace of a matrix? Trace of a matrix is basically the sum of diagonal elements right of a matrix right. So, if it is  $A$  square.

So, I have already mentioned in the last lecture that  $A$  square would give you the pairs which are connected through 2 hops right and we are actually looking at 2 hops from  $A$  to  $B$  and  $B$  to  $A$ . So, if I start from  $A$  and through 2 hops if I come to  $A$  again right then it should be it would be basically be  $A$  reciprocal edge right. So, if I multiply  $A$  with itself adjacency matrix  $A$  with itself and if I look at the diagonal elements say 1 1 in this  $A$  square metric.

If this is 1 it basically indicates that there is a reciprocal edge ok and the things are same right you then divide it by mod  $E$  by 2. Take an example you just you know take a dummy network and you calculate this quantity and this quantity we will see that both of them are same ok.

(Refer Slide Time: 12:25)

## Measuring Structural Equivalence

### Common Neighbors

Number of common neighbors shared in the neighborhoods of the nodes  $a$  and  $b$

$$s_{CN}(a,b) = |N(a) \cap N(b)|$$

### Jaccard Similarity

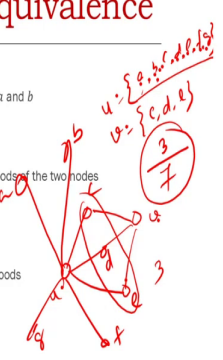
Normalizes the common neighbors by the combined size of the neighborhoods of the two nodes

$$s_{CN}(a,b) = \frac{|N(a) \cap N(b)|}{|N(a) \cup N(b)|}$$

### Cosine Similarity

Normalizes the common neighbors by the individual sizes of the neighborhoods

$$s_{CN}(a,b) = \frac{|N(a) \cap N(b)|}{\sqrt{|N(a)||N(b)|}}$$



Now, let us look at the other quantities something called structural equivalence. Structural equivalence basically says that whether 2 nodes are equivalent in terms of their neighbours let us say; let us say you have  $u$  and  $v$  right. So, if  $u$  and  $v$  have all common neighbours then  $u$  and  $v$  are equivalent right. So, how do we do that? We take the neighbourhood intersection of  $a$  and  $b$ . So,  $N(a)$  is the neighbours of  $a$   $N(b)$  is the neighbours of  $b$  you take the intersection right.


Now this intersection it may not be enough because it may happen that  $u$  has a lot of other neighbours right, but  $v$  has only 3 neighbours. So, the intersection will give you 3 right, but  $u$  and  $v$  are not equivalent because  $u$  has a lot of other neighbours also. So, we need to normalize it, a simple way to do it do this thing is that you take something called Jaccard coefficient. What is Jaccard coefficient? In Jaccard coefficient you take intersection by union right.

Say node  $u$  has  $a, b, c, d, e, f, g$  and. So, node  $u$  has  $a, b, c, d, e, f$  and  $g$  these neighbours and node  $v$  has  $c, d, e$ . So, what is the intersection size of the intersection? Intersectional size would be 3 and what is the union size? Union size would be one remember in a union its not a multi set in the union we will remove the duplicates. So, it would be this one. So, the intersection the union would be the size of the union would be 1, 2, 3, 4, 5, 6, 7 right.

Of course, you can also normalize. So, this is just a normalization factor you can normalize it in other ways you can actually you know take the you know root over right square root of the

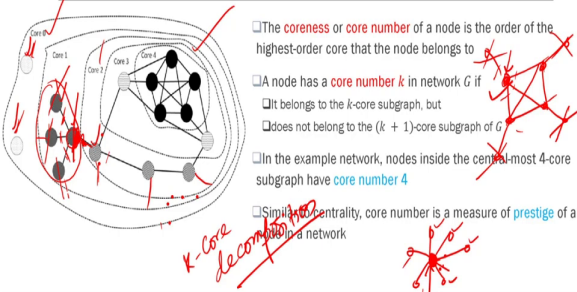
cardinality of both  $N_a$  and  $N_b$  and there are many ways through which you can measure the structural equivalence ok right.

(Refer Slide Time: 14:36)




## Degeneracy: Core Number

*Core-periphery analysis*



- The **coreness** or **core number** of a node is the order of the highest-order core that the node belongs to
- A node has a **core number**  $k$  in network  $G$  if
  - It belongs to the  $k$ -core subgraph, but
  - does not belong to the  $(k + 1)$ -core subgraph of  $G$
- In the example network, nodes inside the central-most 4-core subgraph have **core number 4**
- Similar to centrality, core number is a measure of **prestige** of a node in a network

*k-Core decomposition*



So, now let us look at the another important concept this is called degeneracy. Now, what is degeneracy? Now this has some relation with the assortativity that I mentioned earlier. So, it is basically saying that I am trying to understand the peripheral structure of a network and the core structure of a network ok. What does it mean? So, peripheral nodes of a network are basically those nodes which have degree 1 ok.

If you try to visualize it right you have fragmented nodes which are you know orchestrated around the peripheral component of a network. So, if you keep on you know moving towards the center, you will see that you basically you can get core components of a network. So, degeneracy structure is explained through something called core periphery analysis; core periphery analysis right idea is very simple.

So, let us look at this network ok. What you do? You first identify those nodes whose degree is 0. So, say these two nodes you see forget about this lines ok. These broken lines only look at the nodes and links. So, this node, this node right they have degree 0 you remove them and those nodes constitute the core 0 set. Now you look at nodes with degree 1.

So, this node, this node and this node will degree will have degree 1 you remove them. So, when you remove this and this the associated edges will also be removed that would after



removing this you will see that some other nodes will have degree 1. For example, if you remove these 3 nodes and edges this node will also have degree 1 you also remove this node.

So, you keep on removing nodes with degree 1 you remove first you remove nodes with degree 1 that would result in some other nodes with degree 1 you also remove them you keep on removing nodes until you see that there is no node with degree 1 right. So, if you remove this node you will see that you will have only this node and this node you see that this node will have degree 2 other nodes will also have degree 2. So, these 4 nodes will constitute core 1 set.

Now you again keep on removing nodes with degree 2 right. So, this node will be removed this will be removed this will be removed and so on right that would constitute core 3 sorry core 2 right. And you keep on removing nodes with higher degree and you will see that after certain point the network will vanish and you stop there.


So, you have core 0 which is the most peripheral part then you have core 1 the outermost the second layer of periphery core 2 the third layer of periphery and then say core n which is the nth layer of periphery nth layer of periphery which is the core part of the network. So, and then you assign this core value to the nodes. So, you say that these two nodes have core 0 these two nodes have core 1 so on and so forth.

So, its core values may not be same as the degree let us let us look at this one right let us look at this one. So, what is the core value of this node the central node right? The core value of central node is 1 why? Because, although it has lot of degree right the central node. But if you remove these 3 nodes, this 5 nodes, the 6 nodes the 6 peripheral nodes you will see that this node will become isolated you will also need to remove that. So, degree and coreness may not be correlated.

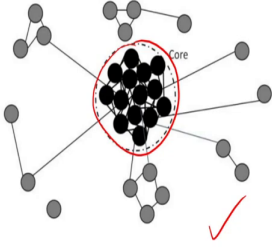
So, if they are sometime correlated, but they may not be correlated now think of another network like this ok. So, when you remove these nodes you will get core 1 but these nodes the other the remaining nodes will have core 3, core 2 there is nothing core 3 you will have all these 4 nodes ok. Although degree wise this node has degree 1, 2, 3, 4, but this node has degree 1, 2, 3, 4, 5, 6, but this node has high core value core. This is also called k-core decomposition k-core decomposition right core k-core decomposition core periphery analysis they are kind of synonymous.

What basically they are trying to say is that we identify you know we keep on removing nodes and try to identify the core part of a network this core part of a network are the central parts. If you want to destabilize a network you want to break the break the network you should attack this core nodes in a network rather than attacking the peripheral nodes ok.


(Refer Slide Time: 20:18)



### Degeneracy: Core-Periphery



- Real-world networks often consists of
  - A dense and connected core, and
  - Surrounding the core by disconnected and scrambled periphery
- The structure above is termed as the **core-periphery structure** of the network



So, the example that I have given you see here this is a core part of a network and that would basically you know give you ideas about the actual bonding in a network. So, the core part. So, there are studies you know in fact, in our group we showed that you know in this fraud kind of network where you have users some users are fraud some users are genuine this fraud accounts fraud users are controlled by a very small set of you know core nodes.

And those core nodes are hidden within a network its not very easy to detect them, but if you do core periphery analysis you will see that those nodes are prominent those nodes have higher core values ok. So, that is about the network measure. So, in summary we have understood how to quantify you know a network different parts of a network edges, nodes, we have discussed degree distribution centrality PageRank right other measures like clustering, coefficient, assortativity, reciprocity, degeneracy and so on and so forth alright.

So, I hope you have you know fair bit of ideas about network measures. In the next chapter we will discuss a kind of a mathy you know more of a math related stuff which is on network model right we will discuss different types of models which basically will tell you how a

network grows over time ok. How do we simulate a network, how do we synthetically generate a network ok and so on. With this I stop here.

Thanks.