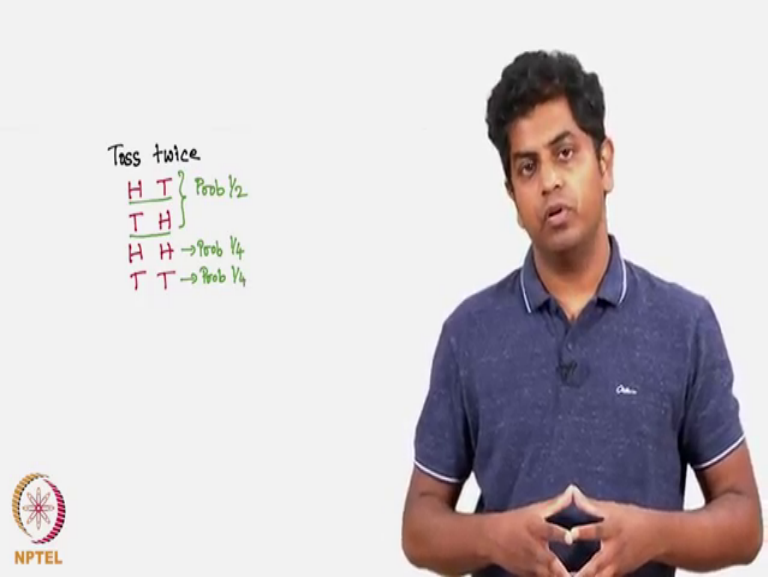
Social Networks
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar

Lecture – 44
Strong and Weak Relationships (Continued) & Homophily
Homophily – Definition and Measurement

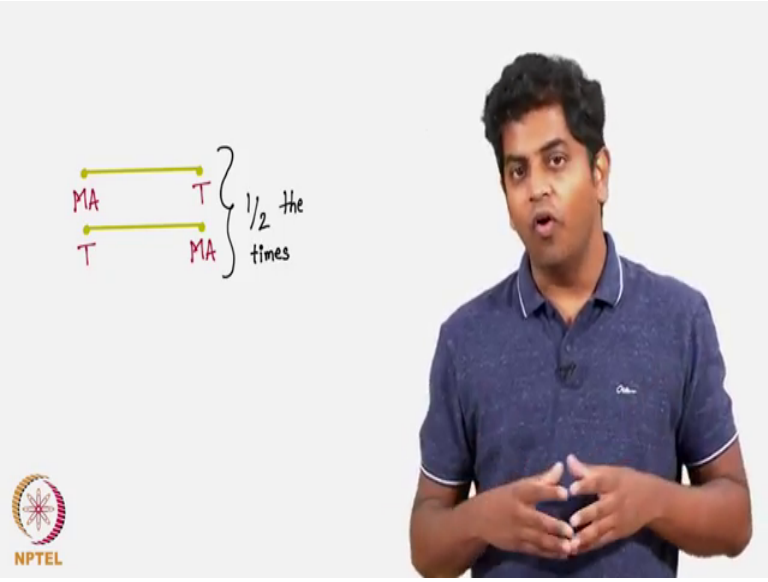
Let us now consider an example assume there is a party going on with 50 teenagers and 50 middle aged people. Now there is some kind of a friendship that gets formed between these two people these two groups of people, and now if I pick a friendship let say uniformly it random I should see on an average one person should be the middle aged person, another person should be the teenage person right. So, this is analogous to you toss a coin twice you will expect a head and a tail or a tail and a head or will be head and a head and tail and a tail, but you observe that head and a head comes with a probability with 1 by 4 head and a tail comes with a probability.

(Refer Slide Time: 00:51)



1 by 2 by that I mean a head and a tail and tail and a head comes with probability 1 by 2 a tail and a tail comes with probability 1 by 4.

(Refer Slide Time: 01:06)

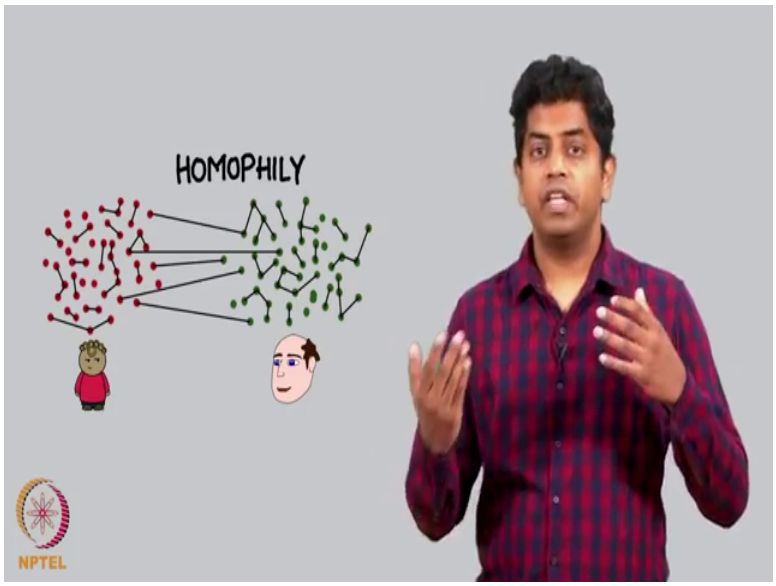


So, it is very likely that if you pick some friendship out of this 100 people, that friendship will have one middle aged and one teenage person half the times that is how I think mathematically it appears to us.

(Refer Slide Time: 01:21)

But you all know the teenagers may want to talk to only teenagers, and middle aged people.

(Refer Slide Time: 01:31)



May want to talk to only middle aged people this is called homophily again the code that I was telling in a previous lecture videos.

(Refer Slide Time: 01:38)

Birds of the same feather flock together.

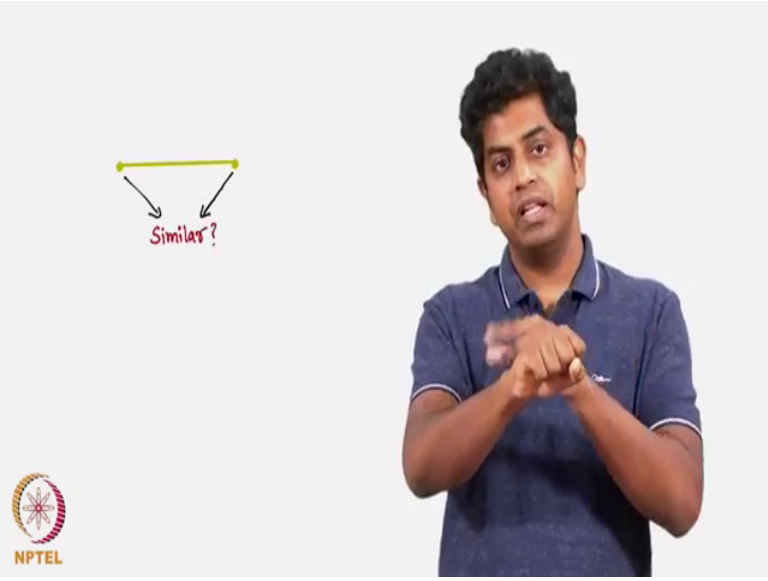
(Refer Slide Time: 01:42)

Now, let me ask this question do you think this is true if there is a bunch of people let say 100 people of the same age group and they are sort of there are 50 men and 50 women half men, and half women and they are sort of in a dating spry they want to talk with each other then definitely across genders you will observe that any friendship will have opposite genders definitely for sure.

But then the previous example of middle aged men and teenagers you saw that tend to cluster middle aged men would be friends with middle aged men, teenagers would be friends with teenagers correct.

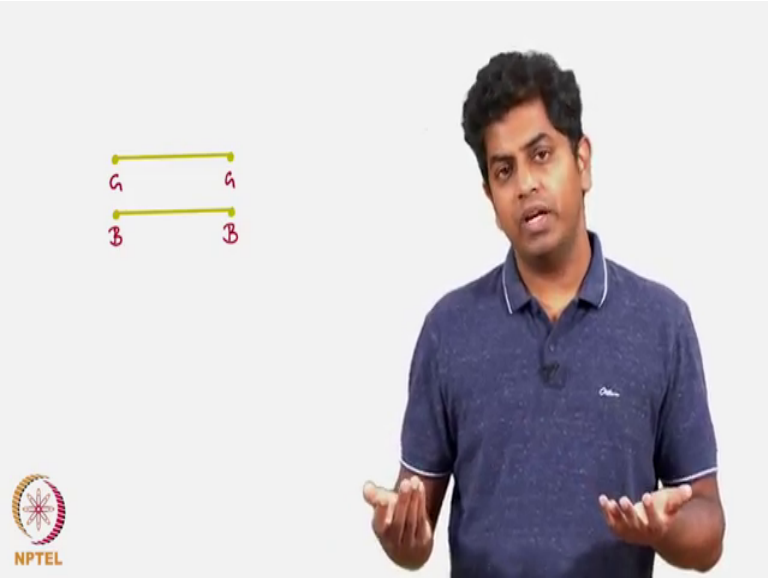
(Refer Slide Time: 02:27)

(Refer Slide Time: 02:38)



So, what will do, what will now do is given a network will observe, does it exhibit homophily what do you mean by this? By this I mean given a network if you pick a friendship an edge across the edge are people similar.

(Refer Slide Time: 02:47)



For example in a class room if you go and observe that girls and girls are basically friends with each other boys and boys friends with each other girls and boys are also friends, but quite comparatively rare right unless it is a romantic relationship.

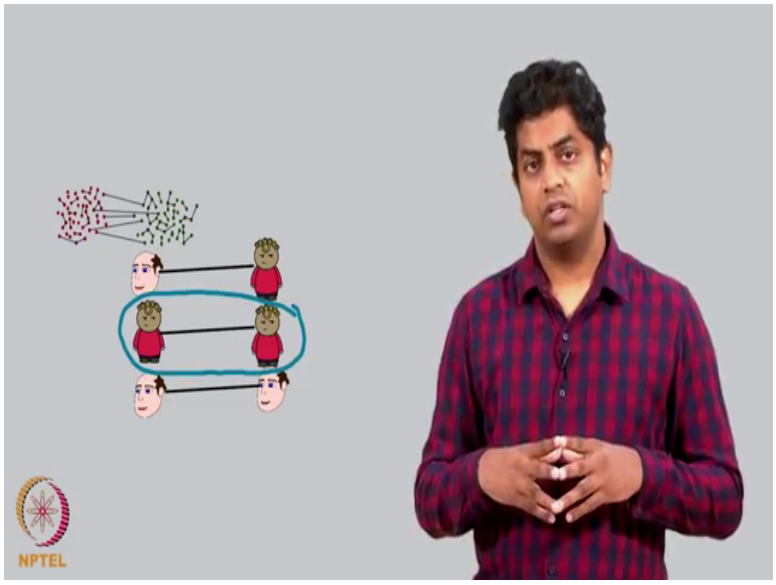
(Refer Slide Time: 03:01)

So, we will now try to quantifies; what do I mean by quantifying this? First question that we can ask is does the network exhibit homophily this is like asking do you have milk in the vessel the answer is yes or no. If the answer is yes then you ask another question how much milk do you have in the vessel. Similarly you ask this question does the network

exhibit homoplily if the answer is yes you ask the next question to what extent does it exhibit.

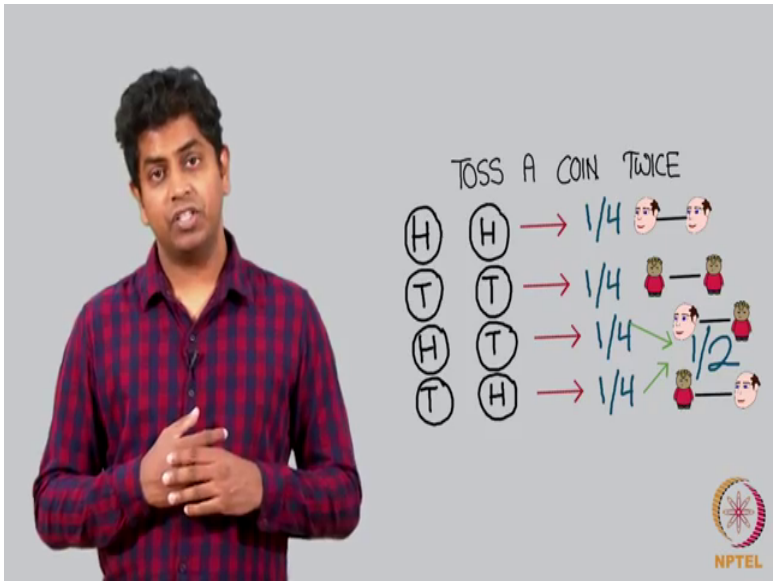
Let us now define what is homophily rigorously? Let me use the same example assume there are 100 people in a party hall, 50 of them are middle aged people and 50 of them are teenagers. Now there is friendship between these 100 people which is actually classified as teenagers and middle aged people. Now what do I expect? I expect that all of them talk to each other become friends with each other irrespective of whether they are teenagers or middle ages.

(Refer Slide Time: 04:17)



Let me take one such friendship and then observe whether this friendship I pick a random friendship and I see whether this friendship is across a middle aged person and a teenager or between two teenagers or between two middle aged people. Now what do you suspect would be the answer for this? If I ask you what is the possibility that if you pick a friendship, uniformly at random, that friendship is actually between two teenagers, it is in fact easy to see that this is very similar to a coin tossing experiment.

(Refer Slide Time: 04:53)



I toss a coin twice let say and I get what are the possibilities that I will get a head and a head a tail and a tail a head and a tail or a tail and a head right all these 4 possibilities are equi probable

Now what is the possibility that I get head and head 1 by 4 pretty obvious, tail and a tail 1 by 4 a head and a tail and a tail and a head both put together is 1 by 4 plus 1 by 4 which is 1 by 2. Now is in this what is happening in our example where I said a party with some 100 people half of them are teenagers, half of them are middle aged people and friendships between them is what we are observing; half the friendships should be between middle aged and teenager people, and quarter of them should be between teenagers and quarter should be between middle aged and middle aged.

Now, here are two contrasting examples.

(Refer Slide Time: 06:03)

One example is middle aged and teenagers in a party hall, what do you expect? You will definitely expect there is a whole lot of friendships between teenagers and between middle aged people by that I mean the expected probability of half which is the friendships between middle aged and teenagers across that is half right that will not divisible infact that will be a lot less it will be some 18. If the friendships across these two parities are less than half, then we say we observe homophily here this makes perfect sense you see let me rephrase.

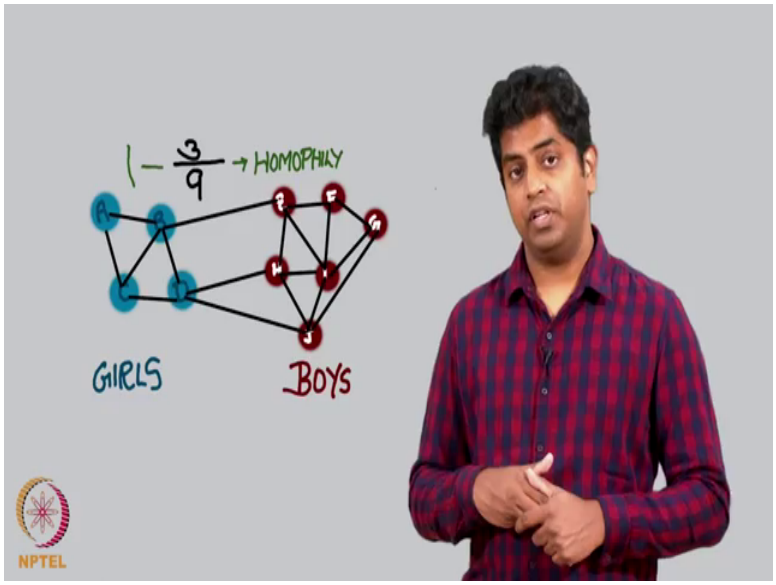
Party with 100 people half of them 50 of them are teenagers, 50 of them are middle aged people you pick friendships and observe what kind of friendships people have, you will observe that the friendships between teenagers and middle aged people are a lot less what do you mean by a lot less? You expect it to be half the friendships out of all the friendships half of them should be between middle aged and teenagers, but you observe that less than half is between middle aged and the teenagers, which means there is homophily happening here, there is the that the same old court I have been repeating birds of the same feather flock together is what is happening here.

(Refer Slide Time: 07:43)

Now, let us look at the contrasting example; the contrasting example is this assume there is a party hall full of Bollywood hero and heroines right. So, a 50 of them are Bollywood heroes 50 of them are Bollywood actress. Now what kind of friendships will happen here? I suspect that the actors have the tendency of going and talking to the actress more than actors themselves. An actress have the tendency of talking to an actor than a fellow actress. Now if I look at the kind of friendships that happen in a party hall between a Bollywood actors and actresses what would you observe think for a moment. You will observe the exact opposite of the previous example right.

Now, there will be very less friendships within the actresses and there will be a very less friendships within the actors there will be a lot of friendships across. Now what is happening here? There is no homophily here absolutely if anything there is very less homophily here correct. Why? Pick an edge out of all friendships that happen in this party hall of Bollywood actors and actresses, pick all the friendships and see how many of them are actually between the actresses and between the actors, you will observe that a lot less is between the same type of people, a lot more will be across right which means every friendship has different types of people across the edge, which means homophily is a lot less.

(Refer Slide Time: 09:25)



So, let us now look at a nice example; look at this network of 10 people let say in a class of 10 there are four girls namely a b c and d and then 6 boys e f g h i and j, and this is the network between them there is some friendships there is a friendships fine done. Now let me see what is the nature of friendships? Let me look at one such friendship let say between a and b as I can see friendship between a and b is present and a and b are both girls. If I randomly pick a friendship and look at who are on the other side of the friendship do I mostly see two boys or two girls or do I also see a boy and a girl now let me quantify this properly. Let me first count the total number of edges on this network as you can see there are 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 edges here, which means 18 friendships are there out of this, 18 friendships I expect half of them as we are being discussing I expect half of them to be between boys and girls.

Now, let see how many of the friendships are actually between boys and girls. I observe that 1 2 3 and 4 these are the only four friendships between boys and girls, no not four actually it is three one two and three just 3 friendships between boys and girls, the rest is either between boys or between girls.

So, I now try to define what is homophily. By Homophily I mean take look at this fraction in the numerator you put what are the actual number of friendships across the divisions boys and girls that is in the numerator, in the denominator I put the expected number of friendships between boys and girls. Since there are 18 friendships I expect the

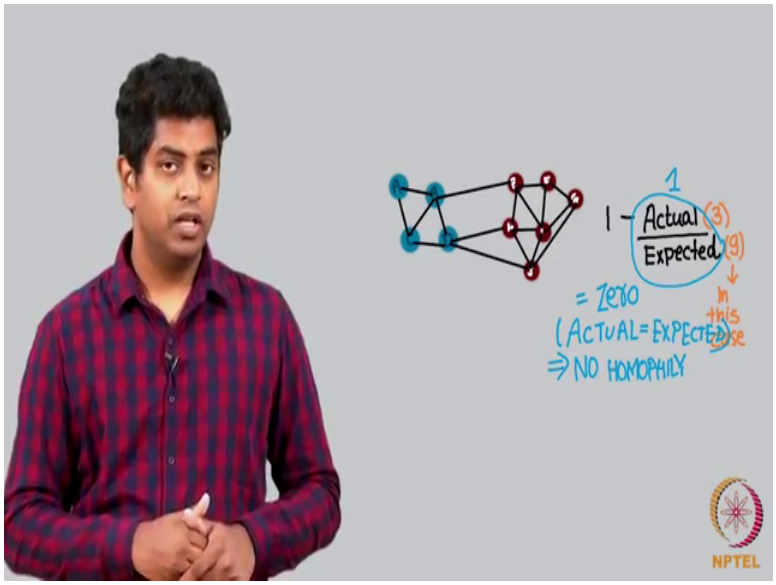
number of friendships between boys and girls to be half of 18 which is 9. Under the numerator is the actual friendship, across boys and girls which is 3. So, 3 by 9denotes homophily correct. Lesser this fraction more is the homophily think about it, more this fraction lesser the homophily why is that? If you think the numerator is very small then the fraction becomes very less. What do you mean by numerator is very small? By numerator very small I mean that the actual number of friendships across boys and girls is very small, which means there are lot of friendships happening within them which means there is a lot of homophily.

Now, what we do is a typical trick that we do in math physics, computer science in general is try to subtract this from one. So, what I do is I take 1 minus this fraction, actual number of edges divided by expected number of edges what does this denote? This simply denotes that as this number goes higher there is more homophily, as this number goes lower there is lesser homophily.

(Refer Slide Time: 13:03)

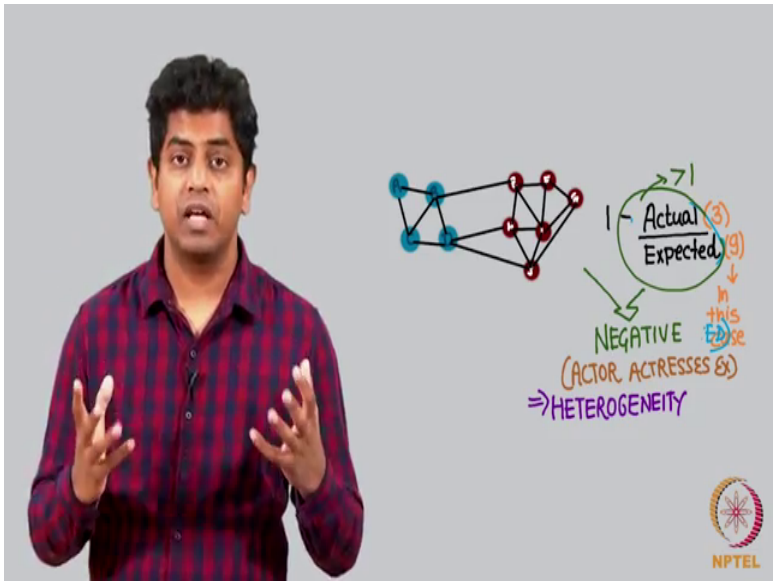
Let us look at this number do you look at the fraction? The fraction can actually become equal to 1 and hence 1 minus 1 can be 0 what does 0 denote here?

(Refer Slide Time: 13:10)



0 denotes that the number of edges across these two parities boys and girls, is the expected number of edges which means homophily is not happening here. Please note there is also possibility that this big number this entire number 1 minus this fraction can become negative; when is it negative? It is negative when the actual number of edges are more than the expected number of edges.

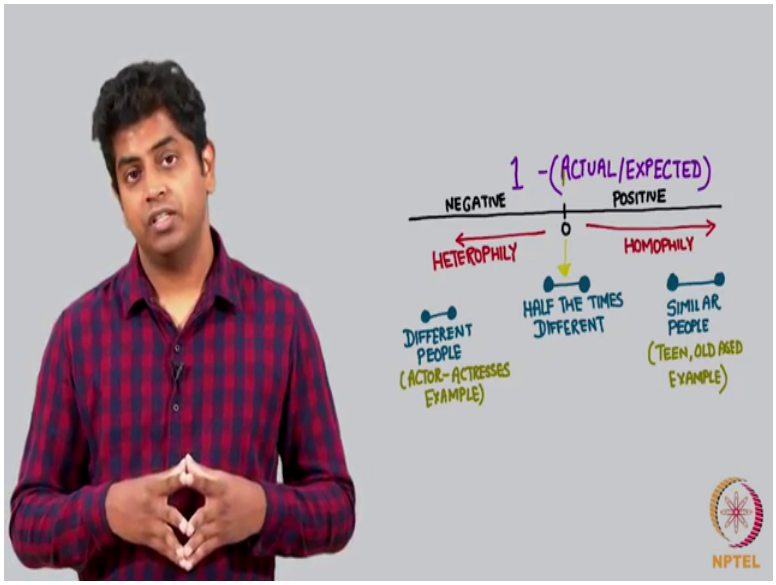
(Refer Slide Time: 13:36)



Do you see the actor actresses party example that I gave you, where friendships are mostly between two people of the opposite gender. Now what will happen? The expected

number in the denominator and the actual number in the numerator the actual number will be more than the expected number in that case 1 minus of this will become negative; and the negative number denotes the heterogeneity in the network, where friendships are actually between two people who are of opposite types, and a whole lot of friendships are like this.

(Refer Slide Time: 14:31)



Now let summarize if the friendships are such that across the friendship two people are similar more often, then you call such a network as homophily the network exhibits homophily. If they are half the times different then you say the network looks very random, then this number is 0, 1 minus actual by expected is 0. When almost always the types of people across this edge are different like in the actor actresses example party example, then you say such a network exhibits heterogeneity and is no way hemophilic.