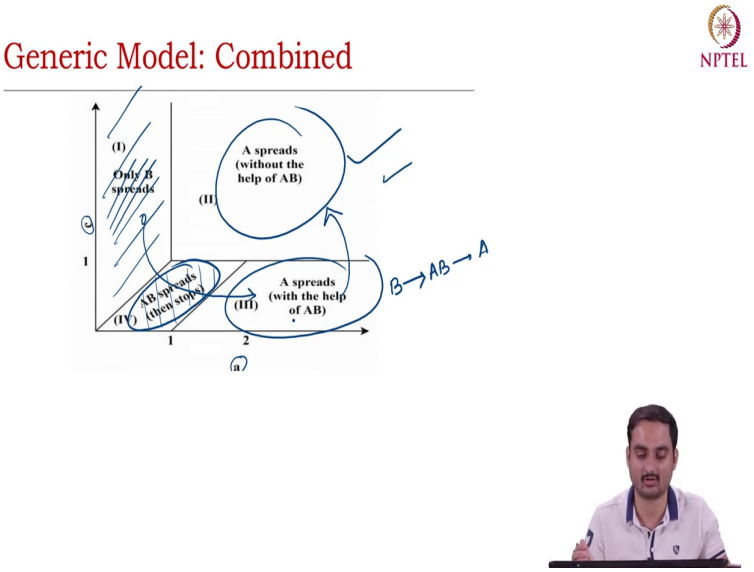
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 07
Lecture - 04

Alright. So, in last few lectures we have been discussing about you know information cascade. Particularly we have started you know discussing on decision based model and we have seen different cases where we choose a single strategy or multiple strategies and so on and so forth right.

(Refer Slide Time: 00:50)



So, if you remember last lecture we have stopped here. Where we have shown that in the space of A versus C, where a is the payoff when you adopt strategy A and c is the cost, when you adopt both A and B strategies at the same time how would the space look like right.

And we have seen that particularly cases like here. Where you basically you know you are sticking to a particular strategy B, which was your earlier strategy, but gradually you yeah a new product a new strategy penetrates the market and you are moving from this region to this part of the regions and then when you have actually capitalized on the inter market.

You have convinced your customers to retain the strategy A you would essentially increase the cost right with better products better quality and customers will also be happy. Because

the quality of the strategy, the quality of the product is essentially very good right. That is why the customers will not switch to any other strategy.

(Refer Slide Time: 02:05)

Now, let us look at a you know a real application. A real case study of a decision based model ok. So, this case study and this was basically a movement that happened way back 2011 in Spain. In Spain you know Spain witnessed one of the highest unemployment rates across Europe in 2011.

And that was around 21 percent unemployment and the entire crowd of Spain they were extremely unhappy they started protesting against the government right. Government also started taking many many activities for example, you know government increase the retirement age of working population, government gave access to cheap labors access of cheap labors to business owners, but none of the strategies actually worked out properly ok.

So, you know all these strategies were not acceptable to the citizen of Spain. And they started you know started a movement and this movement actually started on social network ok. And we will show that how strategically you know people recruited citizens in Spain or again outside Spain. So, that mass protest can I mean could happen against the government and gradually how you know how things became normalized right.

So, in 2011 right at that time Twitter was also quite new right, but they chose Twitter Because at that time Twitter was not that crowded and most of the government personalist, most of the

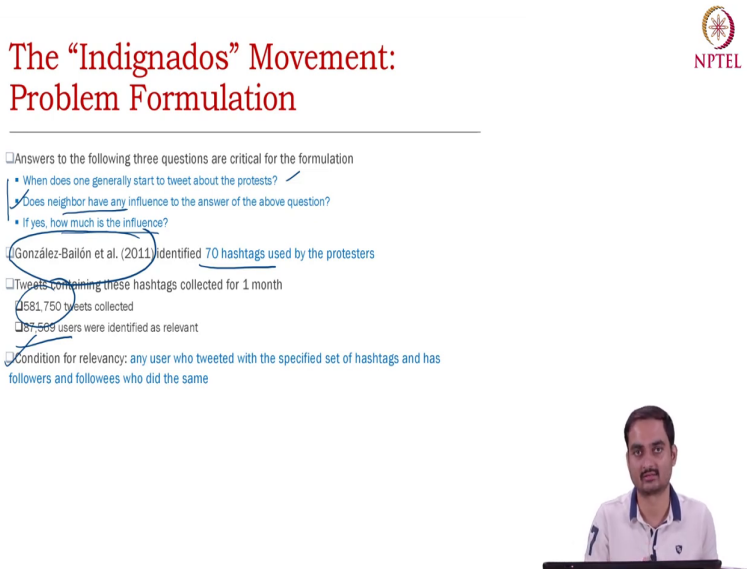
most of the political party leaders they were part of Twitter. And at that time Twitter was not that noisy right. In the sense like if you look at Twitter these days you see you would see a lot of garbage things right. Fake news misinformation and a lot of other you know crap activities.

But that time Twitter was quite new and then people start people respected the platform and you know joined there to consume information. Therefore, the protester choose Twitter, as one of the one of the social medium social media platforms and you know and then this protest happened and this protest is called as in indignado movement ok. This is a Spanish word and the meaning of this word is I think this is basically out outrage ok. And this movement is also this movement is also called as anti austerity movement right.

And this is quite popular if you Google it, you would get a lot of information about this indignado movement. And its it started around I mean officially started on 15th of May 2011. So, in this particular research paper right. In this particular research paper basically the author script data scripted tweet data related to this movement and tried understanding how recruitment happened and how actually things moved right.

For example basically the idea was to understand cascade effect. Particularly whether you know a citizen joined this movement due to social pressure or due to his or her own you know intention right. And we will see a series of such series of metrics, that we will use to understand the this particular movement on social media ok.

(Refer Slide Time: 06:30)



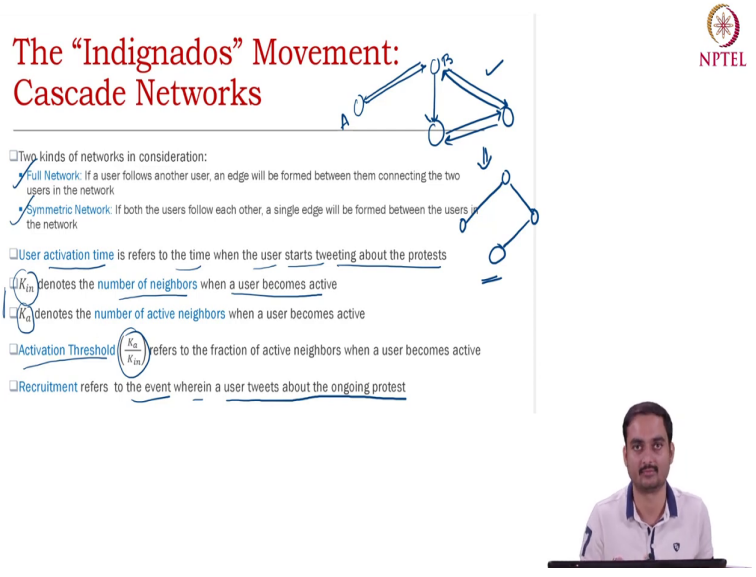
So, these are the following questions that we are trying to answer here. When does one generally start to tweet about the protest ok? Remember here we are looking at online protest. So, say for example, if I tweet about this one, I also become a protester right. If I tweet against the government right, I also I basically become a protester right.

So, the second question is does neighbor have any influence to the answer of the above question. For example, whether I am motivated by my neighbors or I join you know by you know by myself without any social pressure right. And how much influence a neighbor actually can make on a particular citizen right.

So, this was a study. So, what they actually did? They collected 70 such hashtags related to the protest. And they collected around you know 5 lakhs such tweets related to this protest. And around 87000 users related to this protest ok. So, what is the condition for relevancy and any user who tweeted within the specific set of hashtags and has followers and followees who did the same thing right.

Now, these are these would become the population which we would basically try to analyze. So, we look at those users who have tweeted who had tweeted about this one with one of these 70 hashtags or those who are followers or followees of somebody who had shown protest right alright.

(Refer Slide Time: 08:22)



So, we start off by creating a network ok. So, we look at two types of networks. One is called a full network. In full network we look at follower followee. So, users right they are linked by this follow followee links, directed links and this is a full network. The second one is more of a symmetric network.

In a symmetric network, we only consider those edges which are bidirectional in nature. Meaning we will consider those edges where A follows B and B also follows A right. So, if you have a network like this. For example, right. So, this is a; this is a full network and in the symmetric network we will consider this edge and this edge. So, this is kind of a kind of an undetected network, because we are considering symmetric edges, but this is a directed network ok. And for each user we define something called the activation time.

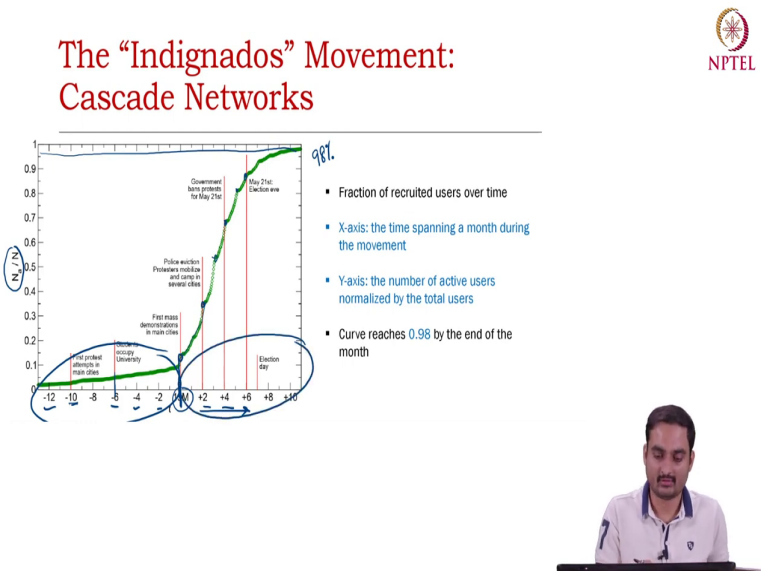
So, the activation time of a user is a time, when the user starts tweeting about the protest. The first tweet wrote the first tweet written by the user the time of that tweet is the activation time of the user ok. And then we define these two quantities one is Kin, Kin is the number of neighbors right.

When a particular user becomes active. So, when I become active, when I write the first tweet what is the number of what is the number of neighbors ok? That is the that is Kin and what is Ka? Ka is the number of active neighbors, when I join the protest.

So, Kin is the total number of neighbors out of them Ka number of neighbors are active when I join ok. So, we define something called activation threshold. Activation threshold is the fraction of active users, when I join the movement remember this when I join the movement I calculate this. This one activation threshold Ka by Kin ok. And recruitment refers to the event where in a user tweets about the ongoing protest ok.

So, if a new user starts tweeting about the protest. It means that user has been recruited ok. Now, we will try to understand the relationship between recruitment of a user; meaning the activation time of a user and the activation threshold ok.

(Refer Slide Time: 11:14)

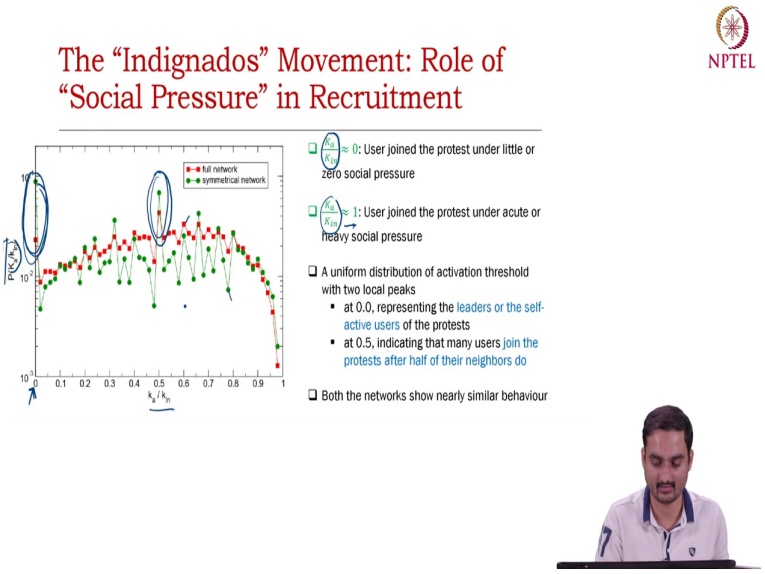


So, now this is the tentative kind of timeline. If you look at here this is 15th of May, when this movement officially started and this is 2 days before this, 5 days before this, 6 days before the movement, 12 days before the movement and the right side is 2 days after the movement, 4 days after the movement and so on and so forth.

And if we just plot Na by N, Na is a number of active users and N is a total number of users. If you look at this fraction of active users, you see this kind of plots ok. And you see certain some sweet spots right. Some you know some peaks some slopes and so on right. And this peaks right or this sudden increase in slope, these indicates different different you know different points. For example, this is the point when the first mass demonstration started.

This is the point 6 days before this 15h of May when students occupied the university. 10 Days before this first protest attempted in the main city and so on and so forth. So, if you look at it, it is around 0.98, the total that the fraction. So, 98 percent population were active during that time active protestor during that time ok.

(Refer Slide Time: 12:51)

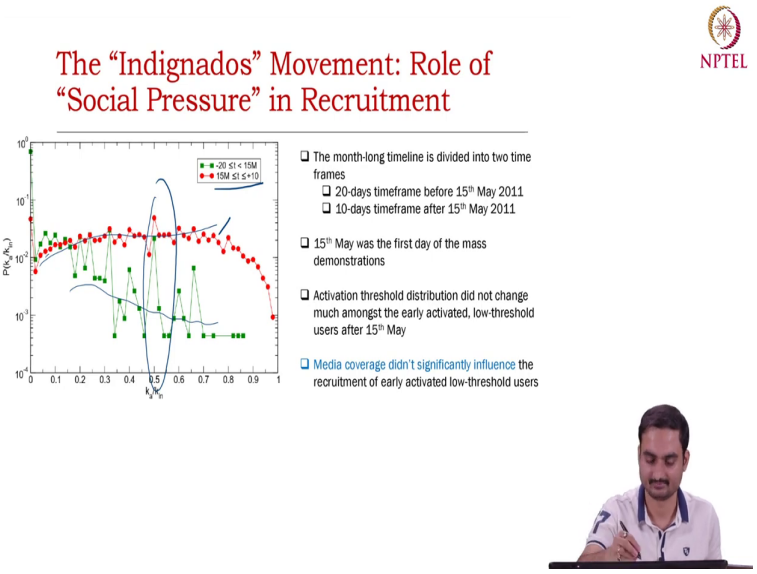


Now, let us try to understand this activation threshold, Kn Ka by K in. So, K a by Kin if it tends to 0, what does it mean? It means that I have a lot of neighbors who are not active right. Whereas, Ka by Kin if it tends to 1, it means I have a lot of active users I have a lot of active neighbors right. So, what we do here, we simply plot the distribution of this activation threshold ok. x axis is the Ka by Kin and y axis is the you know fraction of users with a given K by Kin number right.

And this red line corresponds to the full network and the green line corresponds to the symmetric network. We consider two networks separately ok. So, you see that in this curves there are 2 peaks. The first peak is here and the second peak is here ok. So, the first peak is at 0th time, where this protest essentially started right. And this peak is basically at the value of activation threshold of 0.

So, essentially this set of populations are those who join this movement by themselves without any social pressure ok. So, for this set of populations Ka by Kn is 0. Meaning there is no social pressure still they join because they were motivated to join. They are basically those users those citizens who started the movement ok. And another peak you see at this point, where Ka by Kin is around 0.5, meaning when 50 percent of my neighbors became active I joined. Now, this is because of the social pressure ok. Majority of my neighbors became active therefore, I joined ok alright.

(Refer Slide Time: 15:22)



Now, let us look at the time also. So, we study the same thing the distribution of activation threshold during this time and during this time separately ok. You see here this red line corresponds to the time after this 15th of May and green line before 15th of May. You see that the pattern is more or less same with some I mean more or less uniform this curve is also uniform.

This is also uniform kind of uniform, it means that and also peak around this time right. So, basically says the pattern is more or less same before the activation, because the before this protest started and after the protest started right.

(Refer Slide Time: 16:21)

So, now we look at another quantity, which we called burstiness. So, burstiness is defined by. So, what happens is that sometimes we do not look at we only I mean we do not get motivated by only the number of active neighbors.

We get motivated when suddenly, we see a lot of my neighbors you know started protesting and that too within a very short time alright. Now, let us look at another quantity which is called burstiness. So, in the last slide we have seen that you know what is the social pressure. I mean what is the fraction of active neighbors for which I join you know this particular movement. Now, let us look at, now this quantity may not be enough right.

What is even important is to look at how rapidly those active neighbors join the movement ok. So, say if I suddenly wake up and see that most of my neighbors joined the movement versus in a I see gradually my neighbors joined the movement and then I joined right. There is a difference right. In fact, in the first case, the first I think the first you know situation would motivate you more to join the movement right. So, we define the burstiness, burstiness is the.

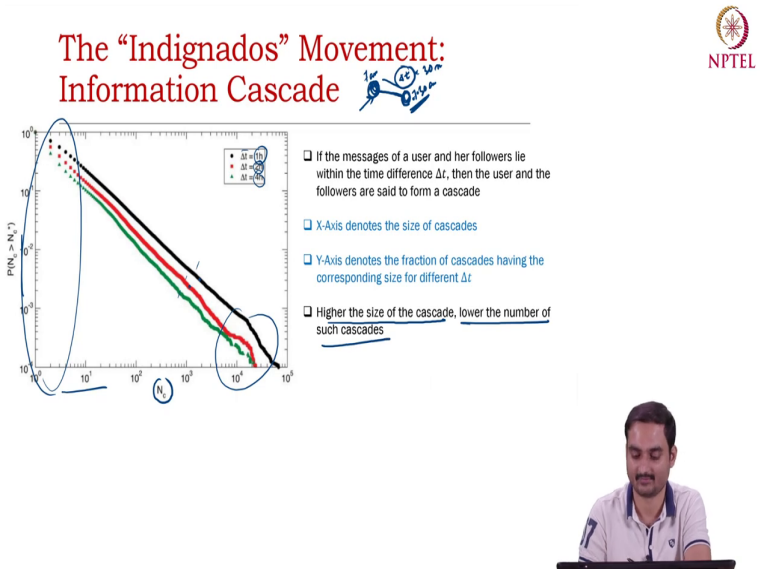
So, burstiness takes into account two consecutive time stamps at time t at time t plus 1. At time t number of active users at time t plus 1 number of active users and then we look at the difference ok. And that is normalized by the active neighbors at time t. So, these are all active neighbors ok, not active users. So, if you plot this delta Ka by Ka right with respect to ka by

kin right. So, what you see is that for those users whose activation threshold is quite low, for them also this burstiness energy is low.

They are actually motivated you know protesters. They join the protest without any social pressure. Whereas, if you look at the change in burstiness, after 0.5 right you see a steady increase right. So, it is essentially means that this is not only about activation threshold, but also the burstiness ok. Suddenly you see a lot of your neighbors joined the protest that motivated you to that motivated you to join the protest as well ok.

So, high activation threshold users are more likely to join the protest ok. If you see a sudden increase ok.

(Refer Slide Time: 19:34)



So, now let us look at you know the cascade. So, we have already discussed what you mean by cascade. So, here we look at a follower followee relation and if a follower tweets the or retweet or tweet whatever, the same content that one of his or her followees has retweeted right. And that too within a certain time gap delta t we will assume that this guy has got motivated by this guy ok.

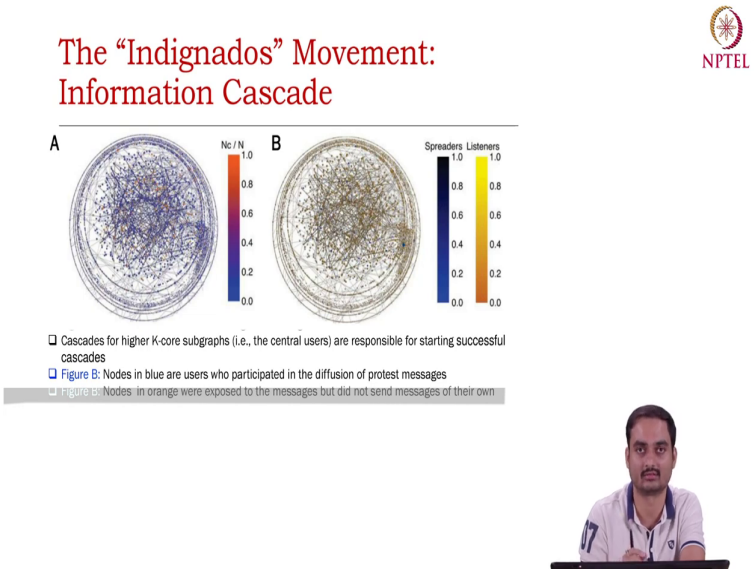
Say this guy has tweeted something about the protest at 7 am and this guy retweeted the same thing at 7.30 am and the window that we are considering delta t is 30 minutes ok. So, in that case we will assume that this guy has got motivated by this guy, therefore, retweeted. So, it

means that there is a cascade from this node to this node ok. So, in this way we can check the cascade the depth of the cascade right.

So, if we check the depth of the cascade right. And if we basically plot the cumulative distribution of the not the cumulative distribution basically the distribution of the length or depth whatever of the cascade and that too in a log scale. We see that it basically follows a power law right.

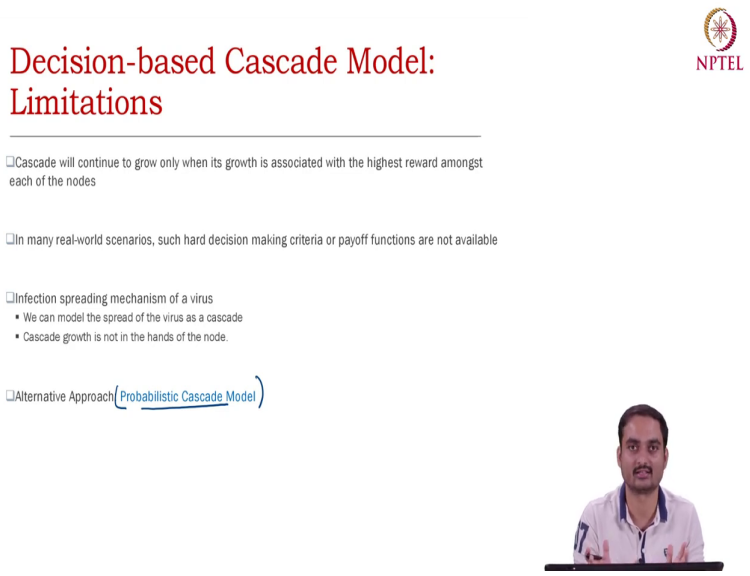
Now, why these three lines? The black line corresponds to this time gap, which is 1 hour, the red one is 2 hours and the green one is 4 hours ok. You see that the length actually follows power law. Meaning that there are lot of cascades with small length. And there are very few cascades with larger length right. So, higher the size of the cascade, lower the number of such cascades ok.

(Refer Slide Time: 21:48)



The last observation is that. So, we look at the core periphery structure ok. And try to understand whether the initiators are core nodes in the core periphery structure and you know this figure may not be that indicative, but you know I took this figure from the paper, but if you look at the findings it basically indicates that indeed the initiators were the core users of that network. Both the full network and the symmetric network ok.

(Refer Slide Time: 22:29)



So, that is about that protest. You can actually you know come up with many such interesting studies with such social media protest based on the evidence. Now let us look at the limitations of decision based cascade models right. So, the first limitation is to define all this payoff right, payoff values, cost values and so on and so forth. Now, cascade is a continuous cascade grows continuously right. So, defining the reward and other parameters that would be difficult for a certain information diffusion setting right.

And in many cases it is not about whether you want to decide your strategy. It is also about whether you accidentally get affected by a certain disease or certain protest or certain activities right. So, therefore, this decision based model may not be good enough for cases like COVID-19 spread or say Ebola spread or any disease spreading right.

Therefore, we will discuss in the next lecture something called probabilistic cascade model. Where we will not where the nodes will not have the choice to choose which strategy to choose right. It basically it depends on the spread of the disease it depends on the way people interact, it depends on other social restrictions like mask wearing mask lockdown and so on and so forth. So, we will discuss this thing in the next lecture.

Thank you.