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

Lecture – 43
Strong and Weak Relationships (Continued) and Homophily
Interplay between Selection and Social Influence

(Refer Slide Time: 00:08)

So, we saw the big question, what is the difference between social influence and selection; which one of the two is happening, are both happening. Now, it is very difficult for us to go ahead and look at the real-world dataset of friendships and ask this question. If we ask this question of, can we have some very concrete dataset where this experiment can actually be conducted, and the facts be unraveled.

(Refer Slide Time: 00:48)

There was one such data set which gave some very promising observations. So, our motive for this video chunk is to understand this piece of research on Wikipedia.

(Refer Slide Time: 00:57)

Let me now motivate the concept by explaining; how Wikipedia works, number one and number two by explaining what one means by similarity measure; these two are important for us to go any further.

(Refer Slide Time: 01:08)

(Refer Slide Time: 01:12)

So how does Wikipedia work? Wikipedia basically is a blank slate; you can basically anyone can go write anything on Wikipedia. For example, you can open the Wikipedia entry of India; edited. If you think some information is incorrect there, delete it add the right content; that you think is right and if someone else thinks; what you have entered is wrong, they will come and edit it and change it to what they think is right this goes on and on.

(Refer Slide Time: 01:42)

One frequently asked question is, can we really trust such a database, where anyone can come and enter anything. Very intriguing, but it is known to be true that when a lot of people get together and start discussing and debating, when they have conflicts of their ideas; true knowledge is known to emerge.

Initially of course, there will be some false information here and there but a lot of people watching it, will converged to the right answer. So, Wikipedia is actually believed to be very trustworthy these days; that is about Wikipedia.

(Refer Slide Time: 02:14)

But what happens in the background of Wikipedia? There are many people who will be coming and editing it.

(Refer Slide Time: 02:20)

If there are conflicts on their belief of what should be put and what should not be put, they talk in the background with each other and the background talking is also visible for all of us; they are called user talk page. One can click on this tab and then take a look at what all discussions they have about the content.

Now, this big dataset is available to our disposal and one can conduct any kind of research on this data set.

(Refer Slide Time: 02:57)

I am now going to explain one such research which actually helped us understand the inter play between social influence and selection.

(Refer Slide Time: 03:09)

As I told you, just now I helped you understand how Wikipedia works; that was the first prerequisite. The second prerequisite is understanding what one means by a similarity measure. So, let me motivate what is the similarity measure now, so let me define what is the similarity measure; it is actually quite intuitive. Let me motivate it with a good example.

(Refer Slide Time: 03:22)

Assume, I like the following dishes I like upma, let us say idli, dosa, pizza and let us say fruit punch; this is the five things that I like upma, idli all those things are South Indian dishes.

(Refer Slide Time: 03:46)

(Refer Slide Time: 03:56)

And let us say my friend Priya she likes pizza, burger let us say pasta and idli; these four things. As you can see, we do not have a lot of commonalities, but we have; we like pizza and idli that is common between both of us. So, what I do is in the denominator I put the total number of items that we like, whatever I like and whatever Priya likes; both put together as you can see is 7 items, out of which we both like 2 items.

(Refer Slide Time: 04:35)

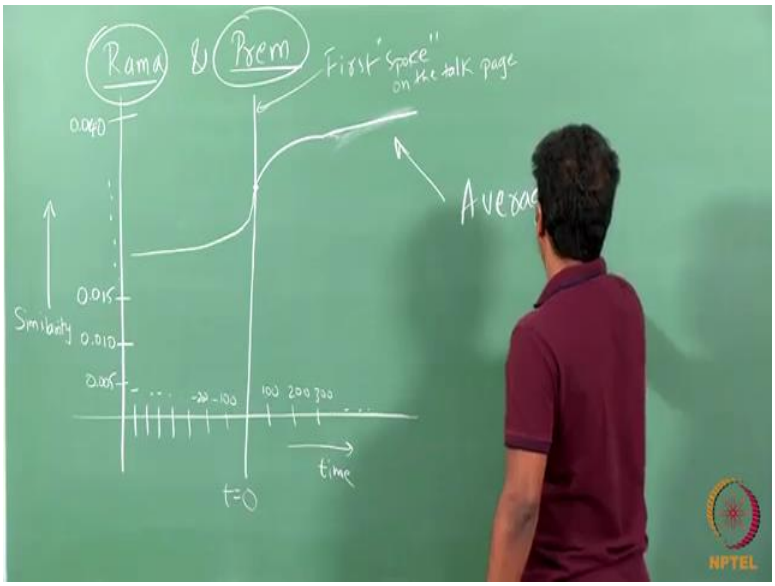
So, I say the similarity measure between me and Priya is 2 by 7; this is what we mean by similarity measures. Now, I am going to use this on my Wikipedia dataset that I was

telling you. So, what do I do? There are two people two authors who are editing Wikipedia pages randomly here and there let us say; I am editing the Wikipedia entry of India and let us say Chennai floods and Wikipedia entry on iPhone 8 and things like that; different Wikipedia pages I am editing. And a friend of mine let us say Peter is editing Wikipedia pages, again some other Wikipedia pages, some ten other Wikipedia pages.

The similarity measure between me and Peter is the total number of Wikipedia pages that we both are editing, total number of unique Wikipedia pages that we both are editing in the denominator; and in the numerator we write those Wikipedia pages we both have edited, just like the similarity measure that I defined between me and Priya. The similarity measure between me and Peter on Wikipedia is total number of Wikipedia articles that we both have edited in the numerator; divided by the total possible unique Wikipedia pages that we have edited, both put together; that is in the denominator. This gives a good measure of how similar we both are on our Wikipedia editing transactions.

So, now that you understand what is; how Wikipedia works, and you also understand the definition of similarity measure. I am going to use these two things in answering the big question of social influence versus selection. So, here it goes; I will do the following, I will look at two people who have spoken to each other on Wikipedia. I told you people do talk in the talk pages, in the background; if I observe that two people have spoken to each other, please note I have a dataset here; in that data set I observe if two people have spoken to each other.

(Refer Slide Time: 06:52)



If let us say two people; I will say Rama and Prem have spoken to each other on the wiki talk page, then what I do is I plot this. What is the plot? The plot is the following, this is my Y axis, this is my X axis, I will see when exactly they spoke. Let me call that time T equal 0 and I will observe, what were they doing before and what were they doing, what did they do after this time equal 0, What do I mean by that?

Let me define what is the X and Y axis properly; on the Y axis, I will put their similarity measure. Their similarity measure is given by let us say 0.005, 0.010 so on and so forth; 0.015 so on up to let us say 0.040; I am doing this on the Y axis. This is my similarity and this is my time; I repeat Rama and Prem have started editing some Wikipedia articles and their similarity measure, I am seeing when they first spoke; this line denotes the time when they first spoke; by spoke let us be clear.

I mean first time they spoke on the talk page and I observed what was their similarity measure before and what is their similarity measure after, it was observed that the plot looks something like this; it looks like this. Let me correct this, it just goes steep it just goes like this that is it. Point to note is the steep increase and further increase and then it becomes sort of constant. What is happening here? Let us just pause and observe; let us revise time, similarly, measure Rama and Prem.

Their similarity measure was; so, this is time when they started let us say some few days before and the first time they met, this probably some minus something minus 100,

minus 200 or whatever is the scale and the time when they first met and then time t equals 100 and then 200 and then 300 so on. Whatever you want, 200 minutes or 200 seconds or you can have any scale here, but this is the time when they first spoke to each other.

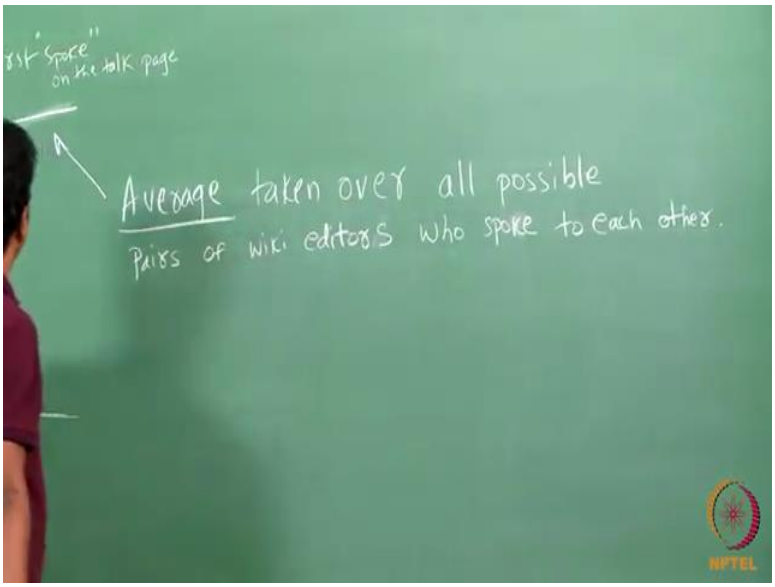
Do you see something happened here? Something fascinating, what does the curve tell you. The curve tells you that before they first spoke to each other, there was a sudden surge of similarity between them. These two people were similar because of which they spoke maybe, because of which they had a talk on the spoke on the talk page maybe. Do you see; after they spoke there was increase in their similarity measure. What does this translate to? This translates the following because that their interest were common, that is what similarity measure means because that their similarity was high, their interest were common; they actually spoke to each other, after speaking there was further increase in their similarity.

This translates the following: two people become friends, if they agree on lot of; let us say dishes in my previous example of me and Priya; maybe if we have 2 by 7 similarity, we may not become friends. If we agree on a lot of food interest, maybe we will become friends; I do not know whether if this is true or not, but assume that people become friends because they have a lot of common food interest. What will happen after we become friends? Once we become friends, we tend to have all the more common food interest, we will probably will hang out together and eat more often the same kinds of dishes, give each other our testimonials of this kind of food in this kind of place and we may want to try these things in each other's presence.

So, the fact that we are similar makes us talk; the fact that we have spoken right now makes us become all the more similar. Fine, this plot was given to you for reasons of better explaining things to you, but the fact here is; if you look at the textbook, this plot is the plot of the average. By average, I mean they have looked at all possible people who have actually spoken and they observed; what was their similarity measure before and what was their similarity measure after. This is the average plot that you see in the text book; which says that on an average people tend to become more similar and then talk and then become all the more similar not just this.

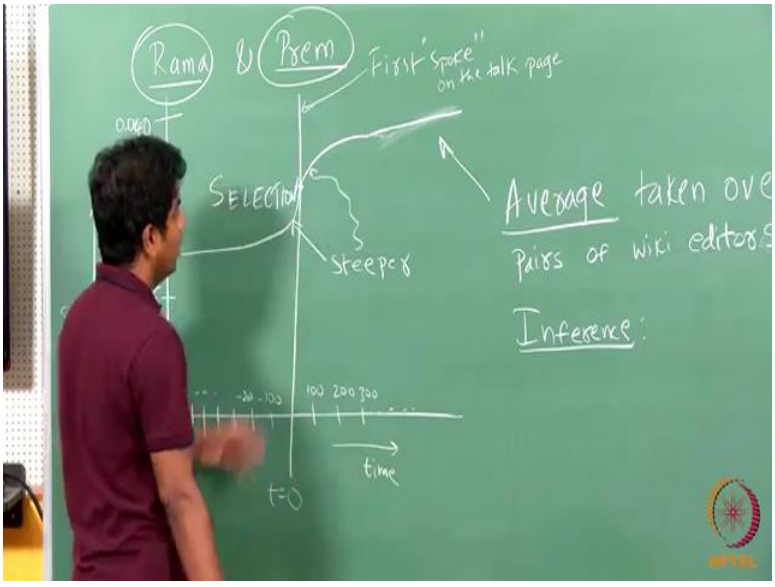
This; what is coming next is what makes the research all the more interesting. They observed; I repeat on an average people become similar; people because this is average, I told you.

(Refer Slide Time: 12:59)



This is average taken over all; let me write that down, average taken over all possible pairs of wiki editors; who spoke to each other, sorry who spoke to each other and we observed that; when they speak at this time, there is sudden increase in their similarity measure before speaking and after speaking; this is the average curve.

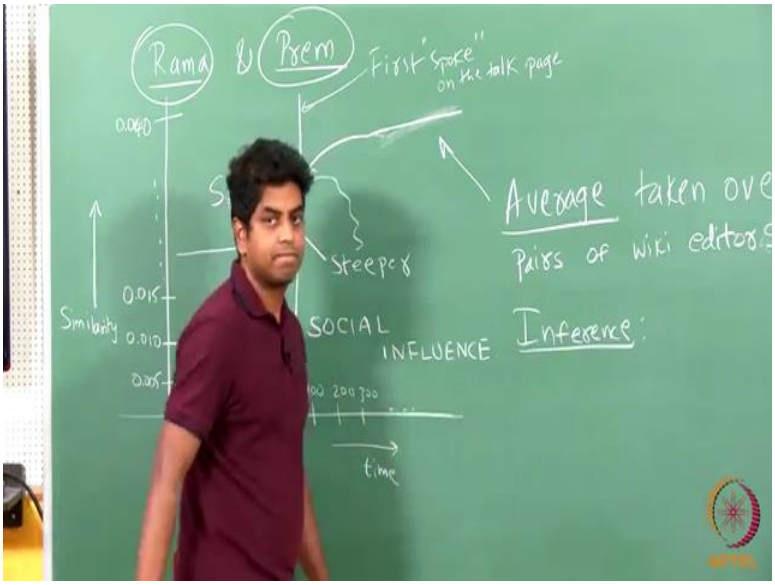
(Refer Slide Time: 13:49)



A mega inference that one can make here is that this is steeper than this, now that is the climax point in the piece of research. This is steeper than this; let me write that, this is steeper as compared to this. This is steeper; than this, what does that mean? That means, that you become similar and then you talk and then you continue to become better and better similar. So, what is this and what is this; take a minutes pause and observe, what is this and what is this in our language that we are motivating from the past few minutes.

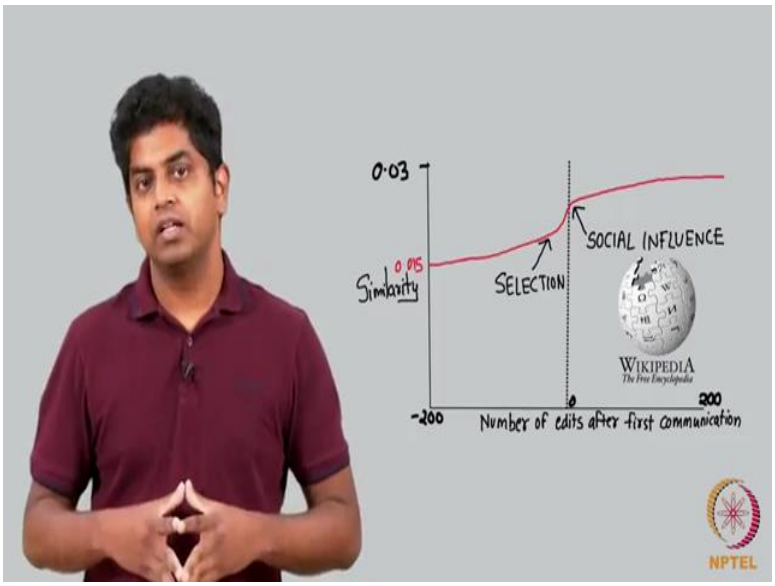
Which one is social influence, and which one is selection? When do we select? We select when people are similar. So, I am sure you would have figured out the answer right now, two people select each other to talk when they are similar. This is selection here happening on this part of the vertical T equals 0 line selection and on this part; once they spoken to each other, maybe they will share more common interest and they will started eating pages together, that is what is happened, that is what this means; there is steep increase in the curve here as well and this is going to be our social influence.

(Refer Slide Time: 15:21)



So, we just now saw the plot and we observed the interplay between selection and social influence on a particular type of dataset.

(Refer Slide Time: 15:34)



Now, let us get a little critical and ask this question; fine the plot is very clear, it says that initially there is a lot of selection happening here and then social influence is also happening, selection seems to be more in play than social influence while both actually happening; that is with this dataset. Maybe this kind of a observation is very context dependent, dataset dependent.

(Refer Slide Time: 16:16)

Now, instead of taking Wikipedia dataset, if I were to take some other data set, let us say the way smokers become friends with each other. Maybe there is a lot of social influence

factor there, the un-selection factor correct and in fact, obesity research says otherwise whatever we saw here it is the reverse of that. Here, there is more selection than social influence, while in the obesity research what we inferred was there is more social influence than selection.

This is a very context dependent observation the people have made in fact; this opens up brand new questions namely. Which is more in play in the society? So, be it with for bad habits, is it more of social influence than selection. For good habits, is it selection and less of social influence; one can ask many such questions and thankfully we do have some datasets on which we can experiment these questions and then find our answers.