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

Lecture – 14
Handling Real-world Network Datasets
Introduction to Datasets

(Refer Slide Time: 00:20)

So, here we are with a brand-new chapter on datasets.

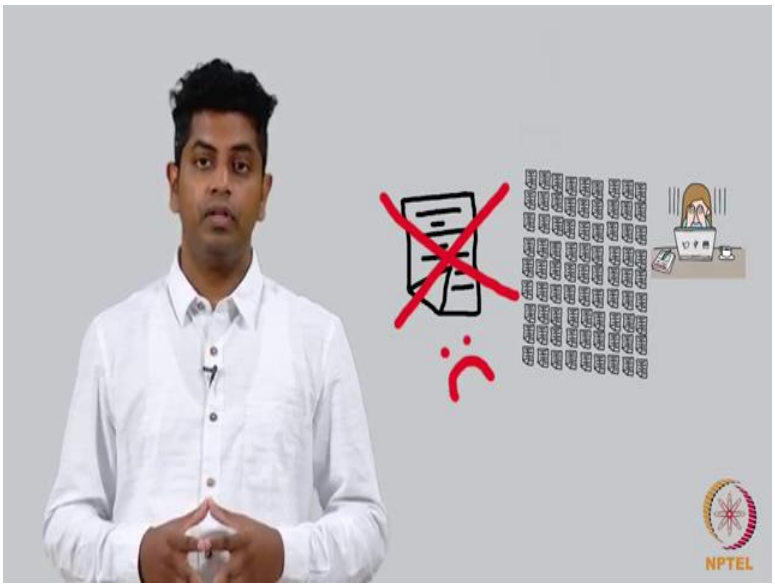
(Refer Slide Time: 00:26)

So, we are going to teach you the importance of crunching datasets and how to go about crunching them.

(Refer Slide Time: 00:30)

So firstly, what are datasets? Especially datasets that are relevant to social networks.

(Refer Slide Time: 00:38)



You see there was a time when we did not have any data and today we are in that era where it is all about data we have so much of data that we do not know how to make sense out of this data that a complete contrast over it was let us say some 30-40 years back.

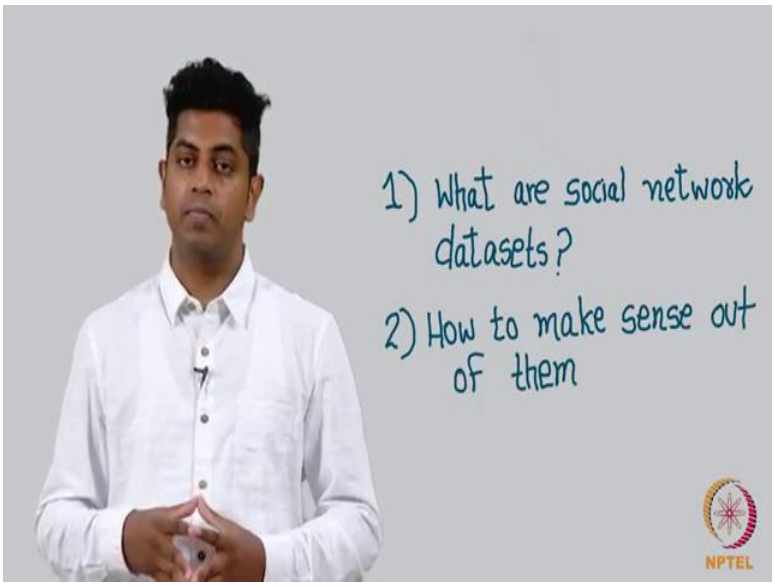
(Refer Slide Time: 00:53)

For example there was an; there was time when people look at the friendship network on 30 to 35 nodes, this is called to the Zachary Karate network, it had some 32 nodes and some edges on it, it is very popularly studied, very widely studied and are very popular social network, but today; back then we did we probably did not have a way in which we could get a bigger internet.

(Refer Slide Time: 01:28)

But today with the advent of internet and higher computing and storage facilities, we can crunch a whole lot of data whether we can crunch or not we have a whole lot of data.

(Refer Slide Time: 01:42)



So, this week this a entire module is all about understanding what are social network datasets and how one can make sense out of these things.

(Refer Slide Time: 01:52)

Please ensure that you have understood Python and networkx and another APIs on Python really well, especially the things that we thought in the last week.

(Refer Slide Time: 02:03)

I am sure you all know of this Play-Doh; Play-Doh is where you have this sand like thing that is sort of you can make any shape out of it and you realize that the options are multitude, you can go on making a small box out of it or a cute animal out of it, I mean we are in with limit is just in your imagination.

(Refer Slide Time: 02:28)

So, similarly when people got hold of this big network data set they realized that one can actually asked the whole lot of questions on it. Let us say first question that one can ask

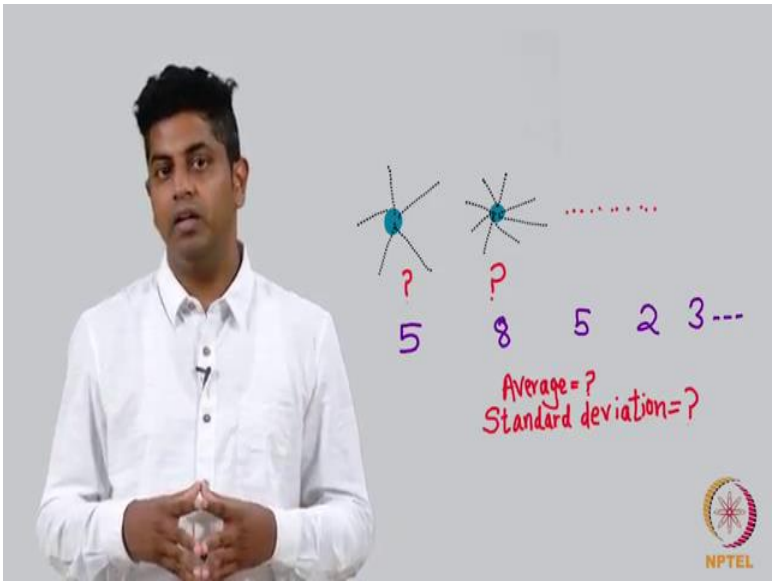
on our friendship network, there is a classroom full of 100 students and data is available to me of their friendship network, who is friends with whom?

What kind of questions can we ask the obvious question is the graph connected we have discussed in a on this topic before most all graphs are connected, fine?

(Refer Slide Time: 03:06)

Second question; what can you say about the degree of the nodes?

(Refer Slide Time: 03:10)



By degree you mean; pick a person look at how many friends he has, pick another person look at how many friends he or she has so on and so forth, make a list of all these friendships of individual people, what is the average here and what is the standard deviation of this list what are the mean by this? By this I mean is the average sort of do everyone agree with the average or is it very widely spread of what uses is such that kind of research?

(Refer Slide Time: 03:45)

Data sets gave us this ability to question random stuff like this just like we would build a random stuff with Play-Doh and then played around with his datasets and arrived at fantastic conclusions. So, now, what we are going to do is we are going to pick some random datasets for a discussion right now. I am going to go one by one go through some 8 to 10 datasets and explain what these datasets and what kind of question are can one ask and answer.