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

Lecture – 26
Handling Real-world Network Datasets
Summary to Datasets

This week we saw the side of with introduction to datasets we saw why datasets, the fact that this error has helped us collect a whole lot of datasets and even we have good computational facilities to make sense out of these data sets. We should you using python networks x like APIs, how one can write a piece of code and infer make inferences from these data sets. We also told you a powerful tool like Gephi where you input a graph and it beautifully shows you the different nodes and edges of the graphic we can visualize it in different ways, it appears very nice.

Gephi is not just used to visualize graphs it is also used to analyze graphs especially for people who are not from programming background. If you were to analyze a given a graph quickly and you want to make sense out of it in let us say no time, python is not your destination it has to be Gephi like tool. But python again gives you the flexibility because to write your own code the way you want a Gephi does not it is sort of standardized. Nevertheless we saw a tool like Gephi which we realized is both powerful and fun to use and then we moved on and saw a piece of puzzle which we have been discussing from the first weeks first lecture which is the emergence of connectedness. We saw how a graph becomes connected when you take a bunch of nodes without any edges and you start putting edges one edge at a time and you will see that there is a movement when it becomes connected, when does it become connected? If you remember it is when you put n log n number of edges.

Then we also saw that it is not just mathematically true that its n log n our program also says it is something to do with a little more than n number of edges for a graph with n vertices to become connected and that talks about the power of what is called synthetic data sets. Synthetic means you synthesize a dataset of your choice for example, you have a huge dataset of 2000 nodes or 2 million nodes let say, but you want to observe what happens on a graph of 100 nodes say you would like to create your own network of 100 nodes. We this the experiment with which we concluded this week was a typical example

of how one can do this. We took, you saw the plot of as you take vertices and you look at the emergence of a connectedness and the plot looks plot has something to tell you right. So, that is a good example of how you can make your own data sets as and when required.

There are many ways to do this we will cover that in the forthcoming chapter. To put the entire chapter in a nutshell we learnt about datasets we saw the emergence of connectedness and we saw the importance of synthetic data sets that us with this week. So, next week onwards we are going to shift our gates completely, the past two weeks is only been playing around with python and understanding some basics of a network x and crunching datasets and things like that. Next week onwards we are going to see some hard core results in network science. So, all that I told you in the first weeks lectures on do you think you should watch your company, how does Google work, how do you predict who is your next friend so on and so forth all those things are coming next.

So, next week especially we are going to start off with a clip of Harry Potter and we will see what has one to learn from this clip.