My course title is introduction to

data analytics and today we are

learning about social networks.

Mining social network graphs.

So outline for today's topic is

introduction to social networking.

Then how to view a social network as a graph.

Clustering of graphs and direct

discovery of communities in the graphs.

At the end of this session,

students will be able to understand

the term social networking,

visualize the social network as a graph,

and understand clustering of such graphs.

So when we talk about social networking,

it is a Internet based social media platform

to stay connected with their friends,

families and our peers.

So we have used many social media

networking websites like Facebook,

Instagram, Twitter.

Here the main goal of this networking.

Site is to stay connected with our

friends and families so we have

to form a network social network.

How it is visualized or how actually

the communities are identified on

social network that we are going

to learn in today's session.

Social networking here.

Social media can help to connect

the people and businesses and

help from what brand events?

So an important question that we have to

answer is how to identify communities.

So one way of identifying the

communities on social media

networks are we can use clustering

algorithms to identify communities.

So,

but when we think of our

clustering algorithms,

the club,

the communities which will be

identified will be based on the

set of nodes which we which we are

going to find out in the network.

In clustering algorithm what happens

is the communities usually overlapped,

so we will not get the actual

communities which are related.

So if there is any relationship between

two members of the two communities

it will consider that as a one.

Community which shouldn't happen

so that we have to eliminate in

clustering algorithms so that we will

see later how it is eliminated next.

Weaving our social networks as a graph.

So here the most important term that

we are learning in this is a locality,

so locality is nothing but it is a

property of social network that says

that nodes and edges of the graph

tend to cluster in the communities.

So when we think of a social network,

we always intervenes us about the

various social media networks like Facebook,

Twitter and other website and that

is called as a social network.

So this kind of network is a

representative of broader class

of networks or social.

So the essential characteristics which

are required for any social networks

are first important characteristics.

Is entity,

so network is visualized as

collection of different entities

that participate in the network.

So these entities,

maybe it can be people,

or they can be something else also.

Second important characteristics of

any social network is a relationship

between the entities which are

present in the network.

For example, on Facebook,

when we say relationship,

we say that two entities or

two persons are related

when they are friends with each other, right?

So that that is about your Facebook.

When two Friends 2,

two people are friends on Facebook,

we say that there exists a

relationship between those two people.

OK, those two persons next is sometimes

the relationship is all or nothing.

There will be some relationship or there

won't be any relationship between two people.

The third essential characteristics

of a social network is locality

that is based on the relationship.

We tend to cluster if entity is related

to both B&C then there is a higher

probability then average that B& are.

Related, so locality plays an important.

Locality plays an important

place in social network.

Social Network is a graph.

We can model the social networks

as a graph which we sometimes

refer to as a social graph.

So here the entities in this

graph will be the nodes,

and an edge connects two nodes

if there exists a relationship

between those two people.

So if there is a degree

associated with the relationship,

then that degree will be labeled

on the edges.

OK, so often social graphs are undirected.

As for the Facebook,

there will be a friends graph.

Correct?

So social networks is a graph you can

see in this example of a social network.

Here you can see that there are altogether

abcdefg 7 nodes and there exists a

relationship between two adjacent nodes.

For example, there exists a

relationship between A&B&C&C&B.

So if we talk about communities,

ABC forms of 1 community.

OK, so your node B is with his.

Is friend with A and friend

with C and friend with D.

So there exists relationship

between A&B C&D and D&V.

So how to identify communities?

So for identification of communities we

use clustering of social network graphs.

So one method of clustering is

hierarchical clustering in hierarchical

clustering of social network graphs,

we start by combining some two

nodes that are connected by an edge

and successively edges that are

not between two nodes of the same

cluster would be chosen randomly.

To combine the clusters to

which they are two nodes yellow

so the choices could be random

because all the distance is

represented by an edge other say.

So for example,

if you talk about the graph which

we have seen in Figure 10.1 here,

there exist two communities,

2 broad communities.

We can say one is ABC.

I'll go to that graph.

This is ABC community and there is

another community which is DEFG.

So broad categories or broad

communities are ABC and EFG.

So out of that your DFG can be

splitted as DF and qFG 2 separate.

Communities,

right?

So these two subcommunities they are

overlapping into their two of the members.

So what is the problem with

hierarchical clustering?

Is that at some point of time we are

likely to choose we are likely to

choose to combine B&D because there

exists a relationship between B&D.

So we may add it as A1 cluster.

So the reason we are likely to combine

BND is that D and any clustering

containing 8 is as close to V and

any clustering containing it is is

as close to A and C are to be. OK,

so that's why we make cluster B&D together,

which shouldn't happen actually.

So that is the problem with

your hierarchical clustering.

So next method of identifying communities

is direct discovery of communities.

So indirect discoveries your vertices

in the network are often found to

cluster into tightly knit groups with a

high density of within group edges and

lower density of between group edges.

So example of a real world community

network is shown in this graph.

So we we identify tightly need groups

by identifying groups which have.

Find density within group edges and low

density between between group edges.

So this is the method which is

better than your clustering method

of identifying communities.

These are my references.

Thank you.