

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

Chapter - 01

Lecture - 01

Hi everyone, welcome to the course on Social Network Analysis. My name is Tanmoy Chakraborty and I will be the instructor of this course. This is going to be a very fun course because of the obvious reason, as the name suggests this is on social network and we all are part of you know some kind of social networks, Facebook, Twitter, Instagram, WhatsApp what not right. And we browse social networks every day and we know how to react to certain post, how to like, how to share, how to communicate, how to increase the friendship, how to follow somebody.

But we may not know that how you know social networks are generally operated right. For example, say you know when you as you know wake up in the morning and start browsing your Facebook profile you certainly see that you know a very interesting post and so why that post is suddenly you know come to your news feed right. For example, or you are you know user feed.

Why suddenly you see a user right, who has been you know recommended by say Facebook as your friend right. So, you see you do not know how such things are operated right. And you know these days we all you know have been hearing about Twitter being you know sold by and any and Elon Musk who is the you know richest man in the world, he has bought Twitter right. So, what is there in Twitter that he has suddenly decided to buy Twitter right, why not in other company?

So, there is definitely something that is there in social network and we need to understand the importance of that, right. So, this course is all about you know understanding the importance of social network, how social networks are being operated right. What are the fundamental you know properties of social networks and you know how you can utilize social networks for different purposes.

(Refer Slide Time: 02:27)

Acknowledgement

<http://web.cecs.pdx.edu/~mm/ExploringComplexityFall2009/>

<http://snap.stanford.edu/class/cs224w-2018/>

<https://web.stanford.edu/class/cs224w/index.html>

<http://snap.stanford.edu/class/cs224w-2015/handouts.html>

<https://cse.iitkgp.ac.in/~pawang/courses/SC15.html>

https://cse.iitkgp.ac.in/~animeshm/course_cri2012.html

https://www.slideshare.net/ssuser66b1fd/overlapping-community-detection-survey?from_action=save

And many online content



So, before starting this you know course, I would acknowledge the content some of the content that I have taken from you know some of the excellent courses offered all around the world. For example, courses offered by urlescopic and Stanford courses in you know different different Indian Institutes as well. So, these are some of the lists that are you know, that I have mentioned from where I have I have taken different content, but of course, the list is endless. So, you know I have just acknowledged a few, but this is endless, ok.

So, I would like to thank you know all the content creators for creating such a wonderful course content and based on that I have tried to design my own course.

(Refer Slide Time: 03:06)

How many of you feel that you are part of a network?



• In fact, you are part of multiple networks!

• Do you feel an inertia/pull/attraction, being part of the networks?



So, let me start by you know asking a interesting question, might be very silly question. How many of you feel that you know you are part of a network? Right. You may you know you may say that what is network right. So, I am a part of Facebook; Facebook it a is not a network right; what do you mean by network? In fact, if you say that you are not a part of any network you are basically lying right.

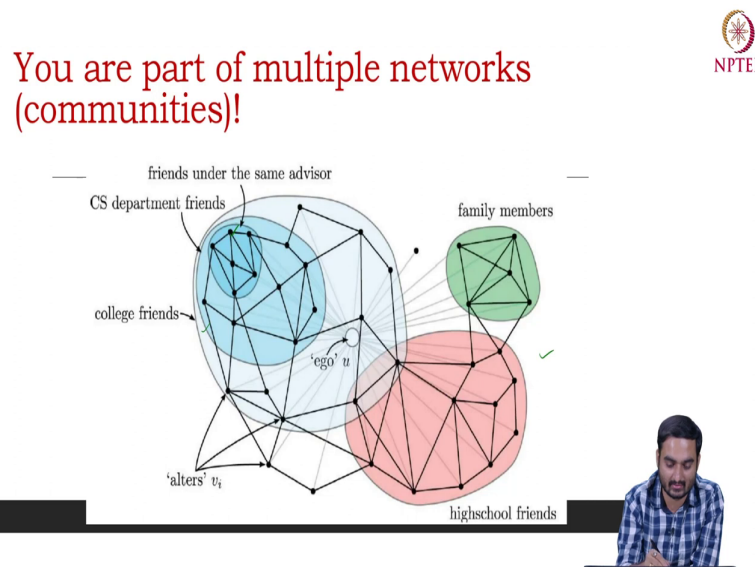
So, you are part of multiple networks not 1 right. And I am pretty sure that when I am delivering this lecture or when you are you know hearing this lecture you are at the same time chatting with somebody on WhatsApp or say you know browsing your Facebook profile and so on and so forth.

So, every time you know you feel some sort of inertia, some sort of you know attraction in one form or the others from this kind of social media, right. So, if you feel that you know this Facebook, I mean when I talk about social network I only refer to Facebook or Twitter or WhatsApp or you know WhatsApp or Instagram, actually not right; I am also referring to social network which is an offline network.

For example, say for example, you are part of a community you are your friends are going to the gym at the same time right, you are part of a music club for example, and you are part of a college right you interact with your friends, your colleagues in a regular basis right that also forms a network right. So, when I say that social network it can be online social network, it can also be offline social network right.

Online social network as I mentioned can be WhatsApp right you know you just say you are browsing an internet right, internet itself is a network right. I will discuss why we call internet as a network right. It can be a you know telecommunication network for example, right or it can be simple offline social network user interaction network right.

(Refer Slide Time: 05:14)



So, as I mentioned we all are part of multiple networks and this is a stylized example right for example, say you know a student right in an university a research scholar for example, right. So, he is a part of the departments he is part of the CS Department for example, he or she is a part of you know the college, he or she is a part of the same friends club who are advised by the same advisor for example, right and so on and so forth. He is also part of the same family right.

So, you see that multiple offline social networks also emerge where we are part of in our daily life, day to day basis, right.

(Refer Slide Time: 05:57)



What is a Network?

It is not (only) a computer network.



So, let us start right, so what is network. So, when I say network right, in fact, when I was in my undergrad. So, when I when somebody said that you know do you know what is network, so I feel the I always felt that ok network is computer network right. There is TCP, you know UDP protocols and all these protocols you know packet is being sent from one say one computer to another computer, one router to a computer and so on and so forth right. So, here network does not mean computer network.

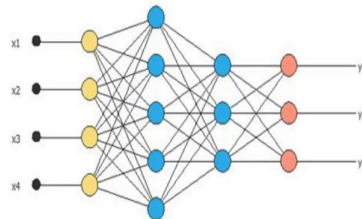
Of course, computer network is an example of the network that we are going to talk about, but this is not the only network that we are talking about here, right.

(Refer Slide Time: 06:41)



What is a Network?

It is not (only) a neural network.



And these days when I ask a student that you know what is network, so the student says that yes, I know neural network right. So, neural network is of course, a network ah, but this is not the only network that I am talking about here, this is of course, an example of the network that I am talking about here, right.

(Refer Slide Time: 06:57)



What is a Network?

Sounds vague?

A network is a simple (mathematical) abstraction of a complex system.

Networks are a general language for describing complex systems.




Then what is network? So, network is a simple abstraction of a complex system right. What do you mean by that? What is complex system? Right. Let us say, you know let us say our body right, our metabolic system for example, right the way proteins are in proteins interact

you know due to different metabolic process, how neurons interact right, how cells interact and so on and so forth.

This is a complex systems think about say, think about internet right, think about World Wide Web right, it is a gigantic system and this is so complex that you know you may not be able to understand what is happening the at different parts of the network right. So, this network is basically a simple way of abstracting such a complex system right networks are also called as a general language for describing complex systems, right this is you know think of it as a language right.

For example, we use say python or Java to you know as kind of you know way to understand to let the computer understand, you know what we are saying right. Similarly you think of networks as a kind of language to understand the entire complex systems right. If the this definitions sound little vague, do not worry, I will give you a concrete definition of what is a network ok.

(Refer Slide Time: 08:29)




What is a Network?

"Network" is another nomenclature of "Graph"
• aka "Complex Networks"

What is so "complex" about it?

A complex network consists of nodes/vertices and edges/links, exhibiting non-trivial topological features that do not occur in a simple networks such as "lattices" or "random graphs", but often occur in graphs which model real systems.

<http://esebainmoro.org/post/2013-12-21-temporal-networks-whit-and-igraph-updated/>



So, now network is also called as graph, ok. A computer scientist called such structure network whereas, physicist calls a structures graph ok. They are kind of you know same, mathematicians also called the call this structure graph so, but you know networks and graphs are basically you know synonymous.

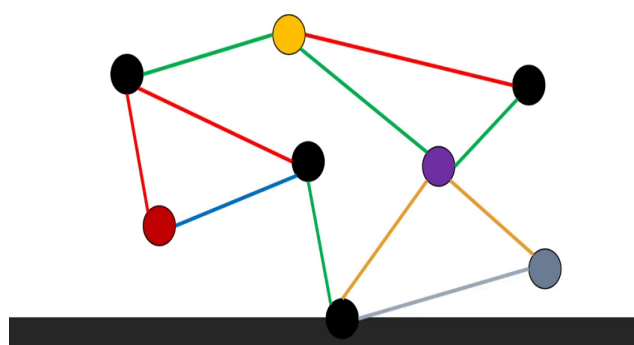
So, but why, so network is the kind of network that we are talking about here they are also called complex network ok. So, why it is called complexity? I mean why it is called complex network? What is the complexity behind the network? Right. So, you know the kind of network that I am talking about, I am going to talk about in this whole lecture series. So, this network size is huge right, you can think of it as you know a billions of nodes trillions of edges right.

This is not as simple as say a star right. Not as simple as a star like this ok or not as simple as a line like this or not as simple as say for example, a grid like this right. We are talking about networks which are of billions of size right, the 1 billions in nodes right, 1 billion edges right. So, the topological property makes this kind of network so nontrivial that we will not be able to understand the structural property of these networks by you know by simply looking at it on you know blind eyes right. This is not as simple as a structure like star or grid right.

So, a complex network consists of say nodes and edges right, these are the two fundamental components of a network or a complex network right. Exhibiting non-trivial topological properties. Now, these terms might sound little vague, do not worry I will explain what do I mean by topological feature, topological structure non-triviality and so on, that do not occur in simple network like a lattice or a random graph. Again random graph that is this terminologies may sound little vague, but I will discuss all these things right.

(Refer Slide Time: 10:56)

Network = Individuals + connections



So, in general when we talk about a network or when we define a network, we generally you know say that there is a set of nodes or vertices right or entities and entities are linked through edges right or relations or links. As you see here in this figure there are this circles of different colors right, these are basically nodes or vertices right or individuals for example and they are connected through different edges or links, connections right.

Now, this is the you know the bookish definition of a network. So, you basically define a network by a tuple of nodes and edges and that is all, right. So, is if you see this network right the nodes are of different colors, edges are of different colors; why so? So by the color what I mean to say is that nodes can be of different types ok, edges can also be of different types.

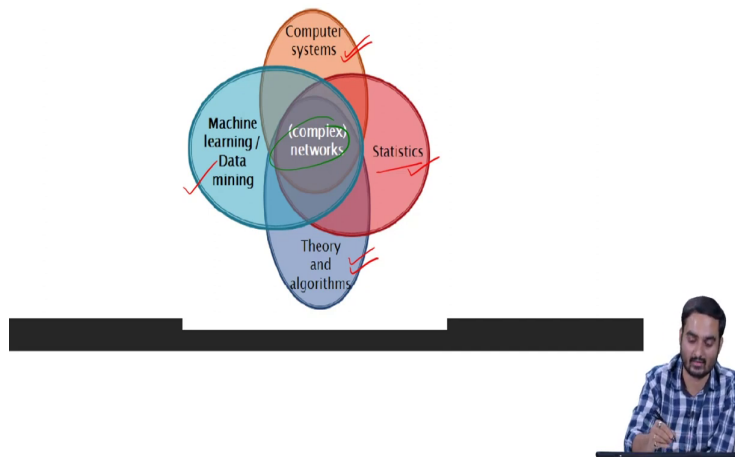
For example, says in social networks, say let us say Twitter right, you can think of nodes as users, you can think of nodes as tweets for example. Links can be a follower following link, one is following another user, it can be a retweet link right, retweet link again it can be, I mean if you can basically connect two users using a retweet link if somebody retweets some others tweet you can connect two users through this retweet link.

In fact, a tweet between a tweet and a user you can also connect right through say something called posting right, if a user has posted a tweet you can connect these two entities right. So, nodes can be of different types, edge edges can also be of different types, nodes can have properties right. For example, users you can characterize a user based on his or her occupation right the location right the gender, the race and so on and so forth. You can also characterize ages in different ways we will talk about all these things in the, in this lecture.

So, but what I wanted to say is here is that you know, when I talk about a network do not you know, do not think that nodes are of same type. Edges can be of same type, it can be; it can be they can be different, edges can be different nodes can be different right.

(Refer Slide Time: 13:24)

Network Science & Computer Science



So, if you look at networks or the complex networks right, the position of this particular subject right, you see that you know it basically falls at the intersection of say machine learning, data mining or statistics and you know some sort of algorithms, computer systems and so on and so forth.

So, throughout this lecture we will see that you know a lot of data mining techniques are being used in a regular basis to understand the properties of a network right. We will use whole bunch of statistics to understand the properties right, we will use algorithms to understand the you know different tools or different applications that people use these days based on this network. And of course, you can design real world systems based on the analysis of the network.

So, all these branches are important to understand complex networks or networks in general right.

(Refer Slide Time: 14:30)

Network: Definition



- Network: An ordered pair $G(V, E)$, where V is a set of nodes/vertices, and E is a set of edges/links.
- Type of the network:
 - ✓ Undirected (also called asymmetric, or irreversible), directed (also called symmetric, or reversible) edges, or hyperedges, etc.
 - ✓ Nodes and links with attributes/features like weights, timestamps, textual features, etc.
- Self-loops: Edges with same node as end nodes.
- Parallel edges: More than one edge joining a pair of nodes.

The diagrams show: 1. Simple graph: A set of 6 nodes connected by 7 undirected edges. 2. Weighted graph: A set of 6 nodes connected by 7 undirected edges, with some edges labeled with numbers (2, 3, 4, 5, 6, 7, 8). 3. Directed graph: A set of 6 nodes connected by 7 directed edges. 4. A graph with loops and parallel edges: A set of 6 nodes connected by 7 edges, including a self-loop and two parallel edges between a pair of nodes.



So, let us now look at the definition. We have already you know discussed that network is defined by an ordered pair $G(V, E)$, where V is a set of nodes or vertices and E is a set of edges or links right. Now, depending on the you know topological structure of the network you can think of two different types of you know networks, the network can be of undirected type.

For example, right there is no direction be a direction of edge right between two nodes. For example, if you see you know say this network you see here nodes are connected, but edge edges do not have any direction right. Example, if you think of say friendship network on Facebook right, there is no direction because if I am a friend of a person, right a person x ; x is also a friend of mine right. This is basically bi-directional right and bi-directional is equivalent to unidirectional edge undirected edge right.

Similarly, a network can be of directed in nature. So, edges can have directions right for example, follower following network, where if I follow a person x it is not necessary that x will also follow me. So, you can see a direction from say from y to x right, but as you see here x is not following y . So, there is no opposite direction like this ok. Again, it all depends upon how you basically model a particular complex system.

So, Twitter in this case is a complex system or Facebook in this case is a complex system right. Even you can think of something called hyperedge, this is even a you know more broader version. I mean a kind of a generic way of defining a network ok we will discuss

what is hyper edge, hyper graph later ok. Nodes and edges can be of you know can have attributes as I mentioned earlier right, edges can also have a weights.

Now, these weights you can think of weights as the strength of an of an edge right. Say for example, if two users are interacting very very frequently right you can think of the edge connecting these two users has you know having higher weights right. So, as you see here in this particular example, all the edges have some sort of weights right weights are positive, but of course, depending upon the application you can think of negative weights as well, ok.

Edges can also edges can also have time stamps right for example, again follow followee, network; say I am following, I have already followed another user x at time stamp t_1 , I am following you know user y at timestamp t_2 . So, this t_1 , t_2 these are basically timestamps associated with edges ok. Edges can also have features right and when you talk about these kind of graphs or networks right, some of you may have learned graph theory the basic graph theory although this is not that needed right.

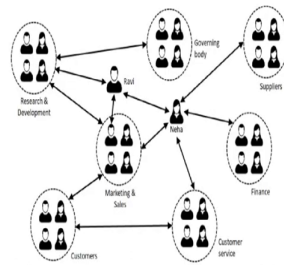
But it is always good to you know brush up some of the skills on graph theory. For example, you know you have heard about something called self loop right; what is self loop? Self loop is basically an edge which starts from say node x and also ends at node x right. So, the starting edge starting vertex and the ending vertex they are all same. As you see here this edge this is basically a self loop right. You can also have parallel edges.

Parallel edges meaning you know say there are two nodes and you know there are two edges or some three edges connecting these two nodes right. And this is possible again let us say a Twitter network right. Say one edge indicating follower followee another is indicating retweet right, whether x has retweeted y , y 's tweet. Another edge say indicating indicates whether you know, this user has liked some of the post of user y and so on and so forth right.

So, but when we talk about network, we generally avoid such graphs where self loop and parallel edges are there right. So, when I say that let us say a network G right, V comma E I will assume that self loop and parallel edges are not available. If needed I will explicitly mention that these two things are available ok.

(Refer Slide Time: 19:09)

What is Social Network Analysis?



Network

An abstract representation of relations among entities

Social Network

A simplified representation of the social structure characterized by actors and ties

Social Network Analysis

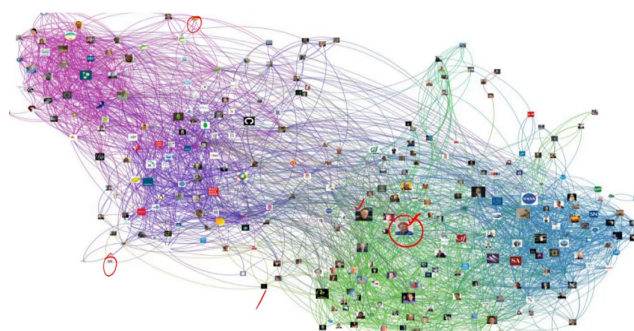
The application of networks and graph theory to analyze the relations present in a society



So, right so we have discussed what is a network, but this course is about social network right. So, what is social network? So, social network. So, network is an abstraction of a complex system, social network is a simplified representation of a social structure right between users, among users.

It can be online social network, offline social network and what is social network analysis. This is basically an application of networks and graph theory to analyze the relations present in a society right. The society can be online society or offline society ok.

(Refer Slide Time: 19:47)



A sample of Twitter follower-followee network


(image source: <http://allthingsgraphed.com/2014/11/02/twitter-friends-network/>)



Now, this is an you know an example of a very very tiny portion of a Twitter network, this does not this is not at all the Twitter network these days right, but this is a portion of Twitter network. As you see here nodes are you know nodes are indicated by you know different users, you see here bill gates and you know different dignitaries, you also see you know normal users like us right.

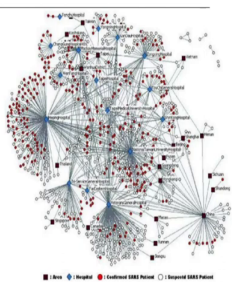
And all these links are present between users right, some if you look at here bill gates right a lot of edges are connected with this node right. Whereas, user like this one right or user like, this one there are three four edges connected with this node right. So, we will understand, we will analyze you know how this kind of network is formed right. How do we characterize nodes based on these edges and so on.

(Refer Slide Time: 20:47)




SNA Applications: Healthcare

- Modeling the spread of infectious diseases
- Contact tracing during epidemic outbreak to identify possible patients
- Identify and isolate super-spreaders
- Planning lockdown schedule
- Identify vulnerable population during vaccination
- Planning vaccination schedule, etc.



2003 SARS contact Network in Taiwan
https://doi.org/10.1007/978-1-4419-5382-0_15



Now, let us look at the application, some of the important applications that we can think of. So, you know the first and foremost application is health care. Now, this is of our priority because of you know the kind of pandemic through which we are going. So, you can actually use social network analysis to understand or to model different things. For example, we can understand how you know how an infectious disease actually spreads you know in our society right.


We can actually model it, I am pretty sure you have heard about many such network, many such models the data-driven models which basically try to mimic the way, say Covid-19 virus

spreads right you may also have heard about Ebola for example, right. And people basically modeled how this kind of virus spread over you know offline network right.

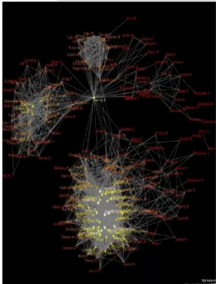
We will also talk about you know say, if we understand how such you know things spread you can also come up with policies. For example, lockdown policy right, we cannot afford you know unlimited lockdown. So, what would be the typical lockdown policy that a government should take to you know to stop the spread of the virus right. In fact, if the vaccine is available how do you efficiently or effectively vaccinate people?

So, that the virus is not able to spread from one part of the network to the other part of the network; one part of the society to the other part right. Whole bunch of things can be possible through social network analysis and we will discuss about this in a separate chapter altogether ok.

(Refer Slide Time: 22:32)




SNA Applications: Social Media



- Identifying friendship structure
- Recommending friends, and groups, or pages to follow
- Identifying information propagation patterns
- Designing Humanitarian response applications after large-scale crises, etc.

An Example Map of a Facebook Friendship Network
<https://mathconceptions.wordpress.com/2012/01/16/application-snippet-friendship-and-influence-in-social-networks/>



Of course, social media is you know an important part where social network analysis is important. For example, you know in Facebook, Twitter kind of setting friendship recommendation, follower, followee you know follower recommendation right, all these things. How information spread say for example, misinformation, fake news, spreads over social network, how can we understand the pattern, how can we come up with techniques to stop the spread and so on and so forth. These things are also will be discussed in the subsequent chapters.

(Refer Slide Time: 23:06)



SNA Applications: E-Commerce

- ❑ Customer profiling to Know the customers
- ❑ Product/Service recommendation based on customer profile
- ❑ Instigating viral marketing by pinpointing influential players
- ❑ "People like you buy", "Frequently bought with this", or "Frequently browsed", "Trending" are common buzzwords



Influential Communities in Social Network
<https://towardsdatascience.com/influential-communities-in-social-network-simplified-fe5050d8e5a4>




E-commerce service, another important application right. We have seen like applications, I mean we have seen E-commerce services like Amazon, Flipkart, right eBay where products are being recommended to users right say for example, if you buy a mobile right, you will be also recommended to buy say a headphone right.

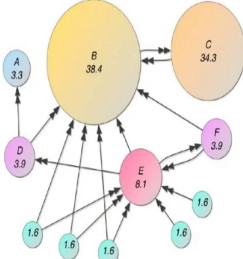
If you buy a laptop, you will be also recommended to buy a mouse for example, right. So, you may wonder why right, you may wonder why these items are being recommended right to me right. You in fact, it is also possible that you suddenly you have wake up and see that some very weird item is you know is recommended to you. So, you feel that, well I mean I have never seen this product before, but why suddenly this is recommended right.

You may be surprised to hear that you know all the users are being tracked right, who are basically browsing E-commerce services. And depending upon their you know buying pattern the way they browse through the e-commerce services right, the clicking patterns everything are being analyzed in a systematic way and based on that you are recommended products. Again, we will discuss this kind of recommendation systems right in general in all the chapters.

(Refer Slide Time: 24:26)





SNA Applications: Web Search Optimization



Toy Example of Google's PageRank
<https://en.wikipedia.org/wiki/PageRank>

- Ranking the webpages based on hyperlinks
- Ordering of displayed pages based on ranks
- Displaying webpages based on user search profile






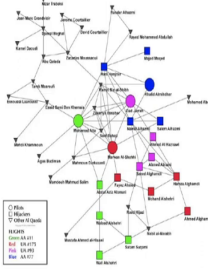
Web search optimization, web in general is a massive network right, I mentioned World Wide Web right. So, there are web pages, web pages are linked through hyperlinks right when you open a web page you see there are you know many hyperlinks, you can basically click on one hyperlink you move to another web page and so on and so forth. So, the entire you know the entire web pages and their links through hyperlink this is basically a network right and how do you know the importance of a web page right.

It all depends on how we browse, how we browse through different pages right how we click on a particular page, how we move from one page to another page we will discuss an algorithm called page rank, which was responsible for you know for a creation of a company like Google right. Way back 1998 1990s 1999, when this page rank algorithm was proposed, was published and company like Google was formed based on the simple idea ok.

(Refer Slide Time: 25:29)




SNA Applications: Cybercrimes & Terrorism



9/11 Terrorist Network
<http://www.orgnet.com/hijackers.html>

- Online fraud
- Fake news propagation
- Cyber bullying/trolling,
- Sharing pornographic materials


- Terrorists often use social media to communicate as well as to brainwash innocent people




And this is very interesting you can also analyze you know how criminals, how terrorists you know interact, how terrorists recruit, how criminals recruit you know new individuals right through social network. In fact, there is a very interesting study which actually said that if we had analyzed social network before, we could have stopped the 9 11 attack right and that was really you know terrible finding right.

So, we can also think of you know detecting fraud users for example, right fake news, cyber bullying all sorts of you know harm online harms through social network analysis right.


(Refer Slide Time: 26:12)



SNA Applications: Scientific Research & Academic Collaboration




Web of Science®
The Concept of Citation



- ❑ Scientific authors cite (refer to) the works of other authors
- ❑ Various networks of scholarly articles may be formed exploiting this relationship
 - Paper-paper citation network
 - Paper-paper co-citation network, etc.
- ❑ Scientific authors collaborate with one another
- ❑ Various scholar networks may be formed using these relationships
 - Author collaboration network
 - Author citation network
 - Author co-citation network

<https://library.bu.edu/citedreferences>

https://scholar.google.com/citations?view_op=view_authors&authors=asci



Social network is also useful for scientific research mining right. Let us say; let us say you have whole bunch of scientific articles or patents right, papers, journals, conference papers right available, how do you mine such data sets? Right. You can also think of a network out of it where you know nodes can be scientific papers, scientific articles and links and these nodes can be connected through citations for example, right.

Citations between papers, so say paper x is citing paper y this is definitely a directed network as you can imagine right. And you know if you know the citation network, you can do you know you can unfold whole bunch of things. For example, you can see how a knowledge basically propagates from one community to another community, how such an interdisciplinary field emerges you know over time how to two fundamental you know communities.

For example, say AI and say you know database right for example, interact and interact through citations and you know and that leads to an emergence of another new field altogether. You can also track you know a scientific field of a particular user, let us say you have a successful researcher, you have not so successful researcher, if you look at that career trajectory of both these researchers you will see a clear trend, you know that how they how two researchers published publish their papers, which communities right what are their citation patterns and so on and so forth.

And based on that you can do whole bunch of things, you know each index imperfect these are of course dependent on citations, but if you have the citation network it will give you a holistic view of the publication data set in general right. So, citation network is something which is very important. And another network you can also create out of this publication data set is something called co-authorship network, where nodes are researchers and two researchers are connected when they wrote a paper together for example, or they filed a patent together right.

So, in this courser ship network this is also called collaboration network, you can also see you know different communities, you can clearly see that you know physics community and you know computer science community, they do not often interact right, but sometimes when they interact they create magic, right for example. So, whole bunch of things can be possible if you model this scientific publication data sets through networks right.

So, in the next lecture, I will try to give you more motivation more applications behind you know behind network science, social network analysis in general. And hopefully you know I will be able to convince you why this is an important you know course altogether.

Thank you.