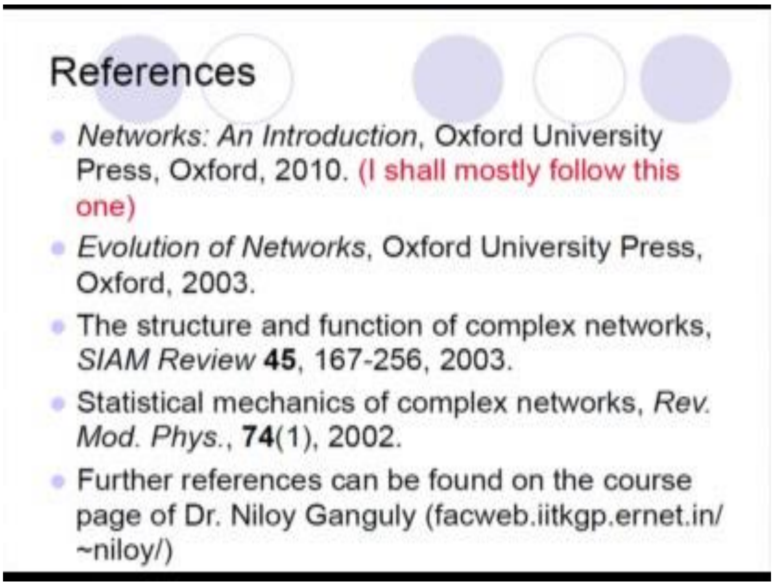
Complex Network: Theory and Application
Prof. Animesh Mukherjee
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 01
Introduction

Welcome to this introductory lecture on Complex Networks. So, today the first thing that I am going to do is to give you idea of the like, resources which I will be mostly following during the course.

(Refer Slide Time: 00:30)



So, the first book that I suggest; Networks: An Introduction, Oxford University is kind of the book the text book for this course. But like whatever we will be doing much of it has been covered in this course, but that is not like everything, we will also have materials from other resources, and as we go forward we will explore. I will let you know which parts you need to refer to other resources.

How many of you are aware of this picture? I guess most of you have seen it at one point in time, this is the famous Gangnam Style video, snap shot of that video and that the point that I am trying to illustrate through this picture is that, there has been remarkable thing that has happened as soon as this video was released. And that remarkable thing is what we show in this slide.

(Refer Slide Time: 01:40)

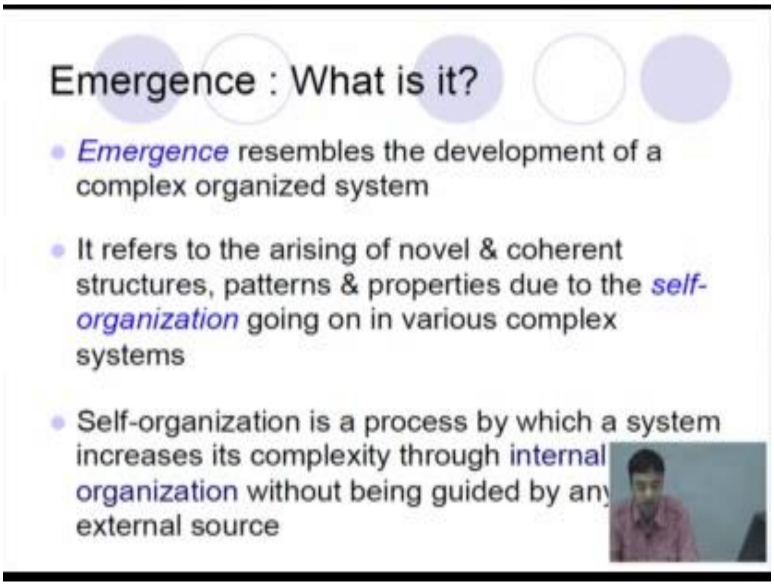
So, what you see is that the YouTube viewership of this video has exponentially grown over times. So, it is now like 800 million views like, 1 billion likes on Facebook 150 million followers in Twitter. So, the question is like what actually made this video go, if you say viral. What was the reason why this video went viral? So, and these are the kind of questions that we will you know try to answer through this course, we will try to make theoretical developments as to why this kind of things happen in a, in a network settings.

So, as we go forward, we will understand what is a network? What is actually meant by a network and things like that, but this is one of the very first motivation to study these kind of a subject like what made, so the question that you can start off with is like what made Gangnam style popular or viral? And the rest of this course is kind of a way to answer this question.

(Refer Slide Time: 02:56)

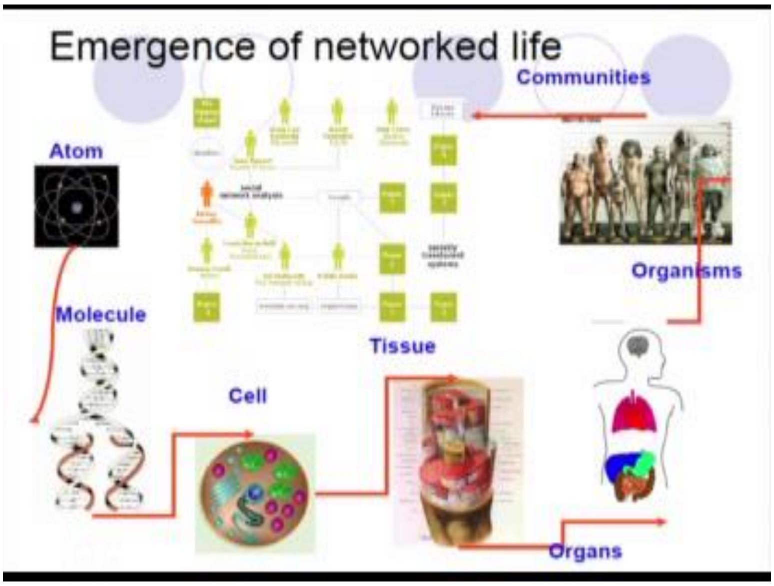
Another example where you see like complexity or complex things is this picture. So, this picture it terms out it looks like a tree trunk, but it is actually not that and it resembles or it is actually the termite “cathedral” I guess sort of your aware of a termite cathedral. So, these termites are work like tiny engineers you know. So, they bring in mud particles from different areas and they accumulate it in a in a place and like they do it for quite some time and suddenly a mound of this structure gets found. So, this is striking, this is another striking example of a complex system.

(Refer Slide Time: 03:44).



So, this brings us to the very basic idea of what a complex system is, and what it means bites emergence. So, the emergence resembles the development or the growth of a complex system, but this growth is not externally controlled. There is no external coordination to this growth. It is mostly self organize that is the particles organize themselves; they interact among themselves and organize themselves to result into complex structure as you have seen like termite cathedral.

(Refer Slide Time: 04:25)

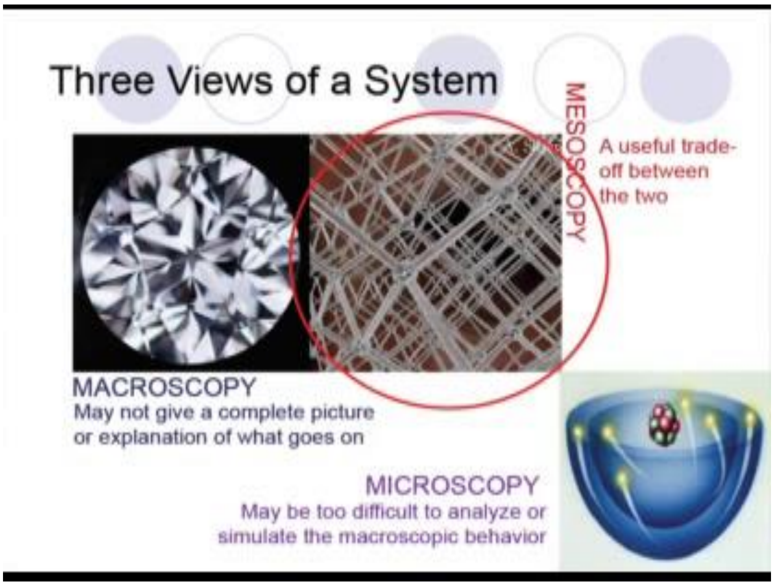


Example: Complexity actually you can find complexity examples of complexity everywhere starting from our life, say for instance I have given I have illustrated one example here in this slide where, you see atoms those atoms combined to form molecules and these are like molecules in your DNA chains.

If you come across them, these DNA chains actually contribute in the formation of cells; these cells then have complex interactions to form what you call tissues. Tissues interact with each other to result in to organs; organs interact with each other to form what you say organisms, the way we are. And these organisms then interact with each other to form what you call what you see today in Facebook's and Twitter social networks.

So, the point that I am trying to make here is that complexity is all progressive. So, you find complexity in every segment of development and complex networks the course itself tries to provide us with a tool to understand the emergence or the organizing principles of this complexity.

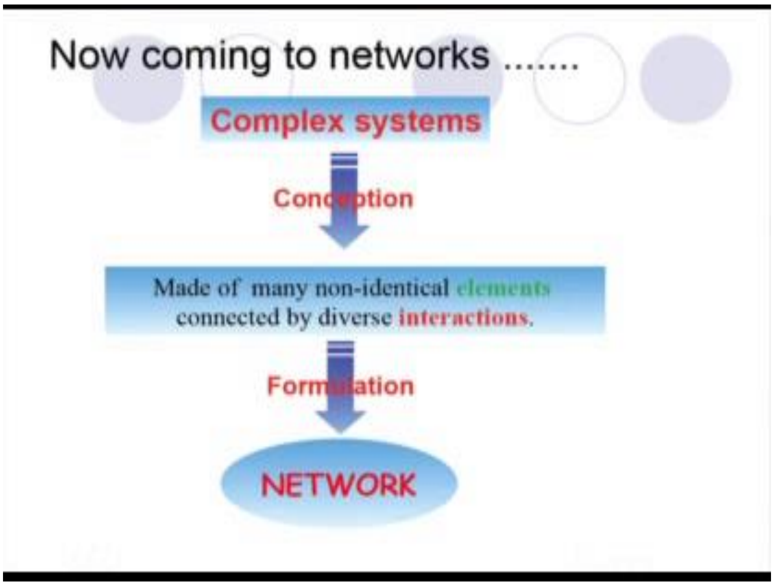
(Refer Slide Time: 05:40)



So, physicists actually view complexity from three different in a. So, they divide a complex system and they look at a complex system from three different perspectives. So, one perspective is the complex system as a whole, it is called the macroscopic level of view. So, for instance in this slide I have given an example of a diamond. So, there is the macroscopic view of the system on the other extreme is the microscopic view where you have all this tiny atoms and interactions between these atoms to give shape to that diamond that you see there at the Microscopic leave. And some here in between is another level which is the Mesoscopic level.

So, the Mesoscopic level as you see here which I circle out in red, in this slide is a Mesoscopic level is where you see this all this carbon bonds which actually result into this hard diamond material that you see at the microscopic level. So, this is actually study complex system from all these 3 different perspectives because, all these 3 different perspectives have to tell you interesting stories. So, all these 3 different perspectives have their own story and if you untangle them you get to understand the organizing principles far better.

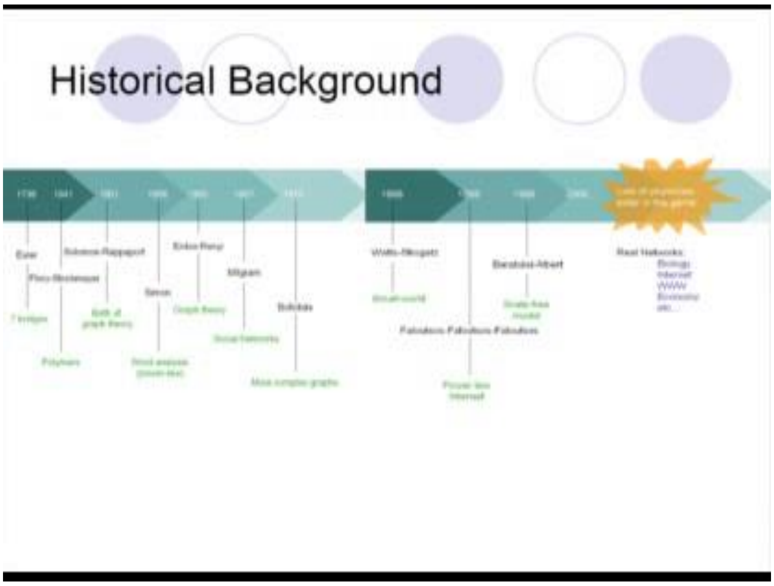
(Refer Slide Time: 07:11)



Now coming to networks, all these time we have been only talking about complex systems. So, the question then is; like where you fit in network or complex networks in this whole gammet of complex systems. So, the idea is very simple, complex systems can be conceptualized as a network. So, network is a formal structure or an abstraction which has two basic entities; nodes and interactions among nodes.

So, each unit in the complex system or each entity in the complex system can be thought of as a node and interactions between entities between heterogeneous as well as homogeneous entities inside the complex systems, inside a complex system can be thought of as edges. Edges of a network, so you have two entities in a network the nodes and edges the nodes model the entities of a complex system and the edges model the interactions between the entities. So, that is how you formulate a network or a complex network from complex system.

(Refer Slide Time: 08:28)



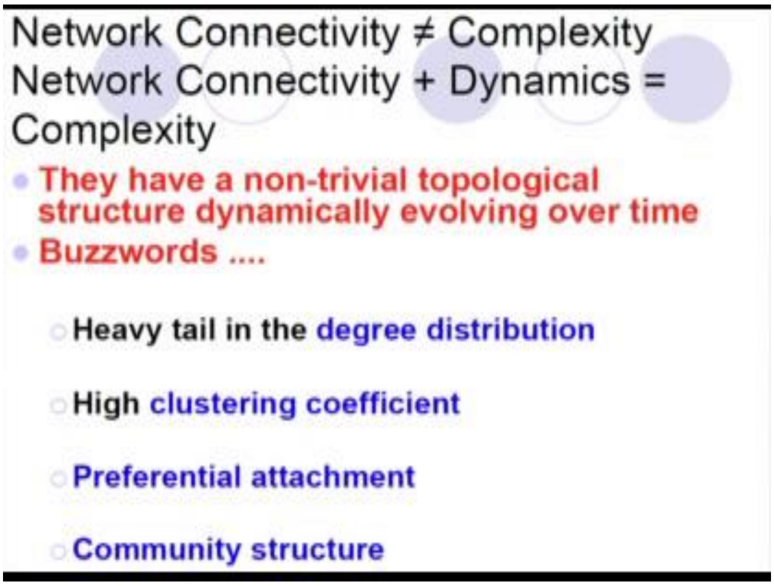
So, complex system studies have actually. So, this idea of net formalizing complex systems as a network or a graph, this saw it is boom only lately, but people have been studying formalizations or formalisms of graphs for very long time. So, many of your probably aware of the Euler's 7 way bridges experiment that he started in 1736 then there was, there were people from chemistry, there were people from physics, there where people from mathematics, there were people from sociology, who were using graph structures or network structures to study various interesting problems.

Now, suddenly somewhere around 1998 to 2000 if you look at the time line, suddenly somewhere around 1998 to 2000 lot of complex systems. Lot of statistical physicists actually started taking interest in modeling complex systems as networks or as complex networks. And these saw a very big boom after 2000, when there was this World Wide Web and the internet that came in to big. So, the World Wide Web can also be thought of as a network, where each node in this network is page is a web page and 2 nodes are connected by a hyperlink that is one web page, hyper link in another web page.

So, that is how a World Wide Web graph is conceptualized. Similarly you can think of internet as a complex network. Where each node is a computer, and the interaction between two computers is through a communication line and there are there could be important computers in this network like routers, which we will later on know that these, there will be special names given to this nodes like hubs etcetera. So, as we proceed with the course we will get to know more about these.

But, the basic idea here that I am trying to illustrate is that, it was during this time from 2000 onward that there was this big boom in modeling complex systems like the World Wide Web or the internet into networks.

(Refer Slide Time: 11:02)



So, now, the question is that, So, why we are so interested in these kind of complex networks or modeling complex systems as complex networks. There are two important things. So, what is so complex about this complex networks? If you try to understand this, we have to question that why are we doing this formulation? So, the primary reason that we are doing this formulation is actually 2 folds; one is we want to understand these structural arrangement that has gone in to the network. So, we want to have precise quantitative metrics that will illustrate the structural arrangement of the network. And as a result we will also under we will also get an data feeling of this structural arrangement of the complex system from which the network has been modeled.

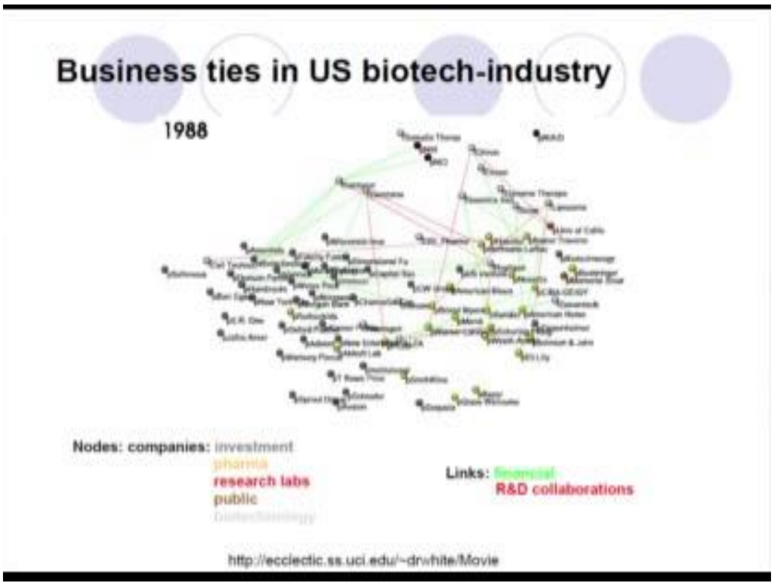
Similarly the other reason that we try to model complex systems has network is the dynamicity. That is these systems are not static, they change over time for as you can imagine the World Wide graph. The World Wide Web graph or the internet graph it is changing every day.

Every day new web page is are coming in and new hyperlinks are built. So, it is changing every day. So, new nodes are entering in to the system new edges are formed and these actually makes the system dynamic and it is easy to model these dynamicity in the form of a network where the dynamicity of particular page entering can be modeled as node entering into the network, and the dynamicity of new edges getting included in the network and be modeled as new hyperlinks being built.

So, now the question is that, how can one endeavor, or how can one try to collect structural property from these large networks try to make some interesting inferences from this large network, that is one question. And the second question is how one can understand the dynamicity; if there are ways to do so. So for these people statistical physicists and computer scientists, together have try to develop bunch of metrics and some of the buzzwords that we will come across through the course are like; heavy tail degree distribution, high clustering coefficient, preferential attachment, community structure. So, these are some of the buzzwords that you will quiet often come across in the rest of these course.

So, as we move along the course, I will define each of this independently and with appropriate examples wherever necessary.

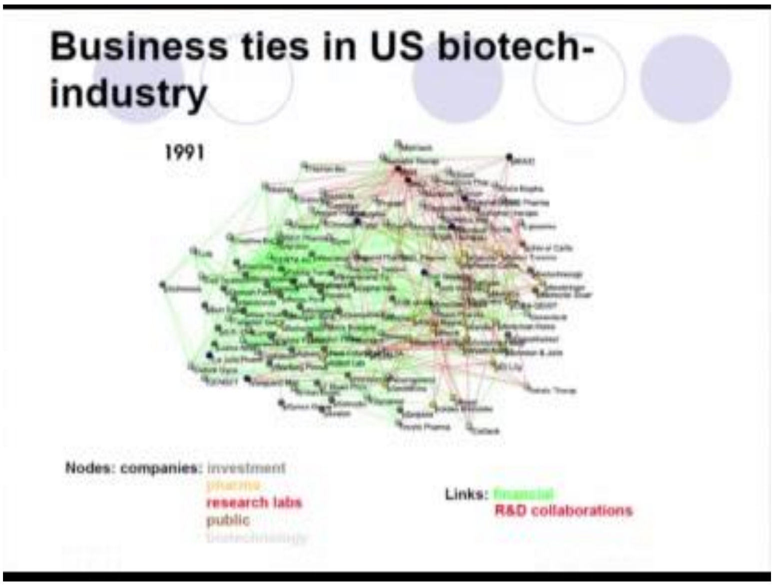
(Refer Slide Time: 14:00)



So, but before we go into all those quantification’s, it is actually good you know to motivate ourselves that really complex networks or complex network structure can give you good indications of the structural properties and the dynamicity of a particular complex system. So, here in this slide I actually give an example of the business ties that were found in the u s biotech industry in the year 1988. So, these are like business relation that the nodes are the different business organizations, and the edges are the different business relations like; it could be financial relations, it could be research and development collaborations.

So, the red links indicate research and development collaborations, the green links indicate financial collaborations and nodes could be like different companies like, investment companies, pharmaceuticals research labs public sector companies, biotechnology labs etcetera. So, as you see. This gives you an, immediately this picture tells you something about the structural organization of this complex system. And this becomes more apparent as we proceed in the next slide you see the previous slide that I showed you was a snap shot from 1988.

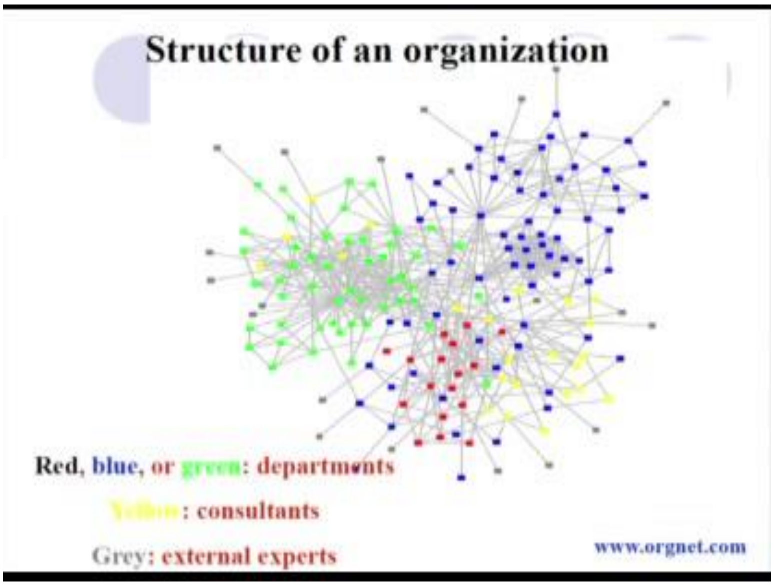
(Refer Slide Time: 15:24)



Now, in the next slide which I show you the say snap shot of the same set up organizations, but taken in 1991. You immediately see differences. So, you first difference that one can immediately identity is that, the number of edges in the network has increased and it has increased manifolds. Also apart from that you clearly see that there are 2 partitions of edges; one is a set of nodes among whom you have mostly R and D collaborations, and another is a set of nodes among which you mostly have financial collaborations. So, you see these two important developments already happening from the 1998 to 1991. 1988 to the 1991 snap shot.

So, that actually tells you like how complex system can be nicely modeled in the frame work a network and interesting topological properties and the dynamics can both be ultimate from such constructions.

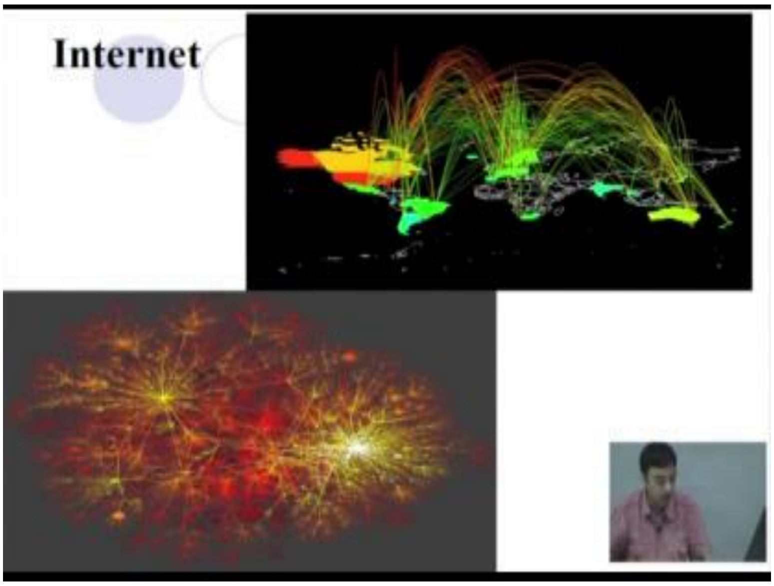
(Refer Slide Time: 16:41)



To motivate further, further I give you some other examples, like there is this other example of a structure of organizations. So, all these red, blue, green are the different departments of the organization. Yellow are the consultants to the organization and grey are some external experts and you see edges running between each of these nodes, each of these entities in this organization. So, again here you can see that the blue nodes organize themselves in one set, the green nodes organized in themselves and another set.

So, in a sense there is a modular structural out there. So, that is already becoming very obvious from this visualization. So, there are certain such interesting topological structures that one can really identify from such networks. So, as I was telling you that the internet can also be conceived as a graph.

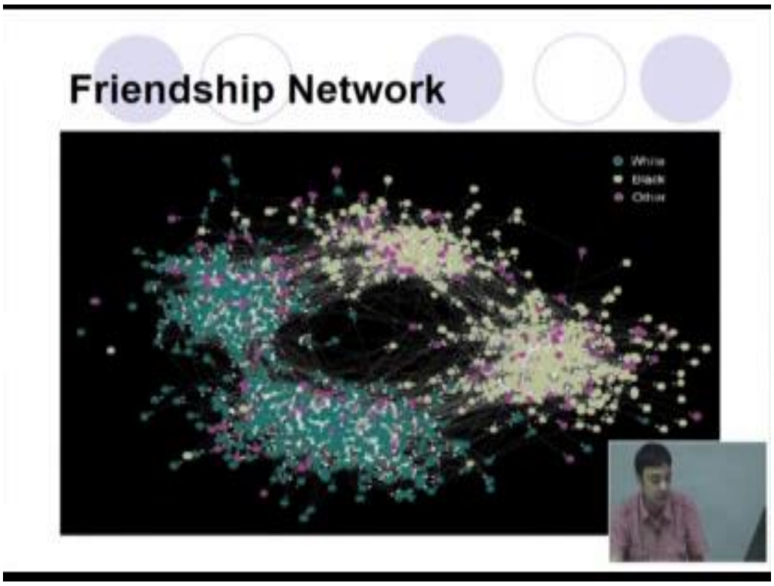
(Refer Slide Time: 17:33)



And here I give you example of such a complex graph or a complex network. So, this was taken like I think in 2000 or something and believe me this network has grown much, much more complex now. So, you can already see that there is a lot of complexity already in this network, but these has become much; this is grown much, much more complex now at this point. So, apart from the complex structure there is one more interesting thing that I was already pointing out in one of the previous slides.

So, you see that the network immediately shows you some star like structures. So, that is some hot spots. So, these are kind of the most important routers or gate ways through which maximum communication takes place. So, that is why they are like star that they are appear as a star structure and in complex networks languages these are mostly called hub nodes. So, the idea is that visualization, actually good visualization of the network already tells you some interesting signatures of the structural properties of a network.

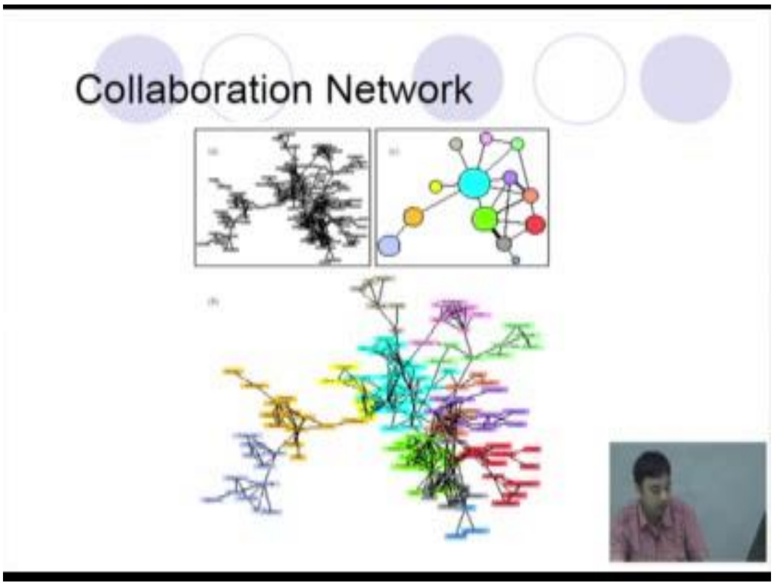
(Refer Slide Time: 18:56)



The next example that I am showing you is of a friendship network that was like collected from a US school where you have 3 different types of entities. The white students, the black students and other students, who could not be classified into any of these groups, these are the blue nodes are the white students, the white nodes are the black students and the purple nodes are the other students and edges in this network represent friendship relationship.

So, as you can immediately see that this network there is a bi partition happening. So, blacks tend to be more friends with blacks with other blacks. And whites tend to be more friends with other whites. Whereas interactions or friendships between blacks and whites in between blacks and whites is much less that actually becomes pretty apparent from this networks. So, again there is this idea of modular organization, where you have basically two distinct groups, one with the black entities and the other with the white entities, and the others the other group is well spread.

(Refer Slide Time: 20:21).



So, another example, another very interesting example is that of the collaboration network. So, this is one of the examples that I like very much. So in this scientific world actually what happens is that; when a scientist work on a particular problem this work needs to be accredited by some organization, and that happens in the form of something called a publication.

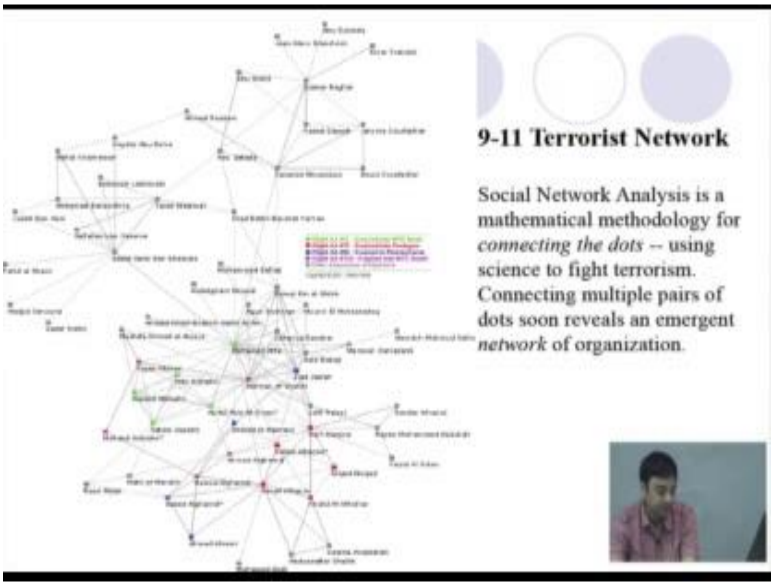
Scientist actually tends to get tend to get their work published in order to get these works recognized by other scientist. So, and while in in the process of doing. So, while paper is written or a publication is being made the paper has to refer to a bunch of other papers. So, this phenomenon is called the phenomena of citations. So, as we go deeper in to the lectures we will discuss in more details the construction of citation net networks, but just give you brief idea.

So the basic idea is that one paper at a particular point in time can refer to older papers in time from which, some of the ideas of the current paper has been borrowed. So, each paper in this network is a node and there could be citation links indicating that particular paper has sighted or referred in it is reference list some other papers from older times. So, this is one form of the network, and another form of the network in this scientific world is the collaboration network, where we say that each node in this network he is a scientist and two scientists are connected by an edge if they have collaborated or co-authored on a paper on a single paper.

So, there is an edge between 2 scientists if they have collaborated in a particular paper. So, and if you look at the visualization of such a collaboration network you again see the same kind of modular structures that I was talking of. So, there is this orange group, the blue group, the green group each of this is a set of scientist working probably on a similar type of a problem or a similar topic or a similar area of research.

For instance in computer science you can think of like the rate people may be working in artificial intelligence while the yellow people might be working in data mining, the orange people might be working in something more hot like machine learning. And you immediately see such groups actually appearing in the network.

(Refer Slide Time: 23:21)



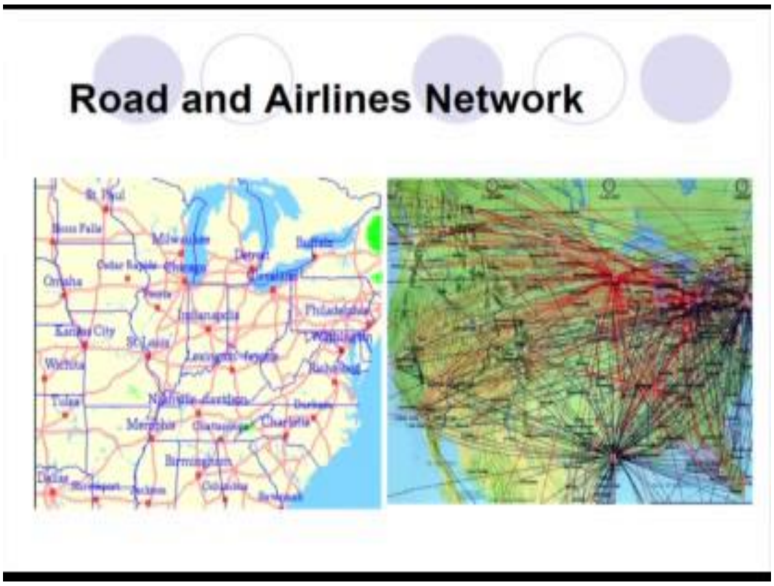
So, this example that I am giving you now is kind of one of the very striking examples because, if we if such kind of studies or such kind of analysis where made before at a terrorist attack then, the attack could have been by passed. So, this is actually the social network of the 9-11 terrorist attack. So, actually each node in this network is actually corresponds to a terrorist and then edge corresponds to a telephone call that the terrorist made to some other terrorist. So, this this is like the trap of thus telephone calls. Now if you look at this network you immediately see some interesting things.

So, there are actually as I was talking in telling you way in the case of internet there was certain hot spots or hubs, in this particular example also you see certain hot spots or hubs. If this kind of a network structure was studied before the attack, and if you could somehow dis-function the nodes which act has hubs in this network you could have stop the 9-11 attack; so actually, this analysis tells you, how practical or how practically important study of networks could be in the context of modeling such social or complex systems.

(Refer Slide Time: 24:55)

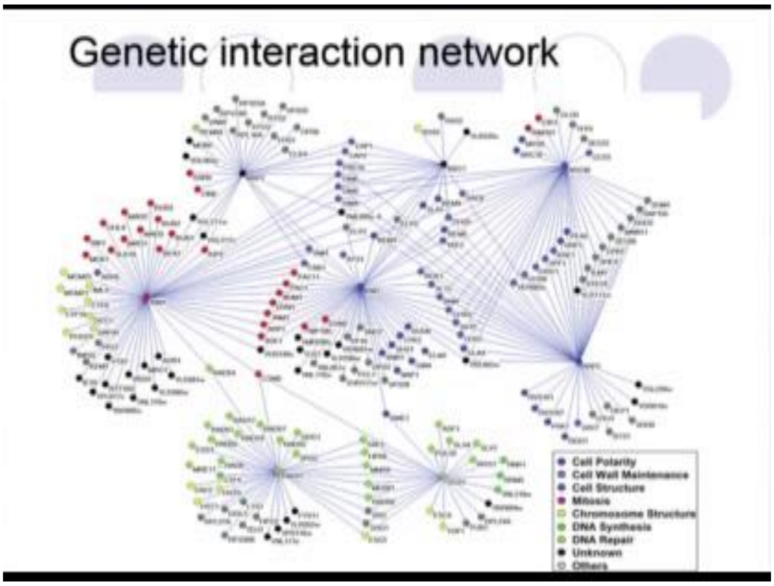
Another example that I would like to point out is about this Swedish sex web. So, we will talk more about this example in the next lecture, but just to give you brief idea. So, here the network is composed of nodes, which are like individual’s males and females. And edges indicate whether a sexual part machine has taken place between pair of nodes and there were around 4781 swedes who, actually participated in this assignment and within the age of 18 to 74 and the response rate was like roughly 59 percent. Like, we will see that, this analysis this network construction and the posterior analysis give certain, very, very interesting indications about the topological structure of the network.

(Refer Slide Time: 25:57)



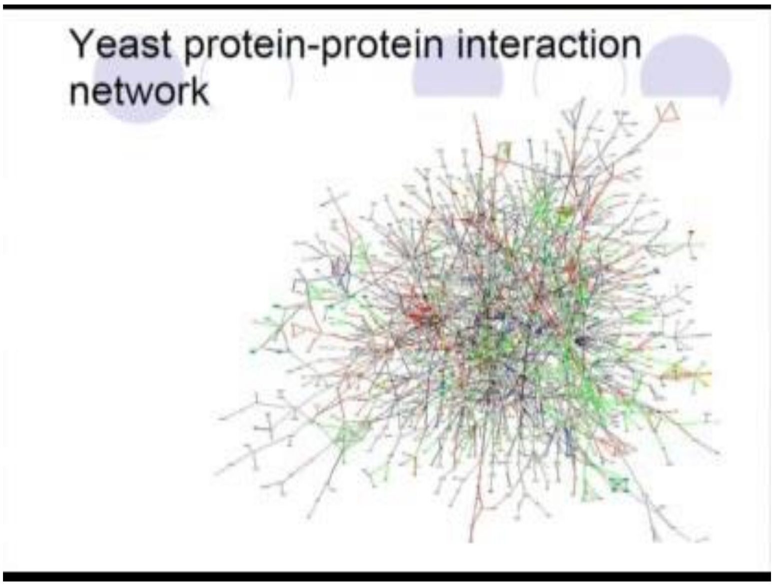
So, similarly you can have road and airlines network. So, air in the airlines network each node could be an airport and then the flights going from one airport to another airport might establish links between the two airports. In this way you can and so here again you can have concepts of hot spots and so, some when some hot spots this function what would be the overall effect of on the network these types of things could be actually studied very nicely in the frame work of this formulation.

(Refer Slide Time: 26:32)



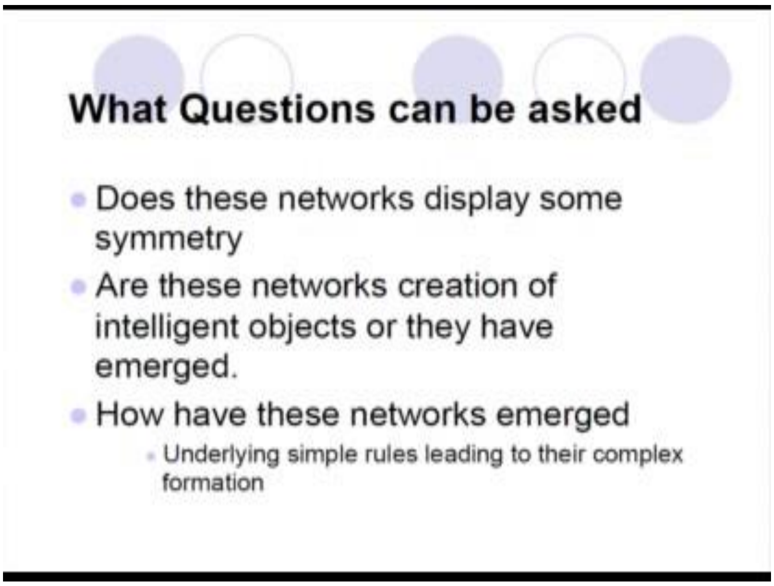
Similarly, in the biological world you have genetic interaction network, these example is also very interesting. Where you see that, this is like the genetic interactions that take place between different genes and what you see here is that, there are a bunch of stars and two important stars are connected among themselves. So, this is this is again very interesting structure. Actually this looks very much like the internet structure the technological networks.

(Refer Slide Time: 27:06)



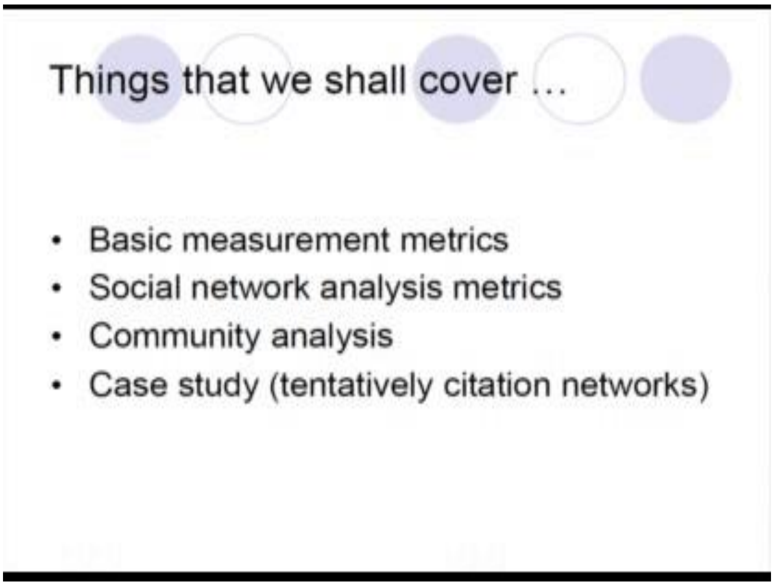
So, similarly there is this yeast protein inter protein, protein interaction network the here again you see modular structures like there are parts which are green, more green there are parts which are more red, there are parts which are more yellow. So, you will look into more details of how to identify these modular structures in an algorithmic fashion later in the course.

(Refer Slide Time: 27:28)



So, the questions that we would like to ask in the rest of these this course is, does this networks display some symmetry. Are these networks creation of some intelligent objects or have they emerged can there be some very simple principles, can there be some very, very simple ways to understand it the some stochastic guide lines which tells you how these networks, how these complex structure have emerged.

(Refer Slide Time: 27:55)



So, some of the things that we will try to cover, that we will definitely try to cover in the rest of the course are some basic measurement metrics, social network analysis metrics, community analysis and case study and citation networks. So, in the basic measurements we will try to look in to degree distribution, clustering efficient, under the way basic properties which people usually study to understand the topological structure of the network. In the social network analysis we will basically concentrate on understanding network roles, like what roles people play in a network, how these roles are related to another and things like that.

In the community analysis section, we will mostly study how to identify modular structures from a network. We will try to under. So, this is a combinatorial optimization problem and it is usually known to be a heart problem, but we will try to come up with heuristic solutions and show like in certain cases how well these heuristics perform. And in the last part of the course we will try to look in to one specific example of citation networks which is really growing network.

So, it changes actually over time, new nodes join in new links come in, new links are built in and we will see certain interesting properties of this citation networks and what these properties actually tell you about an entire field like computer science and it is sub fields. So, all these interesting structures, gets revealed through the analysis through an appropriate quantitative analysis of citation networks that is what we will see in the last part of this course.

Thank you very much.