SIX DEGREES OF SEPERATION: MEET YOUR FAVOURITES 02

Hello guys, welcome to programming screen cast of six degrees of separation so before going to the actual programming screen cast let us see some few pre requisites that is required. The first pre requisite that is required is you must know what is the graph data structure? By graph i don't mean this kind of a graph which you would have been doing in your school physics, chemistry something like that, you take some quantity on your x axis and you take some quantity on your y axis and you observe as you vary this quantity what is the variation in the other quantity, so such thing is what you call as a graph in till your school days, this is not what we are going to say as a graph, we are going to say something like this a set of few points and lines as graph so basically graph is nothing but group of some points and lines that lie between the points that is what we call as a graph and technically we call the points as nodes and lines as edges this is the technical terminology so to illustrate better let me take an example, let us consider the railway network of the city of the country railway network of the country so the nodes or the points let us get used to the terminologies nodes and edges, the nodes in this network is nothing but the cities or the stations and the edges are laid if there is a direct train between the two cities at least one direct train must be there only then we will draw an edge this is how we will construct this railway network. So between any two cities if you take there may be an edge or they may not be an edge depends on there is a direct train or not so for example see two cities Chennai and bengaluru there is an edge between these two cities because there is some direct train between these two cities where as if you see these two cities Chennai and Guwahati or guwahati and Bengaluru there is no edge because there were no direct trains and see this graph is actually dynamic in real scenario because trains maybe introduced every now and then as well as few trains maybe not operated because of various reasons so this graph keeps on changing that i what i been made and dynamic edges maybe newly added or deleted anything may happen so this is just earn example that there existed a direct train then you will put an edge if there is no direct train you don't put an edge that is how it is and if there is an edge you can say that i can reach this node to this node that is from Chennai i can reach bengaluru and from bengaluru i can reach Chennai that is what is what is what is captured if there is an edge but it doesn't mean that if there is no edge we cannot reach this at all so let us see, see there is no edge between Chennai and guwahati but still you can have another city in between Kolkata and you can reach guwahati through Kolkata that is you take a train from Chennai to Kolkata and from Kolkata you can take a train to guwahati this is what you call as a path so path is basically the way which you can reach from one node to another that is what we technically call as a path so there is no edge between Chennai and guwahati but there exist a path so it is not the case that if there is no edge there is no path there may be a path even if there is no direct rain see from Chennai you cannot directly reach guwahati but via Kolkata you can reach guwahati that is the meaning and it is not all the case that there will be just one path there may be multiple paths well as you could see here so if there is a edge you can say that you can definitely reach from this city to this city and if there is no edge that is not the case that there is no way to reach there may be a way to reach even if there is no edge between the two cities that is you

have some cities as an intermediate destination intermediate point through which you can reach your decided destination that is from Chennai to Bangalore there is a direct edge you can directly reach from Chennai to guwahati there is no edge but still there is a path via Kolkata to reach guwahati from Chennai as well as from guwahati to Chennai as well the same path can be followed and as you would see most of the railway network is by direct functional that is you have trains from both the directions, if there is a train from Chennai to bengaluru there is obviously a train from bengaluru to Chennai that's how mostly it works that's why it is undirected graph which we call it as undirected that is directions doesn't matter from Chennai to bengaluru or bengaluru to Chennai both path exist so this is what we call as an undirected graph and we would always be interested in finding the shortest path there may be multiple paths that may be present let me show you, see this graph there may be multiple paths that may be present from Chennai to guwahati see one path is Chennai to Kolkata, Kolkata to guwahati other is Chennai to bengaluru, bengaluru to Kolkata, Kolkata to guwahati there are two paths in this smaller example where may be multiple paths may be present in that case you have multiple options we always go with the shortest path by shortest i mean minimal number of switches of trains in this example there maybe some cases where we have something called weighted graph that is the edges will be having weights so basically you can think of it has a distance between two cities or the time the train takes to travel you can consider anything as your weight so that would be captured here as weight so if for example Chennai to bengaluru let us say the distance is four hundred kilometres this edge would have a label four hundred so something like that the distances would be here as a weight and we would like to find the one the path where the total weights that is the sum of individual weights is minimal that is if you some of all the weights Chennai to bengaluru, bengaluru to Kolkata, Kolkata to guwahati as well as Chennai to Kolkata, Kolkata to guwahati you will find the sum and wherever the answer is minimal you will take that path that is what you call as a shortest path in a weighted graph but here we don't consider weights i just say the number of trains i need to switch that in a undirected case that is what i am considering here in this case the shortest path is from Chennai to Kolkata, Kolkata to guwahati just i need to take two trains but if here i take this path then i have to take three trains Chennai to bengaluru, bengaluru to Kolkata, Kolkata to guwahati so i will have to take three trains so this is not a shorter path in this scenario the shortest path will be generally preferred and whether there will be a path always between any two pair of nodes that is a next question that is of our interest and in some cases there may exist a path in any pair of nodes you take in some cases it may not exist. When in case if a graph is such that between any two nodes you take in the graph there always exist at least one path then you call it as a connected graph see if at least one path exists between any two nodes then you call that graph as connected graph. So out railway network is almost the connected graph because from any city in India you can reach to any other city although you may not have direct trains you definitely have a path that is mostly as much as i know the capital cities are mostly well connected so you can reach from any place to the capital city of the state in from that city you can reach to any city of your of your interest so i fell the railway network of India is connected that's a connected graph so this is now just one place where graphs are used, graphs can be used in many places even people may involved in the graphs that is one such graph were people are involved is something which we all regularly use i guess most of all

used like daily routine facebook so facebook network is a network which involves people so as i said the nodes are people an edge is put if the two people are friends with each other in that case you will put the edge so this is how the facebook network is constructed this is how the facebook analyses the data, the facebook would store the details of your friendships and analyse it will give you recommendations right, you may know this person, you may want to send a friend request to this person such recommendations all these analysis are done for that data has to be stored so the for storing this kind of data oh who is friends with whom facebook uses the graph data structure that i sit denoted people by nodes and edges are put if the people are friends with each other and you can download the facebook data set from this particular link i had already downloaded so i will not be downloading again so the thing that you have to download is there is a text file called facebook combined so basically they have a different versions i guess the discretion is also available in the link you can read through it the thing is they had collected data at various time intervals the various geographical location and they had combine everything into a single text file that is what we have it as facebook combined dot txt just download that particular thing and one more thing that i want to tell is graph is something that i had shown you if you would see this is a pictorial representation of a graph we have nodes and edges this diagrammatical representation is easier on our minds but how do we store this in a computer? We cannot feed in this figure and ask the computer to understand this, we need to store it in some specialised format, there are some formats by which you can store the graph data so there are many formats actually but i will be discussing just one format that would be used in our facebook data that is how we get our facebook data and i am just going to discuss that format. Let us see, see the format is called edge list representation so just consider a sample graph there are nodes one two three four and this is how the edges are present there is an edge between one to two, there is an edge between two and three, there is an edge between three and four. So you just i had said one two two three three four one space two, two space three, three space four this is how you store it so this is what you call as an edge list. Edge list is one of the representation for graph, there are many other representations you can look it up as well as what are the other representations there are some advantages and disadvantages associated with each representations you can take a look at it as an advanced material in graphs.