Lecture 23 Connected Components

The lecture is about the connectivity of graphs and how we can make sense of graphs without relying on datasets. The example of a classroom of 100 people and friendships between them to explains the intuition behind the connectivity of graphs. Can there be a situation where a graph is not connected, where there are groups of people that have no connections with each other. An example of shooting randomly within a rectangle and say that it is very unlikely that all bullets will fall outside a quadrant of the rectangle. The same intuition can be applied to the example of the classroom to show that it is unlikely for a graph to be disconnected, and all the connections to be within groups. We can use combinatorics to calculate the total number of possible friendships and the number of across friendships between two groups of people in the classroom example. It is very unlikely for all the friendships to be within groups and none of them to be across groups.

Lecture 24

Theorem: Let G be a graph on n vertices such that the probability that any two vertices are adjacent is p, where $p = c/n$ for some constant $c > 0$. Then, if $c < 1$, G almost surely has no connected components of size greater than $O(\log n)$, whereas if $c > 1$, G almost surely has a unique giant component containing a positive fraction of the vertices.

This theorem essentially tells us that if the probability of adding an edge between any two vertices is small (i.e., c is small), then the graph will almost surely not have any large connected components. On the other hand, if the probability is large (i.e., c is large), then there will almost surely be a giant connected component containing a positive fraction of the vertices. Now, let's relate this theorem to the question we are trying to answer. We start with n isolated vertices and keep adding edges uniformly at random until the graph becomes connected. Let p be the probability of adding an edge between any two vertices, so $p = x/nc2$, where x is the number of edges we have added so far. Note that p increases as we add more edges.

Now, suppose we have added x edges and we want to know the probability that the graph is still disconnected. By the theorem, the size of the largest connected component in the graph is $O(\log n)$ if $p < 1/n$, and it is a positive fraction of n if $p > 1/n$. Therefore, if the graph is still disconnected after adding x edges, then the size of the largest connected component must be smaller than $O(\log n)$ if $x/nC2 < 1/n$, or it must be larger than $O(\log n)$ if $x/nC2 > 1/n$. Simplifying the inequalities, we get $x < (n-1)/2$ if we want the largest connected component to be smaller than $O(\log n)$, and $x > (n-1)\log n/2$ if we want the largest connected component to be larger than $O(\log n)$. Therefore, we can conclude that the graph becomes connected after adding roughly $(n-1)/2$ edges, or after adding roughly $(n-1)\log n/2$ edges with high probability.

To summarize, the answer to the question of when a graph becomes connected as we add edges uniformly at random is roughly $(n-1)/2$ edges with high probability. This means that if we add $(n-1)/2$ edges, the graph will almost surely be connected, and if we add more than $(n-1)\log n/2$ edges, the graph will almost surely have a giant connected component containing a positive fraction of the vertices.