

# **MS&E 235, Internet Commerce**

**Stanford University, Winter 2007-08**

**Instructor: Prof. Ashish Goel, Notes Scribed by Shrikrishna Shrin**

## **Lecture 11: Small World**

The small world conjecture is that we are all connected to each other through a small path of acquaintances. The small world phenomenon is the reason viral marketing campaigns are so successful and social networks are popular.

Viral marketing is marketing in the process of using the product. An example of viral marketing is when you sell a t-shirt with your logo printed on the back to spread awareness about your product. Hotmail was one of the first products to become popular virally. Hotmail added a snippet with an invitation to try Hotmail every time you sent a message. YouTube became viral by allowing people to share videos with others by embedding the video in their websites or blogs. YouTube immediately caught on with the MySpace community and became immensely popular thereafter.

Stanley Milgrom found through an experiment (popularly known as the six degrees of separation) that the small world phenomenon actually exists. He setup an experiment where he gave letters to a few people (source human beings) whose job it was to make sure the letters reached a particular person (destination human being). The source human being only knew the name, city and occupation of the destination human being. It is easy to see how the letter can be sent from source to destination in relatively few steps through high degree nodes. For example, it could be possible to send the letter to the city council of the source city, the city council forwards the letter to the senator, the senator sends the letter to the senator of the destination city, the senator then sends the letter to the city council of the destination city and the city council finds the destination human being and delivers the letter. Senators, city council members etc. typically have a large number of connections and are therefore referred to as high degree nodes. An interesting result of Milgrom's experiment was that the letters often reached the destination in relatively few steps through low degree nodes.

## **Network Models**

### **Line Network**

In the line network model, nodes are connected in a serial fashion. In such a network, the small world phenomenon does not exist as the nodes become further and further apart as the network grows longer.

## Bernoulli Random Graph

In a Bernoulli random graph, there are  $N$  nodes such that any two nodes are connected to each other with probability  $p$ . A Bernoulli random graph is also represented as  $G(N, p)$ .

Since, there are  $N$  nodes and they are connected to each other with probability  $p$ , every node on average is connected to  $Np$  other nodes. The average degree of the graph is therefore  $Np$ .

Theorem: If  $Np > \log_e N$ , then the network is connected with high probability.

Let us assume that there are 1 billion nodes. Then, the network is connected with high probability if the average degree of a node is greater than 20.7. This means that, if each node on average had 21 connections, then there is a high chance that the entire network of 1 billion nodes is connected. Notice that although  $Np$  is a fixed number,  $N$  can be very large when  $p$  is very small.

Intuition: Consider  $N$  nodes ( $A_1, A_2, \dots, A_N$ ) and another node  $X$ . Now, the probability that  $X$  is connected to  $A_1$  is  $p$ . The probability that  $X$  is not connected to node  $A_1$  is  $1 - p$ . The probability that node  $X$  is not connected to any of the  $N$  nodes is  $(1 - p)^N$ . Now, in order for  $N$  nodes to be connected, we have the expected number of isolated nodes should be less than 1. Therefore we have,  $N(1 - p)^N < 1$ . Taking log on both sides we get,  $N \log(1 - p) < -\log(N)$ . Since  $p$  is small,  $\log(1 - p) \approx -p$ . Therefore we get  $Np = \log_e N$ .

Intuitively it appears that the distance between any two arbitrary nodes would be less for the Bernoulli random graph. This is because at each step, we branch out by a huge number and therefore very quickly reach every other node from a given node.

Consider a node  $X$ . Let  $N_i$  be the expected number of nodes within  $i$  hops of  $X$ . Since the degree of node  $X$  is on an average equal to  $Np$ ,  $N_1 = Np$ . Let  $Np = d$ . Now,  $N_2 \approx d^2$ . In general,  $N_k \approx d^k$ . We need to find  $k$  such that  $d^k = N$ . Therefore, the expected diameter of the network is  $\log_d N$ .

One disadvantage of the Bernoulli random graph is that it does not give us a sense of proximity. This means that if we are at a particular node and are searching for another node, there is no way to tell how close we are or where we need to search to find the destination node.

## Small World Model

Consider a network structure similar to the line network described earlier except the nodes are arranged in an ordered fashion. This means that if we are at a particular node, we know which way to traverse in order to reach a given node. An example of an ordering of people (nodes) is arranging them by their technical competence.

Suppose you are at node  $A_i$  and are looking for node  $A_j$ , then you would traverse right if  $j > i$  and left if  $j < i$ . In addition to this structure, two nodes  $A_i$  and  $A_j$  can have a direct link in between them with probability  $\frac{1}{|j-i|}$ .

Both the expected degree and the expected path length are approximately

$\log N$  is such a network. Also, the paths are discoverable. That is, suppose a node  $A_3$  wants to send a letter to node  $A_{40}$ .  $A_3$  has a link to node  $A_{20}$  as well as  $A_{45}$ . Now,  $A_3$  chooses the path through  $A_{45}$  as it is closer to  $A_{40}$ .

The ability to be able to find nodes is important for the small world model. Therefore a social network that built around communities that makes finding people easier is likely to be richer.

If your goal is to have a rich social network where individuals can easily discover others, then, even a small degree suffices (20-25) for connectivity and short paths. However, you also need a notion of proximity (through communities and other features).