MS&E 235, Internet Commerce

Stanford University, Winter 2007-08

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

Lecture 12: Small World and Internet Architecture

Small World Model Continued

Consider a network structure with nodes $A_1, A_2, ..., A_N$. Two nodes A_i and A_j have a direct link in between them with probability $\frac{1}{|j-i|}$. In particular, nodes A_i and A_{i+1} always have a direct link between them.

This means that if we are at a particular node, we know which way to traverse in order to reach a given node. 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 $j < i$.

The expected degree in such a network is $\approx \log N$.

Explanation: Consider a node A_1. The probability that A_1 is connected to A_2 is 1, to A_3 is 1/2 and in general to A_i is $1/(i-1)$. The expected degree of A_i is therefore equal to $1 + 1/2 + 1/3 + ... + 1/(N-1) = \sum_1^{N-1} 1/i$, which is approximately $\log N$.

The expected intermediary path lengths are also of the order $\approx \log_2 N$.

Explanation: Consider a source node i, a destination node $i + l$ and an intermediary node $i + k$. In order to reach $i + l$, node i contacts that node which is directly connected to it, to the left of $i + l$, and closest to $i + l$. The probability that i is directly connected to $i + l$ is $\frac{1}{l}$. Probability that i is not directly connected to $i + l$ is $1 - \frac{1}{l}$. Probability that i is not connected directly to either $i + l$ or $i + l - 1$ is then $(1 - \frac{1}{l}) * (1 - \frac{1}{l-1})$. Therefore, the probability that i is not directly connected to any node $i + k$ where $k \in [1, l]$ is:

$$\frac{k-1}{k} * \frac{k}{k+1} * .... * \frac{l-2}{l-1} * \frac{l-1}{l}$$

$$= \frac{k-1}{l}$$

The probability that i is connected to at least one of these nodes is therefore equal to $\frac{l+1-k}{l}$.

Let $p(k) = \frac{l+1-k}{l}$. Expected number of positions node i jumps over in 1 step $= \sum_{k=1}^{l} p(k) = \sum_{k=1}^{l} \frac{l+1-k}{l}$. When $k = 1$, $l + 1 - k = l$. Let $k' = l + 1 - k$. Now the equation can be rewritten as $p(k) = \sum_{k'=1}^{l} \frac{k'}{l} = \frac{l(l+1)}{2l} = \frac{l+1}{2}$. Therefore, on an average, at each step, you cover half the distance to the destination. Consequently, the expected number of steps to reach any node is $\log_2 N$.

Therefore, the small world phenomena tells us that even a relatively low average degree is sufficient to make a social network connected and rich.

What is more important is the presence of rich connections between the nodes. Taking the social network Linkedin as an example, we find that you are usually just a few degrees away from many people. However, if you are connected to someone through a friend who has a large number of connections, then chances are high your friend may not know that person very well. However, if your friends only has a few connections, then the chances are high that he will be able to introduce/recommend you to any of his connections.

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).

Internet Architecture and Related Businesses

Internet Addressing: Every node on the internet has (or thinks it has) a unique IP (Internet Protocol) address. An IP address consists of 32 bits.

IP Invariant: Every node on the internet knows the next node on the path to every other node on the internet. However, this is a complex problem as there is a significant amount of data that needs to be stored. Also, nodes get added/removed everyday and IP addresses can keep changing.

IP addresses are of the form A.B.C.D. IP addresses are assigned in big chunks to Internet Service Providers. The IP addresses assigned to one big service provider may be for the form 171.*.*.* and all IP addresses that are of this form and belong to this particular ISP might know what the next node is going to be.

The Domain Name Service server is a client server application that translates names to addresses. Every node on the Internet, knows the IP address of at least on Domain Name Server (DNS server).

A special kind of DNS server is the authoritative DNS server which given a request, not only resolves the address but also directs the request to the appropriate server. This can be used for load balancing. For example, requests from the east coast can be directed to a different server from those that come from the west coast. With the help of such servers, for the same name, the request can be redirected to different servers or requests with different names can be redirected to the same server. For example, requests for or.stanford.edu and stanford.edu can be handled by the same machine.