In the previous lecture we asked: how does the internet graph look like?

We looked for a reasonable model having:

1. Small diameter
2. Most nodes having small node degree (i.e., #connections)

We expressed the small diameter requirement as a logarithmic relationship:

\[ d \approx \log N \]

In practice, I can reach a billion nodes in 9 hops. Currently, there are about 3 billion internet nodes, and studies through measurements give about 10 hops maximum. So it is reasonable. See link to Internet whois count & handout on hop-count on our web page.

A first model we examined was a random graph. This has the two properties we want, but it also has one important drawback:

**NO STRUCTURE**

Any node can connect to any other node with the same probability when \( p \approx N \), we get a giant connected component with a small diameter, but it looks like a giant "blob". The "Spaghetti Internet".

The "real internet" has structure, it is a collection of interconnected networks.
So we introduced a notion of *clustering coefficient*. We want a network that has high clustering and small diameter.

One possibility is to start with a regular graph, like a grid, or a ring, and then add random connections between the nodes to effectively reduce the diameter.

Watts-Strogatz model (1998)

**QUESTION:** At the end, how many "long distance" edges will you have?

There will be on average \( P \frac{NK}{2} \) "long distance" edges.

- \( P = 0 \) \( \Rightarrow \) Regular graph
- \( P = 1 \) \( \Rightarrow \) Random graph with node degree \( K \) and every edge occurring with probability

\[
P_{edge} = \frac{NK}{2} \frac{1}{\binom{N}{2}}
\]

This requires a bit of math to explain.

At the end we have \( M = NK \frac{P}{2} \) edges, and a graph \( G(N, M) \) of \( N \) nodes and \( M \) edges chosen at random among all graphs of this type.

In random graph model \( G(N, P_{edge}) \), \( M \approx \binom{N}{2} P_{edge} \)

so it follows that \( \binom{N}{2} P_{edge} = \frac{NK}{2} \Rightarrow P_{edge} \approx \frac{NK}{12} \)
Average path length in WS model is in between

\[
\begin{align*}
\beta &= 0 & \quad & E(d) = \frac{N}{2K} \\
\beta &= 1 & \quad & E(d) = \frac{\log N}{\log K}
\end{align*}
\] (regular ring) (random graph)

As \( \beta \to 1 \) it approaches the second value very rapidly.

The model also has large clustering for moderate values of \( \beta \) and a degree distribution similar to that of a random graph, namely:

\[ P_k \]

\[ \text{peak at } K \]

\[ \text{exponential tail} \]

This means that all nodes have roughly the same degree \( \approx K \). In fact, the only way to change degree is if a node re-wires randomly to the same degree node more than once and for large \( N \) this is very unlikely.

Similar models can be obtained for grid or connecting "long distance" and for large networks. But is this the model of having almost constant node degree a good one?

We would like to have more variability in node degree distribution.

\[ P_k \]

\[ \text{power-law decay} \]