UNIVERSITY OF
Southampton

UNIVERSITY OF Southampton

Graph Dynamics and the Web

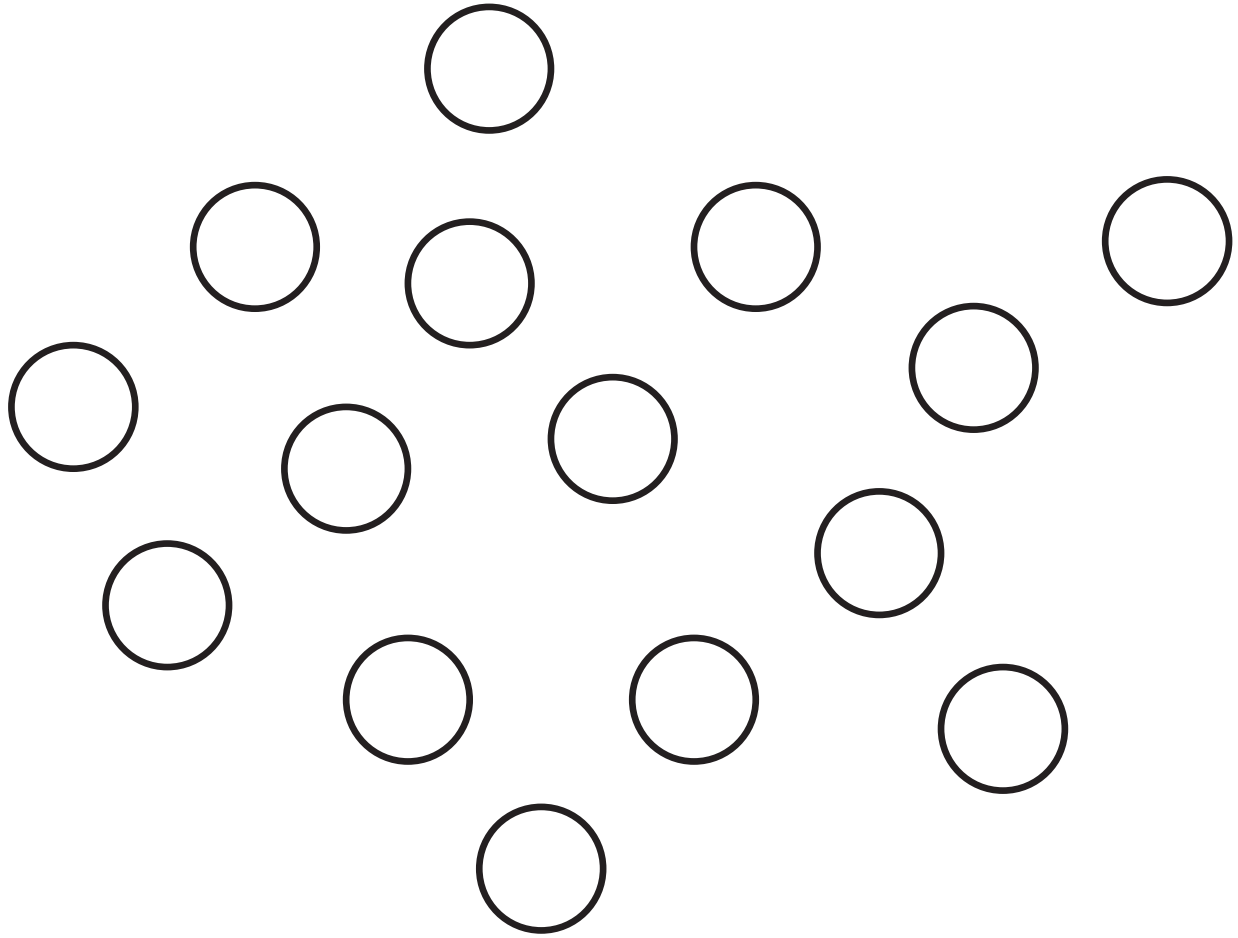
COMP3220 Web Infrastructure

Dr Heather Packer – hp3@ecs.soton.ac.uk

Graph Theory and Graph Dynamics

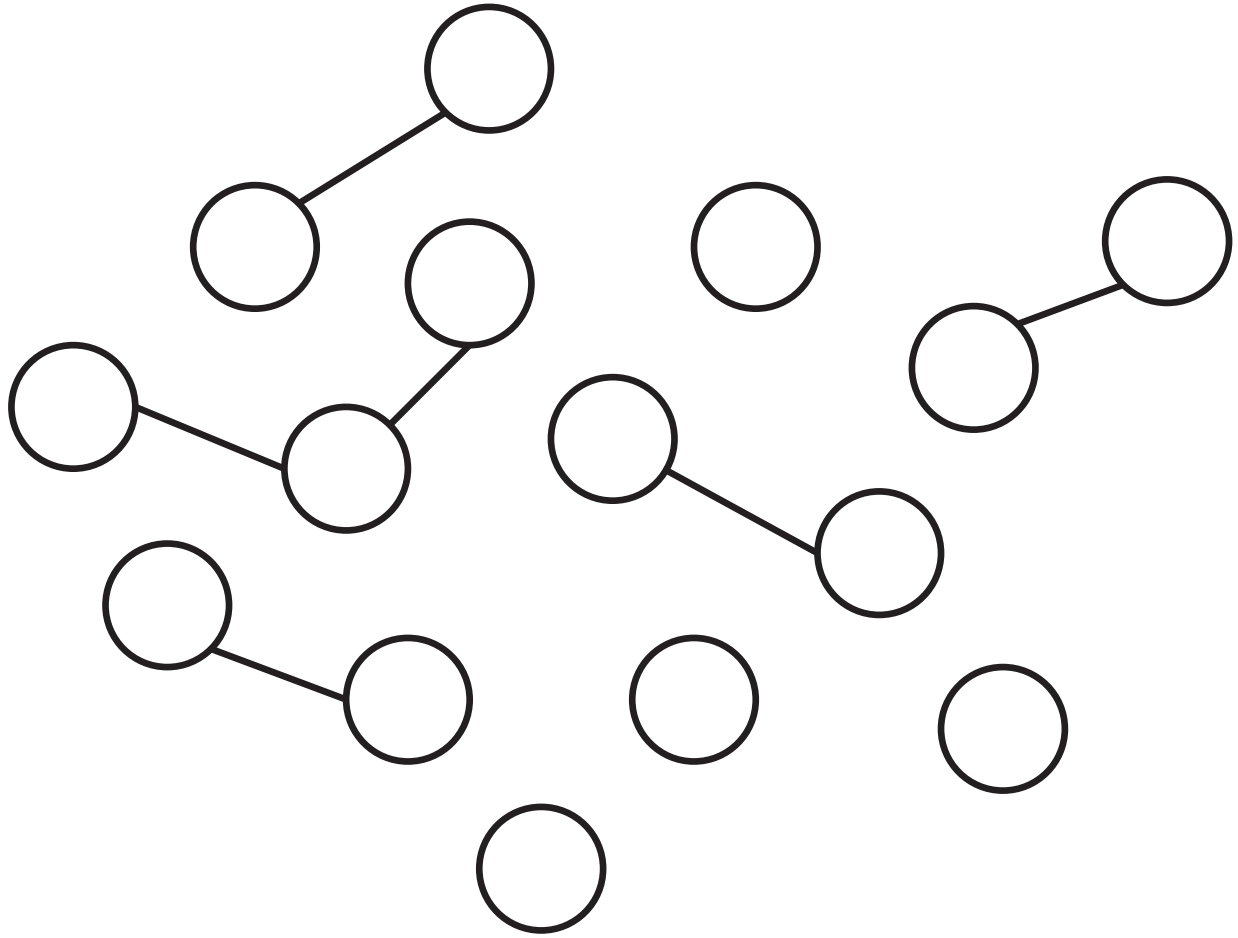
- Graph theory mostly focuses on static graphs
- Almost all real world networks are dynamic
- Evolution over time creates a new level of complexity

Growing a Network



- Simulate growth
  - Number of nodes
  - Number of links

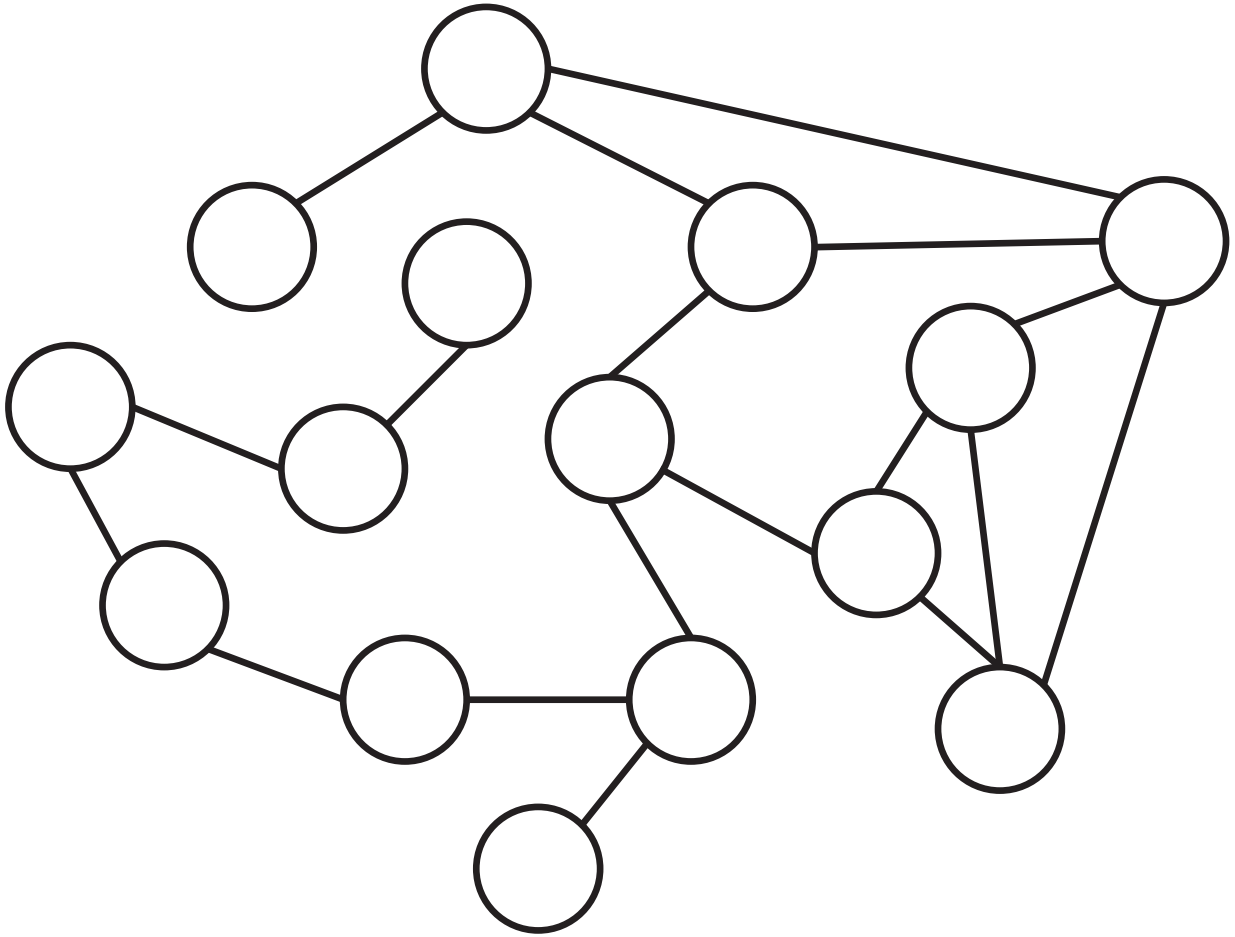
Growing a Random Network



Thresholds

1. When average degree goes above 1

Threshold 1



Small World Network – Erdos 1950s

- Threshold identified by Paul Erdos
- When the number of connections is lower than 1, the network is fragmented
- When it's above 1 it's connected

Small World Network

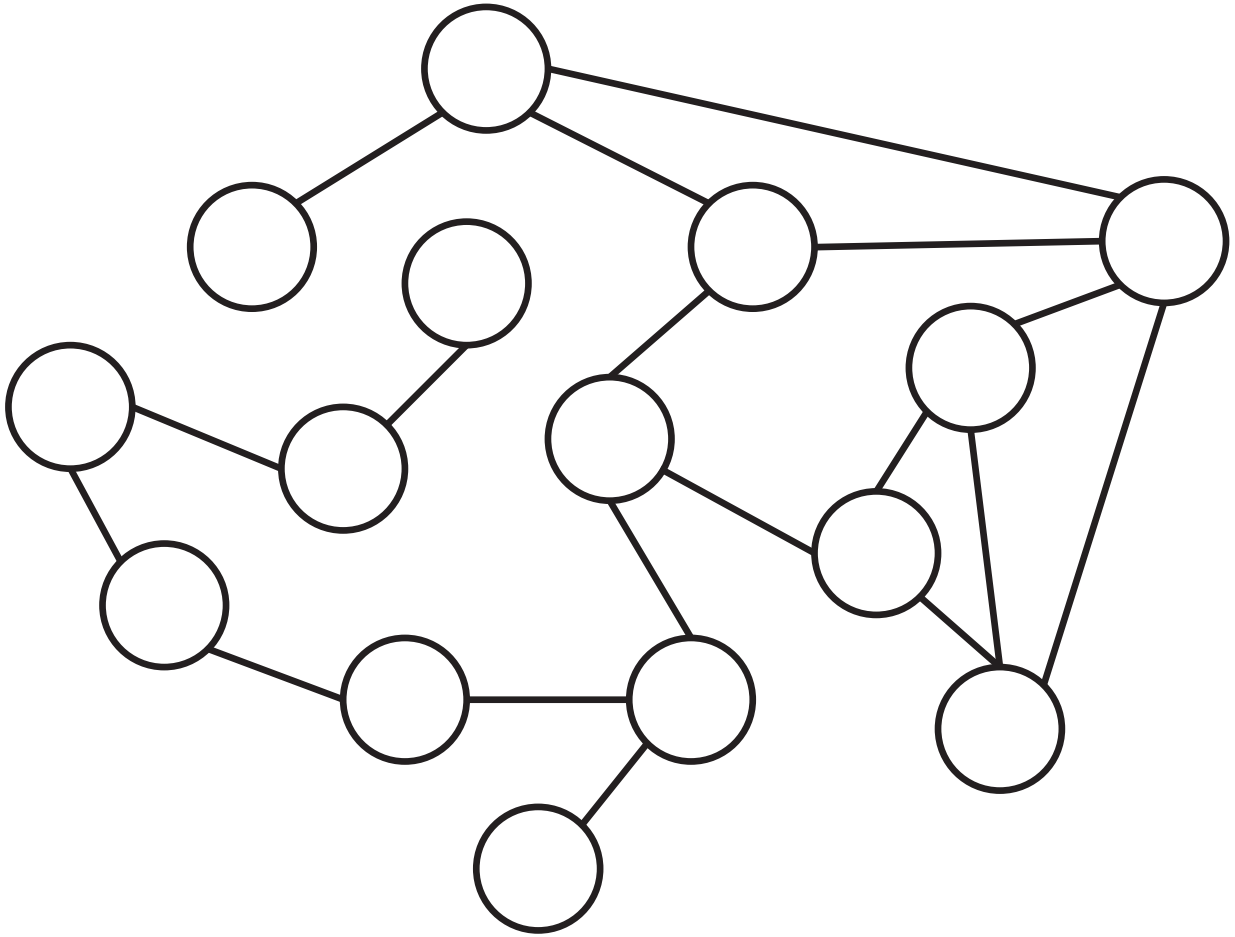
- **Six Degrees of separation** is a theory that everyone is 6 or fewer steps away from any other person in the world
- Experiment by Stanley Milgram in the 1960s
  - 200 letters
  - Sent to intermediary person on a first name basis
  - 64 letters arrived
  - Average of 5.2 intermediaries

- **Small World Network** is a graph where most nodes are not neighbours but most nodes can be reached by a small number of connections
  - Website links
  - Social networks
  - Wikipedia
- Many hubs – pages with in links
- Hubs significance can be modeled with degree centrality
- Robust for random node deletions

Thresholds

1. When average degree goes above 1
2. Nodes have an average degree of log(n)

Threshold 2



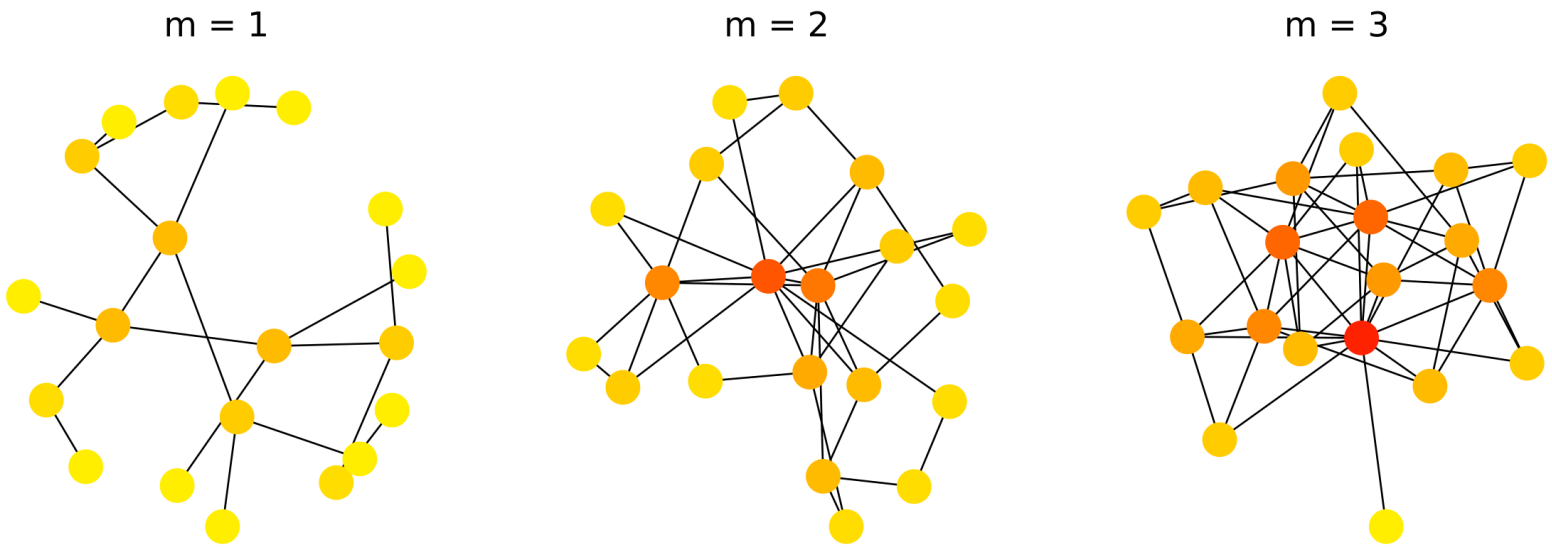
Real World Graphs

- Real world graphs aren't built randomly
- There are lots of social phenomena that effects the links between real world graphs
  - Reciprocity
  - Social Influence
  - Social Capital
  - Homophily
  - Preferential attachment

- Experimental Observations [Barabasi, Albert (1999)]
  - Sparse Graph
  - Power Law

Barabasi and Albert Experimental Methodology

- Intuition: Network grows by iterative process
- Simulate growing a network and then examine it's properties
- Experiment with different models



Graphs created with the Barabasi-Albert model

Preferential Attachment

- The probability of a new node linking to an existing node is proportional to the degree of that node. The probability $\Pi(k_i)$ that a link of the new node connects to node i depends on the degree k_i as:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}.$$

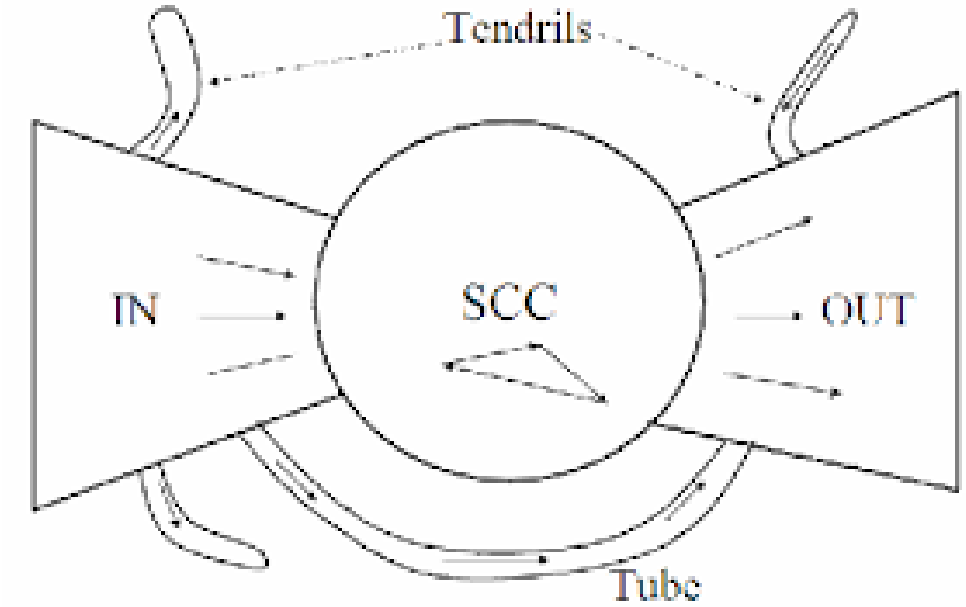
- A new node is free to connect to any node in the network
- For example, if a new node has a choice between a degree-two and a degree-four node, it is twice as likely that it connects to the degree-four node

Why Does Preferential Attachment reflect links on the Web?

Why Does Preferential Attachment reflect links on the Web?

- Quality resources are likely to be popular
- Findability – the more obscure something is the less findable it is, and thus is less likely to be linked to
- Human nature – the more popular a website is, the more likely it is to be shares and thus linked to, similar to PageRank's assumption

The Shape of the Web – The Bowtie



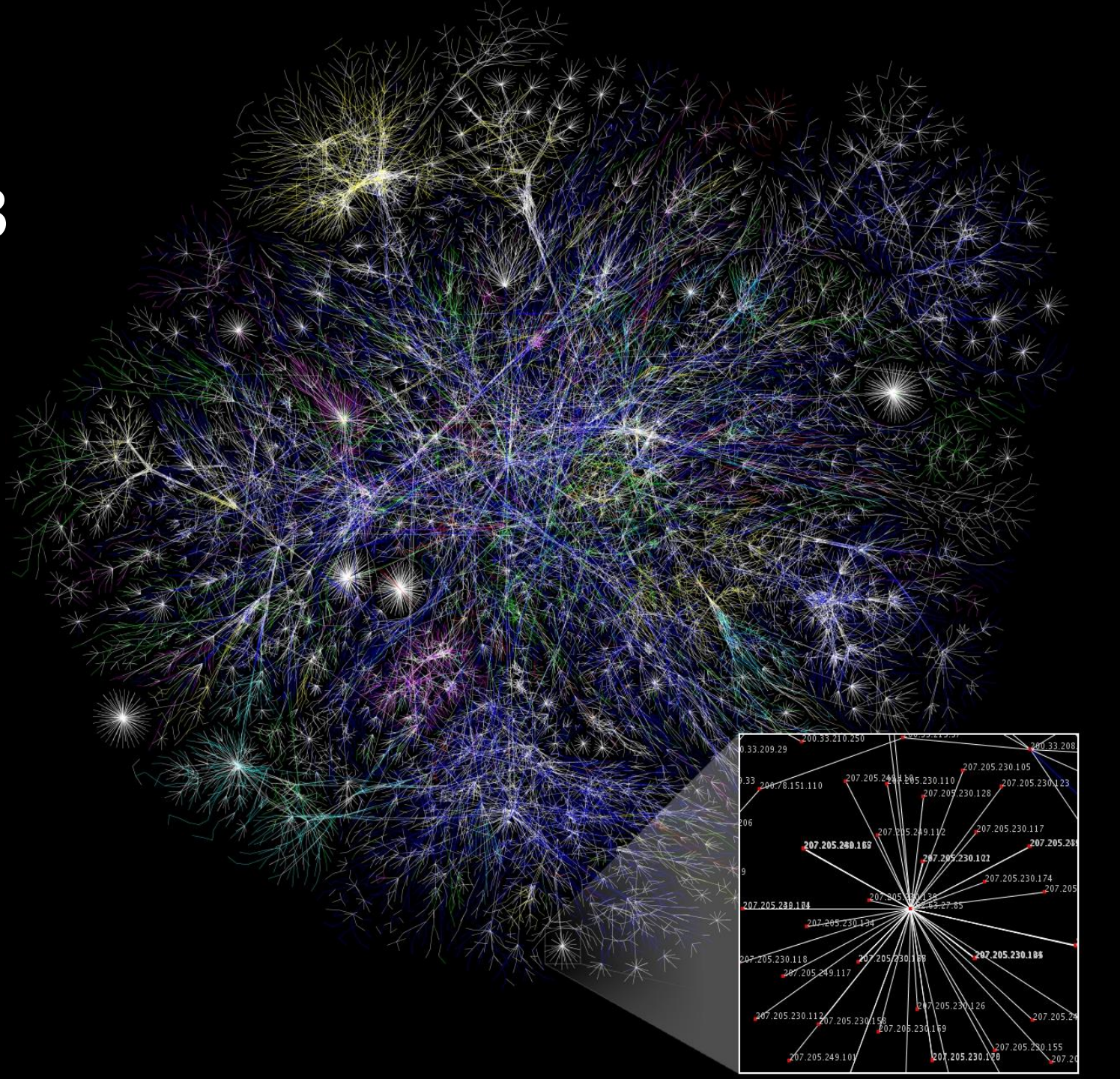
- The Shape of the Web [Broder et al. (2000)]
- Bowtie
  1. Strongly Connected Core
  2. In links
  3. Out links
  4. Tendrils and Tubes

- The diameter of the central core is at least 28, and the diameter of the graph as a whole is over 500
- The probability of a path between randomly chosen pairs is only 24%
  - Average directed path length is about 16
  - Average undirected path length is about 6

The Opte-Project 2003

- Aimed to map the Internet
- 200 million nodes
- 1 billion edges

- Partial map in Jan 2005
- Nodes are IP address
- Length of the lines denote delay



Modern Web Research

- Temporal aspects – how is the Web Graph evolving over time?
- Information aspects – how does new information propergate throughout the web (or blogsphere, or twitter..)
- Finer-grained structure – how to define and computer “communities” in information and social networks

Outcomes

- Describe the properties of the web
- Identify scientific methodologies used to investigate how the web evolved
- Describe the social factors that play a role in the clustering of websites
  - Preferential attachment