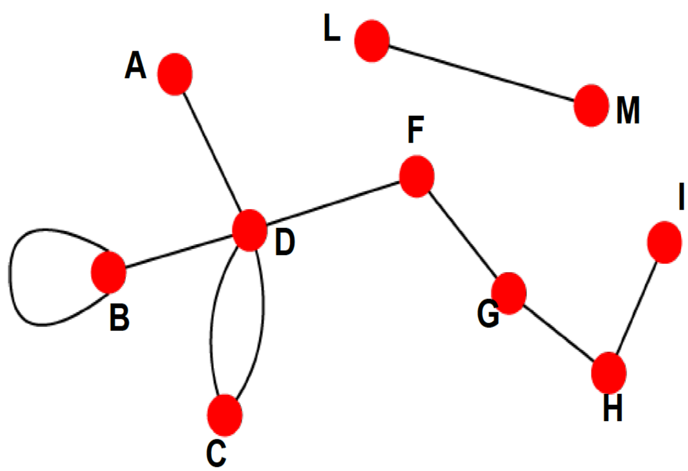
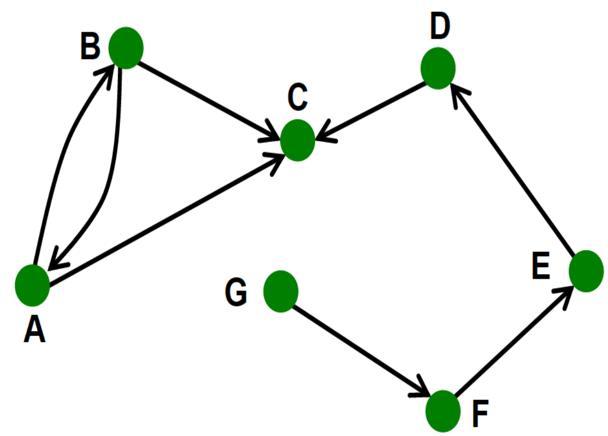
Peer-to-Peer and Social Networks
Fall 2015

Random Graphs

Directed vs. Undirected graphs



Collaborations,
Friendship on Facebook
(Symmetric relationship)

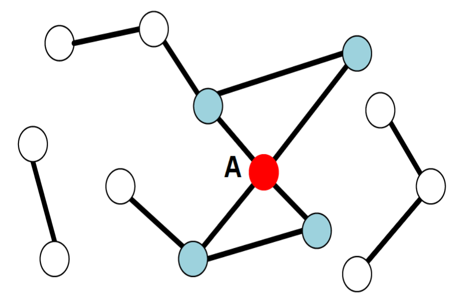


Emails, citations,
Following on Twitter
(asymmetric relationship)

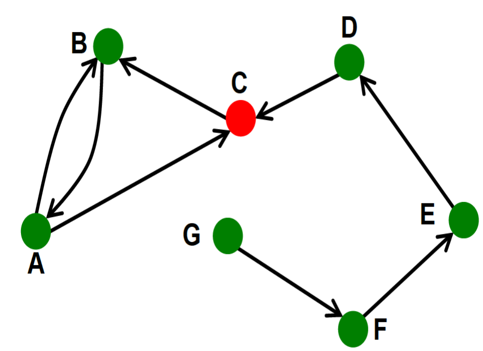
Directed vs. Undirected graphs

Degree of a node
= number of edges incident on it.

Degree of A = 4



The average degree of nodes in an n-node graph G = (V, E) is $\frac{2|E|}{n}$

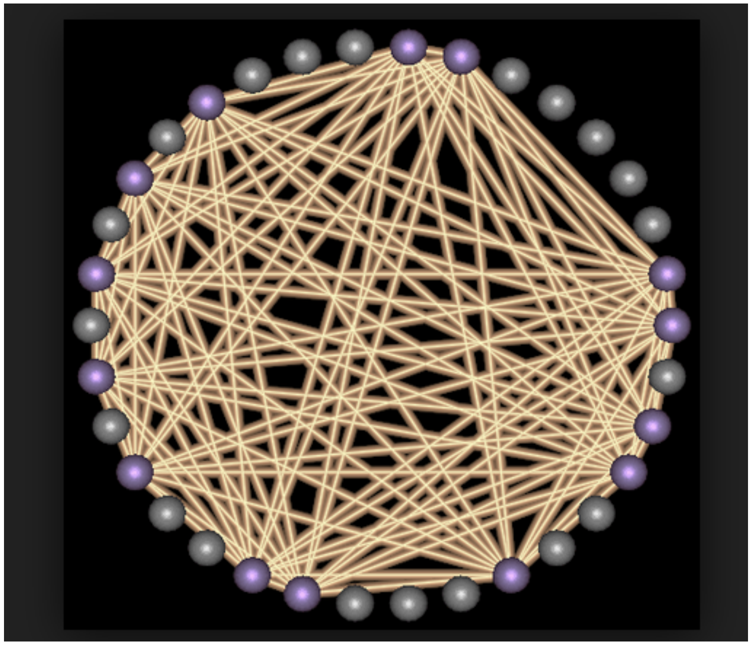


In directed graphs, we have to define **in-degree** and **out-degree** of nodes. What is the *average in-degree* of the directed graph to the right?

Sparse and dense graphs

In a clique with **n** nodes, the degree of each node is (**n-1)**. In **dense** graphs, the average node degree is **O(n)**. In **sparse** graphs, the average degree is much smaller.

Are most real world social networks dense or sparse?



A clique

Real-world graphs

Most real world graphs **are sparse**, i.e. the average degree is **much less than n-1** for an n-node graph. Some examples are

LinkedIn N=6,946,668	Average degree = 8.87
Co-authorship (DBLP) = 317,080	Average degree = 6.62

Graphs can be **weighted** or **un-weighted**.

Is the co-authorship network weighted or un-weighted?

Random graphs

<u>ERDÖS-RENYI MODEL</u>

One of several models of real world networks

Presents a theory of how social webs are formed.

Start with a set of isolated nodes $V = \{0,1,2,...,n\}$

Connect each pair of nodes with a probability p $(0 \leq p \leq 1)$

The resulting graph is known as $G(n,p)$

Random graphs

ER model is different from the $G(n,m)$ model

The $G(n,m)$ model randomly selects one from the entire family of graphs with n nodes and m edges.

Properties of ER graphs

Property 1. The *expected* number of edges is $\frac{n(n-1)}{2}p$

Property 2. The *expected* degree per node is $(n-1).p$

Property 3. The *expected* diameter of $G(n,p)$ is

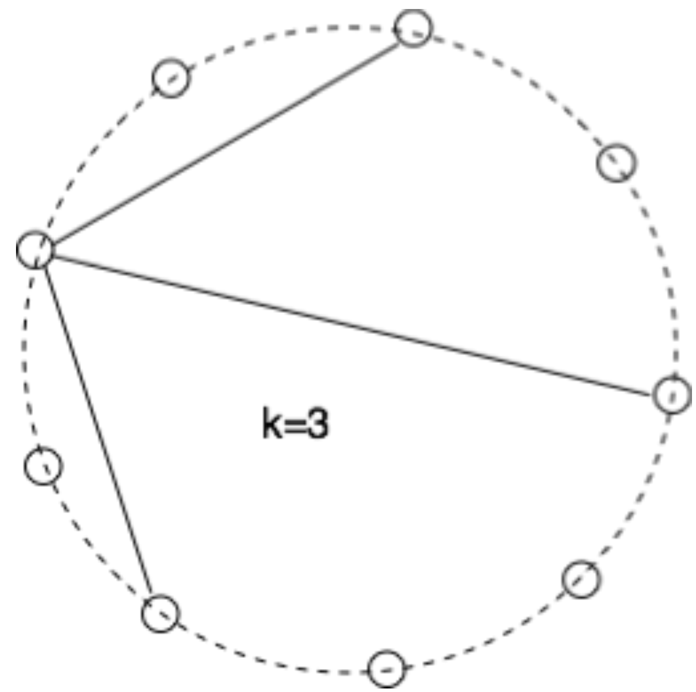
$$\log_{\deg} n = \frac{\log n}{\log \deg} = \frac{\log n}{\log\,(n-1).\,p}$$

[deg = expected degree of a node]

Degree distribution in random graphs

Probability that a node v connects with a given set of k nodes (and not to the remaining remaining $(n-k)$ nodes) is $p^k.(1-p)^{n-k}$

One can choose k out of the remaining $(n-1)$ nodes in $\begin{pmatrix} n-1 \\ k \end{pmatrix}$ ways.

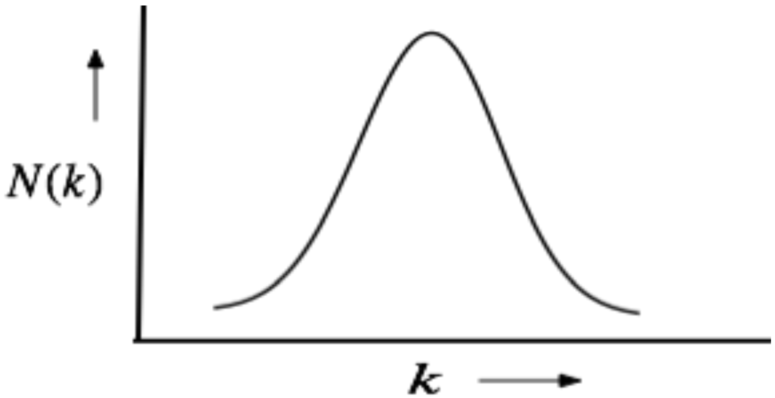


So the probability distribution is

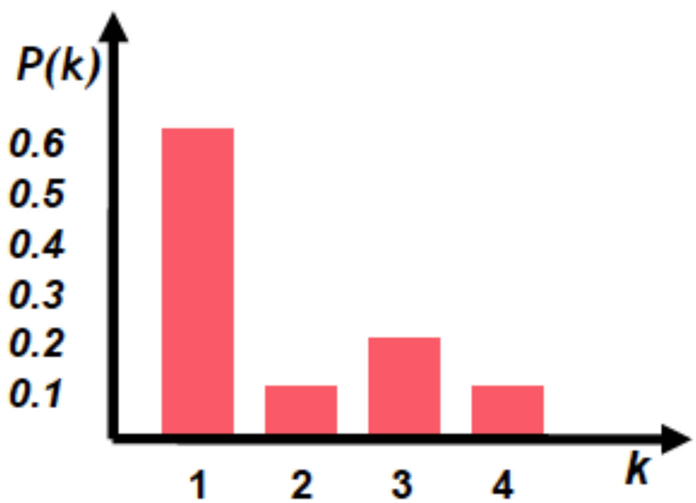
$$P(k)=\begin{pmatrix} n-1 \\ k \end{pmatrix}.p^k.(1-p)^{n-1-k}$$ (binomial distribution)

(For large n and small p it is equivalent to Poisson distribution)

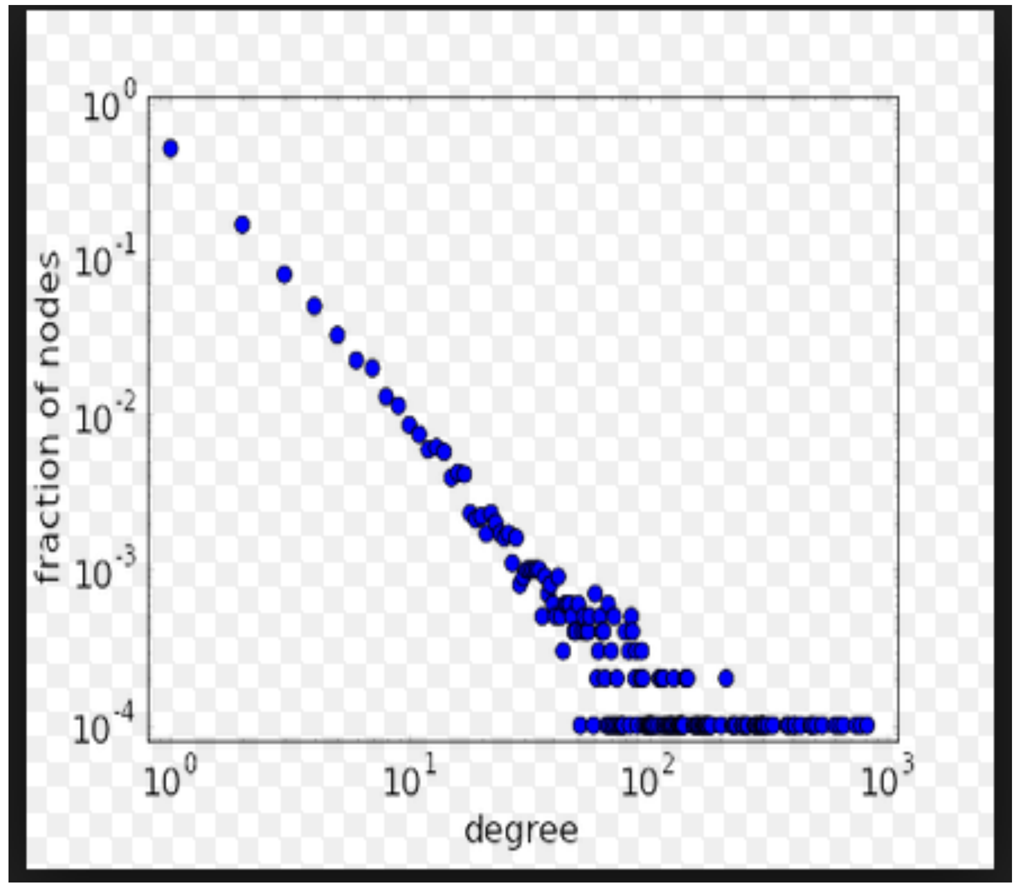
Degree distribution



Binomial distribution



P(k) = fraction of nodes that has degree k



What distribution is this?

Important network properties

Degree distribution: $P(k)$

Path length: h

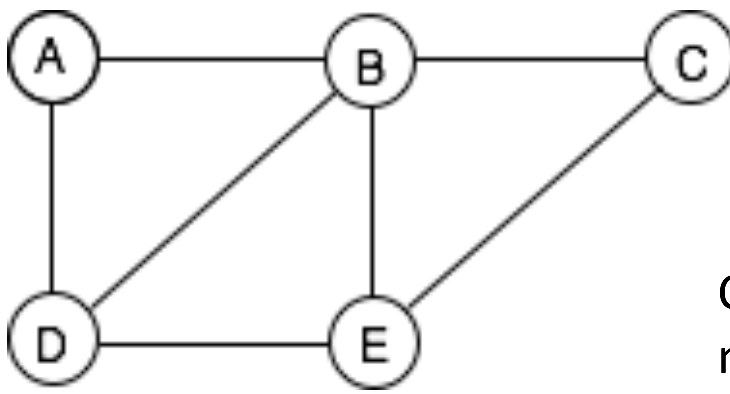
Clustering coefficient: C

Here *path length* means **average path length** between pairs of nodes

Clustering coefficient

For a given node, its ***local clustering coefficient* (CC)** measures what fraction of its various pairs of neighbors are neighbors of each other.

B's neighbors are {A,C,D,E}. Only (A,D), (D,E), (E,C) are connected



CC of a graph is the mean of the CC of its various nodes

CC(B) = 3/6 = ½ CC(D) = 2/3 = CC(E)

The **global CC** is the average of the local CC values.

Properties of ER graphs

-- When $p << \frac{1}{n}$, an ER graph is a collection of disjoint trees.

-- When $p = \frac{c}{n} (c > 1)$ suddenly one giant (connected) component emerges. Other components have a much smaller size $O(\log n)$ [Phase change]

Properties of ER graphs

When $p = \frac{c \log n}{n} (c > 1)$ the graph is almost always connected

why?

(i.e. connected with high probability)

These give "ideas" about how a social network can be formed.

But a social network is not necessarily an ER graph! Human society is a "clustered" society, but ER graphs have poor (i.e. very low) **clustering coefficient.**

The clustering coefficient of an ER graph = p (why?}

How social are you?

Malcom Gladwell, a staff writer at the New Yorker magazine describes in his book The Tipping Point, an experiment to measure how social a person is.

- He started with a list of 248 last names
- A person scores a point if he or she knows someone with a last name from this list. If he/she knows three persons with the same last name, then he/she scores 3 points

How social are you?

(Outcome of the Tipping Point experiment)

Altogether 400 people from different groups were tested.

(min) 9, (max) 118	{from a random sample}
(min) 16, (max) 108	{from a highly homogeneous group}
(min) 2, (max) 95	{from a college class}

[Conclusion: Some people are very social, even in small or homogeneous samples. They are **connectors**]

Connectors

Barabási observed that connectors are not unique to human society only, but true for many complex networks ranging from biology to computer science, where there are some nodes with an anomalously large number of links. Certainly these types of clustering cannot be expected in ER graphs.

The **world wide web,** the ultimate forum of democracy**,** is **not** a random network, as Barabási's web-mapping project revealed.

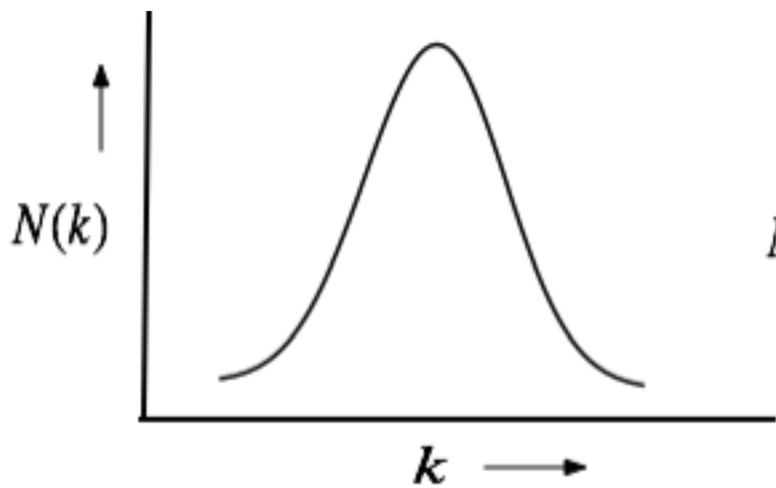
Anatomy of the web

Barabási first experimented with the Univ. of Notre Dame's web.

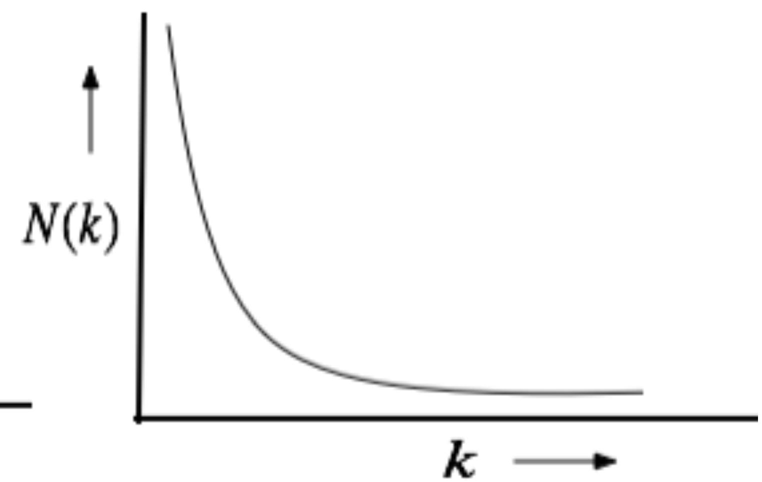
- 325,000 pages
- 270,000 pages (i.e. 82%) had three or fewer links
- 42 had 1000+ incoming links each.
- The entire WWW exhibited even more disparity. 90% had ≤ 10 links, whereas a few (4-5) like Yahoo were referenced by close to a million pages! These are the hubs of the web. They help create short paths between nodes (mean distance = 19 for WWW obtained via extrapolation). **(Some dispute this figure now)**

Random vs. Power-law Graphs

The degree distribution in of the webpages in the World Wide Web follows a power-law

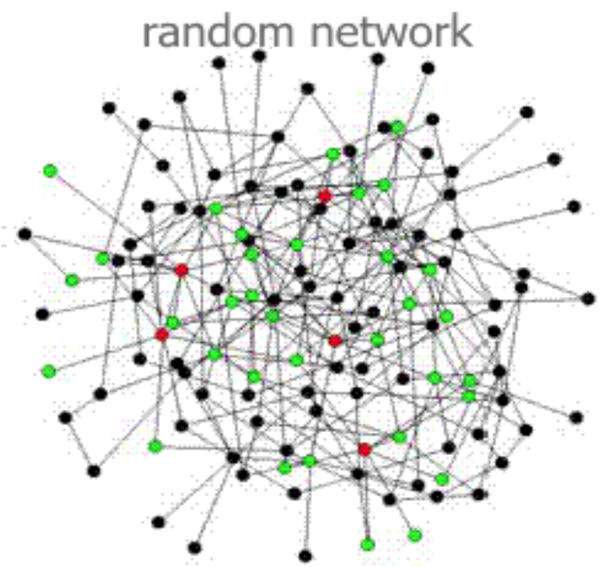


Random graph



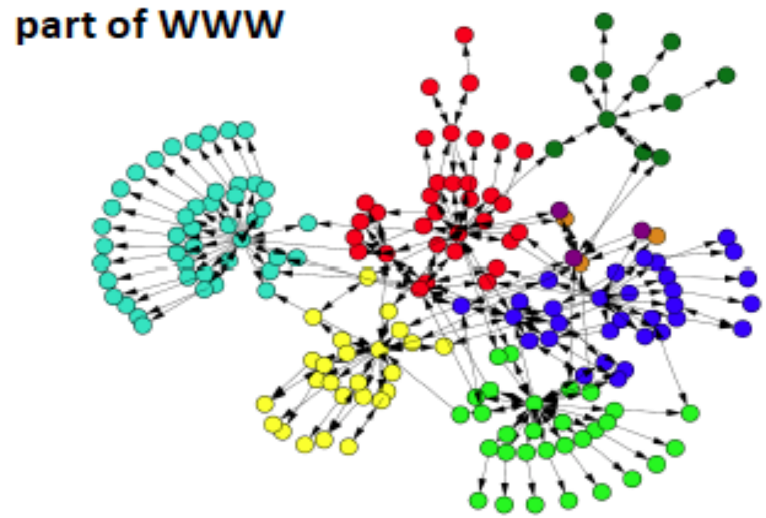
PowerLaw distribution

Random vs. Power-law Graphs



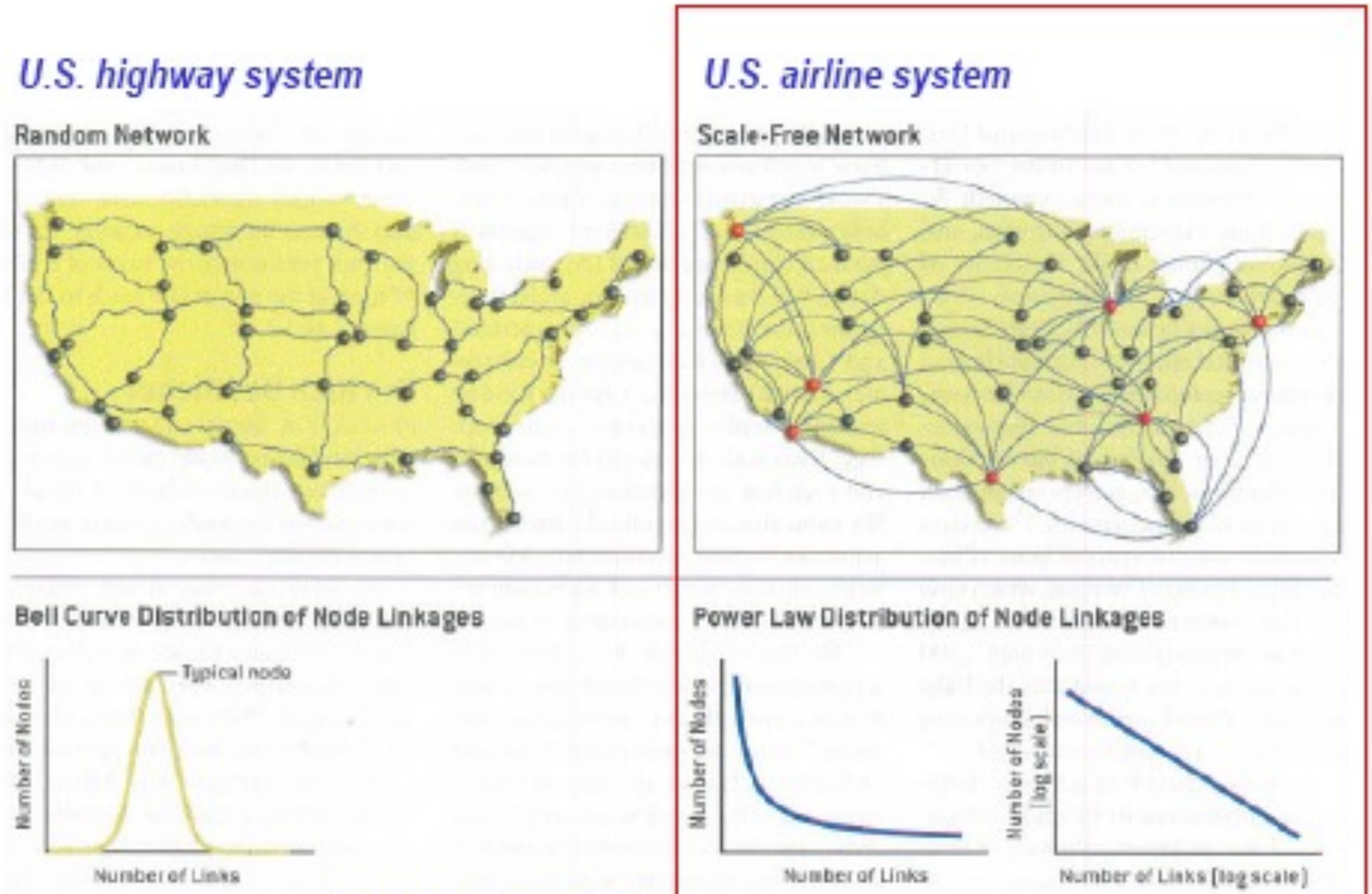
Typical structure of
a randomly connected
network

http://www.dichotomistic.com/images/random%20network.gif



Typical structure of
World Wide Web
(nodes = web pages, links = links between pages)

Random vs. Power-Law networks

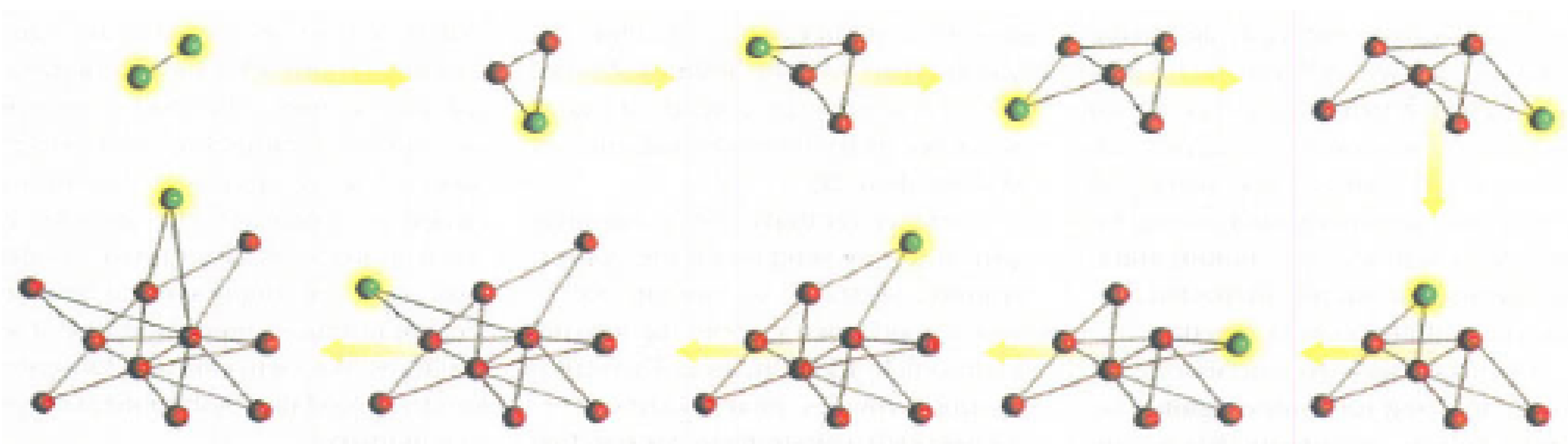


(Barabási & Bonabeau, 2003)

Evolution of Scale-free networks

- the *Barabási-Albert model*, reproduces the scale-free property by:
  - growth and
  - (linear) preferential attachment

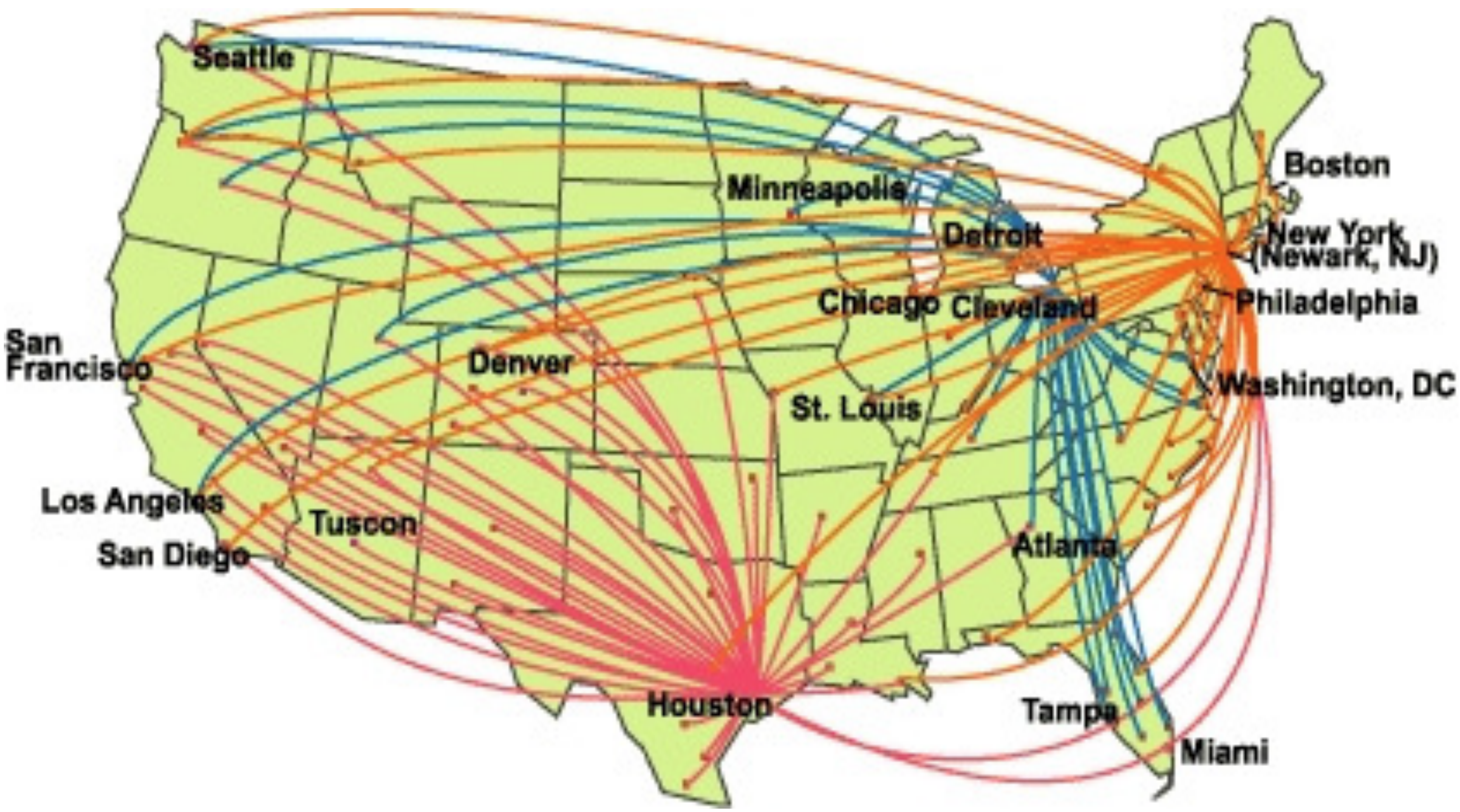
- growth: a node is added at each step
- attachment: new nodes tend to prefer well-connected nodes ("the rich get richer" or "first come, best served") in linear proportion to their degree



Growth and preferential attachment creating a scale-free network

(Barabási & Bonabeau, 2003)

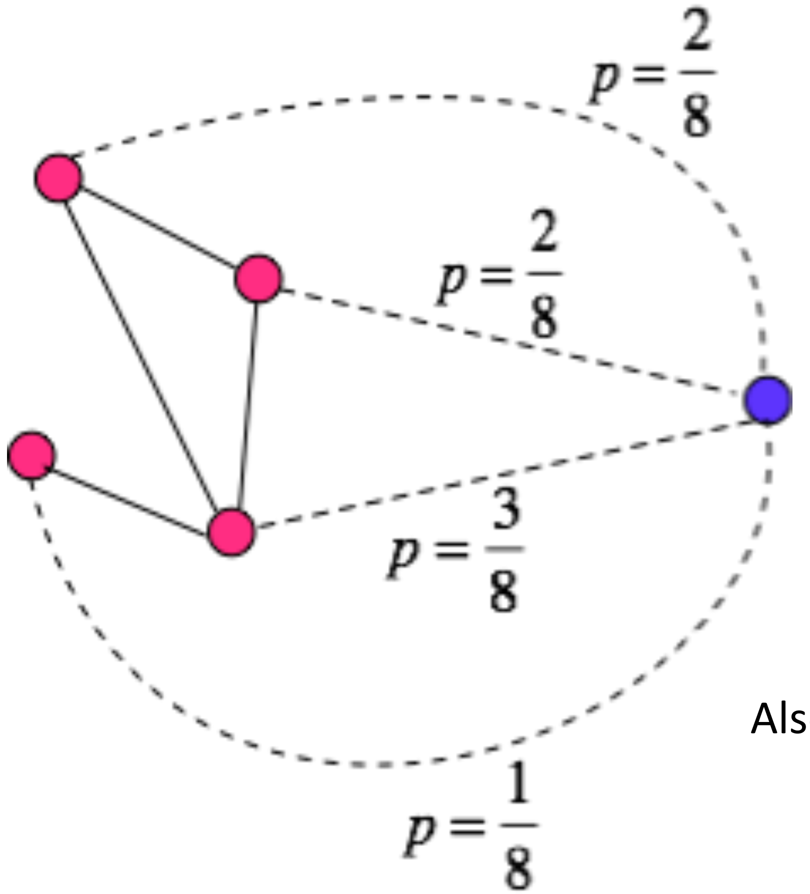
Example: Airline Routes



Think of how new routes are added to an existing network

Preferential attachment

Existing network



A new node connects with an existing node with a probability proportional to its ***degree***. The sum of the node degrees = 8

New node

Also known as “Rich gets richer” policy

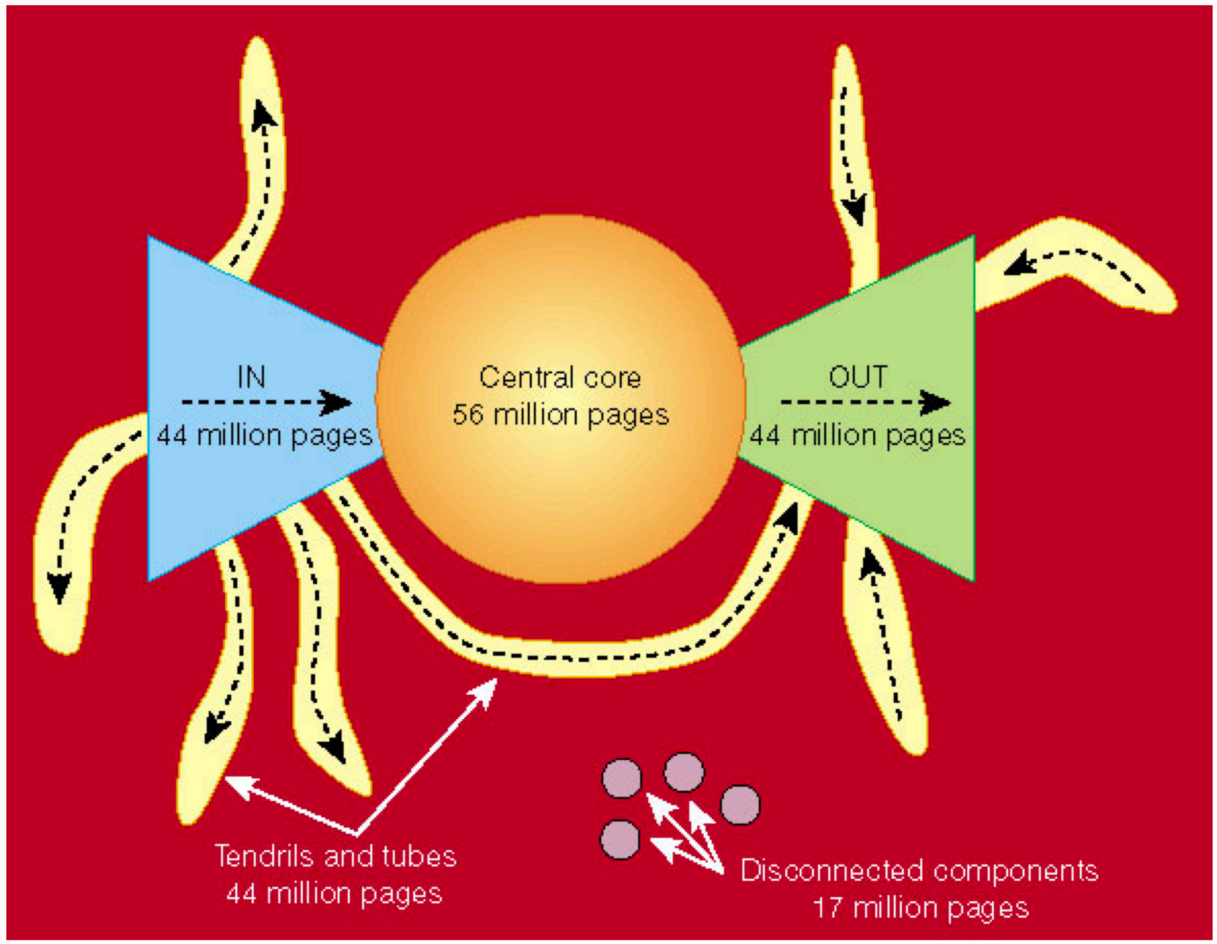
This leads to a power-law distribution (Barabási & Albert)

Anatomy of the web

For a given N, d follows a gaussian distribution so $\langle d \rangle$ can be interpreted as the diameter of the web, a measure of the shortest distance between any two points in the system. Despite its huge size, our results indicate that the web is a highly connected graph with an average diameter of only 19 links.

Albert, Jeong, Barabasi: Diameter of the World Wide Web. (Brief Communication). ***Nature*** 401, 9 Sep 1999

The web is a bow tie



Reference: Nature 405, 113(11 May 2000)

Power law graph

The degree distribution in of the webpages in the World Wide Web follow a **power-law**. In a power-law graph, the number of nodes $N(k)$ with degree k satisfies the condition $N(k) = C.\frac{1}{k^r}$

Also known as **scale-free graph**. Other examples are

-- Income and number of people with that income

-- Magnitude and number of earthquakes of that magnitude

-- Population and number of cities with that population