Social Networks: Models

Advanced Computer Networks

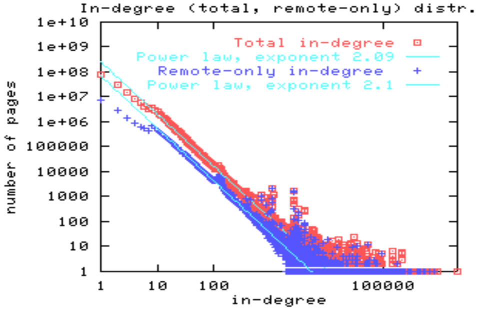
Summer Semester 2013

Recap: Power-law Distribution

- The fraction of node degrees in a social network follows power-law distribution
  - Let f(k) be the fraction of items have value k

$$f(k) = zk^{-\alpha}$$

  where α and z are constants
  - α is the power-law exponent, typically 2< α<3
- Testing for power-law distribution
  - If we draw k and f(k) in "log-log" scale, it shows a straight line
- Power-law distribution describe the popularity of nodes in a social network, where a small number of node have a large proportion of connections.



computer NETWORKS

Recap: Small-world Network

- How far apart are nodes in the network?
  - Network diameter
- How close a set of nodes connect with each other?
  - Clustering coefficient
- Milgram's experiment (1967)
- A small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops. [wiki]
- Properties: high clustering, low diameter
  - Clustering efficient: much larger than random network
  - Diameter: almost equal to random network
- Most of social networks are found to be small-world network.

<table>
<tr><th>Properties</th><th>Regular Network</th><th>Random Network</th><th colspan="2">Social Network</th></tr>
<tr><td>Degree Distribution</td><td>Constant</td><td>Normal distribution</td><td colspan="2">Power-law distribution</td></tr>
<tr><td>Path Length (Diameter)</td><td>High</td><td>Low</td><td>Low</td><td rowspan="2">Small-world</td></tr>
<tr><td>Clustering Coefficient</td><td>High</td><td>Low</td><td>High</td></tr>
</table>

Network Models

Why Model Networks?

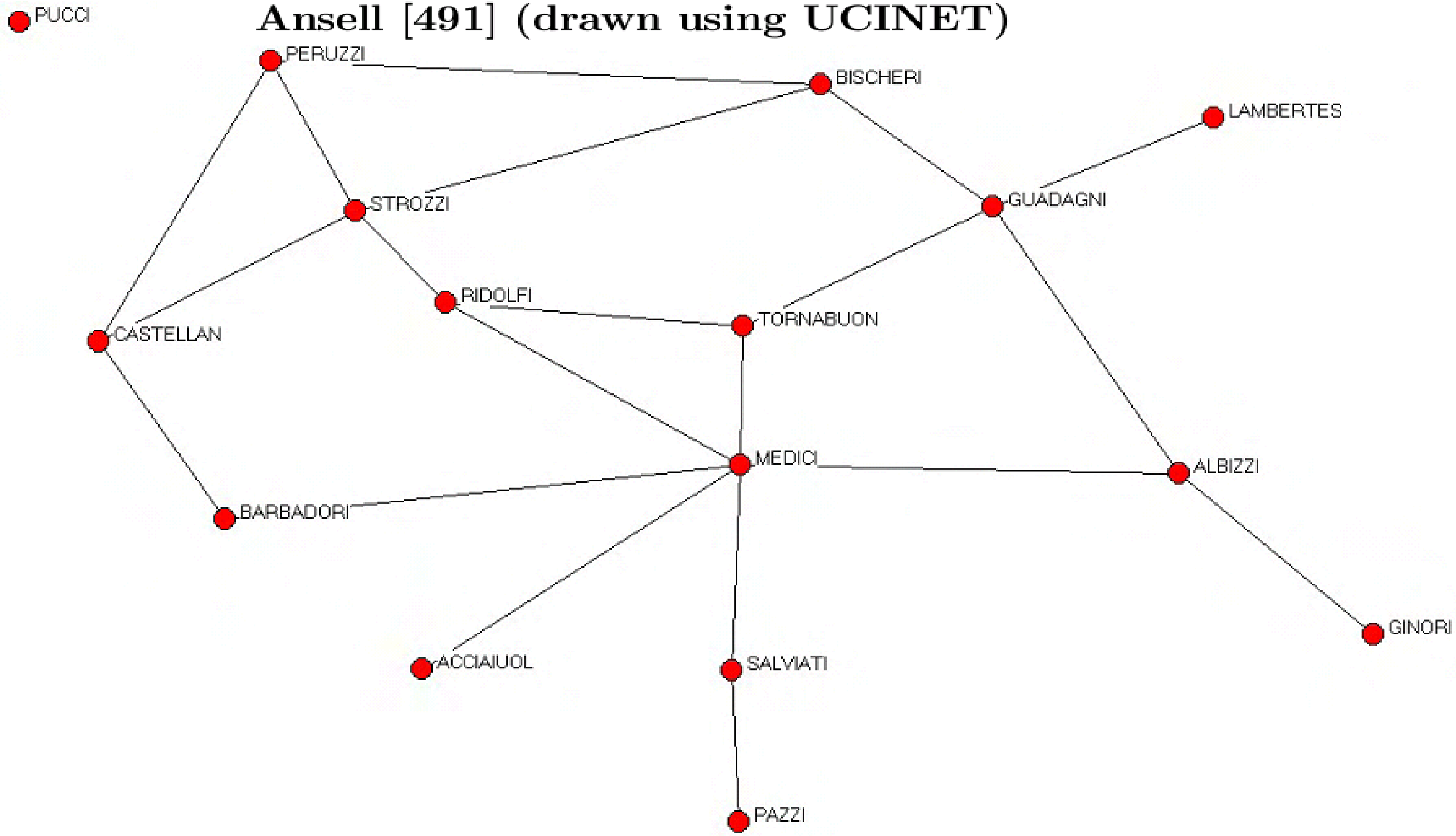
- Social Networks play a central role in
  - transmission of information
  - the trade of goods and services
  - diseases spread
  - which products we buy, which languages we speak, how we vote, …, etc.
- With various social network analysis techniques, we intend to understand
  - how social network structures impact behavior
  - which network structures are likely to emerge in a society

An Example: Florentine Marriages

- In the early fifteenth century, Florence had been ruled by several powerful families.
- The Medici have been called the “godfathers of the Renaissance”.
- Previously the Strozzi had both greater wealth and more seats in the local legislature, and yet the Medici rose to eclipse them.
- The Medici family consolidated political and economic power by leveraging the central position of the Medici in networks of family inter-marriages.
- To understand the rise in power of Medici, Padgett and Ansell [Padgett 1993] provide analysis to the social network structure.

Padgett, J.F. and C.K. Ansell (1993) "Robust Action and the Rise of the Medici, 1400- 1434." American Journal of Sociology, 98: 1259-1319.

15th Century Florentine Marriges Data from Padgett and Ansell [491] (drawn using UCINET)



Degree Centrality: $d_i(g)/(n-1)$

- n=16
- DC(Strozzi)=4/15
- DC(Medici)=6/15

Betweenness Centrality:

$$Ce_i^B(g) = \sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$$

# of shortest paths between k and j that i lies on

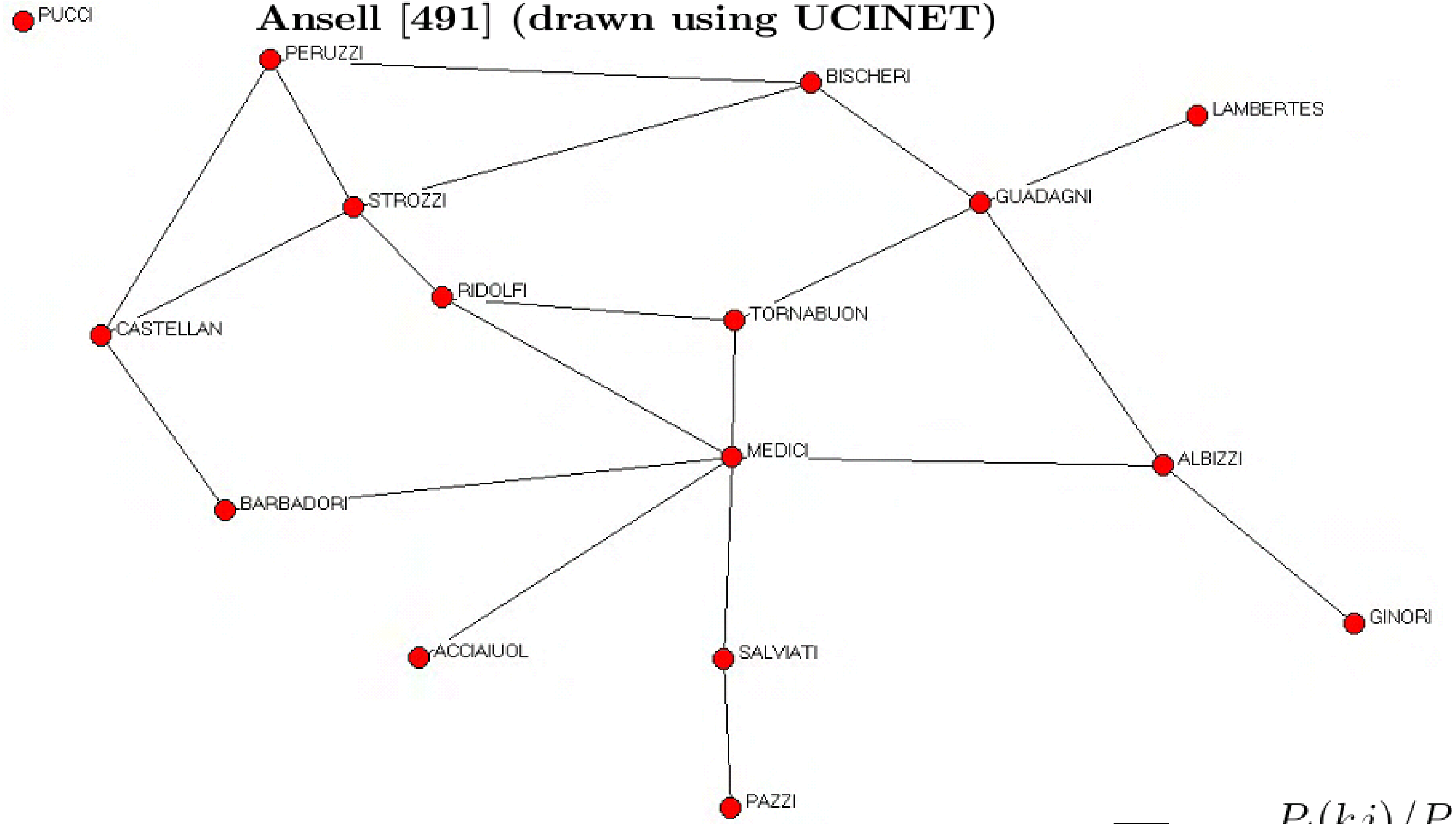
# of shortest paths between k and j

# of node pairs except i

Average fraction of shortest paths pass the node

computer NETWORKS

15th Century Florentine Marriges Data from Padgett and Ansell [491] (drawn using UCINET)



Betweenness Centrality:

$$Ce_i^B(g) = \sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$$

- n=16
- BC(Strozzi)=0.103
- BC(Medici)=0.522
- The Medici lies on over half of the shortest path!

- This analysis shows that network structure can provide important insights beyond those found in other political and economic characteristics.
- The example also illustrates that the network structure is important beyond a simple count of how many social ties each member has, and suggests that different measures of betweenness or centrality will capture different aspects of network structure.

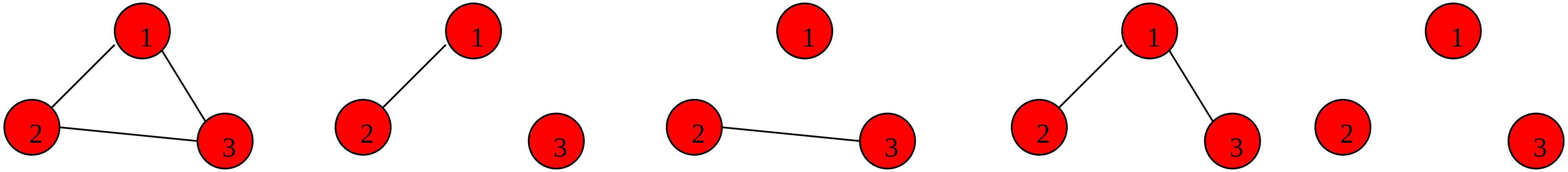
How Networks Form?

- Random Network
- Small-world Network
- Scale-free Network

The Random Network

- Erdös-Renyi Random Graph Model
  - A random process
  - Fix a set of n nodes, N={1,2,…,n}.
  - Each link is formed with a given probability p (0<p<1), and the formation of links is independent.
- For a given network with m links, the probability that it is formed is $p^m(1-p)^{\frac{n(n-1)}{2}-m}$

For example, when n=3



- Degree distribution of random Graph
- The probability that any given node i has exactly d links is

$$\Pr(d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}.$$

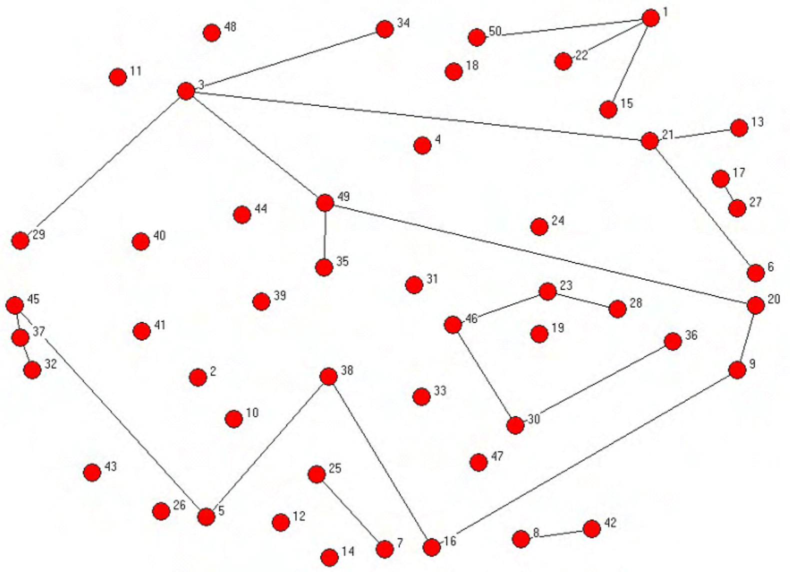
which is a binomial distribution.

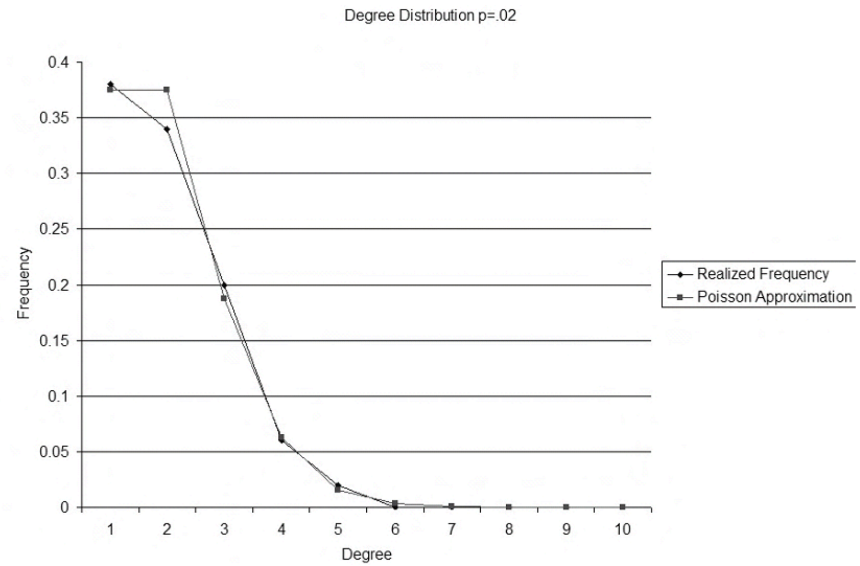
- For large n and small p, this binomial expression is approximated by a Poisson distribution

$$\Pr(d) \approx \frac{e^{-(n-1)p}((n-1)p)^d}{d!}.$$

- Such networks are also called Poisson random networks

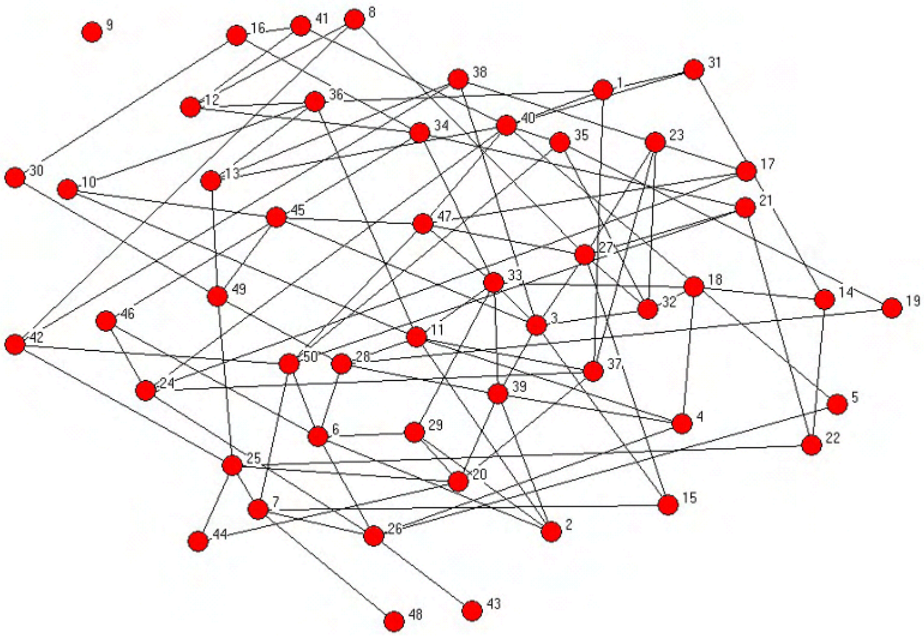
An example of Poisson random network

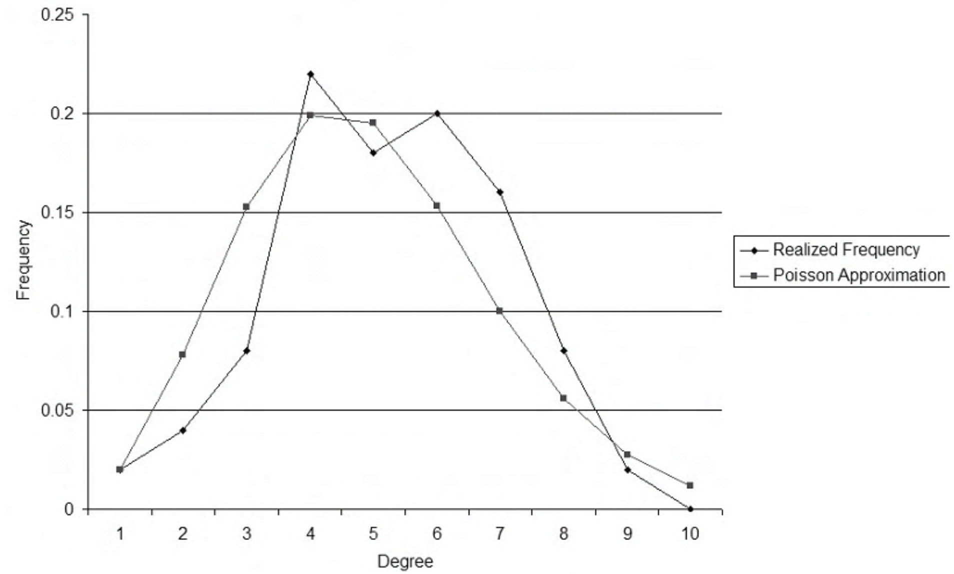




$$\Pr(d) \approx \frac{e^{-(n-1)p}((n-1)p)^d}{d!}$$

- n=50, p=0.02
- This network exhibits a number of features that are common to this range of p and n.
  - Pr(0)≈$e^{-49*0.02}$=0.375.
  - We should expect some isolated nodes. About 37.5 percent of the nodes expected to be isolated (i.e., have d = 0). 19 nodes observed.
  - The network is a “forest”, there are no cycles in the network (the probability is very low)
  - There is a giant component (with 16 nodes).





Pr(d)≈ $\frac{e^{-(n-1)p}((n-1)p)^d}{d!}$

- n=50, p=0.08
  - Pr(0)=$e^{-49*0.08}$=0.02
  - We should expect about 2 percent of the nodes to be isolated (with degree 0), or roughly 1 node out of 50.
  - The rest of the network is connected into one component.

Connectivity of Random Graph

- Consider what fraction of nodes are completely isolated; i.e., what fraction of nodes have degree d = 0?

$$\Pr(d) \approx \frac{e^{-(n-1)p}((n-1)p)^d}{d!}$$

- $\Pr(0)=e^{-(n-1)p}$
- If there is one isolate node, then

$$\Pr(0)=e^{-(n-1)p}=1/n$$

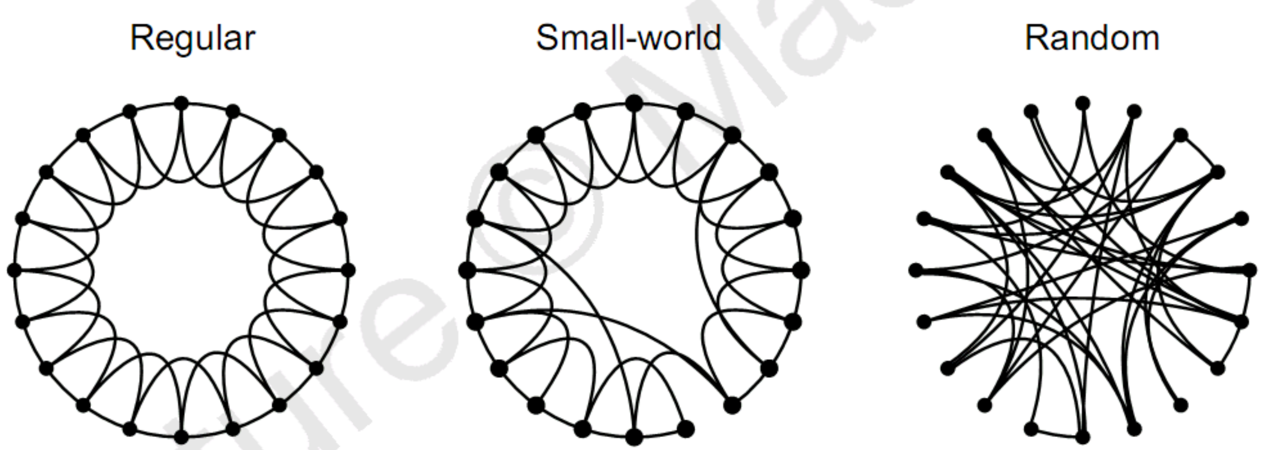
- Solving this equation yields: $(n-1)p= \log(n)$
  - The average degree $(n-1)p$ is $\log(n)$.
  - When $(n-1)p>\log(n)$, i.e., $p>\log(n)/(n-1)$, the whole graph is connected with high probability.
  - For n=50, the threshold $\log(n)/(n-1)$ is 0.079.
  - When $p<\log(n)/(n-1)$, the graph will have several components.

Example: Giant Component

- Is the global friendship network connected?
  - Not necessary, some nodes may have no friends
  - Large complex networks often have a giant component, a connected component that contains a significant fraction of all the nodes
  - Why only one?
    - If there are two, there must not be a single connecting link between nodes in the two components, which is unlikely.

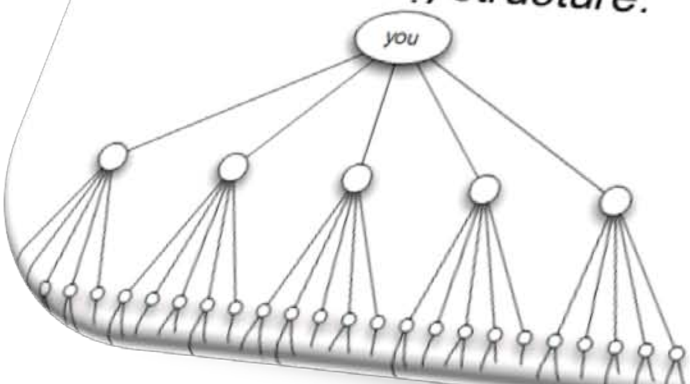
The Small-World Model

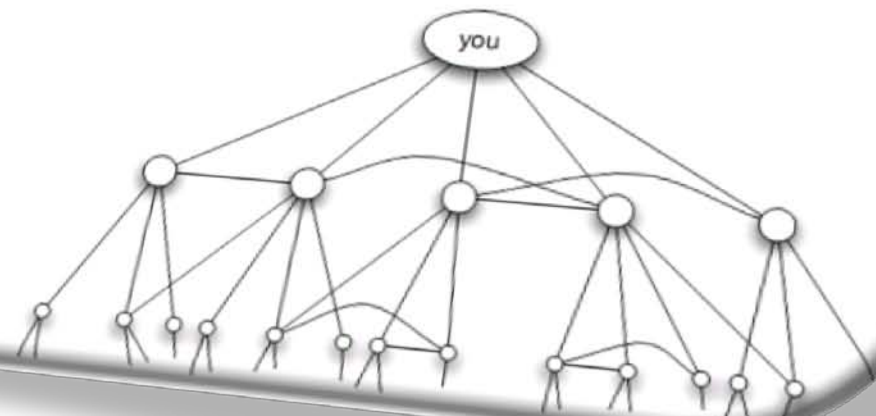
- Can we make up a simple model that exhibits both of the features: many closed triads, but also very short paths?
- **One-deminsional Model (Watts-Strogatz)**
- Starting from a ring lattice with n vertices and k edges per vertex.
  - Regular network with high clustering coefficient
- We rewire each edge at random with probability p ($0 \le p \le 1$).
  - p=0: regular network
  - p=1: random network
  - Randomizing the network, lowering average path length



A Simple Explanation

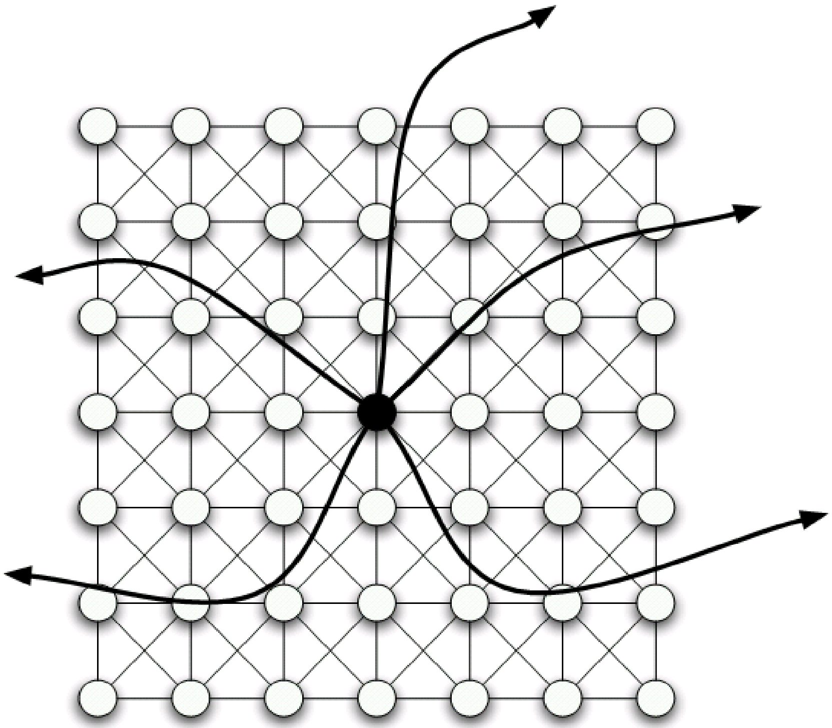
- Suppose each person knows 100 other people on a first-name basis
  - Step 1: reach 100 people
  - Step 2: reach 100*100 people
  - …
  - Step 5: reach 100^5= 10 billion people
  - Ref: the world population is 7.019 billion (Wiki, 2012)
- The numbers are growing by powers of 100
- But it is not true for real network!!!
  - Triadic relationships are common
  - Social network is highly clustered, not the kind of massively branching structure.

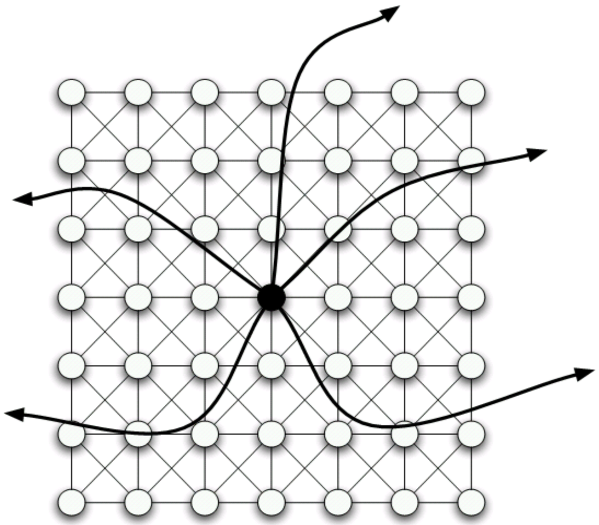




The Watts-Strogatz Model

- **The two-dimensional model: grid**
- Two kind of links
  - Regular links: Links to the other nodes within a radius of up to r grid steps
  - Random: Links to k other remote nodes

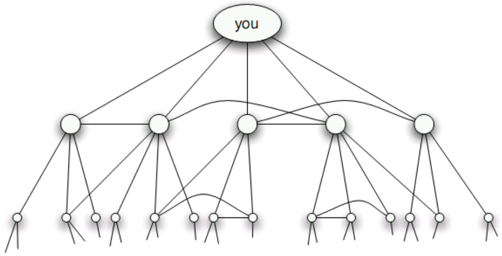




- High clustering

$$C_i \geq 2*12/(8*7) \geq 0.43$$

- Low diameter: short path exists with high probability
  - Since the k remote nodes are random and they barely know each other
    - For each step, at lease k new nodes are reached
    - The numbers are growing by powers of k
  - Still, short path achieves, the diameter is O(logn)



Extension

- Short path still exists even for very small amount of randomness
- For example, instead of allowing each node to have k random friends, we only allow one out of every k nodes to have one random friend
  - We can conceptually group k*k subsquares of the grid into “towns”
  - It will be similar: each town links to k other towns
  - Short path in towns -> short path in people

Small World: Summary

- A network between regular network and random network
- It has high clustering and low diameter
  - Clustering efficient: much larger than random network
  - Diameter: almost equal to random network
- The Watts Strogatz Model
  - Introducing a tiny amount of random links is enough to make the world small, with short paths between every pair of nodes.

The Scale-free Network

- A scale-free network is a network whose degree distribution follows a power law
- Several models for the forming of Scale-free networks
  - Rich get Richer Model
  - Preferential attachment [Barabási 1999]
  - Fitness Model [Bianconi 2001, Kong 2008]

Barabási, A.-L.; R. Albert (1999). "Emergence of scaling in random networks". *Science* **286** (5439): 509–512.

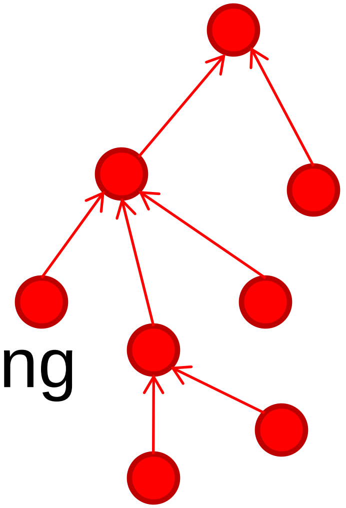
Bianconi Ginestra and Barabási A.-L., 2001a, *Europhys. Lett.* 54, 436.

J.S. Kong, N. Sarshar, V.P. Roychowdhury, "Experience versus Talent Shapes the Structure of the Web", Proceedings of the National Academy of Sciences of the USA, September 16, 2008; 105 (37)

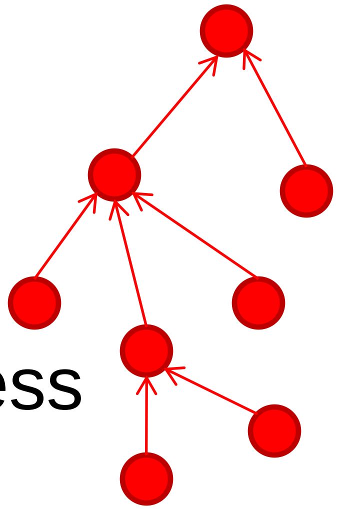
Rich Get Richer Model

- New nodes are more likely to link to nodes that already have high degree
- Power-laws arise from “Rich get richer”
  - We can provide a simple power-law models from consequences of individual decision-making
  - We assume simply that people have a tendency to copy the decisions of people who act before them

A Simple Model

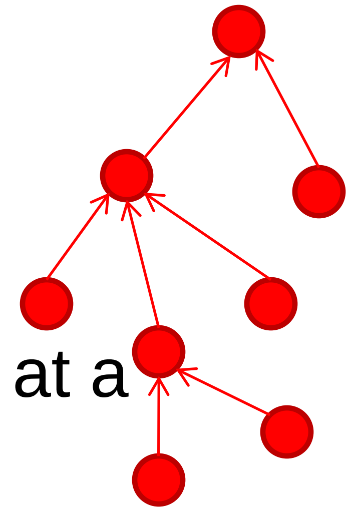


- A simple model for the creation of links among Web pages
  - Pages are created in order, and named 1; 2; 3; …;N.
  - When page j is created, it produces a link to an earlier Web page i according to:
  - 1) With prob. p (0<p<1), j links to i chosen uniformly at random (from among all earlier nodes)
  - 2) With prob. 1-p, node j chooses node i uniformly at random and links to a node i points to.
    - Note: this is same as saying, with prob. 1-p, node j links to node u with prob. proportional to the degree of u



- The model specifies a randomized process that run for N steps (as the N pages are created one at a time)
- Question: can we determine the expected number of pages with k in-links at the end of the process? (Or analyze the distribution of this quantity?)

computer NETWORKS



- Definition: the number of in-links to a node j at a time step t (t≥j) is a random variable $X_j(t)$
- Facts:
  - Initial condition: Since node j starts with no in-links when it is first created at time j, so $X_j(j) = 0$
  - The expected change to X_j over time
    - In step t+1, node j gains an in-link if and only if the link from the newly created node t+1 points to it
    - With prob. p, node t+1 point to the previous t nodes at random, thus the prob. of linking to node j is p/t
    - With prob. (1-p), node t + 1 links to node j with probability Xj(t)/t
    - So the overall probability that node t + 1 links to node j is

$$\frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

- In step t+1, the probability of a new link

$$\frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

- How to calculate $X_j(t)$?
- Approximation
  - Treat it as a deterministic continuous function $x_j(t)$
  - From step t to step t+1 can be viewed as the differential of the function, thus

$$\frac{dx_j}{dt} = \frac{p}{t} + \frac{(1-p)x_j}{t}$$

computer NETWORKS

- Solving the differential equation

$$\frac{dx_j}{dt} = \frac{p}{t} + \frac{(1-p)x_j}{t}$$

  - Let q=1-p, rewriting it as

$$\frac{1}{p+qx_j}\frac{dx_j}{dt} = \frac{1}{t}.$$

  - Integrating both sides

$$\ln(p+qx_j) = q\ln t + c$$

  - Applying the initial condition $x_j(j) = 0$ yields

$$x_j(t) = \frac{1}{q}\left(\frac{p}{j^q}\cdot t^q - p\right) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$$

computer NETWORKS

$$x_j(t) = \frac{1}{q}\left(\frac{p}{j^q}\cdot t^q - p\right) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$$

- $x_j(t)$ is the degree of node j at time t
- What is the fraction of nodes whose degree ≥k?
  - If a node's degree no smaller than k, it satisfies
    $$x_j(t) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right] \geq k,$$
  - Solving the inequality, we have $j \leq t\left[\frac{q}{p}\cdot k + 1\right]^{-1/q}$
  - Since there are t nodes at time t. the fraction of nodes is:
    $$\frac{1}{t}\cdot t\left[\frac{q}{p}\cdot k + 1\right]^{-1/q} = \left[\frac{q}{p}\cdot k + 1\right]^{-1/q}$$
- What is the fraction of nodes with degree exactly k?
  - Take derivative of the equation, we get $\frac{1}{q}\frac{q}{p}\left[\frac{q}{p}\cdot k + 1\right]^{-1-1/q}$
- So it is a power-law distribution with exponent $1 + \frac{1}{q} = 1 + \frac{1}{1-p}$

computer NETWORKS

Summary

- How Networks Form?
- Random Network
  - Erdös-Renyi Random Graph Model
- Small-world Network
  - The Watts-Strogatz Model
- Scale-free Network
  - Rich Get Richer Model