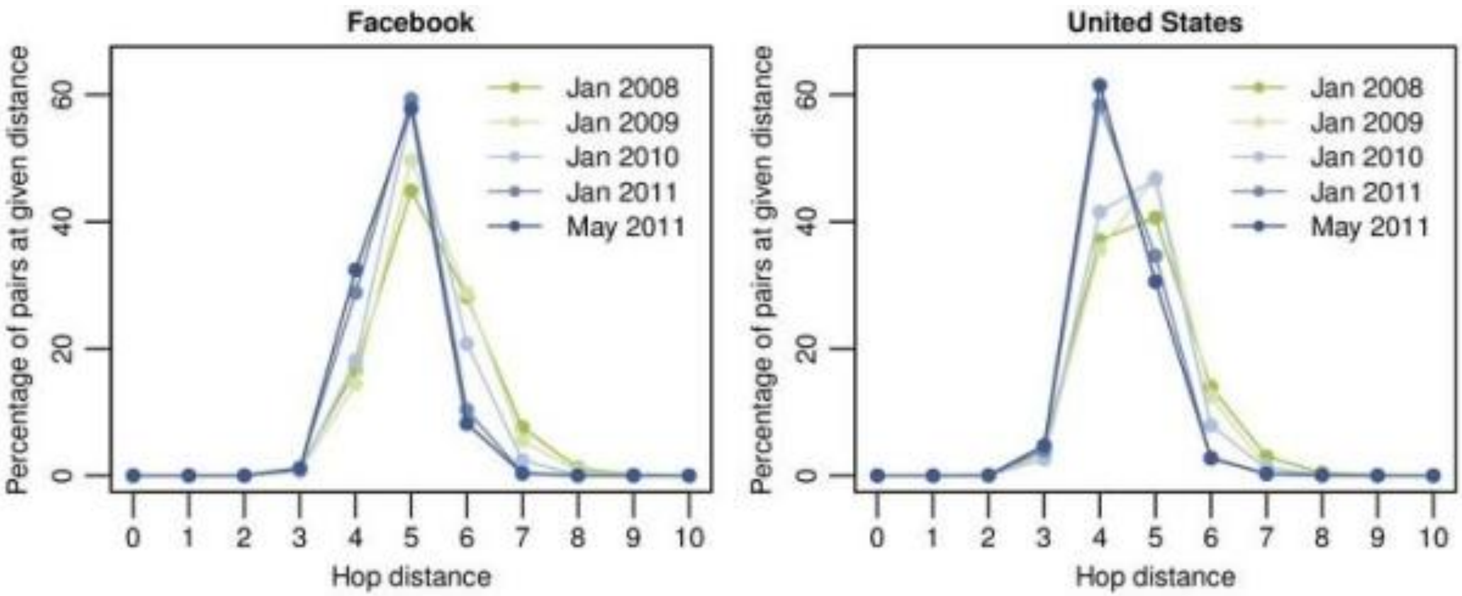
SNA

Small world, clustering coefficient

Small world

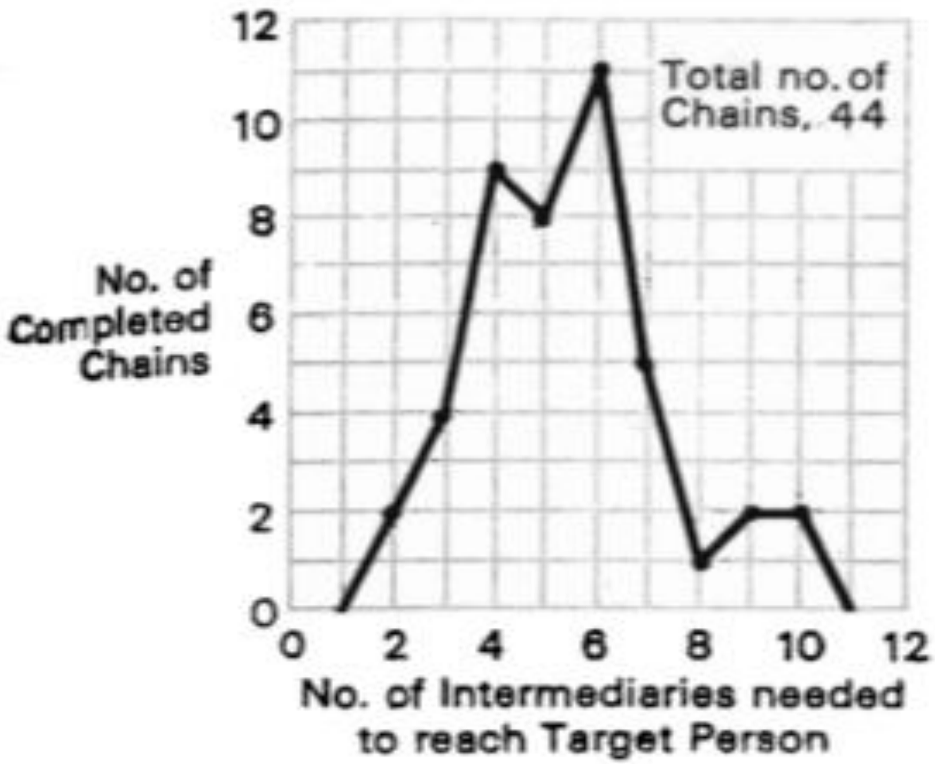
- that even when most of our connections are local, any pair of people can be connected by a fairly small number of relational steps.
  - Six degree of Kelvin Bacon

6 degree on the Facebook



Today it's 3 and half degrees of separation on FB
https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

Milgram’s 1967 “small world experiment



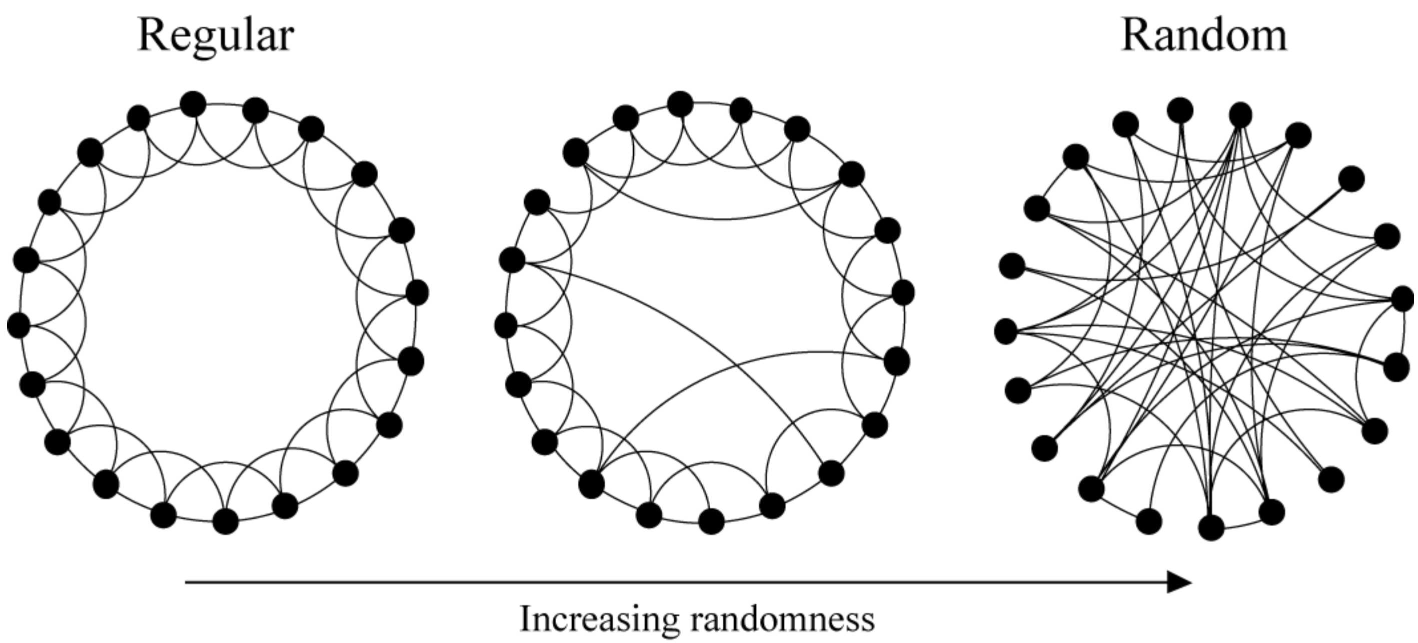
It’s a Small World, After All

- This is essentially the “six degrees of separation” idea—that the number of “steps” or “links” needed to connect any one arbitrarily chosen individual to any other is low
- In Milgram’s 1967 “small world experiment”, individuals were asked to reach a particular target individual by passing a message along a chain of acquaintances.
- For successful chains, the an average (medium?) # of intermediaries needed was 5 (that is, 6 steps)—although note that most chains were not completed. (Easley and Kleinberg, 2010; p. 36, 37)

How to model the small world phenomenon ?

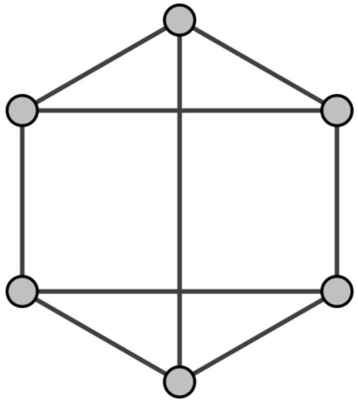
- Two possible graphs (almost at opposite ends of a spectrum) are “random graphs” and “regular graphs”.
- A “small world” can be thought of in-between a random and a regular graph.

Modeling of the small world phenomenon



Regular Graphs

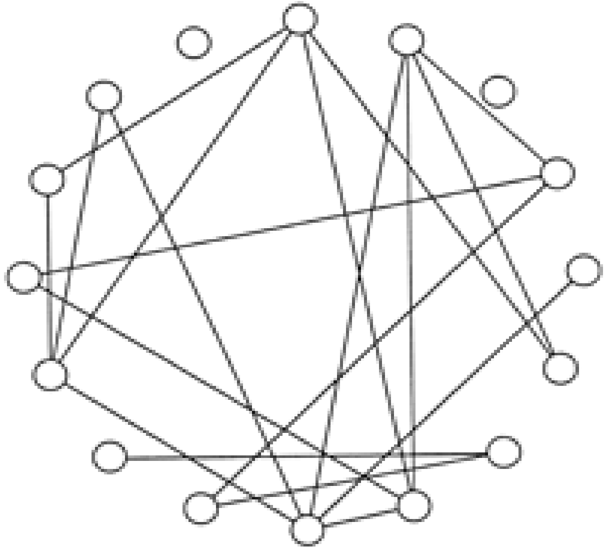
- A **regular graph** is a network where each node has the same number (k) of neighbors (that is, each node or vertex has the same degree k).
- A k-degree graph is seen at the left. k = 3 (each node is connected to three other nodes)



Is it a good model for real social network?

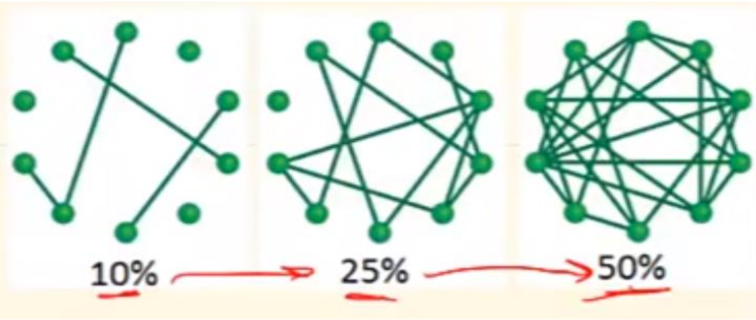
Random Graphs

- This graph has 16 nodes,
  - 120 possible connections, and
  - 19 actual connections
  - about a 1/7 probability that any two nodes will be connected to each other (i.e. density).
- Any two vertices have some *uniform probability* of being connected (i.e. independent event)



Random graphs

- Random graph formation
  - Consider each pair of nodes
  - Establish link with certain probability
  - What will happen if we have a fixed number of nodes and gradually raise the probability



Random Graphs

- Studied by P. Erdös A. Rényi in 1960s
- How to build a random graph
  - **Take *n* vertices**
  - **Connect each pair of vertices with an edge with some probability *p***
- For a network of n nodes, this condition implies that the expected degree (K) of an actor is

Average degree k = p (n-1)

Let n = number of nodes
P = a fix probability that any pair of nodes are connected
E = total number of edges in the graph

E = P * all possible pairs, which is?

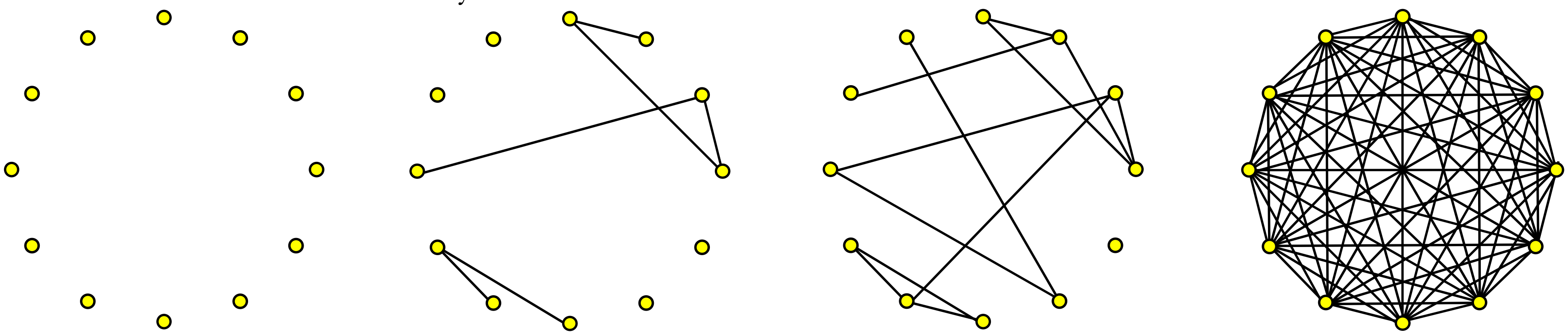
Average degree K = 2E/n (why 2E?)
= P * n (n-1)/n
= P * (n-1)

$$K = \frac{n(n-1)p}{n} = (n-1)p \approx np$$

Random Graphs

Preexistence of a fixed number of nodes (N : number of nodes : 12)
uniform probability of being connected (P : the probability of two nodes being connected).
Average degree, $k = p\ (N\text{-}1)$

*Slides modified from Kentaro Toyama



$p = 0.0$; $k = 0$ | $p = 0.09$; $k = 1$ | $p = 0.181$; $k = 2$ | $p = 1.0$; $k \approx N$

Size of Giant component

1 | 5 | **11** | 12

Let's try generate random graph using Gephi

Phase transition and the critical threshold

- "Random network theory tells us that **as the average number of links per node increases beyond the critical one, the number of nodes left out of the giant cluster decreases exponentially... The more link we add, the harder it is to find a node that remains isolated. "**
- You start with a large number of isolated nodes. Then you randomly add links between the nodes, mimicking the random encounters between the quests.
- When you add enough links such **that each node has an average of on line**, a miracle happens: A unique giant cluster emerges. That is, most nodes will be part of a single cluster such that, starting from any node, we can get to any other by navigating along the links between the nodes. ...The network, after placing a critical number of links, drastically changes.

[Simulation of a phase transition]

Phase transition

*Slide adopted from Kentaro Toyama

Erdős and Renyi (1959)

If $k < 1$:

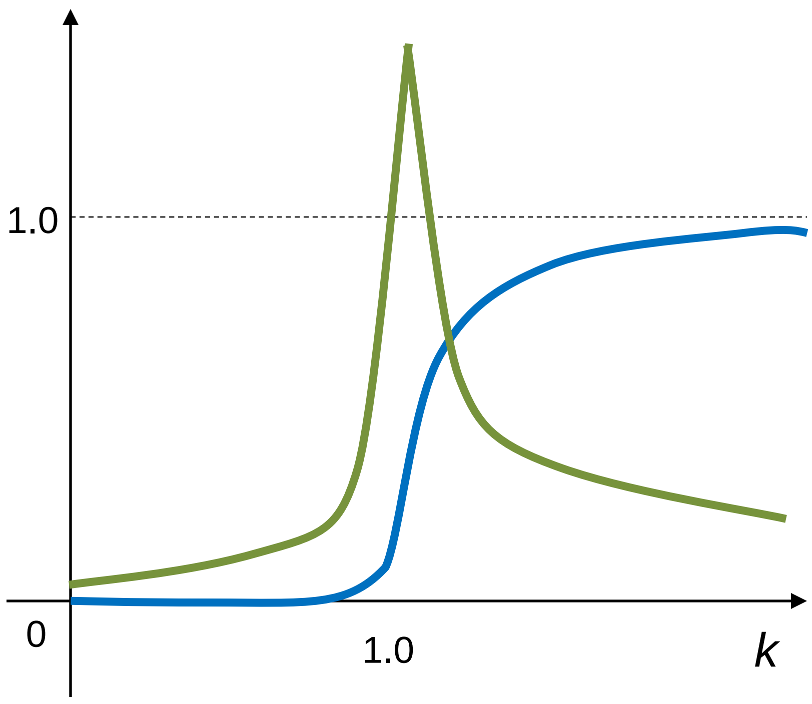
- small, isolated clusters
- small diameters
- short path lengths

At k = 1:

- a *giant component* appears
- diameter peaks
- path lengths are high

For k > 1:

- almost all nodes connected
- diameter shrinks
- path lengths shorten



phase transition

As it is easy for each person to have at least one friend, therefore Small world

Random vs. Real Social Networks

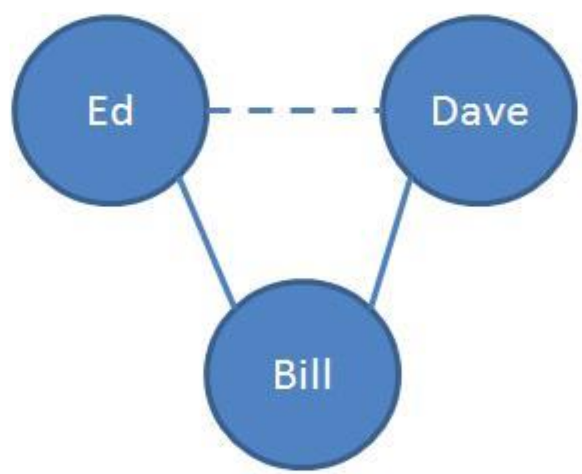
- Two problematic assumption of random network
  - Link formations are independent from each other
    - The presence of a connection between A and B as well as a connection between B and C will not influence the probability of a connection between A and C. (independence)
  - A link between any pair of nodes has a uniform probability
    - The degree distribution following bell curve
    - But, everyone is equally popular?

Clustering

- Most actors live in local neighborhoods where most others are also connected to one another
  - *Preferential attachment*
    - *Easier to make friends with people nearby*

Triadic closure

- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in future.



Real Networks

- Local clustering
  - If A know B, and C, there is higher probability that B also know C
- Supper-connector
  - It was found in Milgram's experiment that approximately 60% of the transmissions passed the same four people
  - Network dynamics

In summary, random graphs

- Do not generate local clustering and triadic closures. ER graphs have a low clustering coefficient.
- Do not account for the formation of hubs. Formally, the degree distribution of ER graphs converges to a Poisson distribution, rather than a power law observed in many real-world, scale-free networks.

Modeling Real Network

- Duncan J. Watts and Steven Strogatz , 1998
  - Clustering
- Barabási, A.-L.; R. Albert (1999).
  - "Scale-free" network and preferential attachment
  - Look into the dynamic of link formation

Small-World Model

- Watts-Strogatz (1998) first introduced small world mode
- In between the regular and random networks
  - Regular Graphs have a high clustering coefficient, but also a high diameter
    - The length of the shortest path connecting two nodes grow very slowly, i.e., in general logarithmically, with the size of the network
  - Random Graphs have a low clustering coefficient, but a low diameter

From Caveman graph to small world

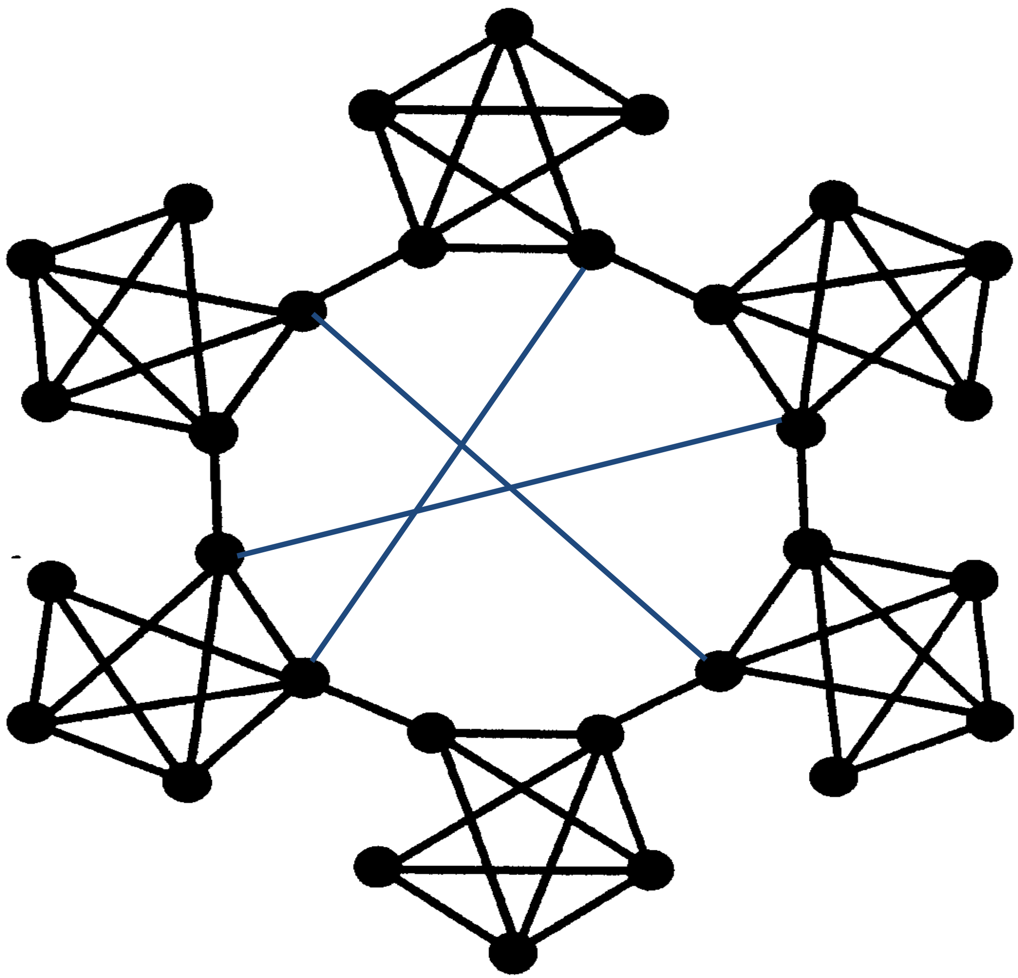


FIG. 1.—Schematic of the connected caveman construction

In a highly clustered, ordered network, a single random connection will create a shortcut that lowers L dramatically

Small Worlds

- Note that it is highly clustered—a higher proportion (than one would expect randomly) of each node's neighbors are actually connected to *each other*
  - *Preferential attachment*
    - *Easier to make friends with people nearby*
- It also has a small diameter, relative to the number of nodes.

Small Worlds

- A graph is small-world if it has significantly higher average clustering coefficient than a random graph constructed with the same number of edges, and if the graph has a short mean-shortest path length.
- These two characteristics are often mutually exclusive in random graphs—but do describe a wide variety of real-life situations.

Local cluster + random shortcuts

Watts-Strogatz's small world model

- Add edges to nodes, as in random graphs, but makes links more likely when two nodes have a common friend
- Add random links reduce diameter greatly

To quantify a small world

- Average path length
  - Measures the average number of intermediaries, that is, the degrees of separation, between any two actors in the network along their shortest path of intermediaries.
- Clustering coefficient
  - How many of an actor's contacts are connected to each other

Clustering coefficient

- Measure how close a node's neighbors are to being a clique
- If A is connected to B, and B is connected to C, then it is likely that A is connected to C
  - "A friend of your friend is MORE LIKELY TO BE your friend"

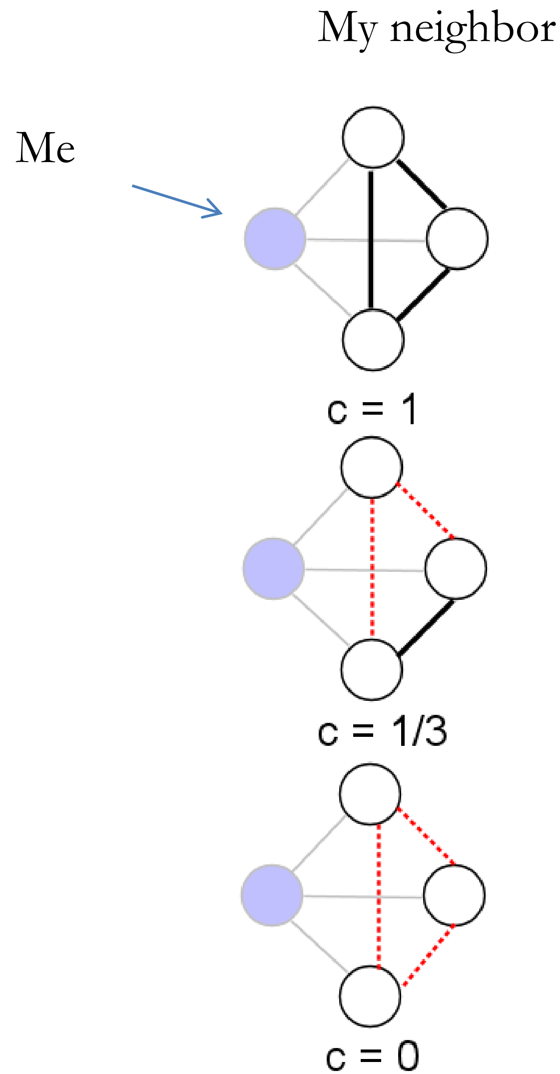
Clustering Coefficients

- As a way to measure how close a node (or vertex) and its neighbors are from being a clique, or a complete graph within a larger graph or network.

- The clustering coefficient of a node is the number of actual connections across the neighbors of a particular node, as a percentage of possible connections.

Actor clustering coefficient

the number of edges connecting the neighbors of the vertex divided by the maximum number of such edges (red dots).



CC of actor i

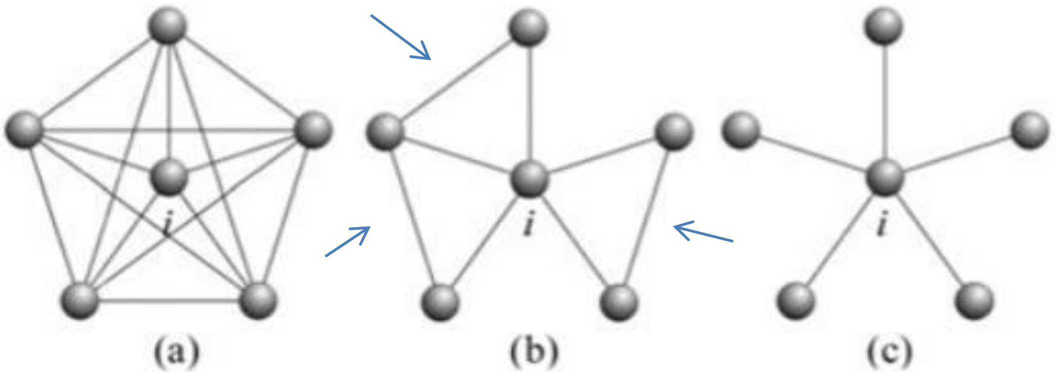


Figure 4 - Example of three networks and respective clustering coefficients (see Eq. (1)). In (a), $cc_i = \dfrac{10(2)}{5(4)} = 1$ (the vertices around i are fully connected), (b) $cc_i = \dfrac{3(2)}{5(4)} = 0.3$ and (c) $cc_i = \dfrac{0(2)}{5(4)} = 0$. The maximum number of edges among the neighbors of i is given by $k_i(k_i - 1)/2$.

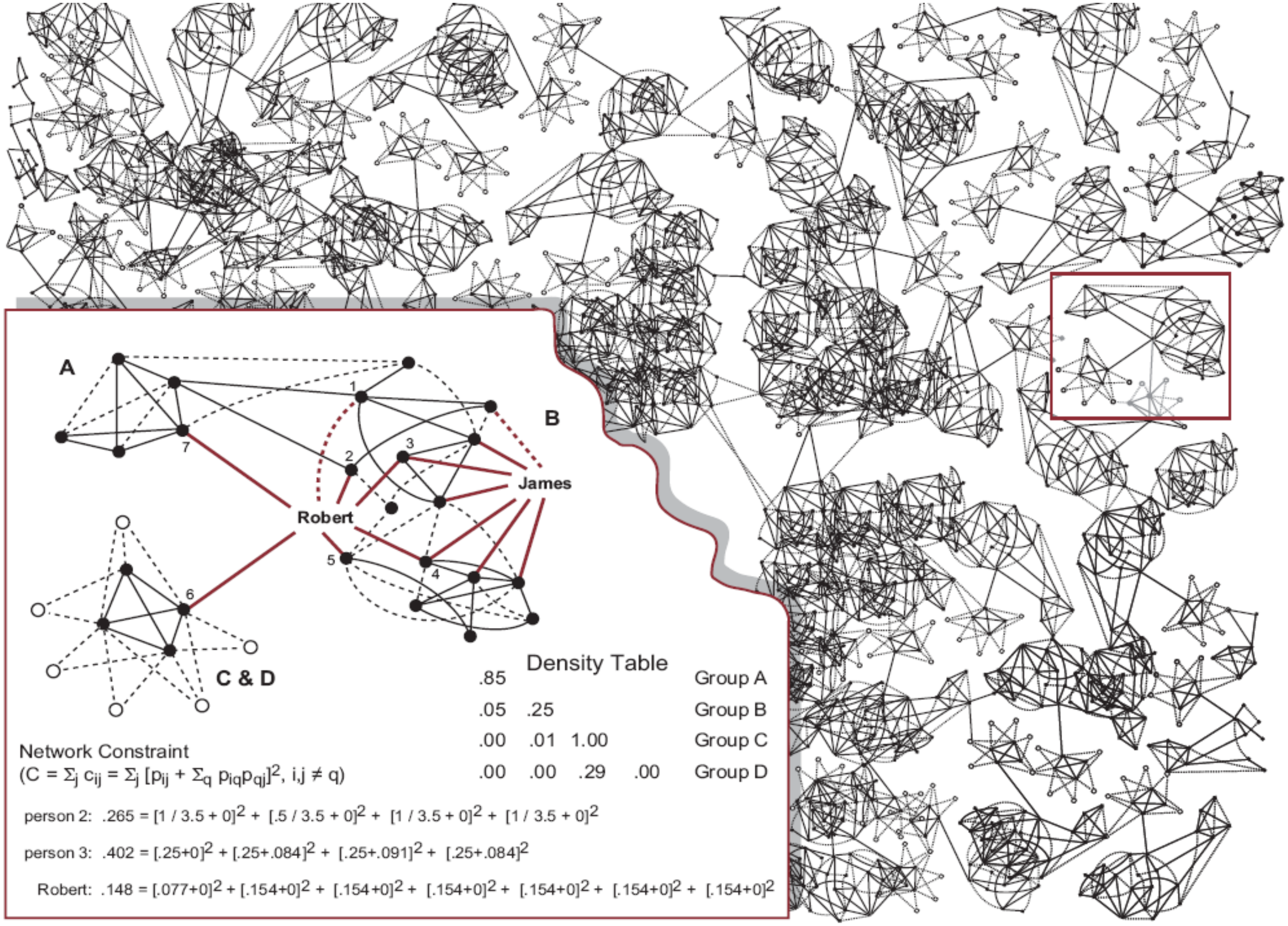


Figure 1. The Small World of Markets and Organizations

Who has a higher clustering coefficient? Robert or James?

Local CF: Average Clustering Coefficients

- This is the formula the clustering coefficient for a group. N=number of nodes. C=clustering coefficient for each node *i*.

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i .$$

Clustering coefficient

- Network>Cohesion>Clustering Coefficient (群聚係數, CC)
- The clustering coefficient of an actor is the density of its open neighborhood.

Clustering coefficient

- To examine the local neighborhood of an actor (that is, all the actors directly connected to ego), and to calculate the density in this neighborhood (leaving out the ego)
- Take the average of the density of the local neighborhoods of all actors (overall approach)
- Weighted approach: actors with larger neighborhoods get more weight in computing the average density

Clustering

- *Network>Cohesion>Clustering Coefficient*

```
Input dataset:                  C:\Program Files\Ucinet 6
Relation: KNOKI
---------------
Overall graph clustering coefficient: 0.607
Weighted Overall graph clustering coefficient: 0.599
```

- Overall: the average of the densities of the neighborhoods of all of the actors.
- Weighted: gives weight to the neighborhood densities proportional to their size.

Weighted clustering coefficient

- Varies from 0 to 1
- A value of 0.65 means that 65% of the triads are closed
- Node level

Node Clustering Coefficients

	1 Clus C	2 nPairs
1	0.667	21.000
2	0.536	28.000
3	0.567	15.000
4	0.733	15.000
5	0.518	28.000
6	0.333	3.000
7	0.514	36.000
8	0.800	15.000
9	0.600	15.000
10	0.800	10.000

- In assessing the degree of clustering, it is usually wise to compare the cluster coefficient to the overall density.
- Why? (strawman)

Expected clustering of a random graph

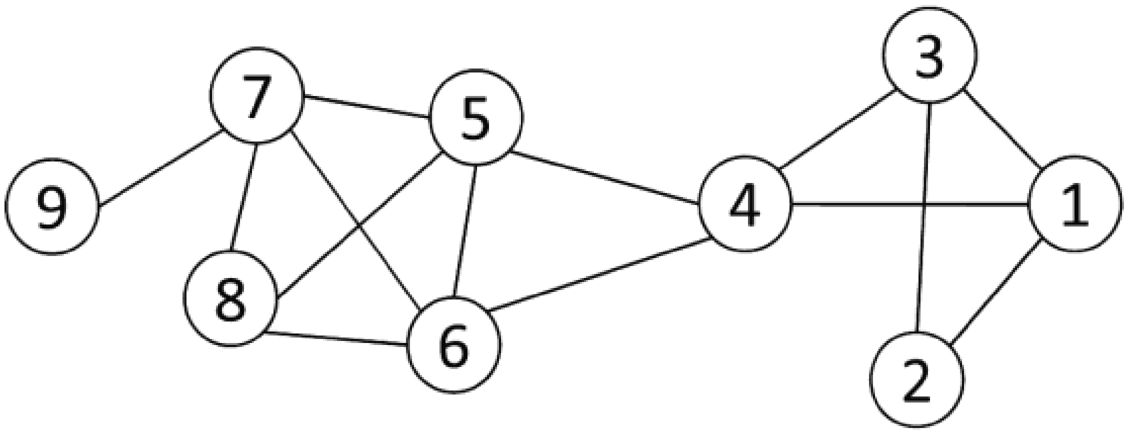
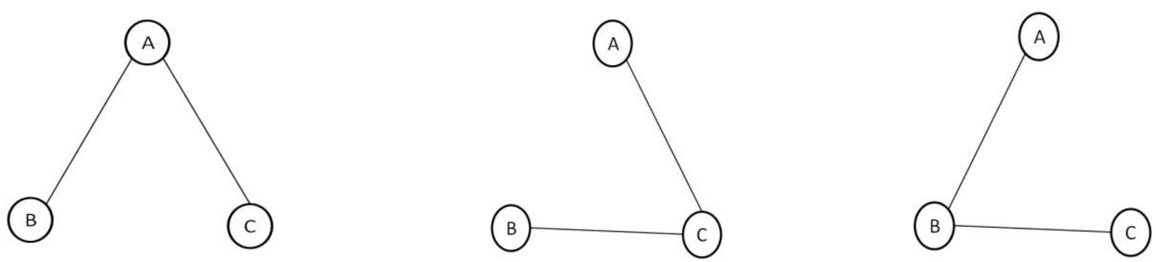


Figure 1.1: A social network of 9 actors and 14 connections. The diameter of the network is 5. The clustering coefficients of nodes 1-9 are: $C_1 = 2/3$, $C_2 = 1$, $C_3 = 2/3$, $C_4 = 1/3$, $C_5 = 2/3$, $C_6 = 2/3$, $C_7 = 1/2$, $C_8 = 1$, $C_9 = 0$. The average clustering coefficient is 0.61 while the expected clustering coefficient of a random graph with 9 nodes and 14 edges is $14/(9 \times 8/2) = 0.19$.

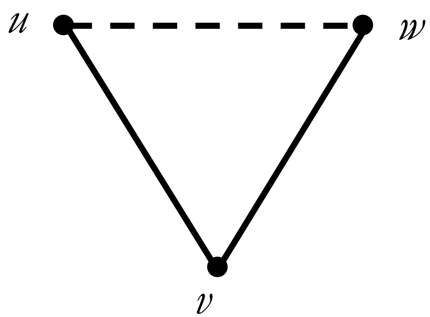
DENSITY

Global clustering coefficient

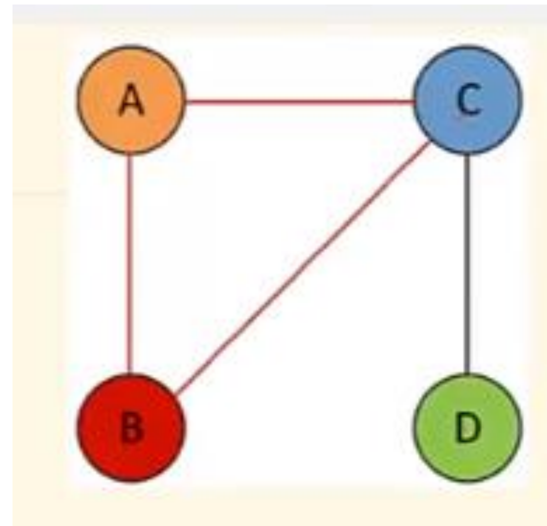
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$



Triad closure

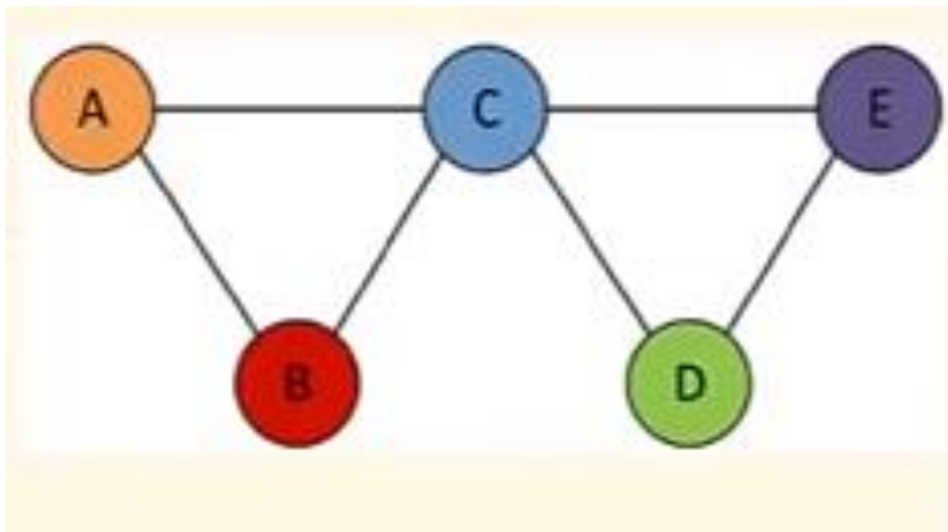


The path *uvw* (solid edges) is said to be closed if the third edge directly from *u* to *w* is present (dashed edge)



How many closed triads?
How many unclosed triads?

Clustering coefficient?



6/10

Ace
Acd
Bcd
Bce

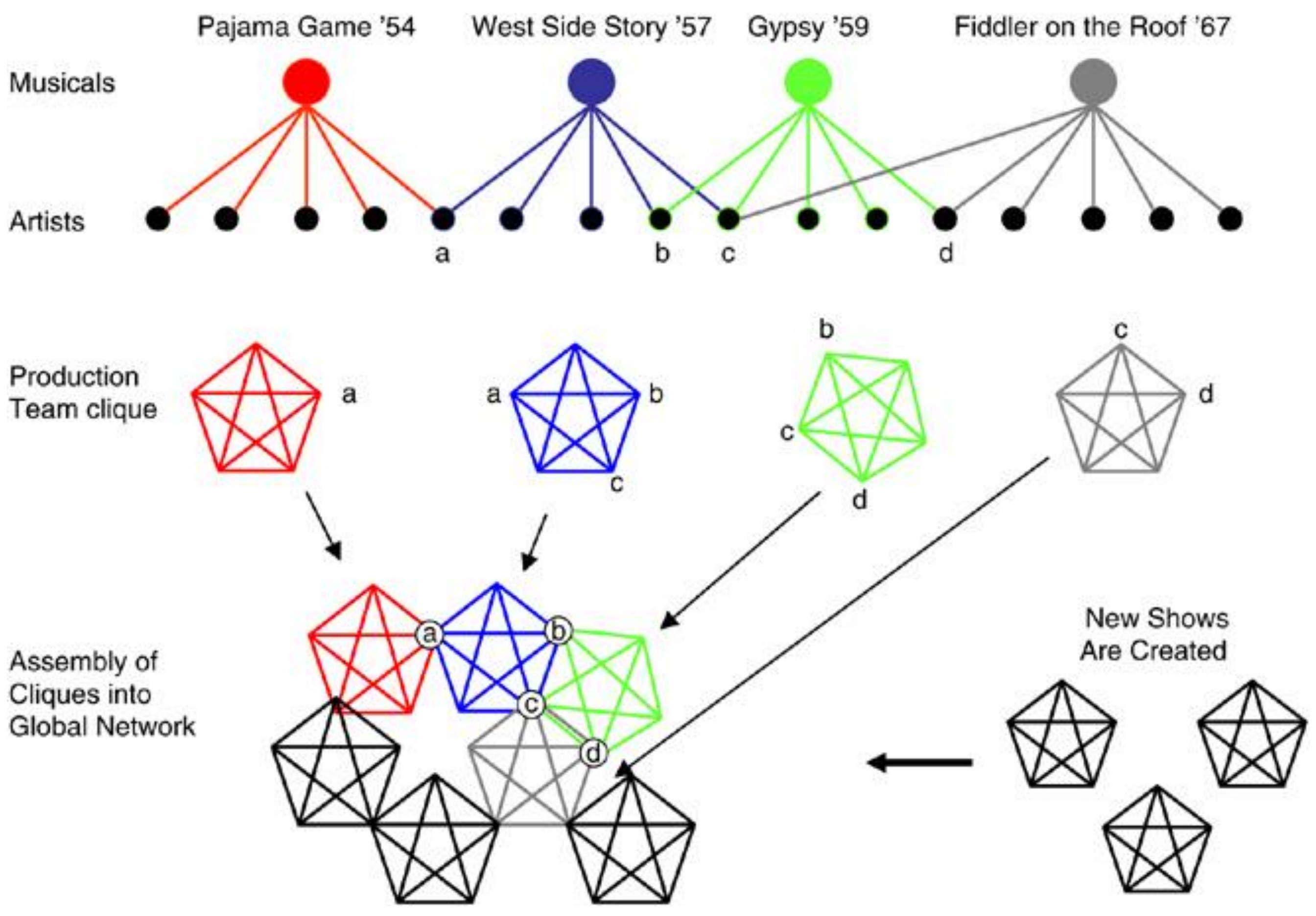
Small world and group creativity

- Good for idea's development and diffusion
  - By engendering trust, cohesive clustering encourages sharing, widespread and lateral communication
- Not so good for the seminal generation of an idea
  - Downside of clustering

To escape this fundamental conundrum:

Cohesive clusters with bridging connections

Real Social Network



Broadway from 1945 to 1989, including shows that were axed before opening night. The final database cataloged 474 musicals and listed 2,092 artists,

Small world quotient

- Short global separation and high local clustering
- Average path length (PL) and high clustering coefficient (CC)
- PL ratio (PL of the actual network/PL of a random graph)
- CC ration (CC of the actual network/CC of the random graph)
- The small world quotient then is

  CC ratio/Pl ratio

Random expectation of L and C

Random expectations:
For basic one-mode networks, we can get approximate random values for PL and CC as:

$$PL_{random} \sim \ln(n) / \ln(k)$$
$$CC_{random} \sim k / n$$

As k and n get large.

Note that C essentially approaches zero as N increases, and K is assumed fixed. This formula uses the density-based measure of C, but the substantive implications are similar for the triad formula.

Broadway performance: financial success

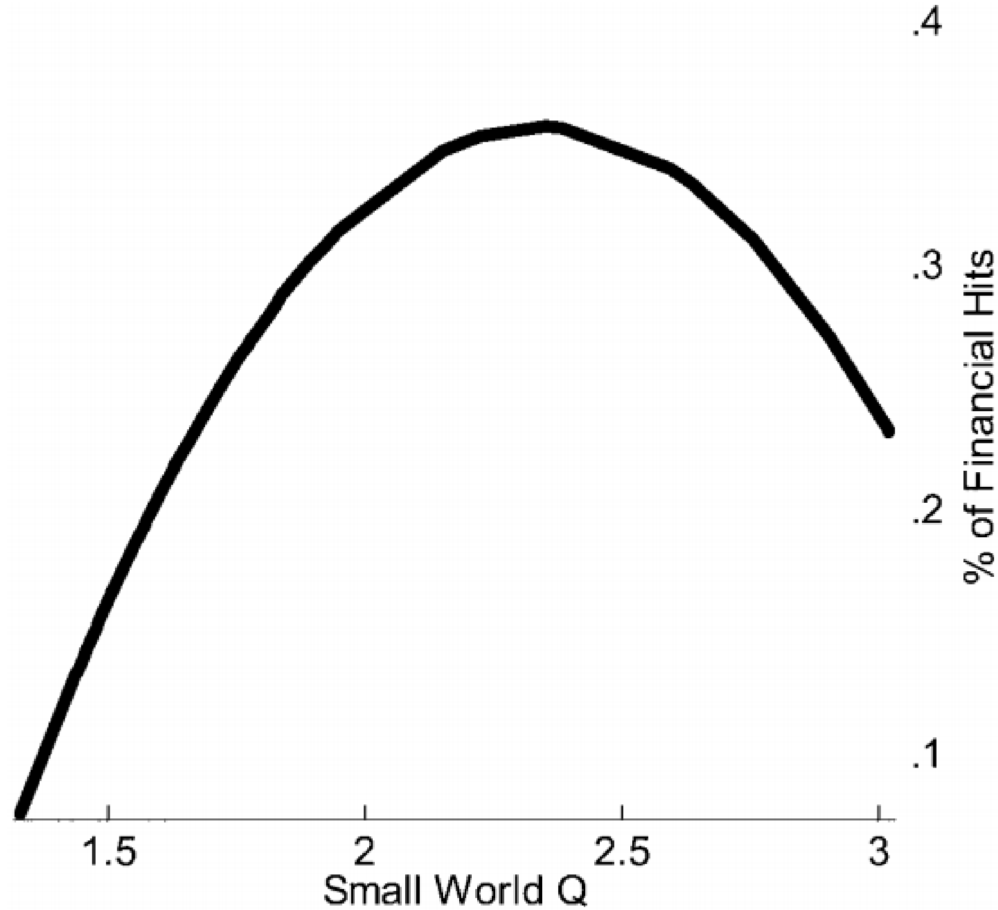


FIG. 6.—Financial success of a season

Broadway performance: artistic success

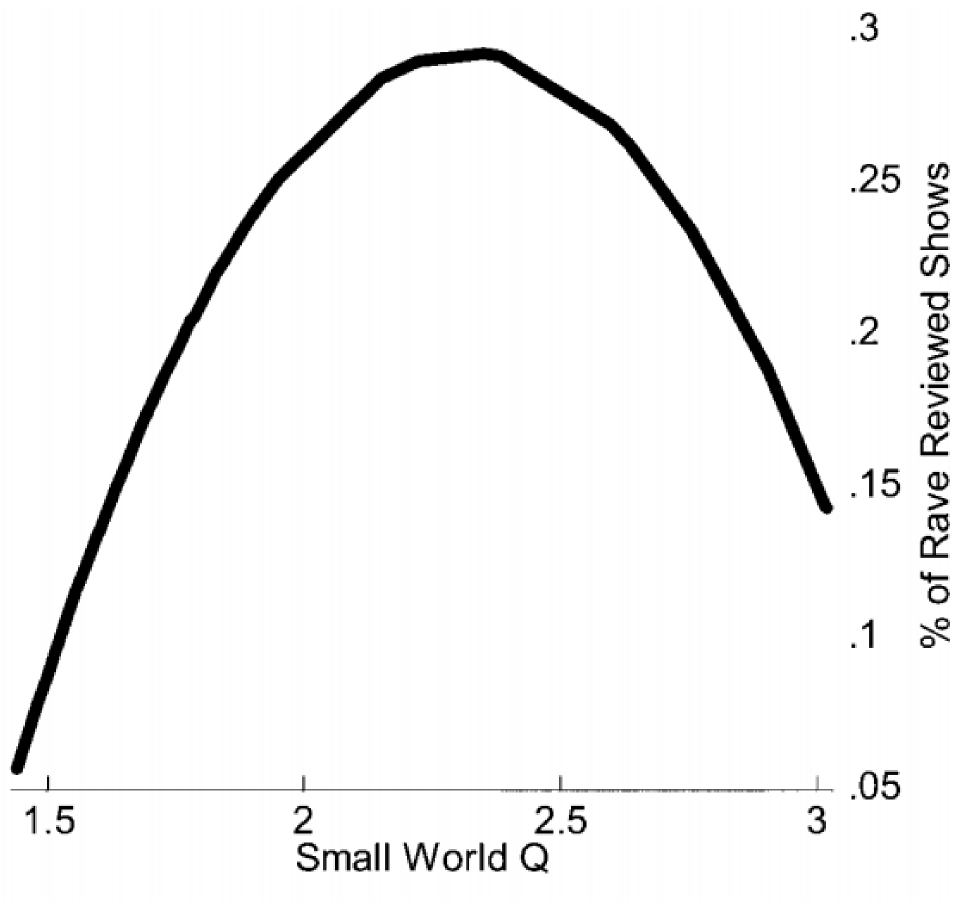


FIG. 7.—Artistic success of a season

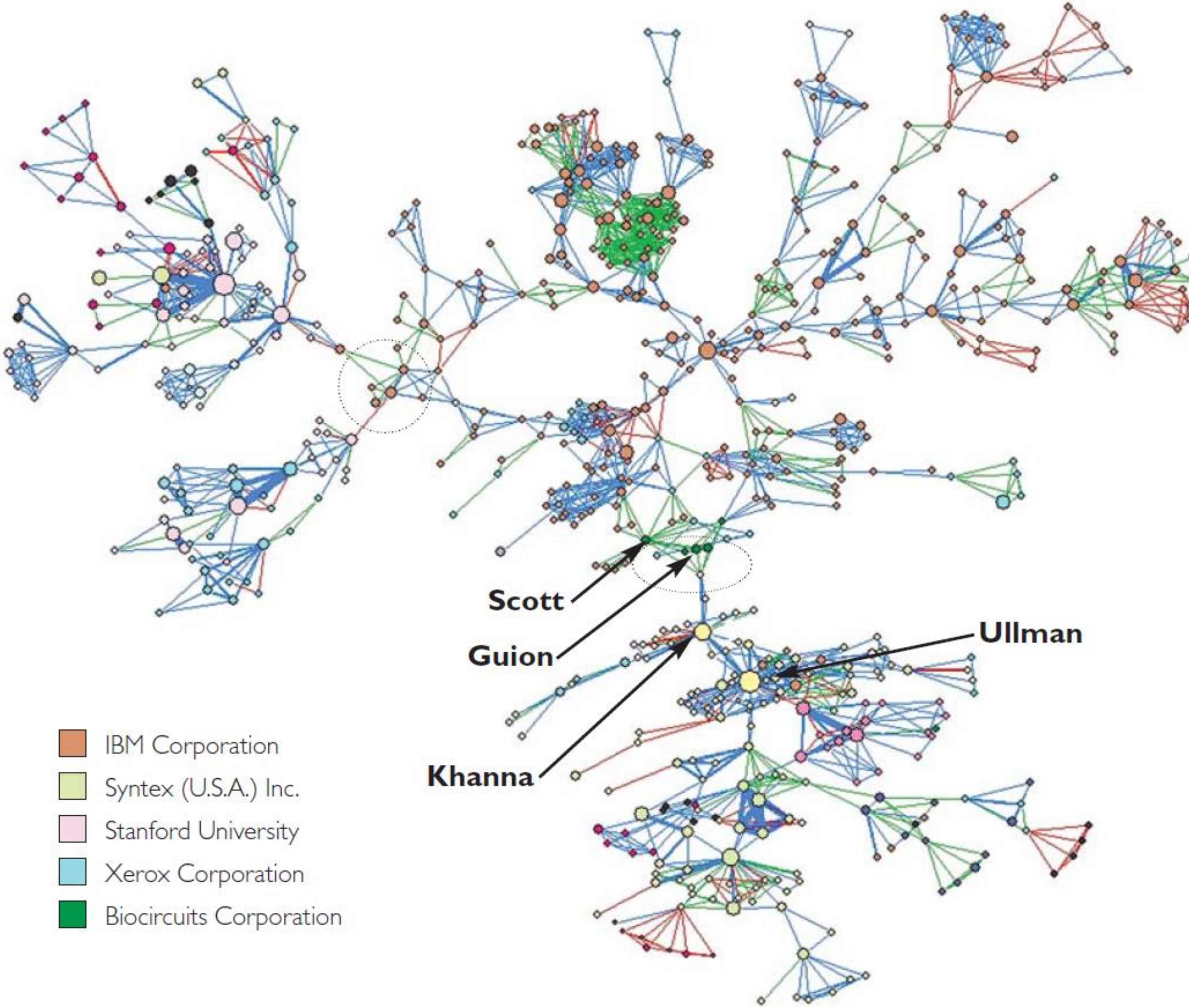


FIGURE 3. Inventors in Silicon Valley's Largest Connected Cluster circa 1986-1990

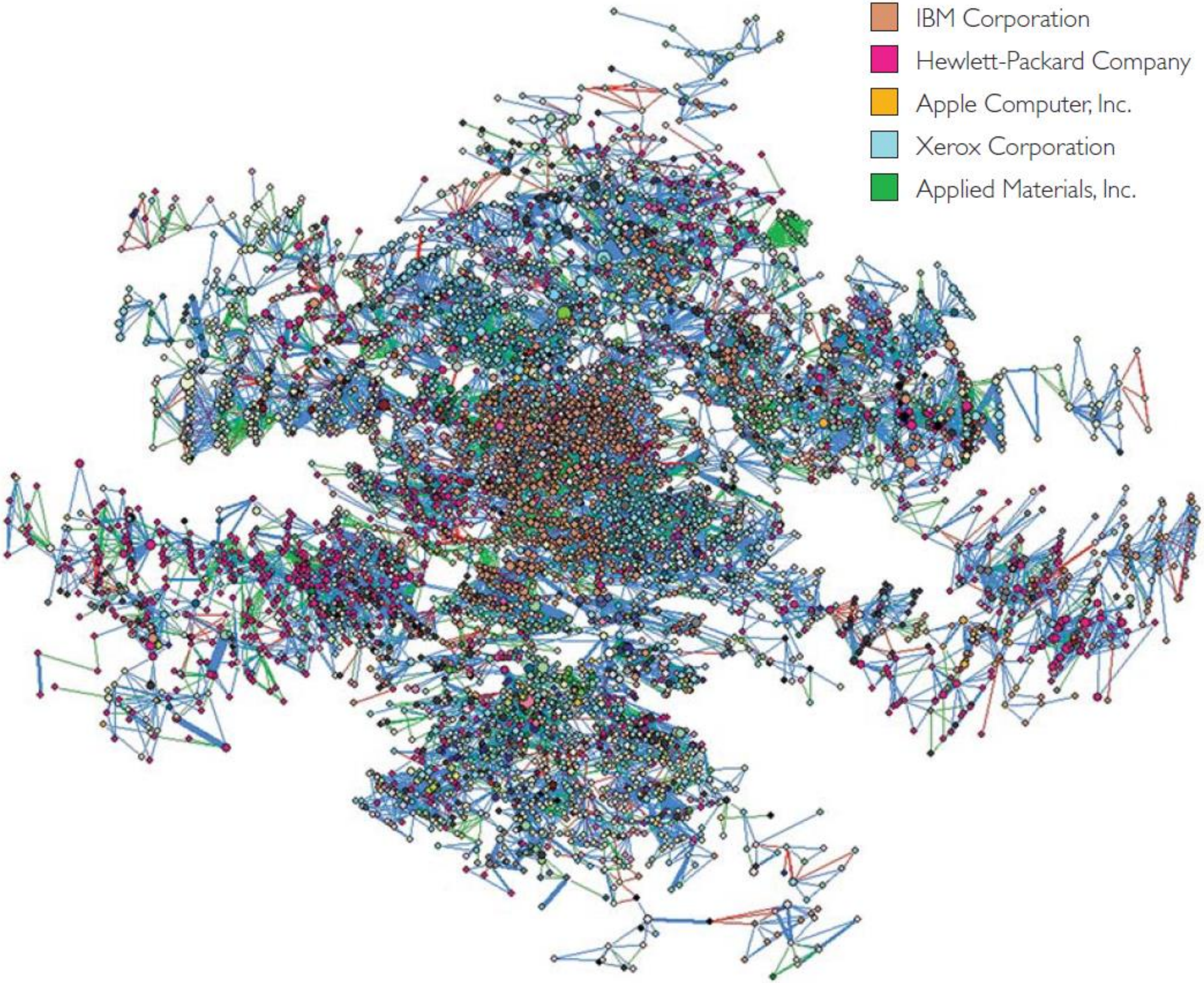


FIGURE 1. The 7,244 Inventors in Silicon Valley's Largest Collaborative Cluster in the Mid-1990s

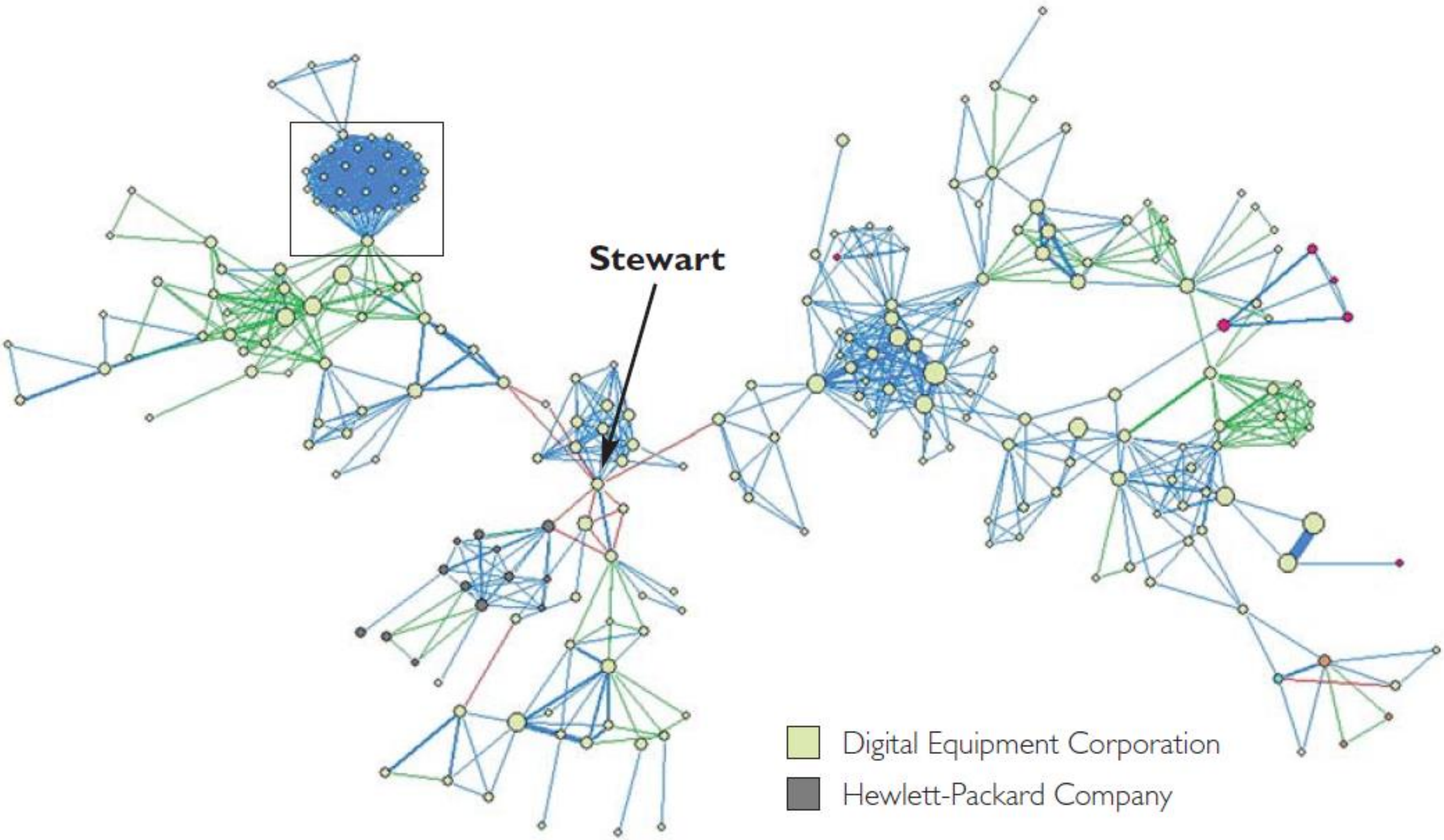


FIGURE 2. Inventors in Boston's Largest Connected Cluster circa 1986-1990

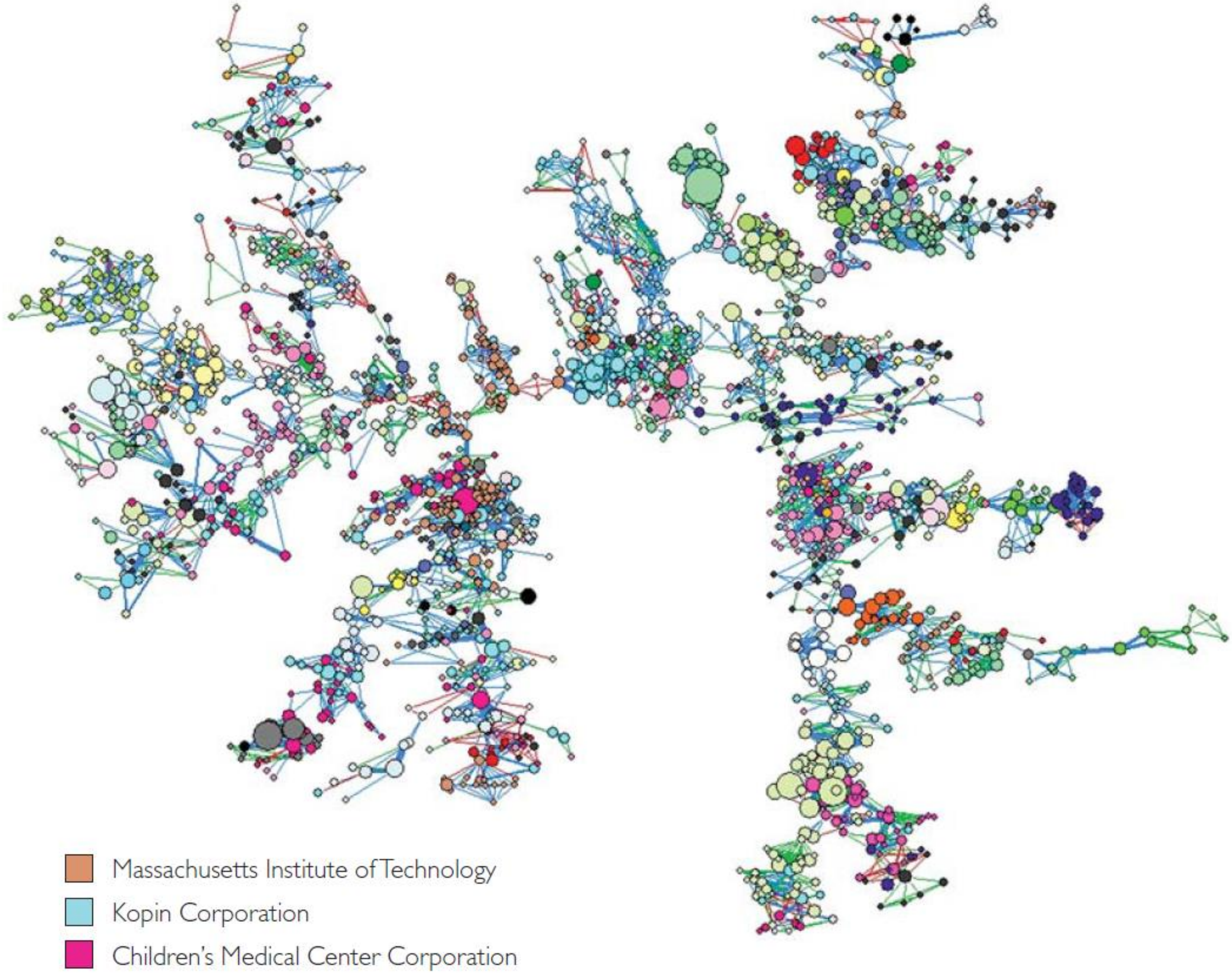


FIGURE 4. Boston's Largest Connected Cluster circa 1989-1993

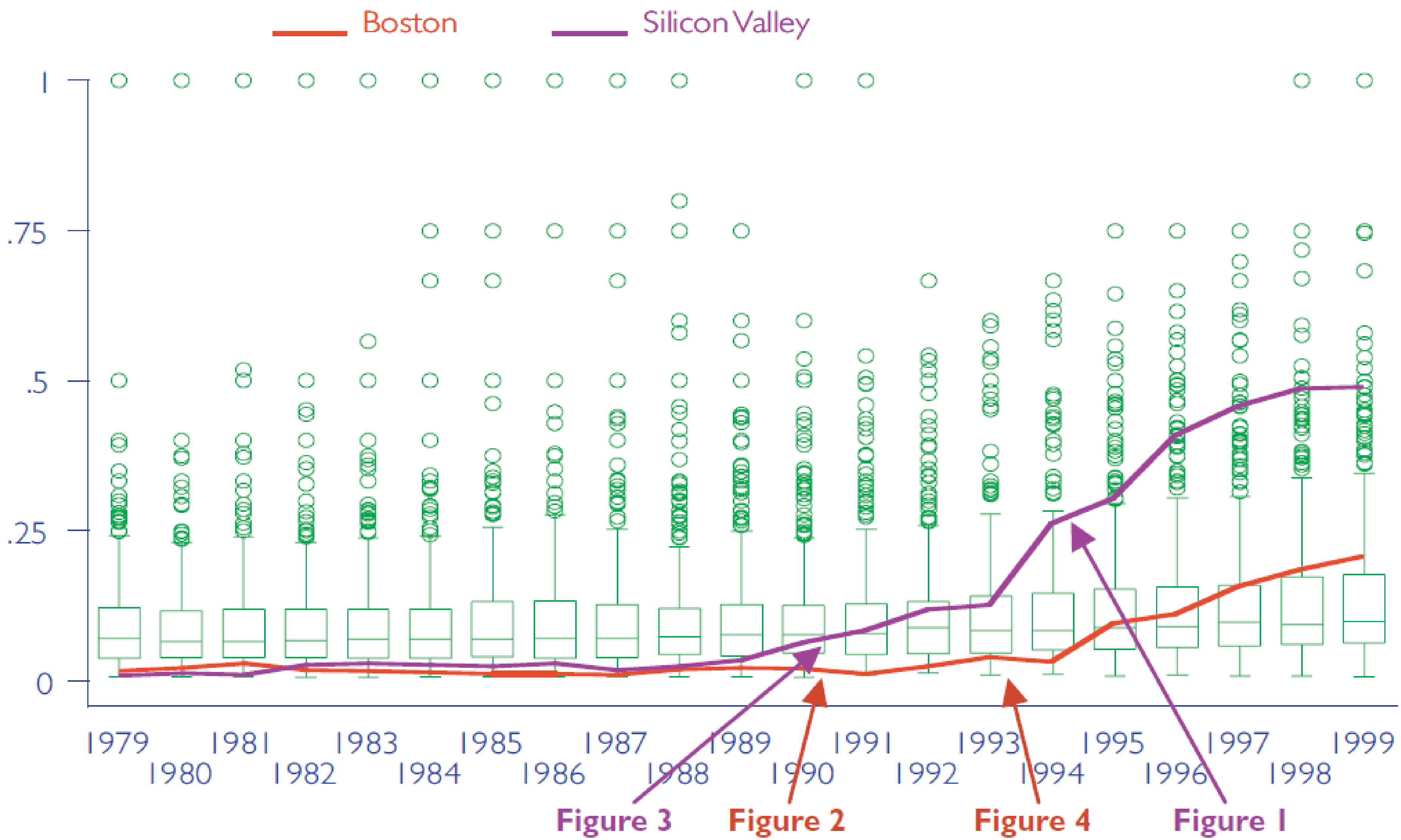
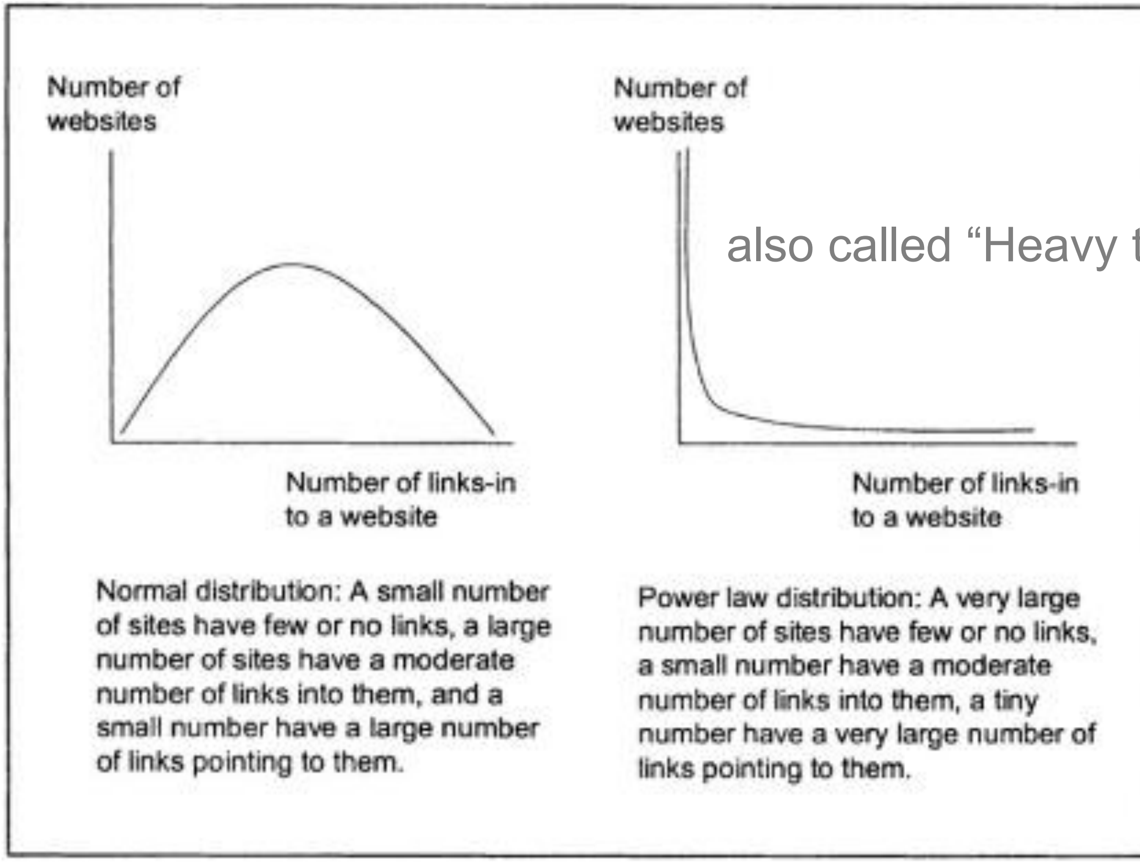


FIGURE 5. Box Plot of Proportion of Inventors within the Largest Collaborative Cluster in each U.S. Metropolitan Statistical Area (MSA)

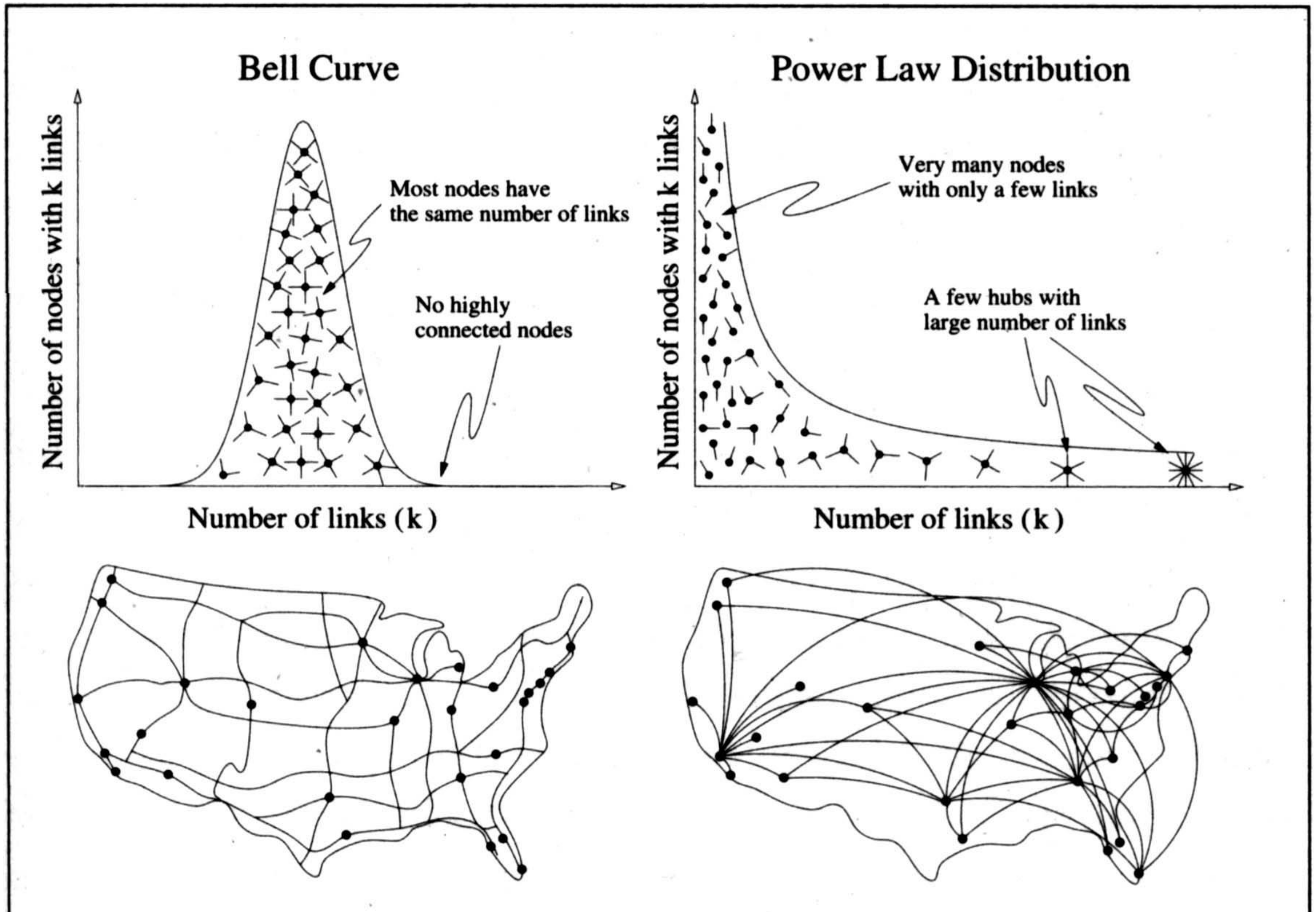
Power law distribution



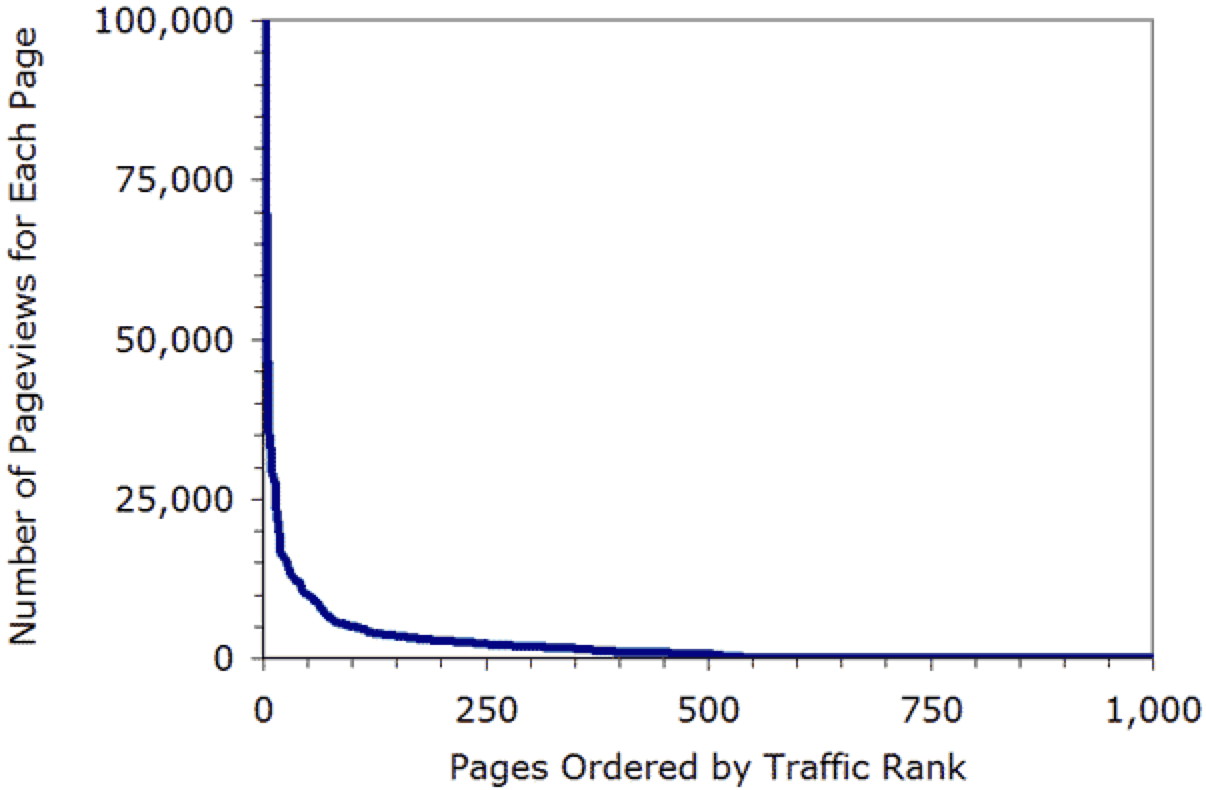
also called “Heavy tailed” distribution

Figure 7.4: Illustration of How Normal Distribution and Power Law Distribution Would Differ in Describing How Many Web Sites Have Few or Many Links Pointing at Them

“Scale-free” distribution



Zipf's law in webpage popularity



Power – law distribution

for evolving self-organized networks

proposed by Barabasi and collaborators

$$P(k) \propto k^{-\gamma}$$

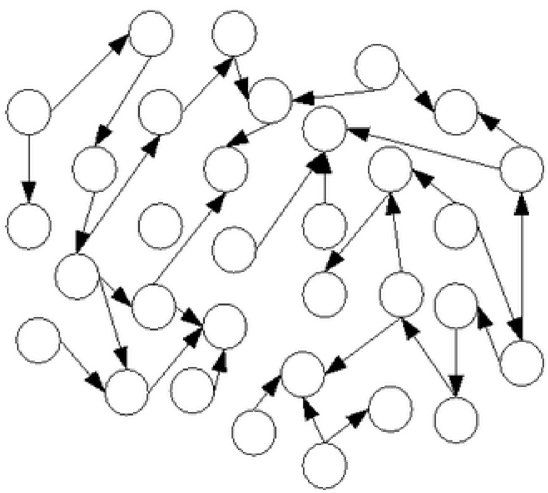
Typical range $2 < \gamma < 3$

Probability of a node having K link is the power of K
Nodes twice this size is four times as rare

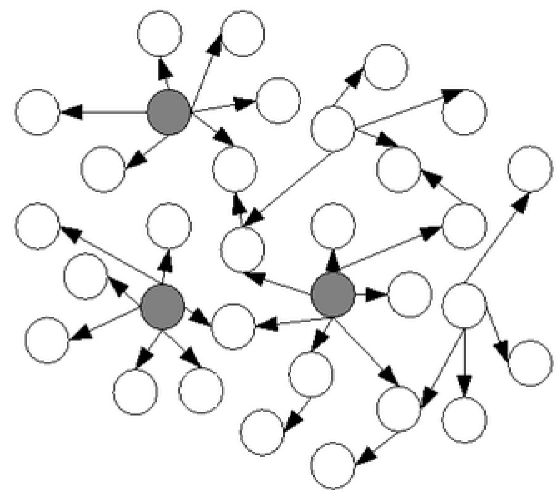
Random vs. Real Social networks

- Random network models introduce
  - A fixed number of nodes
  - an edge between any pair of vertices with a uniform probability *p*

- Real networks are not exactly like these
  - Nodes added to the network with time, it's evolving, instead of static
  - Tend to have a relatively few nodes of high connectivity (the "Hub" nodes)



(a) Random network



(b) Scale-free network

Scale-free networks

- A small number of nodes act has “highly connected hubs” (high degree), but most nodes have a low degree
  - No single node can be seen as “typical”, no intrinsic “scale”
- The scale-free model has a shorter average path length than a random
- Thanks to the hub nodes

The rich get richer

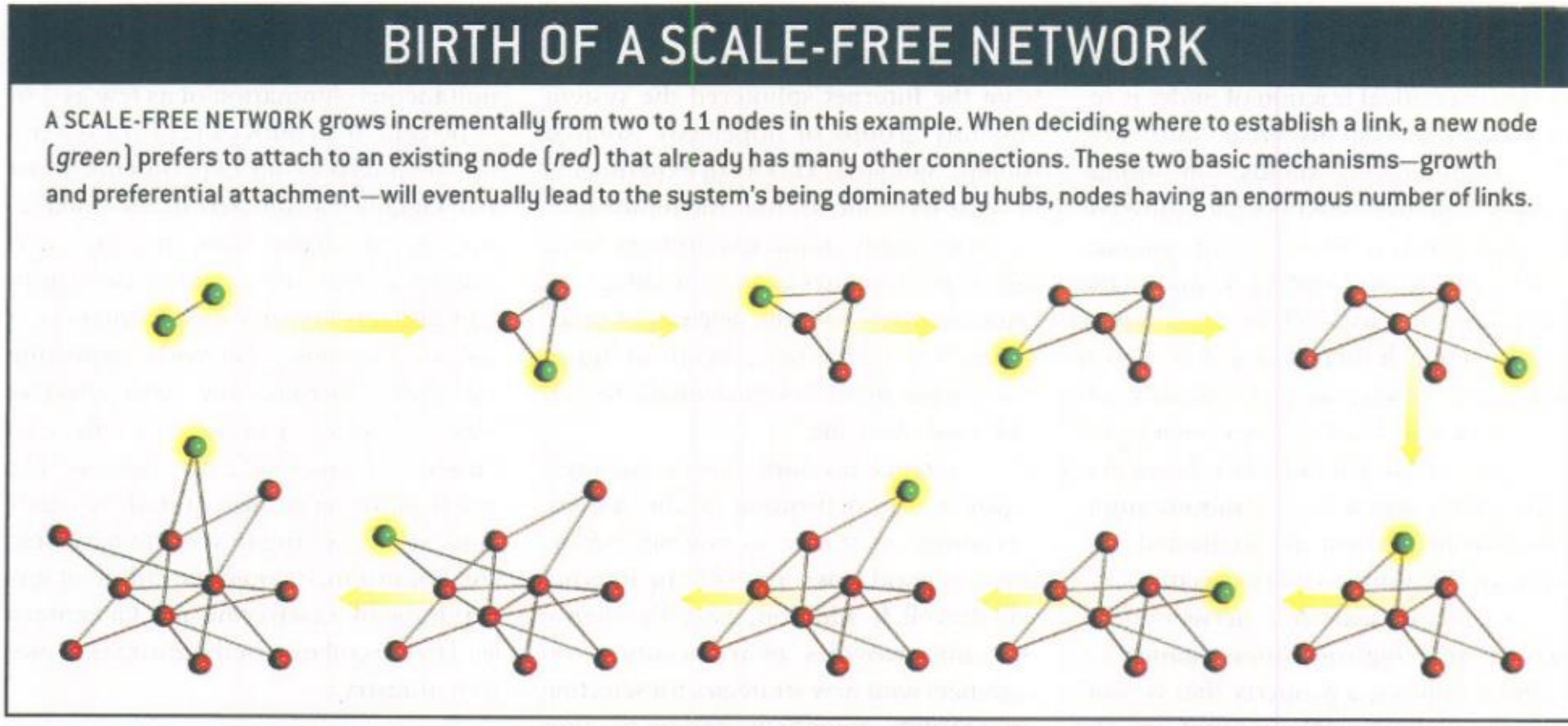
Power-law distribution of node-degree arises if

(but *not* "only if")

- As Number of nodes grow edges are added in proportion to the number of edges a node already has.
- Alternative: Copy model—where the new node copies a random subset of the links of an existing node
    - Sort of close to the WEB reality
- *Preferential attachment*
    - *More incentive to connect with those who own more links*

Preferential attachment (with network effect)

- Network dynamics for evolving self-organized networks



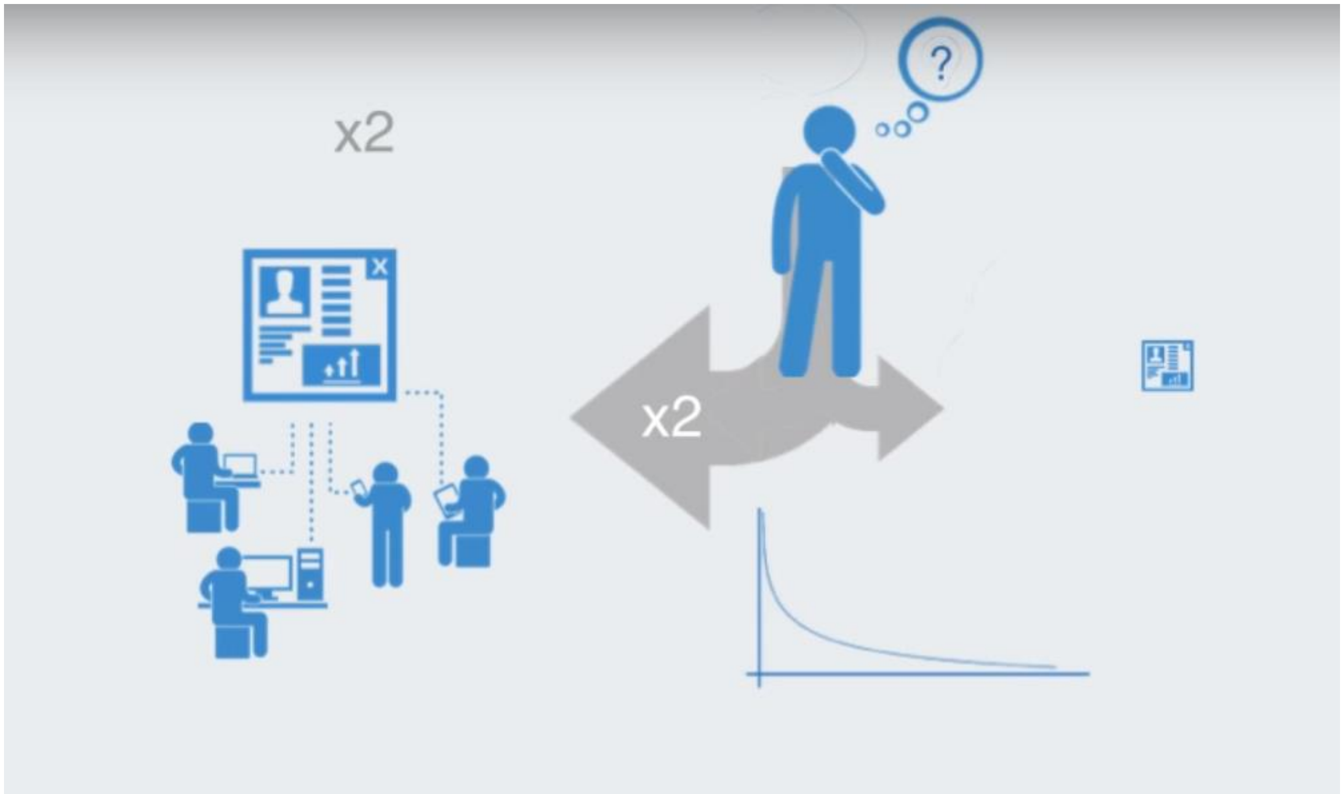
Rich gets richer

Preferential attachment/rich get richer

- that the more connected a node is, the more likely it is to receive new links. Nodes with higher degree have stronger ability to grab links added to the network
- New nodes are added to the network one at a time. Each new node is connected to m existing nodes with a probability that is proportional to the number of links that the existing nodes already have.
- Formally, the probability p_i that the new node is connected to node i is

$$p_i = \frac{k_i}{\sum_j k_j},$$

Preferential attachment (incentive)



[Informative introduction to network dynamics](#)

Matthew effect

- "the rich get richer and the poor get poorer"
  - For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.

    Matthew 25:29